

Computer Science & Information Technology

207

Soft Computing, Artificial Intelligence and Applications

David C. Wyld,
Dhinaharan Nagamalai (Eds)

Computer Science & Information Technology

- 13th International Conference on Computer Science, Engineering and Applications (ICCSEA 2023)
- 4th International Conference on NLP & Artificial Intelligence Techniques (NLAI 2023)
- 12th International Conference on Soft Computing, Artificial Intelligence and Applications (SCAI 2023)
- 14th International Conference on Communications Security & Information Assurance (CSIA 2023)
- 4th International Conference on IoT, Blockchain & Cloud Computing (IBCOM 2023)
- 4th International Conference on Software Engineering and Managing Information Technology (SEMIT 2023)
- 12th International Conference of Networks and Communications (NECO 2023)
- 12th International Conference on Signal, Image Processing and Pattern Recognition (SPPR 2023)
- 4th International Conference on Machine Learning Techniques and Data Science (MLDS 2023)

Published By



AIRCC Publishing Corporation

Volume Editors

David C. Wyld,
Southeastern Louisiana University, USA
E-mail: David.Wyld@selu.edu

Dhinaharan Nagamalai (Eds),
Wireilla Net Solutions, Australia
E-mail: dhinthia@yahoo.com

ISSN: 2231 - 5403

ISBN: 978-1-923107-13-7

DOI: 10.5121/csit.2023.132401 - 10.5121/csit.2023.132423

This work is subject to copyright. All rights are reserved, whether whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the International Copyright Law and permission for use must always be obtained from Academy & Industry Research Collaboration Center. Violations are liable to prosecution under the International Copyright Law.

Typesetting: Camera-ready by author, data conversion by NnN Net Solutions Private Ltd., Chennai, India

Preface

13th International Conference on Computer Science, Engineering and Applications (ICCSEA 2023), 4th International Conference on NLP & Artificial Intelligence Techniques (NLAI 2023), 12th International Conference on Soft Computing, Artificial Intelligence and Applications (SCAI 2023), 14th International Conference on Communications Security & Information Assurance (CSIA 2023), 4th International Conference on IoT, Blockchain & Cloud Computing (IBCOM 2023), 4th International Conference on Software Engineering and Managing Information Technology (SEMIT 2023), 12th International Conference of Networks and Communications (NECO 2023), 12th International Conference on Signal, Image Processing and Pattern Recognition (SPPR 2023), 4th International Conference on Machine Learning Techniques and Data Science (MLDS 2023) was collocated with 12th International Conference on Soft Computing, Artificial Intelligence and Applications (SCAI 2023). The conferences attracted many local and international delegates, presenting a balanced mixture of intellect from the East and from the West.

The goal of this conference series is to bring together researchers and practitioners from academia and industry to focus on understanding computer science and information technology and to establish new collaborations in these areas. Authors are invited to contribute to the conference by submitting articles that illustrate research results, projects, survey work and industrial experiences describing significant advances in all areas of computer science and information technology.

The ICCSEA 2023, NLAI 2023, SCAI 2023, CSIA 2023, IBCOM 2023, SEMIT 2023, NECO 2023, SPPR 2023 and MLDS 2023. Committees rigorously invited submissions for many months from researchers, scientists, engineers, students and practitioners related to the relevant themes and tracks of the workshop. This effort guaranteed submissions from an unparalleled number of internationally recognized top-level researchers. All the submissions underwent a strenuous peer review process which comprised expert reviewers. These reviewers were selected from a talented pool of Technical Committee members and external reviewers on the basis of their expertise. The papers were then reviewed based on their contributions, technical content, originality and clarity. The entire process, which includes the submission, review and acceptance processes, was done electronically.

In closing, ICCSEA 2023, NLAI 2023, SCAI 2023, CSIA 2023, IBCOM 2023, SEMIT 2023, NECO 2023, SPPR 2023 and MLDS 2023 brought together researchers, scientists, engineers, students and practitioners to exchange and share their experiences, new ideas and research results in all aspects of the main workshop themes and tracks, and to discuss the practical challenges encountered and the solutions adopted. The book is organized as a collection of papers from the ICCSEA 2023, NLAI 2023, SCAI 2023, CSIA 2023, IBCOM 2023, SEMIT 2023, NECO 2023, SPPR 2023 and MLDS 2023.

We would like to thank the General and Program Chairs, organization staff, the members of the Technical Program Committees and external reviewers for their excellent and tireless work. We sincerely wish that all attendees benefited scientifically from the conference and wish them every success in their research. It is the humble wish of the conference organizers that the professional dialogue among the researchers, scientists, engineers, students and educators continues beyond the event and that the friendships and collaborations forged will linger and prosper for many years to come.

David C. Wyld,
Dhinaharan Nagamalai (Eds)

General Chair

David C. Wyld,
Dhinaharan Nagamalai (Eds)

Organization

Southeastern Louisiana University, USA
Wireilla Net Solutions, Australia

Program Committee Members

Abdel-Badeeh M. Salem,	Ain Shams University, Egypt
Abdul Ghani Albaali,	King Talal School of Business Technology, Jordan
Abhimanyu Mukerji,	Stanford University, USA
Addisson Salazar,	Universitat Politecnica de Valencia, Spain
Adnan Aldemir	Van Yuzuncu Yl University, Turkey ,
Adrian Olaru,	University Politehnica of Bucharest, Romania
Afizan Azman,	Taylor's University, Malaysia
Ahrar Husain,	Jamia Millia Islamia Faculty, India
Alak Roy,	Tripura University, India
Alex Mathew,	Bethany College, USA
Ali Hussein Wheeb,	University of Baghdad, Iraq
Alireza Valipour Baboli,	University Technical and Vocational, Iran
Almoataz B. Al-Said,	Cairo University, Egypt
Amalina Farhi binti Ahmad Fadzlah,	National Defence University of Malaysia, Malaysia
Amel Ourici,	Badji Mokhtar University of Annaba, Algeria
Amos O Olagunju,	St Cloud State University, US
Ana Luisa V Leal,	University of Macau, China
Anand Nayyar,	Duytan University, Vietnam
Anas Alsobeh,	Southern Illinois University Carbondale, Illinois
Andrea Visconti,	Università degli Studi di Milano, Italy
Angela Lee S.H,	Sunway University, Malaysia
Anirban Banik,	National Institute of Technology Sikkim, India
Antonio Abreu,	Polytechnic Institute of Lisbon, Portugal
Anurag Sewak,	Rajkiya Engineering College, India
Aridj mohamed,	Hassiba Benbouali University Chlef, Algeria
Arunkumar Khannur,	CMR University, India
Ashkan Ebadi,	National Research Council, Canada
Atul Garg,	Chitkara University, India
Ay_kaa,	Oxford University Society Northern California, USA
Azeddine WAHBI,	Ibn Tofaïl University, Morocco
Azlan Mohd Zain,	University of Technology, Malaysia
B.K. Rout,	Birla Institute of Technology & Science, India
Babu Rao,	Abdul Haq Urdu University, India
Badrul Hisham Ahmad,	Universiti Teknikal Malaysia Melaka, Malaysia
Bahaa Al-Musawi,	University of Kufa, Iraq
Bawin Aye,	Higher Education Center, Myanmar

Bencheriet Chemesse ennehar,	University of Guelma, Algeria
Bernd E. Wolfinger,	University of Hamburg, Germany
BhagyaLakshmi,	Chaitanya Bharathi Institute of Technology, India
Bibudhendu Pati,	Rama Devi Women's University, India
Bilal Alatas,	Firat University, Turkey
Bixing Yan,	New York University, United States
BOULOZ Abdellah,	LabSIV Ibn Zohr University, Agadir, Morocco
Bvanss Prabhakar Rao,	Vellore Institute of Technology, Chennai
Carlos Becker Westphall,	Federal University of Santa Catarina, Brazil
Chahinez Meriem Bentaouza,	Mostaganem University, Algeria
Chandra Singh,	Sahyadri College of Engineering & Management, India
Chang-Wook Han,	Dong-Eui University, South Korea
Cheng Siong Chin,	Newcastle University, Singapore
Cherkaoui Leghris,	Hassan II University of Casablanca, Morocco
Christian Mancas,	Ovidius University at Constanta, Romania
Cristian Paul Chioncel,	Universitatea Babeş-Bolyai, Romania
Dallel Sarnou,	Abdelhamid Ibn Badis University, France
Dário Ferreira,	University of Beira Interior, Portugal
Dariusz Jacek Jakóbczak,	Koszalin University of Technology, Poland
Deepa Parasar,	Amity University Maharashtra, India
Deyu Lin,	Nanchang University, China
Dhaya,	KCG College of technology, Anna University, India
Dimiter Velev,	University of National and World Economy, Europe
Dimitris Kanellopoulos,	University of Patras, Greece
Divya Mereddy,	Vanderbilt University, USA
Domenico Rotondi,	Grifo multimedia Srl, Italy
Duc-Trung HOANG,	Clermont Auvergne University, France
E. Fantin Irudaya Raj,	Dr. Sivanthi Aditanar College of Engineering, India
Edward Moreno,	UFS, Brazil
Eliane Maria De Bortoli Favero,	UTFPR, Brazil
Elsaid Mamdouh Mahmoud Zahran,	University of Nottingham Ningbo, China
Elzbieta Macioszek,	Silesian University of Technology, Poland
Erdoğan Dogdu,	Angelo State University , USA
Erfan Babae Tirkolae,	Istinye University, Turkey
Eugénia Moreira Bernardino,	Polytechnic of Leiria, Portugal
F M Javed Mehedi Shamrat,	University of Malaya, Malaysia
F. Abbasi,	Islamic Azad University, Iran
F. Abbasi,	University, Amol, Iran
Fatiha Djemili Tolba,	Badji Mokhtar University of Annaba, Algeria
Fatima Mohammed Rafie Younis,	University of mosul, Iraq
Felix J. Garcia Clemente,	University of Murcia (UMU), Spain
Fernanda Otilia de Sousa Figueiredo,	University of Porto, Portugal
Fernanda Otilia Figueiredo,	University of Porto, Portugal
Francesco Zirilli,	(retired) Sapienza Universita Roma , Italy
Fzlollah Abbasi,	Islamic Azad University, Iran
Gajendra Sharma,	Kathmandu University, Nepal
Gandhiya Vendhan,	International Statistical Institute (ISI), Netherland
Ghasem Mirjalily,	Yazd University, Iran
Gheorghe Grigoras,	Technical University of Iasi, Romania
Gheorghe-Daniel,	Andreescu Politehnica University Timisoara, Romania
Gian Piero ZARRI,	Sorbonne University, France
Goi Bok Min,	University Tunku Abdul Rahman, Malaysia

Gordana Jovanovic,	Dolecek Institute INAOE, Mexico
Gozde Ozsezer,	Ege University - Izmir, Turkey
Gregory Cheng,	Swansea University, UK
Grigorios N. Beligiannis,	University of Patras, Greece
Grzegorz Sierpinski,	Silesian University of Technology, Poland
Guangxia Xu,	Guangzhou University, China
Guoyue CHEN,	Akita Prefectural University, JAPAN
Gustavo Callou,	UFRPE, Brazil
Harir Atimad,	Université Hassan I_Settat, Morocco
Hatim Hafiddi,	INPT, Morocco
Hemavathi P,	Bangalore Institute of Technology, India
Hlaing Htake Khaung Tin,	University of Information Technology, Myanmar
Hyunsung Kim,	Kyungil University, Korea
Ibrahim Hamzane,	Hassan II University of Casablanca, Morocco
Imad Saleh,	Université Paris 8, France
Ireneusz Jozwiak,	Wroclaw University of Science and Technology, Poland
Isa Maleki,	Islamic Azad University, Iran
Issa Etier,	Hashemite University, Jordan
Jagadeesh HS,	APS College of Engineering, India
Jameela Al-Jaroodi,	Robert Morris University, USA
Jawad K. Ali,	University of Technology, Iraq
Jeff Sterling Ngami,	Université de Yaoundé, Cameroon
Jesuk Ko,	Universidad Mayor de San Andres (UMSA), Bolivia
Jianbiao Zhang,	Beijing University of Technology, China
Jianfeng Ren,	University of Nottingham Ningbo China, China
Joao Calado,	Polytechnic University of Lisbon, Portugal
Jollanda Shara,	University "Eqrem Cabej", Albania
Joo Calado,	Polytechnic University of Lisbon, Portugal
Jose Silvestre Silva,	Univ. of Coimbra, Portugal
K. Vinoth Kumar,	SSM Institute of Engineering and Technology, India
Kazem Abhary,	University of South Australia, Australia
Kevin Matthe Caramancion,	Mercyhurst University, USA
Kevin Matthe Caramancion,	University of Wisconsin–Stout, USA
Khalidi Amine,	Kasdi merbah university of Ouargla , Algeria
Khaled Ahmed Nagaty,	The British University in Egypt
Koena Ronny Mabokela,	University of Johannesburg, South Africa
Komalpreet Kaur,	Salem State University, India
Konstantinos Karampidis,	Hellenic Mediterranean University, Greece
Krzysztof Ejsmont,	Warsaw University of Technology, Poland
Lailil Muflikhah,	Brawijaya University, Indonesia
Lakmali Karunarathne,	York St John University, United Kingdom
Layth AbdulRasool Mahdi Alasadi,	University of Kufa, Iraq
Lazhar Khriji,	Sultan Qaboos University, Oman
Leila Ben Ayed,	University la Manouba, Tunisia.
Leo Mrsic,	Algebra University College, Croatia
Ljiljana Trajkovic,	Simon Fraser University, Canada
Loc Nguyen,	Independent Scholar, Vietnam
Loc Nguyen,	Loc Nguyen's Academic Network, Vietnam
Luis Carlos Oliveira Gonçalves,	Federal University of Mato Grosso, Brazil
Luisa Maria Arvide Cambra,	University of Almeria, Spain
Lunjin Lu,	Oakland University, USA
M Tahar Kechadi,	University College Dublin (UCD), Ireland

Magdalena Piekutowska,	Pomeranian University in Słupsk, Poland
Mahdi Bodaghi,	Nottingham Trent University, UK
Mahdi Sabri,	Islamic Azad University, Iran
Mahmoud Hassaballah,	Prince Sattam Bin Abdulaziz University, Saudi Arabia
Mahmoud Rokaya,	Taif University, Saudi Arabia
Maki K. Habib,	The American University in Cairo, Egypt
Mallikharjuna Rao K,	IIT Naya Raipur, India
Man Fung LO,	The University of Hong Kong, China
Manoj Kumar,	University of Wollongong, UAE
Mara del Carmen Carrin Espinosa,	Universidad de Castilla la Mancha, Spain
Marco Javier Suarez Baron,	UPTC, Colombia
Maslin Masrom,	Universiti Teknologi, Malaysia
Maxim Solovchuk,	National Taiwan University, Taiwan
Mehdi Gheisari,	Islamic Azad University, Iran
Mesay Gameda,	Instituto Politécnico Nacional(IPN), Mexico
Michail Kalogiannakis,	University of Crete, Greece
Michail Kalogiannakis,	University of Thessaly, Greece
Min-Shiang Hwang,	Asia University, Taiwan
Mirka Mobilia,	Italian Ministry of Public Education, Italy
MK Quweider,	University of Texas Rio Grande Valley, United States
Mohamed Hamlich,	ENSAM, UH2C, MOROCCO
Mohammad Jafarabad,	Qom university, Iran
Mohammed Mechee,	University of Kufa, Iraq
Mohsen Bahmani,	Technical and Vocational University, Iran
Mounim A. El Yacoubi,	Institut Polytechnique de Paris, France
Mu-Song Chen,	Da-Yeh University, Taiwan
Nadia Abd-alsabour,	Cairo University, Egypt
Nathalie Djiguimkoudre,	University Joseph KI-ZERBO, Burkina Faso
Nikola Ivković,	University of Zagreb, Croatia
Nu Nu War,	University of Co-operative and Management, Myanmar
Olarik Surinta,	Maharakham University, Thailand
Oleksandr Laptiev,	Taras Shevchenko National University of Kyiv, Ukraine
Omar Cheikhrouhou,	University of Sfax, Tunisia
Omar Khadir,	Hassan II University of Casablanca, Morocco
Oroke Kenneth Augustine,	Madonna University Nigeria, Nigeria
P.V.Siva Kumar,	VNR VJiet, India
Parimalakrishnan,	Annamalai University, India
Pascal Lorenz,	University of Haute Alsace, France
Pavel Loskot,	ZJU-UIUC Institute, China
Pawankumar Sharma,	University of the Cumberland, USA
Pedro Torres,	Polytechnic Institute of Castelo Branco, Portugal
Petr Hajek,	University of Pardubice, Czech Republic
Phathutshedzo Makovhololo,	Cape Peninsula University of Technology, South Africa
Piotr Kulczycki,	Systems Research Institute, Poland
Ponnuthurai Nagaratnam Suganthan,	Qatar University, Qatar
Prasan Kumar Sahoo,	Chang Gung University, Taiwan
Prasanna Shete,	Somaiya Vidyavihar University, India
Prokopchina Svetlana V,	Financial University, Russia
Qiang Cheng,	University of Kentucky, USA
QiuJun Lan,	Hunan University, China
R S M Lakshmi Patibandla,	KLEF Deemed to be University, India
Radhakrishna Bhat,	Manipal Institute of Technology, India

Rahul Kher,	University of Canberra, Australia
Rajini Kanth,	Department of CSE-AI&ML, India
Rajini KANTH,	SNIST, India
Ramadan elaiess,	University of Benghazi, Libya
Ramyia Thatikonda,	University of the Cumberland, United States
Rao Li,	University of South Carolina Aiken, USA
Richa Purohit,	Sri Balaji University Pune, India
Rima Tri Wahyuningrum,	Universitas Trunojoyo Madura, Indonesia
Rishabh Garg,	Birla Institute of Technology & Science, India
Rita Yi Man Li,	Hong Kong Shue Yan University, Hong Kong
Robert Bestak,	Czech Technical University in Prague, Czech Republic
Robson Albuquerque,	University of Brasilia, Brazil
Rodrigo Pérez Fernández,	Universidad Politécnica de Madrid, Spain
Rohan Singh Rajput,	Headspace, USA
Rohit Khankhoje,	Independent Researcher, USA
Rushit Dave,	Minnesota State University, USA
Ryspek Usubamatov,	Kyrgyz State Technical University, Kyrgyzstan
S Mary Praveena,	Sri Ramakrishna Institute of Technology, India
S. Gandhiya Vendhan,	Bharathiar University, India
S.M.Emdad Hossain,	University of Nizwa, Oman
S.Rajesh,	Mepco Schlenk Engineering College, India
Saad AlJanabi,	AI-Hikma College University, Iraq
Sadique Shaikh,	AIMSR, India
Safae El Abkari,	Mohamed V University in Rabat, Morocco
Saif Aldeen Saad Obayes Alkadhim,	Xian Jiaotong University, China
Samir Kumar Bandyopadhyay,	University of Calcutta, India
Samir Ladaci,	Ecole Nationale Polytechnique, Algeria
Sanda Florentina Mihalache,	Petroleum Gas University of Ploiesti, Romania
Sanjay Jain,	ITM University Gwalior, India
Sarachandran Nair,	Muscat College, Oman
Satish Gajawada,	IIT Roorkee Alumnus, India
Sayantana Dutta,	Indian Institute of Technology Kharagpur, India
Sd Khalifa,	AL - Hikma University, Iraq
Seppo Sirkemaa,	University of Turku, Finland
Seppo Sirkemaa,	University of Turku,Pori Unit
Sergio Orlando Escalona González,	Las Tunas Medical Sciences University, Cuba
Sesha Bhargavi Velagaleti,	GNITS, India
Seyyed Rohollah Mirhoseini,	Islamic Azad University, Iran
Shahid Ali,	AGI Education Limited, New Zealand
Shahram Babaie,	University at Buffalo, USA
Shashikant Patil,	Atlas Skilltech University, India
Shashikant Patil,	ViMEET, India
Shelly Sachdeva,	National Institute of Technology Delhi, India
Shi Dong,	Zhoukou Normal University, China
Shicheng Zu,	Ericsson Inc., China
Shu Ming TAM,	University of Macau, Macau (China)
Siarry Patrick,	Universite Paris-Est Creteil, France
Siddhartha Bhattacharyya,	Rajnagar Mahavidyalaya, India
Siham Benhadou,	Université Hassan 2 de Casablanca, Morocco
Sikandar Ali,	China University of Petroleum, China
Sirikan Chucherd,	Mae Fah Luang University, Thailand

Siti Hajar Halili,	University Malaya, Malaysia
Sivaram Rajeyyagari,	Shaqra University, Saudi Arabia
Smain FEMMAM,	UHA University France, France
Sofia Dembitska,	Vinnitsia National Technical University, Ukraine
Sompong Liangrocapart,	Mahanakorn University of Technology, Thailand
Sonam Mittal,	B K Birla Institute of Engineering & Technology, India
Sos Agaian,	College of Staten Island, Island
Sridhar Iyer,	KLE Technological University, India
Sriman Narayana Iyengar,	Sreenidhi Institute of Science and Technology, India
Subhendu Kumar Pani,	Krupajal Engineering College, India
Subrato Bharati,	Concordia University, Canada
Suhad Faisal Behadili,	University of Baghdad, Iraq
Sunny Joseph Kalayathankal,	Jyothi Engineering College, Cheruthuruthy
Swathi Ganesan,	York St John University, United Kingdom
T. Arudchelvam,	Wayamba University of Sri Lanka, Sri Lanka
Tahar Kechadi,	University College Dublin, Ireland
Taleb zouggar souad,	Oran 2 University, Algeria
Tamer Zakaria Mohamed Emara,	Damietta University, Egypt
Tanvinur Rahman Siam,	BRAC University, Bangladesh
Taouli sidiahmed,	University Aboubekr-Belkaid, Algeria
Tasher Ali Sheikh,	Residential Girls' Polytechnic, India
Teeb Husein Hadi,	Middle Technical University, Iraq
Tiechuan Hu,	Johns Hopkins University, USA
Tinatin Mshvidobadze,	Gori State University, Georgia
Tofael Ahmed,	Comilla University, Bangladesh
Toufik Bounden,	Jijel University, Algeria
Umesh Kumar Singh,	Vikram University, India
Umesh R. Hodeghatta,	Northeastern University, USA
Valentina Emilia Balas,	Aurel Vlaicu University of Arad, Romania
Varun Shukla,	Pranveer Singh Institute of Technology, India
Vijay Kumar Banga,	Chandigarh University, India
Vijaya Kumar Varadarajan,	University of New South Wales, Australia
Vijayakumar Ponnusamy,	SRM IST, India
Vijaykumar S. Bidve,	Symbiosis Skills and Professional University, India
Vincent Forde,	LJMU, Liverpool, UK
Vivek D,	PSG College of Arts & Science, India
Wanyang Dai,	Nanjing University, China
Wei Cai,	Qualcomm, USA
Xianchuan YU,	Beijing Normal University, China
Xianpeng Wang,	Hainan University, China
Xiaochun Cheng,	Swansea University, UK
Xiaofei Shi,	Dalian Maritime University, China
Xiaoqi Ma,	Nottingham Trent University, UK
Xiaoyan Dai,	KYOCERA Corporation, Japan
Xuemei Li,	Oakland University, United States
Yas Alsultanny,	Uruk University, Iraq
Yashwant Singh Bisht,	Uttaranchal Institute of technology, India
Yew Kee Wong Eric,	Hong Kong Chu Hai College, China
Yousef Abu-Baker El-Ebiary,	University Sultan Zainal Abidin (UniSZA), Malaysia
Yousef Farhaoui,	Moulay Ismail University, Morocco
Youssef Taher,	Center of guidance and planning Rabat, Morocco
Yousuf Nasser Al Husaini,	Arab Open University, Oman

Yuan-Kai Wang,
Yu-Chen Hu,
Yulia Kumar,
Yutao Zeng,
Zahra Pezeshki,
Zakariya Mustapha,
Zayar Aung,
Zbigniew Widera,
Zhang Jianhong,
Zhewei Liang,
Zoran Bojkovic,

Fu Jen Catholic University, Taiwan
Tunghai University, Taiwan
Kean University, USA
Tencent Inc., China
Shahrood University of Technology, Iran
Faculty of Law University of Malaya, Malaysia
National Research University, Russia
University of Economics in Katowice, Poland
North China University of Technology, China
Mayo Clinic, USA
University of Belgrade, Serbia

Technically Sponsored by

Computer Science & Information Technology Community (CSITC)



Artificial Intelligence Community (AIC)



Soft Computing Community (SCC)



Digital Signal & Image Processing Community (DSIPC)



Organized By



Academy & Industry Research Collaboration Center (AIRCC)

13th International Conference on Computer Science, Engineering and Applications (ICCSEA 2023)

Developing Elderly Stress-Relief Service using Personalized Videos and Spoken Dialogue Agent	01-10
<i>Hiro Horie, Sinan Chen, Masahide Nakamura and Kiyoshi Yasuda</i>	
Non-Negative Matrix Factorization based Intrusion Detection System for IoT Traffic.....	11-20
<i>Abderezak Touzene, Ahmed Al Farsi and Nasser Al Zeid</i>	
Employing Large Language Models for Dialogue-Based Personalized Needs Extraction in Smart Services.....	21-33
<i>Takuya Nakata, Sinan Chen, Sachio Saiki and Masahide Nakamura</i>	
Review of Digitalization using Artificial Intelligence Maturity Models: The Case of American Automotive SMES.....	35-44
<i>Dharmender Salian</i>	
Gamified Web Application for Facilitating Zero Carbon Activities by Local Government.....	45-57
<i>Aoi Nagatani, Tasuku Watanabe, Yuya Tarutani, Yoshifumi Kamae, Shun Sato, Marin Shoda and Masahide Nakamura</i>	

4th International Conference on NLP & Artificial Intelligence Techniques (NLAI 2023)

Laughing Out Loud – Exploring AI-Generated and Human-Generated Humor.....	59-76
<i>Hayastan Avetisyan, Parisa Safikhani and David Broneske</i>	
An Intelligent Mobile Application to Facilitate Student's Networking using Natural Language Processing Algorithms.....	77-91
<i>Ruijin Deng and Victor Phan</i>	
Revealing Sustainable Growth for Fitbit: A Data-driven Marketing Approach based on K-Means Clustering and Collaborative Filtering.....	93-108
<i>Akansha Akansha and Stuart So</i>	
An Intelligent Approach to Code-Driven Test Execution.....	109-116
<i>Rohit Khankhoje</i>	
A Comprehensive Mobile Application to Assist the Beginner Snowboarder in Discovering Resources, Aid, Equipment, and Community Support.....	117-127
<i>Licheng Xiao and Ang Li</i>	

12th International Conference on Soft Computing, Artificial Intelligence and Applications (SCAI 2023)

Deep Learning based Zero Watermarking for Authentication of Medical Records.....	129-141
<i>Gurleen Kaur, Bakul Gupta and Ashima Anand</i>	
Quality Challenges and Imperatives in Smart AI Software.....	143-154
<i>Rohit Khankhoje</i>	
DNA Sequence Automatic Classification—Learn the Life Language using Artificial Intelligence.....	155-169
<i>Josephine (Hsin) Liu, Phoebe (Yun) Liu, Joseph (Yu) Liu, Emily X. Ding, Robert J. Hou</i>	
3D Convolution for Proactive Défense Against Localized Adversary Attacks.....	171-191
<i>Henok Ghebrechristos and Gita Alaghband</i>	
Unsupervised Multi-Scale Image Enhancement using Generative Deep Learning Approach.....	193-205
<i>Preeti Sharma, Manoj Kumar and Hitesh Kumar Sharma</i>	
Using Augmented Reality Interfaces for Artificial Intelligence Systems.....	207-222
<i>Büşra Öztürk and Yakup Genç</i>	
Inhance Deep Customizations in a Multi-tenant SaaS Application using the BPMN.....	307-312
<i>Amira Ksiksi</i>	

14th International Conference on Communications Security & Information Assurance (CSIA 2023)

Intrusion Detection in a Stand-Alone 5G Network using Machine Learning Evaluation.....	223-236
<i>Hafiz Bilal Ahmad1 Haichang Gao1 and Fawwad Hassan Jaskani</i>	

4th International Conference on IoT, Blockchain & Cloud Computing (IBCOM 2023)

Exploring DAG-Based Architecture as an Alternative to Blockchain for Critical IoT Use Cases.....	237-244
<i>Ledesma O., Sánchez M.A. and Lamo P</i>	

**4th International Conference on Software Engineering and Managing
Information Technology (SEMIT 2023)**

A Multifaceted Swim Training App on Enhancing Skills and Performance...245-262
Xiuhan(Daniel) Fu

**12th International Conference of Networks and
Communications (NECO 2023)**

**Communication Signals Modulations Classification based on Neural Network
Algorithms.....263-276**
Yahya Benremdane, Said Jamal, Oumaima Taheri, Jawad Lakziz and Said Ouaskit

**12th International Conference on Signal, Image Processing and Pattern
Recognition (SPPR 2023)**

**A Secured Image Communication With Dual Encryption and Reversible
Watermarking.....277-285**
Surya Boppanaa, William Kane and Long Ma

**4th International Conference on Machine Learning Techniques and
Data Science (MLDS 2023)**

**G-KMM: A Flexible Kernel Mean Matching Optimization Method for Density
Ratio Estimation Involving Multiple Train & Test Datasets.....287-306**
Cristian Daniel Alecsa

DEVELOPING ELDERLY STRESS-RELIEF SERVICE USING PERSONALIZED VIDEOS AND SPOKEN DIALOGUE AGENT

Hiro Horie¹, Sinan Chen¹, Masahide Nakamura^{1,2}
and Kiyoshi Yasuda³

¹Kobe University, 1-1 Rokkodai-cho, Nada, Kobe, 657-8501, Japan

²RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, 103-0027, Japan

³Osaka Institute of Technology 5-16-1 Omiya, Asahi-ku, Osaka, 535-8585 Japan

ABSTRACT

Our research group is conducting research of a system to support the lives of the elderly at home. We developed "Rakuraku Video Service" (Rakuraku means easy.), a service that obtains information on the interests and preferences of elderly people and recommends YouTube videos based on this information. The purpose of this service is to help the elderly relieve stress and relax by watching videos. However, it has not been tested yet whether this service has such an effect on the elderly. We conduct an experiment to evaluate the service. In conducting the experiment, we will collaborate with a previous study, "PC-Mei" which aims to watch over the elderly and support their daily lives with a virtual agent. The experiment was conducted to obtain evaluations of watched videos and a questionnaire. From the results, it is clear that the service is useful in relieving stress among the elderly.

KEYWORDS

Elderly at home, watching videos, stress relief, individual adaptive type, spoken dialogue agent.

1. INTRODUCTION

In recent years, Japan's population has been aging, and the country is facing a super-aged society. As a result, the number of people requiring nursing care has been increasing, but a shortage of nursing care workers has become an issue. In response to this problem, the government is working to establish a comprehensive community care system that allows elderly people to continue to live in their own neighborhoods even after they require nursing care. One of these efforts is the establishment of a system of home nursing care and home medical care. As a result, the number of elderly people living at home is expected to increase in the future.

Our research group has been developing a system to watch over elderly people at home [1]

[2] [3] [4]. In our previous research, we have been working on the development of the "Rakuraku Video Service" (Rakuraku means easy.) [5] [6] [7], which aims to relieve the stress of the elderly. Rakuraku Video Service is a service that recommends YouTube videos suited to individuals based on their answers to a questionnaire about their interests and preferences. The screen is designed so

that even the elderly, who are not good at operating digital devices, can easily operate the service. The aim of this service is to provide users with stress relief and relaxation by watching videos that suit their personal preferences. Preliminary experiments with faculty members and students belonging to our research group have confirmed that this service can recommend videos suited to individual users, but the effect on the elderly has not yet been verified.

In this paper, we conduct a demonstration experiment of Rakuraku Video Service for the elderly and evaluate the service. The following research questions are set for the verification experiment.

- RQ1 : Can the service recommend videos that match the interests and tastes of the elderly?
- RQ2 : Is the service easy for the elderly to operate?
- RQ3 : Can the elderly relieve stress by watching videos recommended by the service?

The experiment targets three elderly people in their 70s to 80s, and the duration is two weeks. In order to conduct the experiment, we collaborated with PC-Mei [8] [9], a system that uses a spoken dialogue agent to support the monitoring of daily life. Subjects are asked to invoke and use Rakuraku Video Service by voice operation of PC-Mei. The service was evaluated by evaluating the recommended videos and by a questionnaire after the experiment. The results showed that the system can recommend videos that match the interests and preferences of the elderly, that the operation is easy for the elderly, and that viewing the recommended videos has a stress-relieving effect. On the other hand, it was also found that there is room for improvement in the video recommendation algorithm and operation method. Future tasks are to improve the system and conduct experiments with a larger number of subjects.

2. PRELIMINARIES

2.1. The Elderly and Stress

The world's population aging rate is currently rising; in 2020, it was 9.3%, and it is expected to rise to 17.8% by 2060. By country, Japan currently has the highest rate of aging. 28.9% of the population will be elderly in 2021, and the country is entering a hyper-aged society.

While the number of people requiring nursing care is rising with the aging of society, there is a shortage of nursing care workers. In response, the government is working to build a comprehensive community care system that allows elderly people to continue to live their own lives in their own neighborhoods until the end of their lives, even if they require nursing care. One of these efforts is the establishment of a system of home nursing care and home medical care. Therefore, the number of homebound elderly people is expected to increase in the future. Long time spent at home may cause stress due to loneliness and worries about life and health. While living at home, homebound elderly people need to engage in self-care to relieve their own stress without placing a burden on their families and caregivers.

2.2. Preceding Research: Rakuraku Video Service

Our research group has been conducting research and development of a system to watch over and support elderly people at home. We are developing a "Rakuraku Video Service" (Figure 1, Figure 2) [5] [6] [7] to relieve the stress of elderly people at home. This service recommends personalized YouTube videos based on the user's interests and preferences obtained through questionnaires, and aims to relieve stress and relax the user by having him or her watch them. The screen is designed to be easy to operate, even for those who are not good at operating digital

devices. The service has the following three functions.

Video playback function This section describes a function that automatically plays videos on YouTube sequentially when a list of videos is given. When a list of videos and a user are registered with the service, a URL for playing the videos is generated for each user. Users can access the URL to start viewing the videos. The video playback screen is shown in Figure 1.

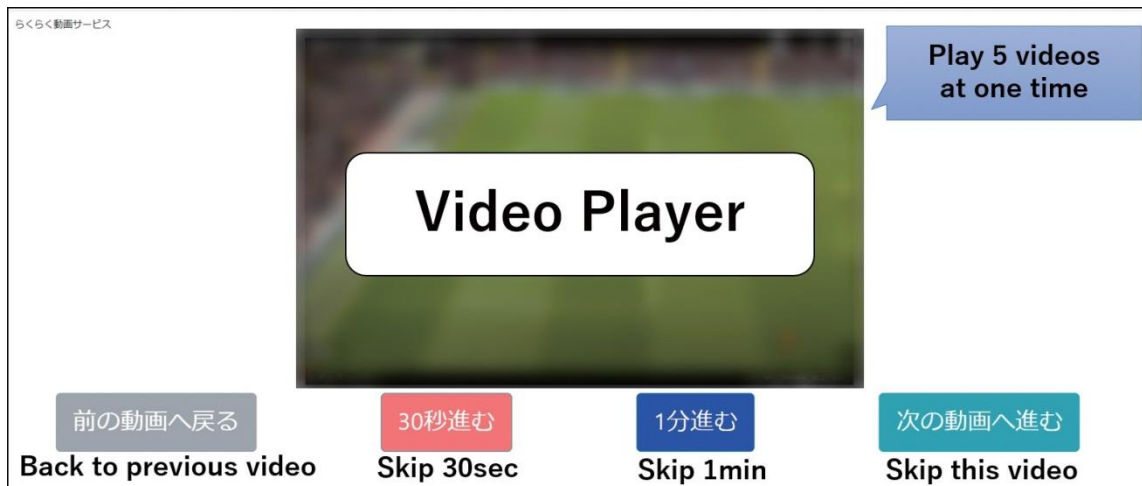


Fig. 1. Rakuraku Video Service (video playback screen)

Video list creation function This section describes the functionality used to create a list of videos tailored to the user's tastes and preferences. First, we obtain information on users' tastes and preferences using paper media and Google Forms. The questionnaire includes "favorite places and places to visit", "favorite music (singers and songs)", "favorite sports and athletes", "favorite celebrities", "hobbies", "favorite TV and radio", and "old favorites and other interests". Each answer is entered into the service as a search word for YouTube videos, and the search program automatically searches for videos that match the individual's preferences. The searched videos satisfy the following conditions.

- Video must be within the top 5 videos related to the search term in order of number of views and relevance.
- Video must have at least 10,000 views on YouTube
- Video length must be less than 20 minutes.

The limitation by the number of views was established because preliminary experiments conducted with faculty members and students in the research group revealed that videos with fewer views tended to be rated lower by users. A limit on the length of videos is set because we believe that elderly people will not be able to concentrate for long periods of time. Searched videos are automatically registered with the service. The service also selects five registered videos and registers them as a list. The videos selected are one from each genre (singers, sports, hobbies, etc.). When a user finishes watching one list, a new list of videos is automatically registered.

Effectiveness measurement function This section describes the function for measuring the impact of video viewing on the elderly. The purpose of this service is to relieve stress, but since it is difficult to actually measure the degree of stress relief, this service obtains subjective evaluations of videos. Immediately after viewing a video, users are asked to rate the video in 5 levels: "Interesting", "Slightly Interesting", "Neutral", "Not Very Interesting", and "Not Interesting" by pressing a rating button. The video evaluation screen is shown in Figure 2. In addition, facial expression recognition during video viewing is used to provide an objective evaluation of the video. Facial expression

recognition captures changes in facial expression by capturing the width of the mouth and the movement of the eyes. However, in this experiment, facial expression recognition is not performed for the sake of subject privacy.

らくらく動画サービス

How did you like this video?
Please evaluate the video using the buttons
at the bottom, then press "Next Video".

この動画はいかがでしたか？

下のボタンで評価をしてから、
「次の動画へ」を押してください

Can't watch Not Interesting Not Very Interesting Neutral Slightly Interesting Interesting

視聴出来ない 面白くない あまり面白くない どちらでもない 少し面白い 面白い

Rate the video on a 5-point scale

Fig. 2. Rakuraku Video Service (video evaluation screen)



Fig. 3. PC-Mei Copyright 2009-2018 Nagoya Institute of Technology (MMDAgent Model "Mei")

2.3. Preceding Research: Pc-Mei

.Our research group has been developing PC-Mei (Figure 3) [8] [9], a system that uses a spoken dialogue agent to support the monitoring of the elderly in their daily lives. In the previous study, we developed a service voice execution function that enables PC-Mei to execute various services on the Web as an extension of this system. The system executes services when an elderly person utters specific keywords. Examples include a search service that performs Web searches using keywords and a weather forecast service that tells the user the weather forecast. This function allows the elderly to easily use Web services by voice without complicated device operations

2.4. Linking PC-Mei with Rakuraku Video Service

To link the previous studies, we register Rakuraku Video Service to PC-Mei's service voice execution function. This allows users to easily access the service startup screen by saying "Rakuraku Video Service" to the PC-Mei agent. The figure 4 shows the overall architecture after the linking.

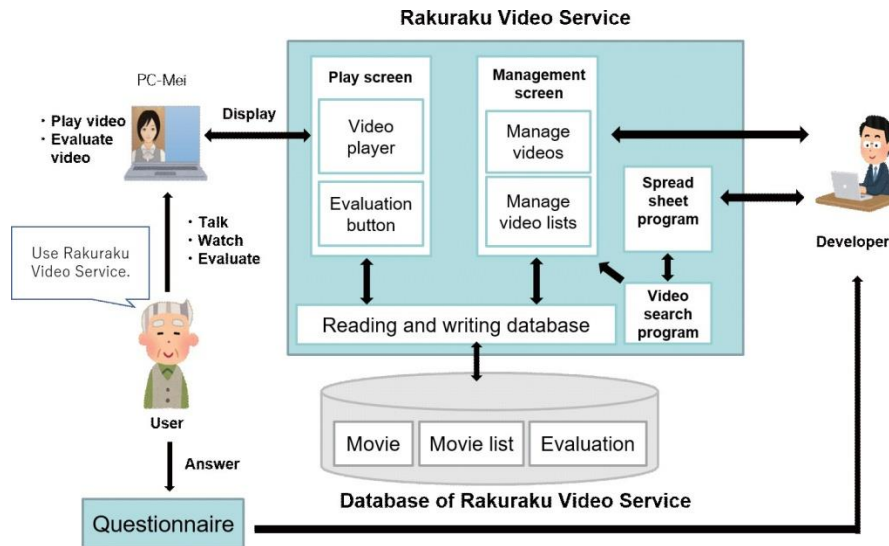


Fig. 4. Overall Architecture

Rakuraku Video Service manages and plays videos, and PC-Mei calls and displays the video playback screen.

2.5. Issues to Focus on

We have conducted preliminary experiments on Rakuraku Video Service with faculty members and students belonging to our research group, and verified that the service can recommend videos that match individual tastes and preferences. However, it has not yet been verified whether the recommendation of preferred videos to the elderly can relieve stress and relax them. In this paper, we propose to investigate the impact of this service on the elderly, and conduct a demonstration experiment on the elderly as a research method.

3. EXPERIMENT

3.1. Purpose of Experiment

The purpose of this experiment is to evaluate Rakuraku Video Service through an experiment targeting elderly people. The following research questions were set up for the evaluation.

- RQ1: Can the service recommend videos that match the interests and tastes of the elderly?
- RQ2 : Is the service easy for the elderly to operate?
- RQ3 : Can the elderly relieve stress by watching videos recommended by the service?

3.2. Method of Experiment

This experiment will be conducted with three elderly subjects in their 70s and 80s, and a PC-Mei system linked to Rakuraku Video Service will be installed in each of their homes. The subject

Table 1. Subject Information

Subject	Gender	Age	Household
Subject A	woman	80's	solitarily living alone
Subject B	woman	80's	solitarily living alone
Subject C	man	70's	elderly only

used the service by saying "Rakuraku Video Service" to PC-Mei at any time during a two-week period.

The experimental procedure is as follows.

- Step 1 :** The subject's basic information (name, etc.) is registered with Rakuraku Video Service and PC-Mei.
- Step 2 :** The PC-Mei (laptop computer) is installed in the subject's home.
- Step 3 :** A questionnaire regarding the subjects' hobbies and preferences is administered
- Step 4 :** Enter the answers to the questionnaire into Rakuraku Video Service and run the video search program
- Step 5 :** The subject says "Rakuraku Video Service" to the PC-Mei agent at an arbitrary time
- Step 6 :** Subjects watch five videos at a time and rate each video on a five-point scale.
- Step 7 :** The service updates the video list when the viewing of a video is finished.
- Step 8 :** Perform Step 5 to Step 7 for 2 weeks.
- Step 9 :** The system is collected and a post-experiment questionnaire is administered.

The post-experiment questionnaire asks the following questions. Except for Question 6, the answers were selected from 1 to 4 (1: not applicable, 2: not very applicable, 3: a little applicable, and 4: applicable). The respondents are also asked to provide reasons for each answer, if any

Questionnaire after the experiment

- Q1: I had a good time using this service.
- Q2 : I felt nostalgic using this service.
- Q3 : I felt the effects of stress reduction and relaxation by using this service.
- Q4 : This service was easy to operate.
- Q5 : I would like to use it again if I have a chance.
- Q6 : Please write down any improvements or comments.

4. RESULTS

The table 2 shows the number of times each subject used Rakuraku Video Service and the number of videos viewed. 3 subjects used the service a total of 21 times and viewed 47 different videos.

Although 5 videos were played back per use, only 47 different videos were viewed as a result because the videos were selected at random from among the videos searched for, resulting in the same video being viewed multiple times.

Table 2. Experimental results: Number of times used

Subject	Number of times used	Number of video types viewed
Subject A	1	5
Subject B	14	25
Subject C	6	17
Total	21	47

The table 3 shows the results of the subjects' evaluations of the videos (the number of each evaluation). When the same video was viewed and evaluated multiple times, the results are shown for the first viewing. 22 videos were interesting, 8 were lightly interesting, 10 were neutral, 3 were not very interesting, and 4 were not interesting for all three subjects.

Table 3. Experimental results: Evaluation for videos(unit: number of videos)

Subject	Interesting	Slightly Interesting	Neutral	Not Very Interesting	Not Interesting
Subject A	4	0	0	1	0
Subject B	10	6	4	1	4
Subject C	8	2	6	1	0
Total	22	8	10	3	4

Table 4. Post-experiment questionnaire results

(Refer to Step 9 of the Method of Experiment for the selection scale of 1 to 4.)

Question	Subject A	Subject B	Subject C
Q1	3	4	4
Reason	It was unfortunate that the publicity ran.	I was glad to see my favorite singer on the program.	It was nice to hear some of the old songs.
Q2	No answer	1	4
Reason	No answer	No answer	The old songs made me feel nostalgic.
Q3	2	4	4
Reason	No answer	I used it when I had time.	It was nice to see old songs and golf videos.
Q4	3	2	4
Reason	No answer	I was anxious to push in the wrong place. The buttons were easy to understand.	I was a little confused when I had to go back, but overall it was good.
Q5	No answer	4	4
Reason	No answer	I would like to use it by all means.	I would like to use it if you call on me.
Q6	It would have been nice if they played karaoke or enka.	Loved to hear the songs.	I would like to see a longer video (about an hour).

5. DISCUSSION

5.1. Discussion of RQ1

In order to examine whether or not the service was able to recommend videos that matched the tastes and preferences of the elderly, we assigned a score to the evaluation given to each video. We assign a score of -2 points to "Not Interesting", -1 point to "Not Very Interesting", 0 points to "Neutral", 1 point to "Slightly Interesting", and 2 points to "Interesting". From the table 3, the

average score of the 47 videos viewed and evaluated was calculated to be 0.87, which was closer to "Interesting" than to "Neutral". Thus, for RQ1, we confirmed that Rakuraku Video Service is capable of recommending videos that match the interests and tastes of the elderly.

5.2. Discussion of RQ2

The results of Q4 of the post-experiment questionnaire are discussed in terms of whether the operation was easy for the elderly. One subject answered "not very applicable" to the question of whether the operation was easy. The subject stated that the buttons were easy to understand, although he felt uneasy when he pressed different buttons (in unrelated places). The remaining two subjects answered "a little applicable" and "applicable," confirming that the current screen design and operation methods are easy for the elderly, although there is room for improvement.

5.3. Discussion of RQ3

The results of Q1, Q2, and Q3 of the post-experiment questionnaire will be used to determine whether viewing the recommended videos relieves stress in the elderly. All subjects answered that Q1 (I had a good time using this service.) was applicable or a little applicable. Regarding Q3 (I felt the effects of stress reduction and relaxation by using this service.), two out of three subjects answered applicable. The above results confirm that using RakuRaku Video Service makes the users feel happy, stress-relieving, and relaxing. However, one subject said that they were disappointed with the advertisements, indicating that the video search program needs to be improved. Q2 (I felt nostalgic using this service.) was set as a question in the post-experiment questionnaire because we thought that if there were things from the past in the answers to the questionnaire about interests and preferences, videos that made the participants feel nostalgic would be recommended. One of the methods of dementia care is reminiscence [10], which aims for psychological stability by reminding people of the past. We expected a relaxing effect of nostalgia based on this method, and it can be said that it had that effect only on Subject C. We believe that this effect can be improved in the future by asking about past events in more detail during the questionnaire.

5.4. Discussion of Video Evaluation

We investigated what genres of videos were likely to receive good evaluations. Here, as shown in the table 5, we consider the survey items related to tastes and preferences as seven genres.

Table 5. Dividing into genres

Questionnaire item	Genre
favorite places and places to visit	Place
favorite music (singers and songs)	Music
favorite sports and athletes	Sport
favorite celebrities	Celebrity
hobbies	Hobby
favorite TV and radio	Program
old favorites and other interests	Interest

The table 6 shows the results (rounded to two decimal places) of calculating the percentage of ratings for each genre. (e.g. The genre "Place" was watched three times in total, and two were rated as interesting. Therefore, the ratio is $2 \frac{3}{3} \doteq 0.67$)

The table 6 shows that music, celebrity, and program were highly rated as interesting or slightly interesting. This is thought to be due to the fact that the answers to the questionnaires for the interests and tastes are concrete names and proper nouns, and therefore, videos that match individual tastes

are easily recommended. On the other hand, a high percentage of respondents rated sport videos as neutral, not very interesting, or not interesting. This is thought to be due to the fact that even when searching for videos on a single sport, there are so many different types of videos that it is unlikely that a video suited to the individual will be recommended. The same can be said for hobby, which have a large variation in evaluation. For other genres, the number of videos viewed was small, so it is necessary to conduct experiments with a larger number of subjects.

Table 6. Percentage of evaluation by genre

Genre	Interesting	Slightly	Neutral	Not Very	Not	Total
	Interesting		Interesting (unit: number of videos)			
Place	0.67	0	0.33	0	0	3
Music	0.75	0.19	0.06	0	0	16
Sport	0	0.13	0.38	0.25	0.25	8
Celebrity	0.67	0.33	0	0	0	3
Hobby	0.38	0.13	0.13	0.13	0.25	8
Program	0.60	0.40	0	0	0	5
Interest	0	0	1.00	0	0	4

Future issues

As mentioned in the discussion above, future issues are to improve the screen design so that even those who are not familiar with the operation of digital devices can clearly understand how to operate them, and to conduct the experiment with a larger number of subjects. In addition, since there were cases in which the same video was viewed multiple times in this experiment, it is necessary to modify the video recommendation algorithm in order to recommend different videos.

6. CONCLUSION

In this paper, we conducted a demonstration experiment with elderly people to evaluate Rakuraku Video Service that helps elderly people relieve stress. From the experimental results, we confirmed that the service can recommend videos that match the interests and tastes of the elderly, that it is easy to operate, and that it can help them relieve stress. We also confirmed that the evaluation of the service is biased depending on the genre of the videos.

The issues found in the experiment are to design a screen that is easier to understand the operation method, to improve the video recommendation algorithm, and to conduct an experiment with a larger number of subjects. The service will be improved based on the results of this experiment. This will enable elderly people at home to easily watch videos that match their interests and tastes, and is expected to help them relieve stress.

ACKNOWLEDGMENT

This research was partially supported by JSPS KAKENHI Grant Numbers JP19H01138, JP20H05706, JP20H04014, JP20K11059, JP22H03699, JP19K02973, and Young Scientists (No.23K17006).

REFERENCES

- [1] M. Chisaki, S. Chen, S. Saiki, M. Nakamura, and K. Yasuda, "Assisting personalized healthcare of elderly people: Developing a rule-based virtual caregiver system using mobile chatbot," *Sensors*, vol. 22, no. 10: 3829, May 2022, <https://doi.org/10.3390/s22103829>.
- [2] S. Chen and M. Nakamura, "Designing an elderly virtual caregiver using dialogue agents and

- webrtc,” in 2021 4th International Conference on Signal Processing and Information Security (ICSPIS). IEEE, 2021, pp. 53–56.
- [3] S. Chen, S. Saiki, and M. Nakamura, “Characterizing quality of in-home physical activities using bone-based human sensing,” *IEICE Technical Report; IEICE Tech. Rep.*, vol. 120, no. 49, pp. 1–6, 2020.
- [4] K. Unigame, D. Takatsuki, S. Saiki, M. Nakamura, and K. Yasuda, “Compass4SL: a service for sharing problems and solutions for the elderly at home,” in 22nd IEEE/ACIS International Fall Virtual Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD2021), November 2021, pp. 62–67, taichung, Taiwan (Online).
- [5] H. Horie, S. Chen, M. Nakamura, and K. Yasuda, “Study of stress relief service by watching personalized videos for elderly people at home,” in 2022 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), July 2022, pp. 130–136, online.
- [6] K. Hirayama, S. Chen, S. Saiki, and M. Nakamura, “Toward capturing scientific evidence in elderly care: Efficient extraction of changing facial feature points,” *Sensors*, vol. 21, no. 20, p. 6726, 2021.
- [7] K. Hirayama, S. Saiki, and M. Nakamura, “Evaluating video playing application for elderly people at home by facial expression sensing service,” in The 22nd International Conference on Information Integration and Web-based Applications & Services (iiWAS2020), November 2020, pp. 21–27, online.
- [8] H. Ozono, S. Chen, and M. Nakamura, “Encouraging elderly self-care by integrating speech dialogue agent and wearable device,” in 8th International Conference, ITAP 2022, Held as Part of the 24th HCI International Conference, HCII 2022, vol. LNCS 13331, May 2022, pp. 52–70.
- [9] H. Ozono, S. Chen, and M. Nakamura, “Study of microservice execution framework using spoken dialogue agents,” in 22nd IEEE-ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel Distributed Computing (SNPD2021), November 2021, pp. 273–278, taichung, Taiwan (Online).
- [10] J. M. Thomas and D. Sezgin, “Effectiveness of reminiscence therapy in reducing agitation and depression and improving quality of life and cognition in long-term care residents with dementia: A systematic review and meta- analysis,” *Geriatric Nursing*, vol. 42, no. 6, pp. 1497–1506, 2021.

AUTHORS

Hiro Horie received the B.E. degree in Information and Intelligence Engineering from Kobe University, Japan, in 2022. He is currently enrolled in a master’s course at the Graduate School of Systems Informatics, Kobe University, Japan. His research interests include the Web service and gerontechnology.

Sinan Chen received a Master of system informatics and a Ph.D. in computational science from Kobe University, Japan, in 2019 and 2021, respectively. He worked at Kobe University as a research fellow (PD) from the Japan Society for the Promotion of Science (JSPS) in the first half of 2022. He is currently an assistant professor in the Center of Mathematical and Data Science Center at Kobe University, Japan. His research interests include the smart home, cloud computing, machine learning, and software engineering. He is a member of the IEEE, ACM, Sigma Xi, CAAI, CCF, IEICE, and IEEJ.

Masahide Nakamura received the B.E., M.E., and Ph.D. degrees in Information and Computer Sciences from Osaka University, Japan, in 1994, 1996, 1999, respectively. From 1999 to 2000, he has been a post-doctoral fellow in SITE at University of Ottawa, Canada. He joined Cybermedia Center at Osaka University from 2000 to 2002. From 2002 to 2007, he worked for the Graduate School of Information Science at Nara Institute of Science and Technology, Japan. From 2007 to 2022, he worked for the Graduate School of System Informatics at Kobe University. He is currently a full professor in the Center of Mathematical and Data Science Center at Kobe University. In 2015, he worked for Universite Grenoble Alpes as a visiting professor. In 2018, he joined RIKEN Center for Advanced Intelligence Project as a visiting researcher. His research interests include the service/cloud computing, smart home, smart city, and gerontechnology. He is a member of the IEEE, ACM, IEICE and IPSJ.

NON-NEGATIVE MATRIX FACTORIZATION BASED INTRUSION DETECTION SYSTEM FOR IOT TRAFFIC

Abderezak Touzene, Ahmed Al Farsi and Nasser Al Zeidi

Department of Computer Science, College of Science, Sultan Qaboos
University, Oman

ABSTRACT

With the emergence of smart devices and the Internet of Things (IoT), millions of users connected to the network produce massive network traffic datasets. These vast datasets of network traffic (Big Data) are challenging to store, deal with and analyse to detect normal or cyber-attack traffic. In this paper we developed an Intrusion Detection System (NMF-IDS) based on Non-Negative Matrix Factorization dimension reduction technique to handle the large traffic datasets and efficiently analyses them in order to detect with a good precision the normal and attack traffic. The experiments we conducted on the proposed IDS-NMF give better results than the traditional ML-based intrusion detection systems, we have got an excellent detection accuracy of 98%

KEYWORDS

Intrusion Detection Systems, Machine Learning, Dimensionality Reduction, IoT traffic.

1. INTRODUCTION

Based on reports published by cybersecurity institutions in several countries worldwide, network cyber-attacks have increased exponentially in recent decades. These days, as we witness the era of the Fourth Industrial Revolution, 4IR and its emerging technologies like the Internet of Things, Quantum Computing, and Artificial Intelligence. Millions of users have become connected to the Internet, and hundreds of millions of devices connected to the network produce millions of large network traffic records datasets. Those massive datasets from network traffic contain millions of cyber-attacks, so it is crucial to analyse this data quickly in real-time to detect attacks.

This paper focus on Nonnegative Matrix Factorization as a dimension reduced technique to efficiently analyses large IoT network traffic. Nonnegative Matrix Factorization (NMF) is an approximation numerical method aiming at decomposing data matrix A into its simpler factors lower-rank matrices H and W . NMF is an unsupervised Machine Learning technique widely used in data mining, dimension reduction, clustering, factor analysis, text mining, computer vision, and bioinformatics, image recognition and recommendation systems to name a few. In contrast to Singular Value Decomposition (SVD) and Principal Component Analysis (PCA), Nonnegative Matrix Factorization (NMF) requires that A , W , and H be nonnegative. For many real-world data, non-negativity is inherent, and the interpretation of factors has a natural interpretation which could be one of the advantages of NMF compared to PCA and SVD.

Formally, Nonnegative Matrix Factorization problem is to find two low-rank factors matrix $W_{m \times k}$ and $H_{k \times n}$ for a given nonnegative matrix $A_{m \times n}$, such that $A \approx WH$. Most of the available optimization techniques include Hierarchical Alternative Least squares HALS, Multiplicative Updates (MU), Stochastic Gradient Descent, and Block Principal Pivoting (ALNS-BPP), which are based on alternating optimizing W and H while keeping one of them fixed.

1.1. Intrusion Detection System Background

This section discusses the background of Intrusion detection systems, including their definition and diverse types. Moreover, it discusses several papers on machine learning IDS algorithms.

1.1.1. Intrusion Detection System (IDS)

An intrusion Detection System is defined as a hardware device or software that observes systems for malicious network traffic or policy violations. The purpose of IDS is to detect various types of malicious network traffic or malicious computer use that a firewall cannot recognize. This is critical to achieving high protection against actions threatening computer systems' availability, integrity, or confidentiality [1].

1.1.2. Types of Introduction Detection Systems (IDS)

There are many classifications for intrusion detection systems. However, this classification has been used extensively in previous studies based on the data collection method:

1. Network intrusion detection system, which observes and analyses data traffic to detect if there is an attack or malicious behaviour (NIDS).
2. The Host-based intrusion detection system monitors and analyses data from log files (HIDS), and based on the detection technique, it can be categorized into three main categories: Specification-based IDS, Anomaly-based IDS, and signature-based IDS.

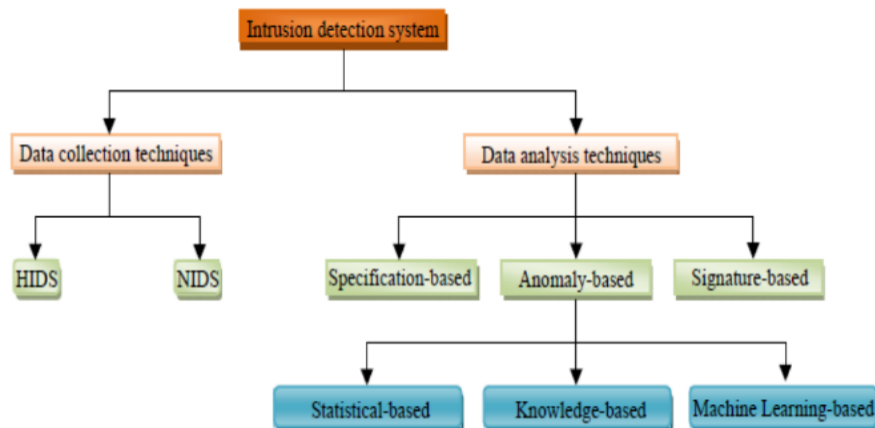


Figure 1: IDS Classification

- **Signature-based (Misuse) Intrusion Detection Systems:** Signature-based intrusion detection analyzes network traffic, searching for occasions or combinations similar to a predefined event pattern that describes a known attack.
Advantages: Signature-based Intrusion Detection Systems effectively detect intrusions without too many false alarms.

Disadvantages: Signature-based-IDS can only identify attacks it knows in advance, requiring constant updating signatures. Signature-based-IDS detectors sometimes are trained very well in detecting specific types of attacks that prevent them from seeing some new kinds of attacks.

- **Anomaly-based Intrusion Detection Systems:** Anomaly-based intrusion detection systems identify abnormal behavior on a network. Commonly attacks differ from regular legitimate network traffic; the detector can detect them by detecting these changes and differences. Anomaly detectors are trained well on normal network traffic from historical data collected. So, they can see abnormal behaviors easily.

Advantages: Anomaly-based Intrusion Detection systems can detect abnormal behavior, so they can detect an attack without any knowledge about it

Disadvantages: Because of the variations in users and network behaviors, Anomaly-based IDS may fire many false alarms. Anomaly detection approaches must be trained in huge datasets of normal behavior activities.

1.2. Motivation

In this paper we aim at overcoming some of the limitations existing in traditional machine learning-based Intrusion Detection Systems (IDS), such as a considerable amount of training and testing on Big data datasets and to detect type of attack in real-time.

In this study we analyse the performance of ML-based IDS using KDD and CIC datasets by applying NMF, which will help reduce the dataset's dimensions into lower-rank matrices that can be used for analysing and testing any new network traffic in real time.

The rest of the paper is structured of the remaining as follows: In section 2, will start with a brief background on Machine Learning based IDS and NMF and introduce the related work on which the proposed solution will be built. In section 3 will discuss the design of the proposed NMF based IDS including the, and detection phase. Section 4 is dedicated to the experimental work, it describes the implementation environment and the datasets used, and the performance evaluation of our IDS. Finally, section 5 will summarize the paper and highlights some limitations along with the future works.

2. BACKGROUND AND RELATED WORK

This section will discuss the background of Machine Learning IDS and background of NMF.

2.1. Machine learning-based IDS

Many recent Anomaly Intrusion Detection Systems (AIDS) is based on Machine Learning methods. There are a lot of ML algorithms and methods used for ML-based IDS, such as neural networks, nearest neighbour, decision trees, and clustering methods, applied to discover the meaningful features from IDS datasets [1] [2].

2.1.1. Supervised learning in Intrusion Detection System

Supervised ML-based IDS techniques that can identify attacks based on labeled training datasets. A supervised ML technique can be divided into training and testing. Training phase, important features are specified and processed in datasets, then we train the model from these datasets. There are many applications of supervised machine learning-based IDS. Li et al. used an SVM classifier with an RBF kernel to classify the KDD 1999 dataset into predefined classes. From a

total of 41 attributes, a subset of features was carefully chosen by using the feature selection approach [3]. K-Nearest Neighbours (KNN) classifier: The k-Nearest Neighbour (k-NN) method is a typical non-parametric classifier used in machine learning [4]. These methods aim to name an unlabelled data sample to the class of its k nearest neighbours. Point X denotes a sample of unlabelled data that needs to be classified.

2.1.2. Unsupervised ML-based Intrusion Detection System

Unsupervised ML can be defined as an ML technique that obtains information of interest from input data sets without class labels. The input data points are usually treated as a set of random variables. A standard density model is then generated for the data set. In supervised learning, output labels are presented and used to train the machine to obtain the desired results for an unseen data point. By contrast, in unsupervised learning, no label is provided. Instead, the data is automatically grouped into different categories through the learning process [5].

2.2. Nonnegative Matrix Factorization (NMF)

Non-negative matrix factorization is an algorithm that takes a nonnegative input matrix $A \in \mathbb{R}^{m \times n}$ and decomposes it into lower rank matrices W and H based in low rank parameter K . NMF is an unsupervised Machine Learning technique commonly used in clustering, dimensionality reduction, factor analysis, data/text mining, computer vision, bioinformatics, image recognition and recommendation systems to name a few. In contrast to Principal Component Analysis (PCA) and Singular Value Decomposition (SVD), NMF requires that A , W , and H be nonnegative. For many real-world data, non-negativity is inherent, and the interpretation of factors has a natural interpretation which could be one of the advantages of NMF compared to PCA and SVD [10] [11].

2.2.1. Foundations of Nonnegative Matrix Factorization

NMF takes a nonnegative input matrix $A \in \mathbb{R}^{m \times n}$ m is number of rows which represent number of features and n is number of column which represent number of samples, and low rank parameter K which is positive integer $< \{m, n\}$, NMF algorithms aims to find two low rank matrices $W \in \mathbb{R}^{m \times k}$ and $H \in \mathbb{R}^{k \times n}$ such that $A \approx WH$.

NMF aims to minimize the following cost function:

$$\min_{W, H} \|A - WH\|_F^2 \quad \text{such that} \quad W \geq 0 \text{ and } H \geq 0 \quad (1)$$

$$W \leftarrow \|A - \bar{W}H\|_F^2 \quad (2)$$

$$H \leftarrow \|A - W\bar{H}\|_F^2 \quad (3)$$

Most of the available optimization techniques include Hierarchical Alternative Least squares HALS, Multiplicative Updates (MU), Stochastic Gradient Descent, and Block Principal Pivoting (ALNS-BPP), which are based on alternating optimizing W and H while keeping one of them fixed.

2.3. Related Work

X. Guan in [6] presented an efficient and fast anomalous intrusion detection model that includes many data from different sources. A new method based on non-negative matrix factorization (NMF) is discussed to characterize program and user behaviors in a computer system. A large amount of high-dimensional data was collected in their experiments. NMF was used and reduced

the vectors to a smaller vector length after that, any simple classifier can be implemented in low-dimension data instead of the entire dataset. After getting low dimension features the model can differentiate between normal traffic and abnormal traffic easily by using a threshold, so any user behavior on that threshold will be considered an attack.

Limitations: Although the implemented NMF-based IDS gives good accuracy, the datasets were nonstandard. Moreover, the threshold technique used in the testing phase could not be applied to multi-class network attacks.

In this work we implement our NMF-IDS and test it using recent standard datasets specialized in IoT traffic such as **KDD99** and **CIC-IDS2017**. We also developed a new technique to detect multi-class attacks.

3. NMF BASED INTRUSION DETECTION SYSTEM

NMF-IDS system consist of three major phases. In phase 1 the network dataset file is converted into a two dimensional matrix $A^{m \times n}$. In phase 2, the matrix $A^{m \times n}$ will be factorized into two low-rank matrices $W^{m \times k}$ and $H^{k \times n}$. Phase 3 consists of the detection phase.

3.1. NMF Factorization Phase

Lee and swing [7] proposed a multiplicative updates algorithm (MU) to solve the NMF factorization problem as follows:

The matrices W and H are updated using the following formulas:

$$w_{ij} \leftarrow w_{ij} \frac{(AH^T)_{ij}}{(WHH^T)_{ij}} \quad (4)$$

$$h_{ij} \leftarrow h_{ij} \frac{(W^T A)_{ij}}{(W^T W H)_{ij}} \quad (5)$$

H^T means the matrix transpose of the matrix H . MU algorithm can be divided into individual smaller sub problems of matrix dot product. In step 1 we update W based on AH^T , HH^T and WHH^T , then in step 2 we update H based on $W^T A$, $W^T W$ and $W^T W H$. See algorithm 1 below.

Algorithm (1)

- $[W, H] = NMF(A, k)$
- $A^{m \times n}$ is the input matrix,
 k is rank of approximation
- (1): initialize random matrix H
- (2): **while** stopping criteria are not satisfied **do**
 - /*compute W given H^* */
 - (3): *computes* $W \leftarrow W_i \frac{A H^T}{WHH^T}$
 - /*compute H given W^* */
 - (4): *computes* $H \leftarrow H^i \frac{W^T A}{W^T W H}$
 - (5): **end while**

The while loop at algorithm 1 will stop if the stopping criteria are satisfied. Either it reaches the maximum number of iterations specified by the user, or it reaches convergence based on the Frobenius norm function $\min_{W, H} \|A - WH\|_F^2$ such that $W \geq 0$ and $H \geq 0$.

3.2. NMF Detection Phase

After MU algorithm reach to convergence or it reach the maximum number of iterations specified by the user, we obtain the factor matrices W and H , that can be used to represent every sample from A as weighted linear combination of columns of W , every column of w called bases where the corresponding h_{ij} called the weights or encoding.

$$\begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \begin{bmatrix} w_1 & w_2 & \cdot & w_k \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ \cdot \\ \cdot \\ h_{1k} \end{bmatrix} \quad (6)$$

$$a_1 = W * h_1$$

$$\begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \begin{bmatrix} w_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} h_{11} + \begin{bmatrix} w_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} h_{12} + \dots + \begin{bmatrix} w_k \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} h_{1k}$$

Now if we have a new sample let's call it i and we want to check if it matches one or more of the samples from the training set, we compute the encoding of it using W :

$$i = W * h' \quad (7)$$

Based on the upper formula we can find h' by:

$$h' = (W^T W)^{-1} W^T \times i \quad (8)$$

Now we check the similarity of this encoding with every encoding that existed in H . The closest match (sample class) is that sample whose encoding is the closest to the new sample (multi-class detection).

We can determine the matching score between the encodings using the following formula:

$$s_1 = \frac{h' \cdot h_1}{|h'| |h_1|} \quad s_2 = \frac{h' \cdot h_2}{|h'| |h_2|} \quad \dots \quad s_n = \frac{h' \cdot h_n}{|h'| |h_n|} \quad (9)$$

Where s_i with the maximum value is the closest encoding to the new input sample i .

4. IMPLEMENTATION OF NMF-IDS AND EXPERIMENTAL RESULTS

4.1. NMF-IDS Methodology

In this study we will test our NMF-IDS on two known datasets (KDD, and CIC) after a pre-processing phase, training phase using NMF factorization eq. 4, 5, and detection testing eq. 7, 8, 9.

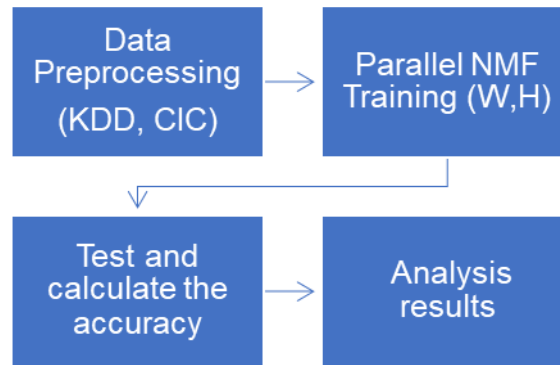


Figure 2: NMF-IDS Methodology

4.2. Datasets

This section will discuss the datasets used in this study in detail and how we pre-process them. To study the performance of the proposed solution in this study, we applied it and analysed its efficiency and accuracy on two different datasets, namely KDD and CIC.

KDD99: KDD dataset is a dataset used in an international competition held at the University of California [8], where the goal of that competition was to build an intrusion detection system that can differentiate between a normal good connection, or a bad connection called an intrusion or attack. KDD dataset is about 5 million connection records that was generated from 7 days' network traffic. It contains 41 features, and it is labelled by either Normal or specific type of attack. KDD contains attacks that can be categorized to Denial of service (DoS), Remote to local (R2L), and User to root (U2R).

CIC-IDS2017: The second dataset used in this study is CICIDS2017, created by I.Sharafaldin [9] from the Canadian Institute for Cybersecurity. It's a benchmark dataset consisting of **2830743 samples** and **78 features**. CIC dataset contains more recent attacks. For example, Brute Force FTP, DoS, infiltration, DDoS.

4.3. Datasets Pre-processing

This section explains the pre-processing methods for KDD and CIC datasets before applying NMF to them to get the best results.

Label Encoding: To apply NMF to any dataset, we must ensure that all elements within the dataset are non-negative numbers. KDD dataset contains some features with text values namely, service, Protocol_type, and flag, so they need to be converted to numeric values using label_encoder from sklearn library of Python.

Normalization: Some features from the datasets contain large numbers. For example, src_bytes and dst_bytes from KDD have large values that can reach thousands. Also, in CIC dataset Flow_Duration contains values reach more than 1 million and Destination_Port can reach thousands, those great values may affect the model's performance as it will be biased to those great values. Therefore, normalization is applied to ensure that all the dataset's values are in the same range. In this study, we apply min-max normalization to make sure that all the values are ranged between 0 and 1 only.

Train/Test Split: We divided KDD and CIC datasets into several training data sizes to apply the proposed parallel NMF on it.

- Training datasets sizes (30K)
- Testing dataset size 3K

The original shape of the datasets was $samples \times features$ to reduce the number of features and to get correct results out from NMF we will transpose the input matrix so it will be on the following shape $features \times samples$.

4.4. Experiments and Results

To conduct our experiments, the computer node is Lenovo Think-System SR630 which is based on CentOS 7 Linux operating system, Dual 14-cores CPUs, 197 GB RAM, and 480TB + 12GB (SSD) storage.

4.4.1. Experiment (1) Find Best Rank K for KDD Dataset

To make the most of NMF Algorithm, we must choose the rank hyper-parameter K carefully to ensure that it gives us good results and simultaneously reduces the dimensions of the dataset. Although we keep the most amount of the features whenever we raise the value of K , but in terms of detection accuracy and performance, it may give worse results and slower training.

Different runs of NMF with different values of K , as shown in Figure 3. After 1000 iterations of NMF updates, it gave varied results. By analyzing the results of the experiments, $K = 10$ was selected as it gives an accuracy rate reaching up to 98% in 1000 iterations.

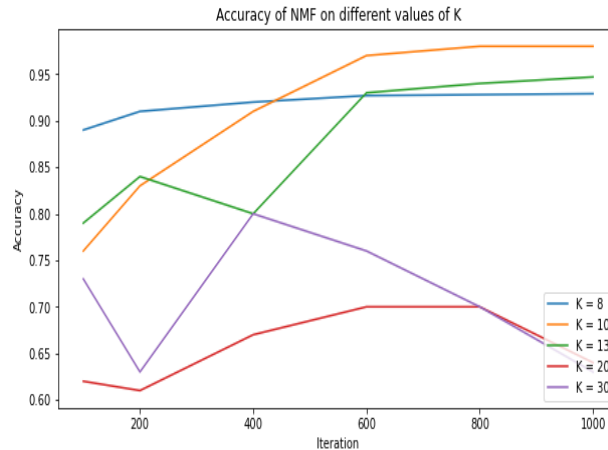


Figure 3: NMF-IDS Best Rank selection for KDD

4.3.2. Experiment (2) Training & Testing on 30K samples KDD

Using $K = 10$ we implemented NMF on 30000 samples of 42 features from KDD dataset. The results are shown in table (1)

Table 1. 30K samples KDD dataset Results

Iterations	Training Time(s)	Accuracy (%)
100	3.9	76
200	7.7	83
400	15.0	91
600	23	97
800	31	98

As shown in the Table 1, increasing the iterations gives better results the accuracy rate is increasing. But we get this at an additional cost of training time (factorization). As it is clear, NMF finished 100 iterations in approximately 4 seconds, with a detection accuracy of 76%, compared to 800 iterations in approximately 31 seconds, but with a detection accuracy of 98%.

4.3.3. Experiment (3) Find Best Rank K for CIC Dataset

In this experiment we run Different runs of NMF with different values of K for 1000 iteration, as shown in the Figure 4. After 1000 iterations of NMF updates, it gave varied results. By analyzing the results of the experiments, $K = 13$ was selected as it gives an accuracy rate reaching up to 90% in 1000 iterations.

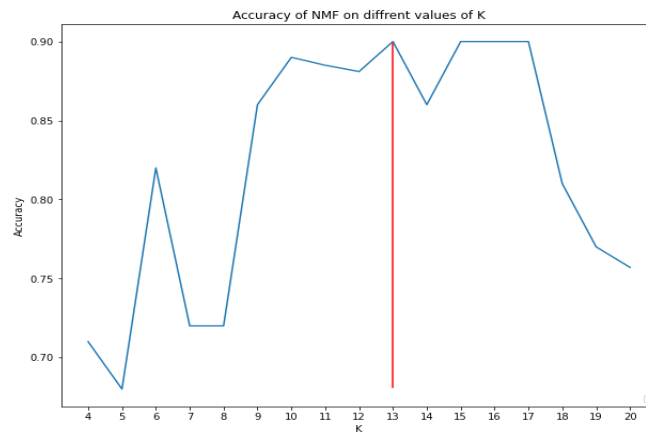


Figure 4: NMF-IDS Best Rank selection for CIC

4.3.4. Experiment (4) Training on 30K samples CIC

Using $K = 13$ based in experiment 4, we implemented parallel NMF on 30000 samples of CIC. Using different sizes of processors, we got the following results shown in Table 2.

Table 2. 30K samples CIC Dataset Results

Iterations	Training Time(s)	Accuracy (%)
100	20.5	66
200	41.0	74
400	82.1	84
600	124.0	87
800	164.0	90

As shown in the Table 2, increasing the iterations gives better results the accuracy rate is increasing but with an additional cost for training. NMF-IDS reaches a detection accuracy of 90% in 164 seconds.

5. CONCLUSION

Millions of users and hundreds of millions of devices connected to the network produce millions of network traffic records. In this paper we proposed an Intrusion detection system based on dimensionality reduction using Non-Negative Matrix Factorization (NMF) to be able to analyze efficiently large IoT traffic datasets. The algorithm used to compute NMF factors (W and H) is based Alternating Update algorithm using Multiplicative Update (MU). Unlike the previous work, our implementation can detect multi-class of network attacks. Experiential results show a detection precision of 98% for KDD datasets and 90% precision for CIC dataset. As a limitation to this work, we noticed that for large datasets the computational cost of the NMF factorization is large using a single computer. Another limitation is the slow convergence rate for MU method in term of number of iterations. In our future work we will extent NMF-IDS using High Performance Computers (HPC) to handle in real-time larger datasets. We will also consider using faster convergence alternative update Hierarchical Alternating Least Squares (HALS).

REFERENCES

- [1] L. Xiao, X. Wan, X. Lu, Y. Zhang, and D. Wu, "IoT Security Techniques Based on Machine Learning: How Do IoT Devices Use AI to Enhance Security", *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 41–49, 2018.
- [2] N. Kshetri and J. Voas, "Hacking Power," no. December, pp. 91–95, 2017.
- [3] Y. Li, J. Xia, S. Zhang, J. Yan, X. Ai, and K. Dai, "An efficient intrusion detection system based on support vector machines and gradually feature removal method," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 424–430, 2012.
- [4] W. C. Lin, S. W. Ke, and C. F. Tsai, "CANN: An intrusion detection system based on combining cluster centers and nearest neighbors," *Knowledge-Based Syst.*, vol. 78, no. 1, pp. 13–21, 2015.
- [5] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges," *Cybersecurity*, vol. 2, no. 1, 2019.
- [6] X. Guan, W. Wang, and X. Zhang, "Fast intrusion detection based on a non-negative matrix factorization model," *J. Netw. Comput. Appl.*, vol. 32, no. 1, pp. 31–44, 2009.
- [7] D. D. Lee and H. S. Seung, "Algorithms for Non-Negative Matrix Factorization," in *Advances in Neural Information Processing Systems*, no. 1, 2000, pp. 556–562.
- [8] S. Stolfo, "KDD-99 Dataset," online, 1999.
<http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html> (accessed Dec. 24, 2022).
- [9] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," *ICISSP 2018 - Proc. 4th Int. Conf. Inf. Syst. Secur. Priv.*, vol. 2018-Janua, no. Cic, pp. 108–116, 2018.
- [10] R. Hedjam, A. Abdesselam, and F. Melgani, "NMF with feature relationship preservation penalty term for clustering problems," *Pattern Recognit.*, vol. 112, 2021.
- [11] T. Masuda, T. Migita and N. Takahashi, "An Algorithm for Randomized Nonnegative Matrix Factorization and Its Global Convergence," *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, Orlando, FL, USA, 2021, pp. 1-7.

EMPLOYING LARGE LANGUAGE MODELS FOR DIALOGUE-BASED PERSONALIZED NEEDS EXTRACTION IN SMART SERVICES

Takuya Nakata¹, Sinan Chen², Sachio Saiki³ and Masahide Nakamura²

¹Graduate School of Engineering, Kobe University, 1-1 Rokkodai-cho,
Nada-ku, Kobe, Japan

²Center of Mathematical and Data Sciences, Kobe University, 1-1
Rokkodai-cho, Nada-ku, Kobe, Japan

³School of Data and Innovation, Kochi University of Technology, 185
Miyakuchi, Tosayamada, Kami, Japan

ABSTRACT

Research concerning the personalization of services encompasses approaches such as machine learning and dialogue agents; however, the explainability of the recommendation process remains a challenge. Previous studies have proposed dialogue-based needs extraction systems utilizing the 6W1H need model, but extracting complex needs using simple natural language processing proved challenging. In this research, we embark on the development of an Application Programming Interface (API) that extracts user needs from natural language by leveraging the rapidly advancing Large Language Models (LLM), and on constructing a dialogue-based needs extraction system using this API. For evaluation, we conducted a verification on 100 needs with the aim of assessing the accuracy and comprehensiveness of the outputs from the needs extraction and restoration API. Through this study, it became feasible to extract needs with high accuracy and comprehensiveness from complex natural language using LLM.

KEYWORDS

Personalization, Need, Large Language Model, Natural Language Processing, Dialogue Agent

1. INTRODUCTION

With the widespread adoption of smart services in society, research on service personalization, which tailors service delivery and functionality to users' preferences, thereby creating added value, has become increasingly prevalent [1,2]. Two primary approaches exist in the study of individual adaptation: recommendations through machine learning and function recommendations and updates through dialogue [3]. Research has also been conducted on conversational recommendation systems, which integrate these approaches [4]. However, a challenge arises in explaining to users the rationale behind the recommendations [5,6].

Traditional research proposed the 6W1H need model, which expresses the user's need regarding how they wish to execute a service through seven elements: where, when, who, whom, why, what, and how [7]. This model aimed to address the challenge of explaining the recommendation process. In this framework, morphological analysis and syntactic analysis were

used to extract 6W1H elements from user needs utterances. However, there remained issues in extracting complex needs and specific elements.

The aim of this research is to address the challenges identified in previous studies and to enhance the diversity and accuracy of needs extraction. A key idea revolves around utilizing the Generative Pre-trained Transformer (GPT) model of Large Language Models (LLM), which has seen an explosive increase in usage in recent years, for the construction of a needs extraction Application Programming Interface (API) and the reconstruction of the dialogue-based needs extraction system [8]. The approach of this research is as follows:

(A1) Construction of the needs extraction and restoration API using LLM.

(A2) Design of the overall architecture.

(A3) Design of the dialogue flow.

For evaluation, the proposed Needs Extraction API and Need Restoration API will be tested with 100 inputs to assess the accuracy and comprehensiveness of the outputs, and whether the 6W1H elements can be sufficiently extracted.

2. PRELIMINARIES

2.1. Personalization

In the digital domain, personalization refers to the process of modifying a system's functionality, interface, and content to enhance its relevance to an individual or a group of individuals. There are two primary approaches to achieve personalization: machine learning and dialogue. The machine learning approach involves deep learning based on a user's past service usage history, aiming to predict and recommend services the user may prefer [9,10]. However, this approach faces challenges in explainability since the recommendation process often operates as a black box. The dialogue-based approach updates the user model, which represents user preferences, through interactions between the user and a virtual agent [11]. Additionally, this approach provides service updates. Dialogue-based personalization has been implemented in devices such as smart speakers and healthcare agents [12]. In recent years, conversational recommendation systems have emerged as an approach that integrates both machine learning and dialogue. These systems allow for user inquiries and feedback through dialogue and recommend service functionalities based on dialogue logs, using machine learning.

2.2. Conventional Research: Interactive Needs Extraction System

In conversation-based recommendation systems, the explainability of the recommendation process, similar to personalized adaptation through machine learning, is a challenge. Previous research aimed to introduce an interpretable need model, situated between dialogue and machine learning, to make the system's interpreted needs explainable to users. Specifically, they proposed a personal adaptation system as shown in Figure 1, with a particular focus on constructing a user needs extraction method. This proposed system extracts user needs through a dialogue-based user needs extraction method, which is then employed in machine learning to recommend services. In the extraction method from previous research, conversational content was gathered through voice dialogue agents, and needs data represented in a 6W1H format, which users could easily comprehend, were extracted and stored using natural language processing. The 6W1H need model consists of the following seven elements that depict how users want to utilize a service:

- where: The location where the service is performed.
- when: The time when the service is performed.

- who: The main entity performing the service.
- whom: The target of the service.

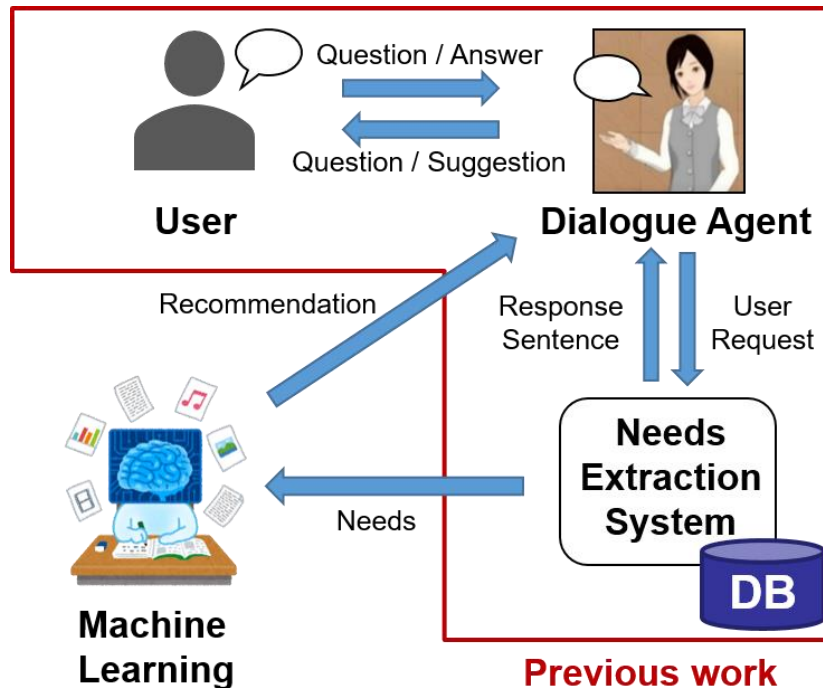


Figure 1. Overall architecture of the dialogue-based needs extraction system

- why: The reason or circumstances for performing the service.
- what: The specific action performed in the service.
- how: The method or tool used to execute the service.

This model employs the 6W1H analytical approach, a common framework for language usage, communication, and information organization, to address user needs effectively. Need data represented in the 6W1H format are referred to as 6W1H need data. For instance, the need “I want the smart speaker to notify my grandfather in the living room with a loud sound at 6:30 every morning if the weather forecast predicts rain” can be represented in 6W1H need data as:

- where: Living room.
- when: At 6:30 every morning.
- who: None.
- whom: My grandfather.
- why: If the weather forecast predicts rain.
- what: Notify with a loud sound.
- how: Smart speaker.

Traditional research used morphological and syntactical analysis to guess each element from the meanings of nouns and particles and to extract 6W1H need data. If some 6W1H elements were missing, the system would repeatedly ask the user questions like, “Where is the location?” to extract as detailed needs as possible from the user during the conversation. Finally, after extracting the needs, the system confirms with the user to ensure the extracted result matches the user’s intended specifics.

2.3. Challenges in Previous Research

The needs extraction method using morphological and syntactical analysis in previous research has its limitations. Firstly, distinguishing between who and whom is challenging. Both elements represent people related to the execution of the service. However, determining whether one is the executor or the target of the service is difficult. Moreover, extracting the why element is problematic. It is challenging to discern the reasons or conditions for wanting to execute a service based solely on grammar. Additionally, extracting needs from complex sentences is demanding. It is not possible to accurately extract the 6W1H elements from utterances concerning multiple service executions. For instance, from the need expression, “I want to watch movies in the living room with a video service and do karaoke every morning with a music service,” it is unclear, based solely on dependency relations, whether the where element pertaining to the living room is related to the video or music service.

2.4. Large Language Model

The LLM is a natural language processing model trained on extensive text data. It can be specialized for specific natural language processing tasks through fine-tuning [13]. GPT, developed by OpenAI, is one such LLM [14]. It is capable of various natural language processing tasks such as text generation, translation, and document summarization. In 2022, the Artificial Intelligence (AI) chat service ChatGPT was introduced, garnering significant global attention. ChatGPT is a service developed based on GPT-3.5 and GPT-4. With GPT, various processes can be executed by creatively designing natural language processing instructions using prompt engineering. The prominent method, Few-shot Prompting, enhances answer accuracy by providing several question and answer pairs as examples [15].

3. PROPOSED METHOD

3.1. Goal & Key Idea

The objective of this study is to enhance the diversity and accuracy of need extraction by extracting 6W1H need data from complex natural language expressions of needs communicated through interactions with virtual agents. The key idea revolves around the construction of a method that utilizes the LLM to extract 6W1H need data from intricate natural language expressions. The approach of this study is as follows:

(A1) Construction of the needs extraction and restoration API using LLM.

(A2) Design of the overall architecture.

(A3) Design of the dialogue flow.

3.2. (A1) Construction of the needs extraction and restoration API using LLM

We developed three distinct APIs utilizing the LLM to achieve the conversion between natural language and 6W1H need data: the Needs Extraction API, the Needs Re-extraction API, and the Need Restoration API. An API is a mechanism that allows programs to be invoked from outside the software. The implementation of these three types of APIs enables the extraction, re-extraction, and restoration of needs from various programs. The Needs Extraction API is designed to transform user needs expressed in natural language into 6W1H need data. The Needs Re-extraction API supplements missing 6W1H elements from previously extracted 6W1H need data using new user need utterances. The Need Restoration API reverts 6W1H need data back into a natural language expression of the need. For the API implementation, OpenAI’s Chat API

was employed. The model used is gpt-3.5-turbo, with a sampling temperature of 0.0. Few-shot Prompting was utilized for the prompt, providing several input-output examples. The programming language used for the API's development is Python.

```
[{
  "where": "",
  "when": "",
  "who": "I",
  "whom": "people who match my interests",
  "why": "",
  "what": "connect and share real-time content",
  "how": "SNS"
}]
```

Figure 2. Example of Needs Extraction API output

```
[{
  "where": "everywhere",
  "when": "every day",
  "who": "I",
  "whom": "people who match my interests",
  "why": "",
  "what": "connect and share real-time content",
  "how": "SNS"
}]
```

Figure 3. Example of Need Restoration API input

Two values are input into the Needs Extraction API. The first is the user's need statement in natural language directed towards a virtual agent, specifying a particular service. The second is a list of service candidates for the how (desired service), derived from preliminary string searches on the need statement. The output yields the 6W1H need data extracted from the need statement in a JavaScript Object Notation (JSON) array format. For instance, when inputting the need statement "I want to connect with people who match my interests on SNS and share real-time content" and providing "SNS" as the how candidate, a result similar to Figure 2 can be expected. The output is not a single JSON format, but a JSON array, which can extract multiple 6W1H JSON elements corresponding to multiple needs related to a service or different services from a single utterance.

The Needs Re-extraction API requires three input values. The first two, the need statement and the how candidate list, are the same as the Needs Extraction API. The third is a previously extracted single 6W1H need data. The output merges the content of the need statement and the previous 6W1H need data, producing a new 6W1H need element in a JSON array format. If, for instance, the provided 6W1H need data lacks where and when elements, and the accompanying statement is "I want to use it every day in the living room", the output might include where as "living room" and when as "every day". However, when multiple 6W1H need data are output, it is necessary to determine if each 6W1H need data updates the input need data or is entirely new.

One potential method for this determination is considering the JSON as a string and identifying the updated need by the shortest Levenshtein distance.

For the Need Restoration API, the only input is the 6W1H need data. The output is a natural language sentence representing the need. For instance, when the 6W1H need data shown in Figure 3 is input, the output might be “Every day, I want to use SNS to connect with people who match my interests and to share real-time content anywhere.”

3.3. (A2) Design of the overall architecture

The overall architecture of the interactive user needs extraction system utilizing LLM is illustrated in Figure 4. The frontend application consists of a virtual dialogue agent and the Google Speech API [16]. The backend application is composed by integrating various systems and APIs, with the interactive needs extraction system at its core. These include the user management system, service management system, Needs Extraction API, Need Restoration API, and Chat API. The 6W1H need data extracted by the system is stored in the needs database through the interactive needs extraction system.

The system’s operations are executed in the order indicated by the arrows in Figure 4. Initially, a user vocalizes their service needs. This voice is recorded by the dialogue agent. The recorded voice is then converted into a text-based needs statement through the Google Speech API. Subsequently, this statement is passed to the interactive needs extraction system. The interactive needs extraction system first queries the user management system to determine whether the user who expressed the need is registered in the system. The text of the expressed need is normalized by converting full-width characters into half-width characters and lowercase letters into uppercase. In case of Japanese, kanji and hiragana characters into katakana are normalized to katakana characters. By comparing the normalized string with the list of registered service names retrieved from the service management system, candidates for the how (service name) that the need targets are extracted. The original expressed need and the how candidates are then fed into the Needs Extraction API or the Needs Re-extraction API, producing the 6W1H needs data as output. The needs extraction (or re-extraction) API internally invokes the Chat API using the GPT model. The extracted 6W1H needs data is accumulated in the MySQL database termed the needs database.

The process for generating a response is intricate, branching according to the content of the expressed need and the results of the needs extraction. To summarize, when a need statement is required for purposes like confirming a need with the user, the 6W1H need data is provided as input to the Need Restoration API, which outputs the need statement. Following that, conditional branching occurs based on the content of the 6W1H need data and the need statement, formulating the response. This generated response is handed over to the dialogue agent. Ultimately, the dialogue agent articulates the response to the user.

3.4. (A3) Design of the dialogue flow

There are two types of dialogue flows: (D1) a response to a new needs statement and (D2) a response to an additional statement made during needs extraction. A single cycle refers to the repetition of dialogue needed to extract one need related to a service. The (D1) flow is applied to the initial user statement of the cycle, whereas the (D2) flow is applied to user statements from the second instance onward. The operational steps for the (D1) response flow to a new needs statement are as follows:

Step 1. Extract the 6W1H need data from the received statement using the Needs Extraction API.

Step 2. If any 6W1H elements are missing, request additional information from the user.
 Step 3. If no 6W1H elements are missing, restore the need statement using the Need Restoration API and confirm with the user if the need has been correctly extracted.

In the (D2) response flow to an additional statement during needs extraction, the same 6W1H need data can be continuously updated using a common dialogue ID to search the needs database. The operational steps for this response flow are as follows:

Step 1. Retrieve previously extracted 6W1H need data from the database based on the dialogue ID.

Step 2. Re-extract new 6W1H elements from the received additional statement and previous 6W1H need data using the Needs Re-extraction API.

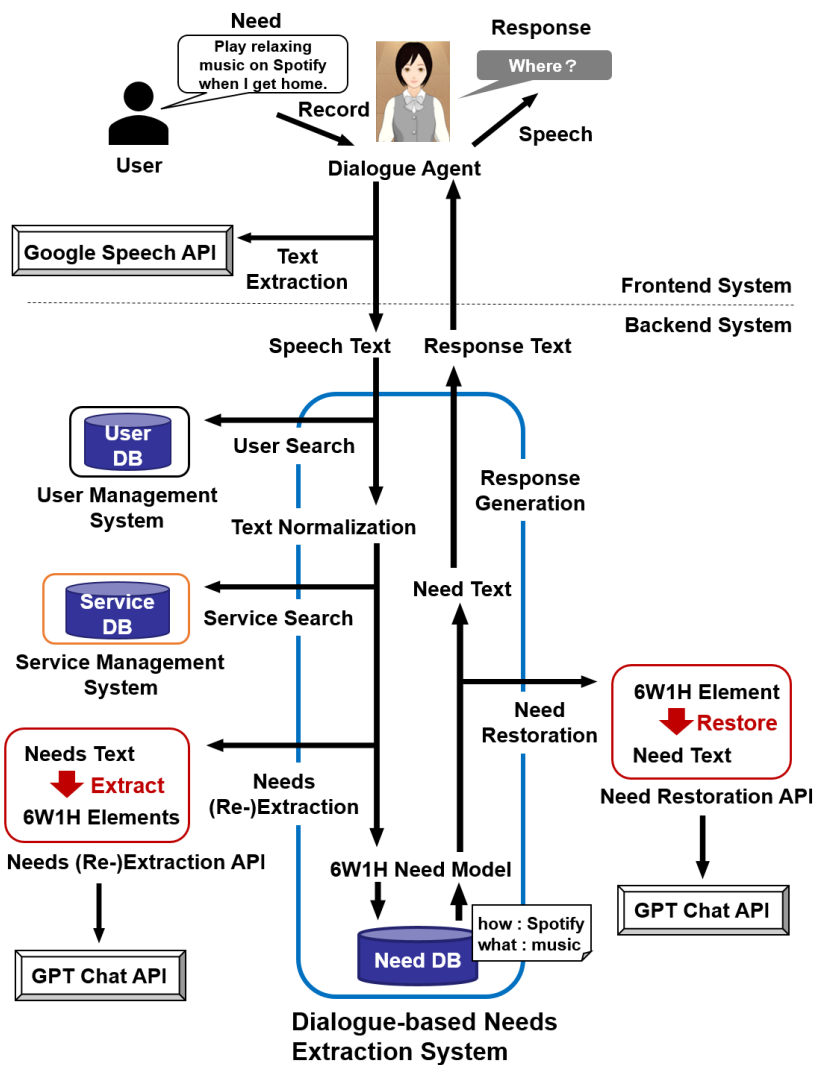


Figure 4. Overall architecture and operational flow of the needs extraction system

Step 3. If multiple 6W1H need data are re-extracted, determine the 6W1H need data closest to the previous 6W1H need data in terms of the Levenshtein distance through calculation. Consider this determined 6W1H need data as the updated data for the current dialogue.

Step 4. Generate the response statement in the same manner as Steps 2 and 3 of the (D1) flow.

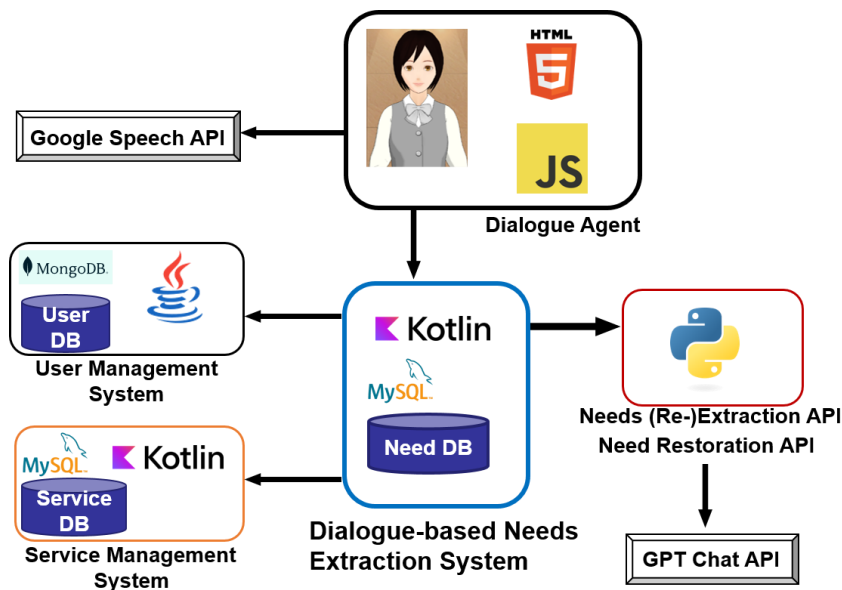


Figure 5. Implementation architecture of needs extraction system

4. IMPLEMENTATION

The implementation architecture is depicted in Figure 5. The Dialogue Agent is implemented using HyperText Markup Language (HTML) and JavaScript, functioning within a browser. The Dialogue-based Needs Extraction System is developed in Kotlin, with the Needs database realized using MySQL. The User Management System is implemented in Java, and rather than using MySQL, MongoDB is employed for the User database to represent complex attributes for each user in JSON. This facilitates the cross-utilization of the database across multiple systems, not limited to the proposed system. However, it is also feasible to implement the database using MySQL if user management is confined within the proposed system. The Service Management System is developed in Kotlin, with the Service database realized using MySQL. The Needs Extraction, Re-extraction, and Restoration API have a simple configuration, consisting solely of sending commands to the GPT Chat API, and are thus implemented on the same Python server. In this implementation, five servers are utilized to clarify roles. This allows for flexible configurations, such as installing servers handling personal information on-premises while placing others in the cloud. Alternatively, given that complex computations are performed using external services via the API, it is possible to virtually realize all five servers on a single server.

5. NEEDS EXTRACTION & RESTORATION API EVALUATION EXPERIMENT

We conducted an evaluation experiment with the objective of addressing the following research questions concerning the needs extraction and restoration API of (A1):

RQ1 Do the Needs Extraction API and Need Restoration API accurately and comprehensively convert needs?

RQ2 Does the Needs Extraction API sufficiently extract the 6W1H elements?

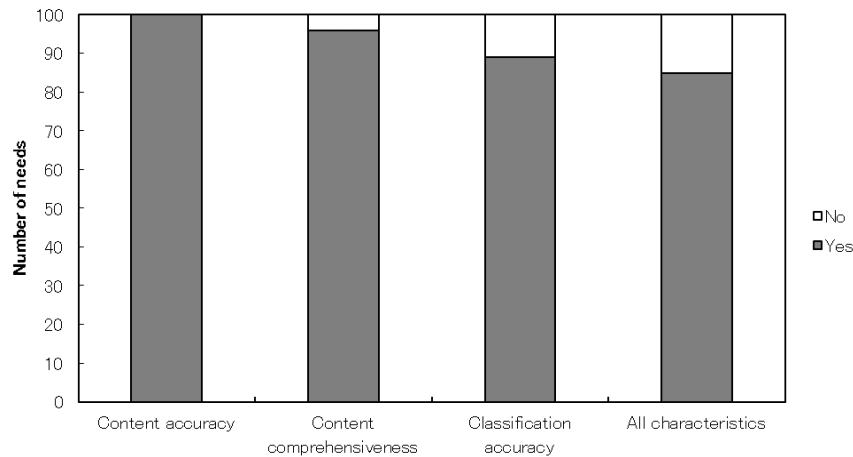


Figure 6. Characteristics of extraction results from Needs Extraction API

5.1. Experimental Setup

In the evaluation experiment for the Needs Extraction API, we generated test data for input to the API using ChatGPT. We created 100 variations of need statements, assuming various perspectives such as students, elderly, and working professionals using ChatGPT. These need statements encompassed a total of 28 different services. To examine the basic quality of the API, each need statement focused on only one service, and we refrained from using need statements that targeted multiple services. The evaluation involved authors analyzing the output results of the API, examining its characteristics and conditions. For the evaluation experiment of the Need Restoration API, we generated 6W1H need data as test input for the API using ChatGPT. The 100 variations of 6W1H need data were crafted by ChatGPT, taking into consideration various perspectives like students, elderly, and working professionals. The how element of the 6W1H need data included a total of 21 different services. The evaluation process consisted of the authors analysing the output results of the API to determine its characteristics.

5.2. Experimental Result

The results of the evaluation experiment for the Needs Extraction API are shown in Figures 6 and 7. The outcomes presented in Figure 6 are analyses related to three characteristics. Content accuracy indicates whether the extracted results contain any false information. Content comprehensiveness signifies whether all crucial information from the spoken sentence has been extracted. Classification accuracy represents whether each extracted need element has been appropriately categorized into one of the 6W1H elements. For instance, if an element, which should be extracted under where, such as “living room”, is categorized under when, then the output result was evaluated as not meeting classification accuracy. The graph also illustrates output results that simultaneously satisfy all characteristics.

The results in Figure 7 provide an analysis of how many types of the 6W1H elements could be extracted. In the section on detailed conditions, we analyzed whether at least one of the where, when, who, or whom elements, which express detailed conditions for service execution, has been extracted. In the section on overall conditions, we assessed whether the why element, which conveys the general conditions for service execution in a sentence, was extracted. The graph also displays the output results that extracted both detailed and general conditions, and those that extracted either one of the two.

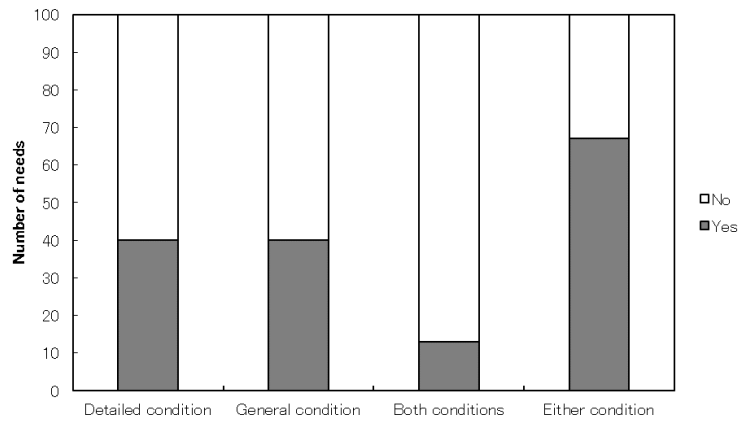


Figure 7. Conditions extracted by the Needs Extraction API

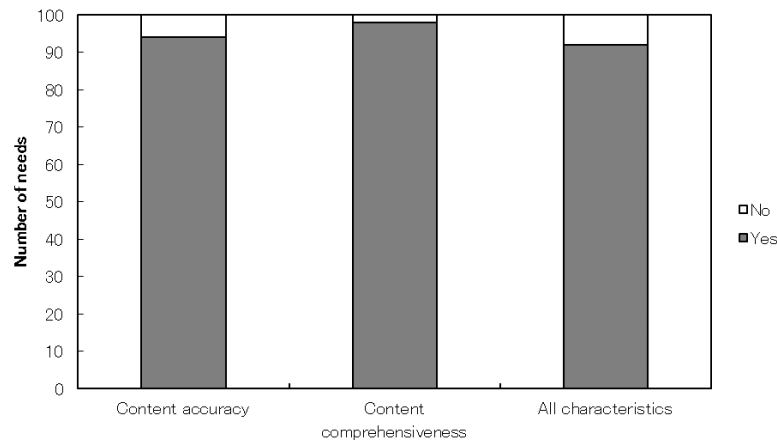


Figure 8. Characteristics of restoration results from Need Restoration API

The results of the evaluation experiment for the Need Restoration API are displayed in Figure 8. Content accuracy indicates whether the restored results contain any false information. Content comprehensiveness signifies whether all essential information from the 6W1H elements has been covered. The graph also illustrates output results that simultaneously satisfy all characteristics.

5.3. Experimental Discussion

We will discuss the results of the experiment with respect to RQ1 and RQ2. First, considering RQ1, we analyze whether the conversion of needs was carried out accurately and comprehensively. In the Needs Extraction API, the content accuracy was 100%, with the extraction results being entirely truthful and accurate. The content comprehensiveness stood at 96%. However, the omitted content was non-core, supplemental information such as “want to shop conveniently” or “wish to have a fun time”. The classification accuracy was 89%. There were many instances where content that should be categorized under detailed conditions like when or under what was mistakenly categorized under why. In the Need Restoration API, the content accuracy was 94%, and there were instances where the content of why and what was reversed in the formulation of sentences. The content comprehensiveness was 98%, and when multiple elements had similar content, some information to be formulated into sentences was missing. Furthermore, 85% in the extraction API and 92% in the restoration API satisfied all the

characteristics. From these results, it can be said that in use cases where users verify the correct extraction of their needs through an interactive needs extraction system, the needs extraction and restoration API possess high accuracy and comprehensiveness.

Next, concerning RQ2, we consider the sufficiency of the 6W1H element extraction. In the Needs Extraction API, 40% could extract detailed conditions, and 40% could extract general conditions. Moreover, 67% could extract either the detailed conditions or the general conditions. From this, it is challenging to claim that sufficient 6W1H elements are extracted in a single extraction. However, the types of 6W1H elements that can be extracted from the need sentence depend on the content of the needs sentence. Furthermore, in use cases where the user is asked about the missing 6W1H conditions during repeated interactions, it can be said that the Needs Extraction API possesses the capability to extract the 6W1H elements adequately.

6. DISCUSSION

By utilizing the LLM for the mutual conversion between needs statements and 6W1H need data, we succeeded in extracting needs from complex natural language utterances. Specifically, we can now extract multiple needs from a single needs statement and have become able to extract the who and why elements, which were challenging in previous research. From the results of the API evaluation experiment, it was determined that needs can be accurately extracted with a single extraction. However, the characteristics regarding the classification and diversity of each element remain uncertain, highlighting the importance of dialogues that inquire about missing 6W1H elements. To realize a needs extraction dialogue, we leveraged a user search system, service search system, and needs extraction and restoration API, implementing the needs extraction system as a distributed microservices architecture. As future challenges, we believe it is essential to undertake performance evaluation experiments of the entire system and evaluate the system quality when used by actual users.

7. CONCLUSION

The objective of this research is to construct a method to extract 6W1H need data from complex natural language needs statements through dialogue with virtual agents, thereby enhancing the diversity and accuracy of needs extraction. As a key idea, we have developed a method to extract 6W1H need data from complex natural language needs statements using the LLM. Specifically, we designed and developed three mutual conversion APIs for extracting, re-extracting, and restoring 6W1H need data from natural language using the LLM. Moreover, we designed the overall architecture of the dialogue-based needs extraction system using the proposed APIs. By incorporating the LLM into the API, we have streamlined and redesigned the complex dialogue flow.

Additionally, we conducted evaluation experiments by inputting 100 test data into the Needs Extraction API and the Need Restoration API. As a result, the extraction API yielded outputs that satisfied 85% for accuracy and comprehensiveness, while the restoration API yielded 92%. Through this research, we have enabled the extraction of 6W1H need data from complex natural language with high accuracy and comprehensiveness using the LLM. For future research, we are considering evaluating the dialogue-based needs extraction system in more practical applications.

ACKNOWLEDGEMENTS

This research was partially supported by JSPS KAKENHI Grant Numbers JP19H01138, JP20H05706, JP20H04014, JP20K11059, JP22H03699, JP19K02973, and Young Scientists (No.23K17006)

REFERENCES

- [1] Fan, H. & Poole, M.S., (2006) “What is personalization? perspectives on the design and implementation of personalization in information systems,” *Journal of Organizational Computing and Electronic Commerce*, Vol.16, No.3-4, pp179-202.
- [2] Hammi, B., Zeadally, S., Khatoun, R., & Nebhen, J., (2022) “Survey on smart homes: Vulnerabilities, risks, and countermeasures,” *Computers & Security*, Vol.117, pp102677.
- [3] Huang, J., Tong, Z., & Feng, Z., (2022) “Geographical POI recommendation for Internet of Things: A federated learning approach using matrix factorization,” *International Journal of Communication Systems*, ppe5161.
- [4] Jannach, D., Manzoor, A., Cai, W., & Chen, L., (2021) “A survey on conversational recommender systems,” *ACM Computing Surveys*, Vol.54, pp1-36.
- [5] Hollis, K., Soualmia, L., & S´eroussi, B., (2019) “Artificial intelligence in health informatics: Hype or reality?,” *Yearbook of Medical Informatics*, Vol.28, pp3-4.
- [6] Zhang, Y. & Chen, X., (2020) “Explainable Recommendation: A Survey and New Perspectives,” *Foundations and Trends® in Information Retrieval*, Vol.14, No.1, pp.1-101.
- [7] Nakata, T., Chen, S., Saiki, S., & Nakamura, M., (2023) “Dialogue-Based User Needs Extraction for Effective Service Personalization,” *HIMI 2023, Held as Part of the 25th HCI International Conference, HCII 2023*, Vol.LNCS 14016, pp139-153.
- [8] Cascella, M., Montomoli, J., Bellini, V., & Bignami, E., (2023) “Evaluating the feasibility of chatgpt in healthcare: An analysis of multiple clinical and research scenarios,” *Journal of Medical Systems*, Vol.47, No.1, pp33.
- [9] Rostami, M., Oussalah, M., & Farrahi, V., (2022) “A Novel Time-Aware Food Recommender-System Based on Deep Learning and Graph Clustering,” *IEEE Access*, Vol.10, pp52508-52524.
- [10] Goldenberg, D., Kofman, K., Albert, J., Mizrachi, S., Horowitz, A., & Teinemaa, I., (2021) “Personalization in practice: Methods and applications,” *WSDM '21*, pp1123-1126.
- [11] Kocaballi, A.B., Berkovsky, S., Quiroz, J., Laranjo, L., Tong, H.L., Rezazadegan, D., Briatore, A., & Coiera, E., (2019) “The personalization of conversational agents in health care: Systematic review,” *Journal of Medical Internet Research*, Vol.21, ppe15360.
- [12] Ozono, H., Chen, S., & Nakamura, M., (2022) “Encouraging elderly self-care by integrating speech dialogue agent and wearable device,” *8th International Conference, ITAP 2022, Held as Part of the 24th HCI International Conference, HCII 2022*, Vol.LNCS 13331, pp52-70.
- [13] Tinn, R., Cheng, H., Gu, Y., Usuyama, N., Liu, X., Naumann, T., Gao, J., & Poon, H., (2023). “Fine-tuning large neural language models for biomedical natural language processing,” *Patterns*, Vol.4, No.4, pp100729.
- [14] Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y.T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M.T., & Zhang, Y., (2023) “Sparks of artificial general intelligence: Early experiments with gpt-4,” *ArXiv*, Vol.abs/2303, pp12712.
- [15] Lu, Y., Bartolo, M., Moore, A., Riedel, S., & Stenetorp, P., (2022) “Fantastically Ordered Prompts and Where to Find Them: Overcoming Few-Shot Prompt Order Sensitivity,” in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, Vol.1: Long Papers, pp8086-8098.
- [16] Speech-to-Text: Automatic Speech Recognition — Google Cloud. <https://cloud.google.com/speech-to-text?hl=en>. [Online]. (Accessed on 10/18/2023).

AUTHORS

Takuya Nakata received Master of System Informatics from Kobe University, and he did Bachelor degree in Engineering. Currently, he is pursuing his PhD in Engineering from Kobe University. His research interests include Service computing, Software engineering.



Sinan Chen received a Master of system informatics and a Ph.D. in computational science from Kobe University, Japan, in 2019 and 2021, respectively. He worked at Kobe University as a research fellow (PD) from the Japan Society for the Promotion of Science (JSPS) in the first half of 2022. He is currently an assistant professor in the Center of Mathematical and Data Science Center at Kobe University, Japan. His research interests include the smart home, cloud computing, machine learning, and software engineering. He is a member of the IEEE, ACM, Sigma Xi, CAAI, CCF, IEICE, and IEEJ.

Masahide Nakamura received the B.E., M.E., and Ph.D. degrees in Information and Computer Sciences from Osaka University, Japan, in 1994, 1996, 1999, respectively. From 1999 to 2000, he has been a post-doctoral fellow in SITE at University of Ottawa, Canada. He joined Cybermedia Center at Osaka University from 2000 to 2002. From 2002 to 2007, he worked for the Graduate School of Information Science at Nara Institute of Science and Technology, Japan. From 2007 to 2022, he worked for the Graduate School of System Informatics at Kobe University. He is currently a full professor in the Center of Mathematical and Data Science Center at Kobe University. In 2015, he worked for Universite Grenoble Alpes as a visiting professor. In 2018, he joined RIKEN Center for Advanced Intelligence Project as a visiting researcher. His research interests include the service/cloud computing, smart home, smart city, and gerontechnology. He is a member of the IEEE, ACM, IEICE and IPSJ.

© 2023 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.

REVIEW OF DIGITALIZATION USING ARTIFICIAL INTELLIGENCE MATURITY MODELS: THE CASE OF AMERICAN AUTOMOTIVE SMES

Dharmender Salian

Department of Information Technology, University of the Cumberlands,
New York, USA

ABSTRACT

The purpose of this study is to review studies related to Artificial Intelligence (AI) maturity models (MM) in automotive manufacturing in a systematic manner. SMEs in the automotive industry must embrace digitalization to remain competitive. SMEs employ a large segment of the USA's workforce. SMEs had not been aggressive in digitalization due to scarce funds, but the benefits of operational efficiency, quality improvement, cost reduction, and innovative culture have made it attractive to consumers. A growing number of operations are being digitalized using Artificial Intelligence techniques. In this paper, AI applications in SMEs are examined through the lens of an AI maturity model.

KEYWORDS

Industry 5.0, Maturity Model, Artificial Intelligence, Maturity level & digitalization

1. INTRODUCTION

In the 1940s, the programmable digital computer was invented, which led to the beginnings of AI. After this, the USA, British, and Japanese governments funded several projects, but it wasn't until the dawn of the 21st century when machine learning applications started generating AI interest and investments. AI can deliver sophisticated solutions to advance manufacturing. Although the changeover will take time, the benefits are likely to be endless. In artificial intelligence (AI), algorithms or computerized systems resemble human mental processes [1]. In the last few decades, the field of AI has made rapid progress in developing decision-making intelligence after the early days were characterized by overpromises and under-delivery [2]. Data is a crucial component of AI, since it is used to train algorithms to detect patterns based on the data collected [3]. Data and computing power are required by AI researchers to build more powerful algorithms [1].

The development of AI technologies was boosted by a summer workshop conducted by mathematician John McCarthy at Dartmouth College in 1956, which looked at some foundational problems. AI techniques like machine learning, computer vision, deep learning, automation, and robotics are now the subject of pioneering research. In 2019, the Artificial Intelligence in Manufacturing Market was worth USD 1.82 Billion, and by 2027, it is predicted to be worth USD 9.89 Billion [4]. One third of global GDP is produced by the industrial sector, and half of global energy consumption comes from it [5]. A new study shows that 58% of manufacturers are positively interested in artificial intelligence, but only 12% are using it [6].

AI will provide the most job security to those who possess creativity and management skills. The number of publications on artificial intelligence in manufacturing has increased exceptionally over the last 40 years, attracting much attention within the scientific community [7]. A growing number of companies are investing in hybrid technology systems to manage inventory, control quality, and optimize production and costs. Expert systems and robot localization, as well as visual surveillance, are less likely to be used by them. Technology can now be valued based on core use cases [8]. AI is used to simulate human reasoning, learning, planning, and other thinking activities, thus solving complex problems that were previously only solvable by human experts [9]. AI is particularly useful for automating learning, acquiring, processing, and using knowledge to perform tasks, enabling human decision-making processes to be improved through improved knowledge.

Achieving the goal of AI rivalling human abilities remains a work-in-progress, and it is uncertain if such a goal can be achieved [10]. Through technological maturity and integration with a variety of technologies, AI has become more relevant. There has been development in computing and chip design as well as neural network algorithms, which have developed into deep learning. There has also been convergence with technology such as augmented reality, robotics, 5G, virtual reality, and the internet of things [11]. There is a growing recognition among companies that they do not want to climb an artificial intelligence mountain. To achieve new heights, they just need to keep taking the right, tiny steps [12]. AI has triggered significant societal concerns, ranging from technological unemployment to the dominance of algorithms at work and in everyday life [13].

The SME sector employs 61.7 million workers in America, which is 46.4% of the total workforce [14]. AI is being funded by governments in advanced economies and large technology companies. Due to previous reviews not adequately addressing AI's use and advancement in engineering and manufacturing, this review is being conducted. The term "SMEs" is used in this paper to describe automotive small and medium enterprises.

MMs exist in different domains, and the goal is to answer the following questions.

RQ1 What is the role of AI maturity models in automotive manufacturing?

RQ2 How does the literature review describe different stages of maturity model?

RQ3 What are the important characteristics and goals of AI maturity models?

The current AI MMs are systematically reviewed to determine the dimensions for assessment. Organizations can use the AI Maturity Assessment to evaluate their current AI capabilities, identify gaps and areas for improvement, and create a guideline to build more successful AI programs [15]. The aim of this study is to illuminate the research gap and to guide future studies that need to consider the dimensions noted above. The rest of the paper is divided into the following sections: Section 2 and 3 presents a SWOT and PESTLE analysis of SMEs. Section 4 provides a literature review. Digital strategy is discussed in section 5. A description of AI MM for the aerospace industry can be found in section 6. Future Challenges is discussed in section 7. The conclusion of this study is in section 8.

2. SWOT FOR SMES

SME digitalization efforts and progress are discussed in this section

Table 1. SWOT

Strengths	Weaknesses
<ul style="list-style-type: none"> • An organization structure that is more flexible and easier to adapt. • The market share of electric vehicles has increased. 	<ul style="list-style-type: none"> • Recruiting skilled workers is difficult • Inadequacies in R&D infrastructure • The cost of investment is high
Opportunities	Threats
<ul style="list-style-type: none"> • Assistance from the government. • Comparatively lower production costs. 	<ul style="list-style-type: none"> • Market competition is intense. • Relocation of manufacturing hubs

- Strengths

An increase in the market share of electric vehicles: EVs have a lot of different components than conventional vehicles, especially the electric drive units, battery packs, and battery modules. Non-American SMEs have an entry barrier since they must start from scratch.

- Weaknesses

There aren't enough skilled employees: long-term training is needed. Investment costs can be a barrier since they are high.

- Threats

The American automobile market is very competitive with many new companies coming in and earlier manufacturing volumes are declining. Automobile manufacturing is shifting to emerging economies like China and India from developed markets.

- Opportunities

Digitalization efforts are being supported by the government. New manufacturing techniques lower production costs compared to conventional methods.

Since the early days of automobile development to the current development of electric and autonomous vehicles, the automotive industry is constantly evolving.

3. PESTLE ANALYSIS

Businesses use pestle analysis to track the environment in which new products or projects will launch (see Table 2).

Table 2 PESTLE

Political	Economical
<ul style="list-style-type: none"> In the United States, the auto industry is growing. 	<ul style="list-style-type: none"> Pandemics hit the automobile industry
Social	Technological
<ul style="list-style-type: none"> Skilled talent is in short supply in the auto industry. 	<ul style="list-style-type: none"> The digital revolution is gaining traction.
Legal	Environmental
<ul style="list-style-type: none"> There are regulations and restrictions in the auto industry. 	<ul style="list-style-type: none"> Impacts of manufacturing on the environment.

4. LITERATURE REVIEW

Over 30 million SMEs with fewer than 500 employees make up 99.9% of all American businesses. Digital tools aren't used favourably by 80% of U.S. SMEs [16]. AI maturity models help organizations evaluate their progress and identify the changes they need to make to be more productive and efficient. The dimensions above are discussed below. The use of AI to mitigate adverse environmental impacts is discussed in Sustainability in AI [17]. To increase sustainability coverage and remit the most profit to stakeholders, organizations need to quantify their environmental, financial, and social impacts [18]. An AI implementation that is connected requires the involvement of a multidisciplinary team with expertise in AI, data science, manufacturing processes, and OT and IT infrastructure. Connecting machines, processes, workplace health and safety, and managing product lifecycles all requires technical infrastructure [19].

Resilience is facing situations and recovering from them, and disruption is something we're all worried about [20]. In today's world, organizations see AI as a strategy to innovate by controlling and commanding, knowing their objectives, and being opportunistic [21]. Leadership needs to leverage AI to gain a competitive edge, and there needs to be inspired and proactive leadership to make AI investments [22]. Supporting customers with AI allows for deeper insights and a better user experience that indirectly builds mass personalization [23]. It's important to have a healthy innovation culture to prosper economically and make technology safer [24]. McKinsey found that 70% of manufacturers use or plan to use AI to improve operations in production, which suggests that many manufacturers are seeing the value of AI and are interested in adopting it [25].

A comprehensive and systematic search was conducted following the steps of identification, screening, eligibility, and inclusion. Identifying relevant records required a title and subject search in various academic databases, including ABI/Inform Global, Springer, ScienceDirect, IEEE, ACM Digital Library, and ProQuest. A manual search using Google and Google Scholar were also used to find papers, conference proceedings, books, and technology reports. In this paper, different search terms were used to identify some keywords based on the research questions such as "Artificial Intelligence", "Manufacturing", "SME", "Maturity Model", and "Artificial Intelligence Maturity Model". A good search strategy involves extracting individual terms from the research question and then using Boolean "ORs" and "ANDs" to perform advanced searches.

Table 3. Overall comparison of AI dimension

Author	Connectivity	resilience	sustainability	Expansive growth	Strategy	Leadership	Customers	Culture	Production
[26]	X	X		X	X		X		X
[27]	X	X	X	X	X	X		X	
[28]						X			X
[29]	X	X	X		X	X	X		X
[30]	X			X	X	X	X	X	X



Figure 1. Dimensions

As a result of the above comparison, the AI maturity model can be improved by identifying the shortcomings (see Figure 1). Since SMEs have lagged other domains in digitization efforts, the research, assessments, and implementation requirements are more intense for them. To save money and gain a competitive edge, digital transformation combines different technologies. SMEs will benefit if an AI maturity model fits their requirements and helps assess AI maturity levels.

5. DIGITAL STRATEGY

Regardless of the type of business, differentiation, leadership, and focus are Porter’s three generic strategies that anyone can use. A digitalization strategy requires the right culture, infrastructure, and capabilities, which can be achieved with tailored transformation measures and step-by-step

transformation procedures. Digital transformation is changing how companies create value and in digital transformation, it's the inversion of a company that makes the most money. This becomes possible when companies move from creating value independently to orchestrating value with other companies. The most successful companies partner with users, developers, and merchants at scale. A high market cap is not achieved through automation or shifting labour to capital, but rather by coordinating external value creation.

6. AI MM FOR THE AUTOMOTIVE SME MARKET

In a MM, the organization's preparedness is assessed, and faults are identified. These weaknesses need to be fixed and corrective actions are needed to move forward. AI MM needs to see how it aligns with Industry 5.0 and forge a more sustainable, human-centric, and resilient industry. Collaboration between humans and advanced technology is the goal of Industry 5.0. While Industry 4.0 is a technology-driven industry, Industry 5.0 is a value-driven one [31]. From total manual production to using operating machines and eventually assembly lines, the previous three industrial revolutions revolutionized our manufacturing industries. AI-related technologies are assessed with the MM, which extends smart manufacturing MMs by adding specific technical and nontechnical competences of these technologies.

It's time to make the organization human-centred, so MM doesn't have to be just about technology. People are the key to a successful digital strategy. Future SMEs need to be resilient so they can handle disruptions and assist vital assets. SME's have realized that sustainability measures can boost their recovery, speed up growth, and make them more profitable [32]. It's going to be tough for future SMEs to gain a competitive edge. During a global shortage of workers and supply chain uncertainty, we're transitioning from predictable internal combustion engines to variable and software-driven next-generation vehicles, including electric vehicles (EVs), hybrids, and autonomous vehicles.

Is it time to refocus our current assessment criteria because of the happening EV market? The digital transformation is the key to success, and EV manufacturers can benefit from a digital backbone [33]. In contrast to established automakers, new EV entrants are leading disruptive change. Electric vehicles (EVs) should be included in the MM assessment to help assess progress to EV and zero-emission vehicles. A vehicle brand is currently known by its engine manufacturer, but in the next ten years, it'll be known by its software provider (including AI-powered autonomous driving, advanced infotainment, etc.).

In the world of intelligent manufacturing, AI technology helps develop new models, system architecture, and technology systems. There are three key technologies that make IoT work: 5G, Big Data, and AI. Data is provided by IoT frameworks, and AI uses it for specific functions. The automotive industry is changing thanks to 5G and AI [34]. In comparison to cabled fibre optic networks, 5G provides a feasible broadband networking option [35]. Industrial IoT needs 5G's high performance and low latency [36]. Factory automation is an example of a latency critical IoT use case that involves real-time control of machines and systems and has latency and reliability requirements [37].

Automotive progress has been helped by electrification and automation, as well as advances in the communications industry, including 5G and 6G. By improving reliability, lowering latency, and guiding hyper automation, 6G will open new growth potential in business. Sixth generation (6G) mobile communications are also essential for Industry 5.0 to unseat Industry 4.0. High-quality services, extensive IoT infrastructure, AI capabilities, and other requirements will be supported by this network.

7. FUTURE CHALLENGES AND OPPORTUNITIES

Enterprises can improve their competitiveness by embracing intelligent transformation, which is the trend of the future [38]. Production is expected to increase by 40% by 2035 thanks to AI technologies. Across different industries, economic development will increase by an average of 1.7%. [39]. Although AI can surpass humans on some very particular tasks, humans still noticeably outperform in all real-world tasks necessitating intelligence. During this era of intelligent work, artificial intelligence has subverted traditional work methods and is widely used in medicine, self-driving technology, image recognition, robotic manufacturing, intelligent assistance, supply chain management, etc., using devices and technologies like cameras, video, light detection and ranging (LiDAR) and motion tracing [40]. Small and medium-sized enterprises have difficulty joining the intelligent manufacturing wave, and their implementation of intelligence is limited [38]. Our pace of invention, however, must keep up. Approximately seventy-five billion IoT-connected devices are expected in use by 2025, an almost threefold increase from 2019, and connecting all these devices with intellectual abilities will be a considerable challenge. [41].

AI's scope is also evolving, as it shifts from being "just a tool" to being a sci-fi creation that threatens mankind [42]. Although AI is still in its infancy, it has reached milestones that a few years ago seemed inconceivable. Christof Koch, a neuroscientist at the Allen Brain Institute and an AI advocate, believes humans will need to enhance their brains to compete with artificial intelligence [43]. Global competition and the need to develop a green economy have forced organizations to implement AI, and IoT in production flows. In the age of intelligent manufacturing, the manufacturing industry needs to lead innovations to compete globally [44].

8. CONCLUSIONS

An uncertain and turbulent environment could result in mandatory closures, logistics bottlenecks, supply issues, and a volatility in consumption trends which would have major consequences for the manufacturing industry. Manufacturers' success and survival are closely linked to their ability to embrace advanced digital technologies, such as artificial intelligence. Scholars and practitioners alike have become increasingly interested in AI over the last decade due to the growing amount of data and information that firms collect and process.

Manufacturing plants can do more than just track operations with machine learning algorithms. It is possible to contain inefficiencies, assess alternative strategies, and conduct new protocols simultaneously without affecting the supply chain [45]. AI and digital transformation are revolutionizing manufacturing. As organizations assess their capabilities, they need a robust AI maturity model that details incremental improvements. Present maturity models have shown deficiencies and therefore need to be addressed as per the AI focused SME maturity model requirements. The lack of technological awareness among SME managers, such as "uncertainty regarding the benefits of AI initiatives" and "insufficient understanding" of AI, is a problem [46].

Manufacturing companies around the world benefit from investing in artificial intelligence (AI) and optimizing their productivity [47]. Several manufacturers do not have the in-house capabilities and are risk-averse to upscaling a factory due to the high cost and lack of skills required for AI adoption [48]. A maturity model for AI needs to be developed to measure current state and developments, and concrete recommendations have been presented for the model.

REFERENCES

- [1] Wang, Y. (2021). Artificial intelligence in educational leadership: A symbiotic role of human-artificial intelligence decision-making. [Artificial intelligence in educational leadership] *Journal of Educational Administration*, 59(3), 256-270. doi: <https://doi.org/10.1108/JEA-10-2020-0216>
- [2] Russell, S. (2019), *Human Compatible: Artificial Intelligence and the Problem of Control*, Penguin, London
- [3] Wang, Y. (2020), "When artificial intelligence meets educational leaders' data-informed decisionmaking: a cautionary tale", *Studies in Educational Evaluation*, doi: 10.1016/j.stueduc.2020.100872
- [4] Artificial Intelligence (AI) Use Cases For The Defense Sector.(n.d.). Qualetics. <https://qualetics.com/artificial-intelligence-ai-use-cases-for-the-defense-sector/>
- [5] BUCHMEISTER, B., PALCIC, I. & OJSTERSEK, R. (2019). ARTIFICIAL INTELLIGENCE IN MANUFACTURING COMPANIES AND BROADER: AN OVERVIEW. DAAAM INTERNATIONAL SCIENTIFIC https://www.daaam.info/Downloads/Pdfs/science_books_pdfs/2019/Sc_Book_2019-007.pdf
- [6] Puitinen, M. (2018). Nordcloud: 10 examples of AI in manufacturing to inspire your smart factory. <https://nordcloud.com/10-examples-of-ai-inmanufacturing-to-inspire-your-smart-factory/>.
- [7] Arinez, J. F., Chang, Q., Gao, R. X., Xu, C., and Zhang, J. (August 13, 2020). "Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook." *ASME. J. Manuf. Sci. Eng.* November 2020; 142(11): 110804. <https://doi.org/10.1115/1.4047855>
- [8] Deloitte Survey on AI Adoption in Manufacturing(n.d.). Deloitte. <https://www2.deloitte.com/cn/en/pages/consumer-industrial-products/articles/ai-manufacturing-application-survey.html>
- [9] Lei, Z. and Wang, L. (2020), "Construction of organisational system of enterprise knowledge management networking module based on artificial intelligence", *Knowledge Management Research and Practice*. doi: 10.1080/14778238.2020.1831892.
- [10] Jallow, H., Renukappa, S., & Suresh, S. (2020). Knowledge management and artificial intelligence (AI). Kidmore End: Academic Conferences International Limited. doi:<https://doi.org/10.34190/EKM.20.197>
- [11] Wisniewski, H. S. (2020). WHAT IS THE BUSINESS WITH AI? PREPARING FUTURE DECISION MAKERS AND LEADERS. *Technology and Innovation*, 21(4), 1-14. doi:<https://doi.org/10.21300/21.4.2020.4>
- [12] Bourne, A. (2019). What's ahead for manufacturing AI. *Industry Week*, <https://www.proquest.com/trade-journals/what-s-ahead-manufacturing-ai/docview/2222885582/se-2>
- [13] Kim, T. W., Fabrizio, M., Katherina, P., Sison, A. J., & Benito, T. (2021). Master and slave: The dialectic of human-artificial intelligence engagement. *Humanistic Management Journal*, 6(3), 355-371. doi: <https://doi.org/10.1007/s41463-021-00118-w>
- [14] Main,K. Bottorff,C. (2022) Small Business Statistics Of 2023. *Forbes Advisor*. <https://www.forbes.com/advisor/business/small-business-statistics/#:~:text=Nearly%20half%20of%20all%20U.S.,even%20have%20employees%20at%20all>
- [15] AI maturity assessment (n.d.). AI sweden. [HTTPS://WWW.AISE/EN/AI-MATURITY-MATURITY%20ASSESSMENT%20IS,A%20MORE%20EFFECTIVE%20AI%20PROGRAM](https://www.aise/en/ai-maturity-maturity%20assessment%20is,a%20more%20effective%20ai%20program).
- [16] The performance of Small and Medium Sized Businesses in a digital world (2019). Deloitte. A report for the Connected Commerce Council <https://www2.deloitte.com/content/dam/Deloitte/es/Documents/Consultoria/The-performance-of-SMBs-in-digital-world.pdf>
- [17] Chierchia, G (2020). Our sustainability frame focus areas – net zero, improving people's lives and caring for our planet. *linkedin*. <https://www.linkedin.com/pulse/our-sustainability-frame-focus-areas-net-zero-peoples-chierchia/>
- [18] Suarez, B., & Caetano, I. (2021). Impact assessment strategy: Materiality, AI & blockchain, and maturity models. Manchester: The International Society for Professional Innovation Management (ISPIM). <https://www.proquest.com/conference-papers-proceedings/impact-assessment-strategy-materiality-ai-amp/docview/2561106811/se-2>
- [19] Gajdzik, B.(2022). Frameworks of the Maturity Model for Industry 4.0 with Assessment of Maturity Levels on the Example of the Segment of Steel Enterprises in Poland, *Journal of Open Innovation:*

- Technology, Market, and Complexity. Volume 8, Issue 2, 77, ISSN 2199-8531, <https://doi.org/10.3390/joitmc8020077>
- [20] Resiliency, the edge, and the future of AI: A conversation with SAS CTO bryan harris. (2022). Cio, <https://www.proquest.com/trade-journals/resiliency-edge-future-ai-conversation-with-sas/docview/2733942446/se-2>
- [21] Hu, L. (2022). Inside AI Maturity Model. Five steps to transform with data-centric AI engineering. Towards Data Science. <https://towardsdatascience.com/inside-ai-maturity-model-3ff645a484b3>
- [22] The AI Maturity Framework. A strategic guide to operationalize and scale enterprise AI solutions (n.d.). Element AI. https://s3.amazonaws.com/external_clips/3430107/AI-Maturity-Framework_White_Paper_EN.pdf?1589551996#:~:text=The%20AI%20Maturity%20Framework%20is,%2C%20Technology%2C%20People%20and%20Governance.
- [23] Kleinings,H.(2023). AI for customer support and why you need it. Levity. <https://levity.ai/blog/ai-for-customer-support#:~:text=AI%2Dpowered%20customer%20support%20enables,even%20the%20generation%20of%20revenue>
- [24] Thiere,A. (2023). Getting AI Innovation Culture Right. Policy studies. Technology and innovation. R street. <https://www.rstreet.org/research/getting-ai-innovation-culture-right/>
- [25] Impact of AI in Manufacturing- Improved Quality and Efficiency (2022). SAXON. <https://saxon.ai/blogs/impact-of-ai-in-manufacturing-improved-quality-and-efficiency/#:~:text=Optimizing%20Production%20Processes%3A&text=Through%20machine%20learning%20algorithms%2C%20manufacturers,suggest%20changes%20to%20improve%20throughput>
- [26] Hu, J. Gao, S. (2019).Research and Application of Capability Maturity Model for Chinese Intelligent Manufacturing, Procedia CIRP, Volume 83, Pages 794-799, ISSN 2212-8271, <https://doi.org/10.1016/j.procir.2019.05.013>.
- [27] Bozic Yams, Nina & Richardson, Valerie & Shubina, Galina & Albrecht, Sandor & Gillblad, Daniel. (2020). Integrated AI and Innovation Management: The Beginning of a Beautiful Friendship. Technology Innovation Management Review. 10. 5-18. 10.22215/timreview/1399.
- [28] Ansari, I., Barati, M., Sadeghi Moghadam, M. R., & Ghobakhloo, M. (2023). An industry 4.0 readiness model for new technology exploitation. The International Journal of Quality & Reliability Management, 40(10), 2519-2538. doi:<https://doi.org/10.1108/IJQRM-11-2022-0331>
- [29] Chen, W., Liu, C., Xing, F., Peng, G. and Yang, X. (2022), "Establishment of a maturity model to assess the development of industrial AI in smart manufacturing", Journal of Enterprise Information Management, Vol. 35 No. 3, pp. 701-728. <https://doi.org/10.1108/JEIM-10-2020-0397>.
- [30] Paschou, Theoni & Rapaccini, Mario & Peters, Christoph & Adrodegari, Federico & Saccani, Nicola. Developing a Maturity Model for Digital Servitization in Manufacturing Firms. 2019
- [31] Xu, X. Lu, Y. <https://saxon.ai/blogs/impact-of-ai-in-manufacturing-improved-quality-and-efficiency/#:~:text=Optimizing%20Production%20Processes%3A&text=Through%20machine%20learning%20algorithms%2C%20manufacturers,suggest%20changes%20to%20improve%20throughput>.
- [32] Lempriere, M. (2022). One in three SMEs investing in EV infrastructure in the next year, says NatWest. Current Briefings.[Online]. Available: <https://www.current-news.co.uk/one-in-three-smes-investing-in-ev-infrastructure-in-the-next-year-says-natwest/>.
- [33] Prock, J. (2022). HOW DIGITAL TRANSFORMATION HELPS EV COMPANIES ACCELERATE PRODUCT INNOVATION. Arena.[Online] Available: <https://www.arenasolutions.com/blog/how-digita-transformation-helps-ev-companies-accelerate-product-innovation/>.
- [34] Contento, M. (2022). 5G and IoT: Emerging Technology with Endless Use Cases. Telit. <https://www.telit.com/blog/state-of-5g-and-iot-current-future-applications/>
- [35] The future of industrial IoT calls for new capabilities, architectures and devices. (2022). RCRWireless News. [Online] Available <https://www.rcrwireless.com/20220308/5g/the-future-of-industrial-iot-calls-for-new-capabilities-architectures-devices>.
- [36] Schulz, P., Matthe, M., Klessig, H., Simsek, M., Fettweis, G., Ansari, J., et al. (2017). Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture. IEEE Communications Magazine, 5(2), 70–78.
- [37] Rehan. (2022). The Future of Artificial Intelligence in Manufacturing Industries. AmyGB.ai. <https://www.amygb.ai/blog/future-of-artificial-intelligence-in-manufacturing-industries>

- [38] Wang, J., Lu, Y., Fan, S., Hu, P., & Wang, B. (2022). How to survive in the age of artificial intelligence? exploring the intelligent transformations of SMEs in central china. *International Journal of Emerging Markets*, 17(4), 1143-1162. doi:<https://doi.org/10.1108/IJOEM-06-2021-0985>
- [39] Kamlani, A. (2022). How hybrid AI improves edge computing, user experience. *Control Engineering*, 69(9), 31-32. <https://www.proquest.com/trade-journals/how-hybrid-ai-improves-edge-computing-user/docview/2729118553/se-2>
- [40] Benaich, N. and Hogarth, I. (2020), *The State of AI Report 2020*, p. 177, available at: <https://www.stateof.ai/>.
- [41] Katz, Y, *Manufacturing an Artificial Intelligence Revolution* (November 27, 2017). Available at SSRN: <https://ssrn.com/abstract=3078224> or <http://dx.doi.org/10.2139/ssrn.3078224>
- [42] Koch, C. (2017). We'll need bigger brains. *The Wall Street Journal*, page C1.
- [43] Badrick, C. (2018). Factories are transforming global industry. *Turn Key Technologies*. <https://www.turn-keytechnologies.com/blog/article/how-todays-cutting-edge-smart-factories-are-transforming-global-industry/>
- [44] Jardim-Goncalves, R., Romero, D. and Grilo, A., 2017. Factories of the future: Challenges and leading innovations in intelligent manufacturing. *Int. J. Comput. Integr. Manuf.*, 2017, vol. 30, no. 1, pp. 1–3. doi:[abs/10.1080/0951192X](https://doi.org/10.1080/0951192X).
- [45] Rajput, M.K. (2022). *The Future of Artificial Intelligence in Manufacturing Industries*. *Yourstory*. <https://yourstory.com/mystory/future-artificial-intelligence-manufacturing-industries>
- [46] Andersen, J.R., Chui, M., Østergaard, H. and Rugholm, J. (2019), *How Artificial Intelligence Will Transform Nordic Businesses*, McKinsey & Company, p. 55, available at: <https://www.mckinsey.com/featured-insights/artificial-intelligence/how-artificial-intelligence-will-transform-nordic-businesses>.
- [47] Harris, A.(2022). *AI in Manufacturing: How It's Used and Why It's Important for Future Factories*. *Redshift by Autodesk*.<https://redshift.autodesk.com/articles/ai-in-manufacturing>
- [48] Artificial intelligence. (2022). In *Wikipedia*. https://en.wikipedia.org/wiki/Artificial_intelligence
<https://qualetics.com/artificial-intelligence-ai-use-cases-for-the-defense-sector/>

AUTHOR

Dharmender Salian is an IT Professional based out of New York.



Gamified Web Application for Facilitating Zero Carbon Activities by Local Government

Aoi Nagatani¹, Tasuku Watanabe¹, Yuya Tarutani¹, Yoshifumi Kamae¹, Shun Sato¹, Marin Shoda¹, and Masahide Nakamura²

¹Graduate School of System Informatics, Kobe University Rokkodai-cho 1-1, Nada-ku, Kobe, Hyogo, 657-8501 Japan

²the Center of Mathematical and Data Science, Kobe Univ.

Abstract. In recent years, Japan has been actively pursuing the realization of zero carbon cities. However, significant challenges persist, including a lack of effective methods for local governments to communicate zero carbon initiatives to their citizens. This has resulted in limited awareness among citizens about how to participate in zero carbon initiatives. To address these issues, the authors develop a gamified application aimed at promoting zero carbon activities in this research. Through a case study conducted in Sanda City, Hyogo Prefecture in Japan, the authors report the progress of its social implementation.

Keywords: Zero carbon, zero carbon city, gamification, web application, local government.

1 Introduction

The concentration of CO₂ in the atmosphere has been increasing since the Industrial Revolution, and global warming is progressing. Various effects of global warming, such as rising temperatures, rising sea levels and the impact on crops, have begun to be observed, and countermeasures against global warming have become an urgent issue on a global scale.

Therefore, in recent years, *zero carbon city* initiatives, in which local governments aim to achieve zero carbon emissions as a countermeasure against global warming, have been expanding in Japan. Efforts to achieve zero carbon city include the substitution of renewable energy sources, energy conservation and a shift to low-carbon transport[1], in addition to these approaches, the participation of citizens is essential.

However, the following problems exist between local governments and citizens with regard to the zero carbon activities.

P1: There is no established method for local governments to disseminate information to citizens on how to initiate zero carbon

P2: Citizens are not clear on how to get involved in zero carbon and activities are not widespread

Therefore, measures are needed to solve these problems.

In order to solve the above problems, this research aims to establish a method for local governments aiming to achieve zero carbon cities to spread the zero carbon activities among citizens and to create opportunities for citizens to engage in zero carbon initiatives. As a key idea to achieve this purpose, the authors develop a web application that introduces *gamification*[2]. Specifically, the feature implementation is based on the following approach.

A1: Realization of the ability of local governments to provide zero carbon information to citizens

A2: Gamification to make it easier for citizens to engage in zero carbon initiatives

First, A1 aims to establish a method for local governments to provide information to promote the zero carbon activities, and proposes features such as F1: Administrator and F2: Article in a web application.

Next, A2 proposes a feature that lowers the barrier for citizens to engage in zero carbon initiatives and allows them to continue to do so. Specifically, the authors propose the following features: F3: Login, F4: Mission, F5: Quiz, F6: Level, F7: Map, and F8: Visualization of zero carbon initiatives.

As a case study, these approaches actually was implemented in Sanda City, Hyogo Prefecture in Japan. Specifically, the authors implemented the *Sanda Zero Carbon Challenge*, a web application that promotes the zero carbon activities among Sanda citizens. The authors also used the implemented Web application to exhibit at events and conduct demonstration tests at the city hall, and based on the user feedback got, the authors examined the effectiveness of the Web application and improve its features.

2 Preliminaries

2.1 Zero Carbon Cities and Their Challenges

In recent years, *zero carbon city* initiatives, in which local governments aim to achieve zero carbon emissions as a countermeasure against global warming, have been expanding in Japan. *Zero carbon* means that CO₂ emissions from businesses and households are reduced to a level equal to or less than the amount of CO₂ absorbed by forests[3]. As the Government of Japan aims to achieve carbon neutrality in 2050, local governments are also mainly aiming to achieve zero carbon cities by 2050[4].

Efforts such as transitioning to renewable energy, energy conservation, and the shift to low-carbon transportation[1] are being made to realize zero carbon cities. However, in addition to these approaches, citizens participation is essential. Due to the nature of projects conducted by local governments, they are closer to citizens and more likely to involve them than projects led by the national government. Therefore, it is important for local governments to encourage citizens to work on zero carbon.

However, the following problems exist between local governments and citizens regarding the zero carbon activities.

P1: There is no established method for local governments to disseminate information to citizens on how to initiate zero carbon

P2: Citizens are not clear on how to get involved in zero carbon and activities are not widespread

Because of these problems, local governments are working towards achieving zero carbon cities, although citizens lack information on how to engage in zero carbon initiatives and find it difficult and intimidating to participate. Therefore, measures are needed to solve these problems.

2.2 Web-based Data Collection

In order to solve the problems between local governments and citizens described in Section 2.1, the authors consider the use of the Web. By using the Web, local governments can provide citizens with information on zero carbon initiatives. In addition, citizens can collect information on zero carbon initiatives from anywhere.

By implementing an appropriate web application, local governments can continuously update their zero carbon contents. Furthermore, by recording the actions of citizens regarding zero carbon, local governments can understand the efforts of citizens regarding zero carbon and utilize the data. For example, by analyzing citizens' zero carbon behavior, local governments can determine what kind of initiatives they should take to promote zero carbon.

In addition, the analyzed and tabulated data can be visualized and fed back to citizens in an easy-to-understand form, enabling them to recognize the current status of the zero carbon activities as a whole and to evaluate its effectiveness.

2.3 Gamification

In order to encourage more citizens to use the Web applications described in Section 2.2, the authors consider the use of *gamification*[2]. Gamification is a method of turning a service into a game by introducing game-like elements, which are expected to stimulate users' curiosity, activate their behavior, and bring positive benefits to them.

In the context of related research that incorporates gamification into web applications connecting local governments and citizens, the example of the *Sanda Machiaruki App*[5] [6] is illustrated. Sanda Machiaruki App is an application for promoting town walking among tourists in Sanda City, Hyogo Prefecture in Japan. The features implemented as an introduction to gamification include a feature to receive points for visiting specific spots in Sanda City using GPS, a present feature using points, and a quiz feature about the spots.

Numerous related research, including the Sanda Machiaruki App, have shown that gamification introduced into Web applications is effective in encouraging user behavior.

3 Proposed Method

3.1 Purpose and Key Ideas

In order to solve the issues in Section 2.1, this research aims to establish a method for local governments aiming to achieve zero carbon cities to spread the zero carbon activities among citizens, and to create an opportunity for citizens to engage in zero carbon activities. As a key idea to achieve this purpose, the authors develop a web application that introduces gamification as described in Section 2.3.

3.2 System Overview

The requirements for a Web application to achieve the purpose of this research as described in Section 3.1 include the following.

R1: Ability to present zero carbon information from local governments to citizens

R2: There must be an easy mechanism for citizens to engage in zero carbon initiatives

Therefore, the authors implement the features based on the following approach in this research.

A1: Realization of the ability of local governments to provide zero carbon information to citizens

A2: Gamification to make it easier for citizens to engage in zero carbon initiatives

By developing the Web application based on these approaches, local governments can easily encourage citizens to participate in zero carbon activities.

3.3 A1: Realization of the Ability of Local Governments to Provide Zero Carbon Information to Citizens

In this approach, the authors aim to establish a method for local governments to provide information to promote the zero carbon activities, and propose features such as F1: Administrator and F2: Article in the web application.

F1: Administrator As described later, this application has various features for providing zero carbon information from local governments to citizens. In order to make appropriate approaches to citizens, it is necessary to have an administrator feature to post, edit, and delete information. The administrator feature should be restricted to only local government officials.

F2: Article This is a feature for local governments to disseminate information on zero carbon to users. By linking to zero carbon contents created by local governments and various useful contents published on the Web, information can be provided to users. In addition, by displaying the title, summary, and image of the linked page, the application can be used as an entry point to create an opportunity for users to get more information on zero carbon.

3.4 A2: Gamification to Make It Easier for Citizens to Engage in Zero Carbon Initiatives

In this approach, the authors propose a feature that lowers the barrier for citizens to engage in zero carbon initiatives and allows them to continue to do so. Specifically, the authors propose the following features: F3: Login, F4: Mission, F5: Quiz, F6: Level, F7: Map, and F8: Visualization of zero carbon initiatives.

F3: Login This is a feature for identifying users who use the application. Users can use the application by entering the required information and logging in. This is a necessary feature for getting information on each user's zero carbon actions and providing various gamification features.

F4: Mission This is a feature that presents missions for citizens to undertake in order to achieve zero carbon cities and encourages users to undertake them in a game-like manner. The purpose of this feature is to create an opportunity for citizens to engage in zero carbon activities by showing them specific measures to participate in zero carbon activities and encouraging them to continue to do so. To this end, the feature presents the significance of the mission, the amount of CO₂ reduction and the amount of money saved associated with the actions related to the mission, and awards points that increase the level described later when the mission is achieved, aiming to improve users' motivation for zero carbon activities. The achievement status of the mission is aggregated for each individual and used to calculate the level described later and visualize the initiatives.

F5: Quiz This is a feature that presents quizzes to test users' knowledge of zero carbon. By answering the quiz correctly, users can get points that increase their level. The purpose of this feature is to learn about zero carbon while maintaining users' motivation by making it a game. Quizzes are presented in different levels of difficulty, allowing users to learn step by step.

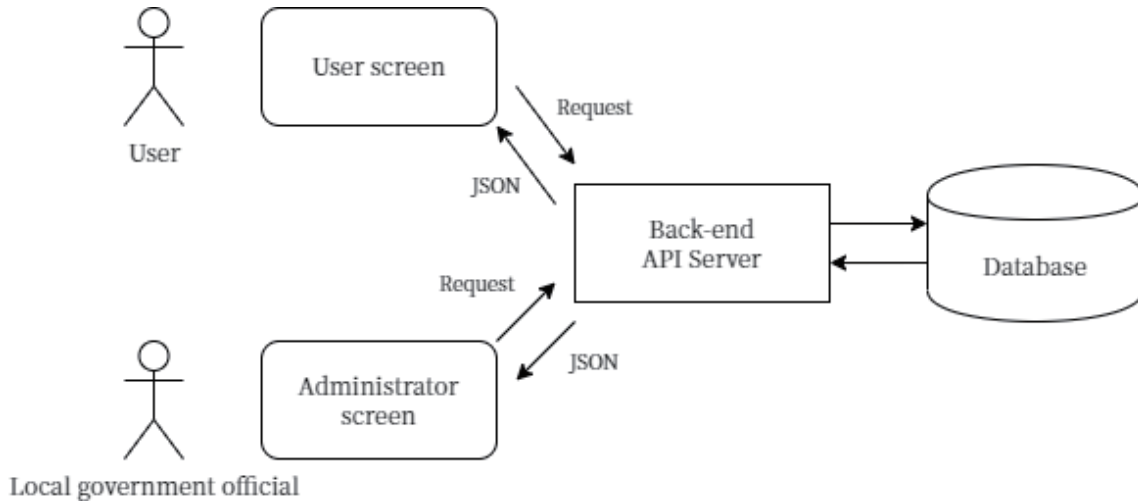


Fig. 1. Architecture of the proposed application

F6: Level This is a feature for visualizing users' zero carbon initiatives. The user's level is calculated based on the user's mission achievement status and the number of correct answers to the quiz. The user's level is used not only to visualize the user's initiatives, but also to expand the viewing range of the map described later according to the user's initiatives, contributing to the improvement of the user's motivation.

F7: Map This is a feature that displays a map that is opened according to the value of the level described above in order to visualize the user's zero carbon initiatives. Users can enjoy game elements such as exploring the map according to their level of zero carbon initiatives, contributing to the improvement of their motivation. By allowing users to explore the map of the local government that provides the service, it is possible to attract users' interest and contribute to a better understanding of the area in which they live.

F8: Visualization of Zero Carbon Initiatives This is a feature that visualizes the amount of CO₂ reduction and the amount of money saved per user aggregated by the mission feature described in Section 3.4. Users can check the amount of CO₂ reduction and the amount of money saved per week for each user, and the purpose is to maintain their motivation by knowing the specific contribution to zero carbon. In addition, by visualizing the amount of CO₂ reduction and the amount of money saved for all users, it is possible to increase users' awareness of their contribution to zero carbon city.

3.5 Architecture

The overall architecture of the proposed application is shown in Figure 1. This Web application is assumed to be operated by local government officials for a long period of time. Therefore, the authors design the database so that the missions, quizzes, and articles, which are elements of gamification, can be changed dynamically.

4 Case Study: Sanda Zero Carbon Challenge

4.1 Implementation

Sanda City, Hyogo Prefecture in Japan, has declared that it will realize zero carbon city status by 2050. Therefore, as a case study, this research implements the proposed method

described in Section 3 for realizing the zero carbon city of Sanda City. Specifically, the authors implement a web application called *Sanda Zero Carbon Challenge* to encourage Sanda citizens to engage in zero carbon activities.

The programming languages and technologies used for implementation are shown below.

- Backend
 - Java
 - SpringBoot
 - MySQL
- Frontend
 - TypeScript
 - React

The authors use these programming languages and technologies to implement the case study in Sanda City.

Prior to the explanation in the next section, the home screen that is displayed when a user logs into this application in the Sanda Zero Carbon Challenge is shown in Figure 2. The menu bar at the top of the home screen exists on each screen, and users can navigate to each feature screen by pressing the buttons on the menu bar.

From here, implementation of each feature is explained.

F1: Implementation of Administrator Feature Administrators can view user information, view, create and edit missions, quizzes, articles and tags on the administrator screen. The tag feature allows users to intuitively add icons to the user screen by setting tags for missions, quizzes, and articles.

F2: Implementation of Article Feature Users can view the article screen by pressing the “Information” button on the top menu bar of the home screen shown in Figure 2. Figure 3 shows the article screen. Administrators can set articles on zero carbon that they want citizens to read, and by clicking on the article, users can access external sites. In addition, administrators can set articles that they want citizens to read in particular as “Important notice” in the administrator screen described in Section 4.1, and display them at the top of the screen. The important notice is assumed to be used to provide information that requires citizens participation, such as announcements of zero carbon events in Sanda City.

F3: Implementation of Login Feature Users can log in to this application by creating an account with a username and password. The actions taken in this application, such as completed missions and quiz answers, are recorded in association with this username.

F4: Implementation of Mission Feature Users can view the mission screen by pressing the “MISSION” button on the screen that transitions to the mission screen by pressing the “Play” button on the top menu bar of the home screen shown in Figure 2. Figure 4 shows the mission screen. The contents of the mission is to encourage users to take actions that can easily reduce CO₂. Six missions per day are displayed in three levels of difficulty, and users can get points by practicing the actions presented and achieving the missions. In addition, administrators can set the points, the amount of CO₂ reduction, and the amount of money saved that can be got when a mission is achieved in the administrator screen



Fig. 2. Sanda Zero Carbon Challenge User Screen: Home Screen

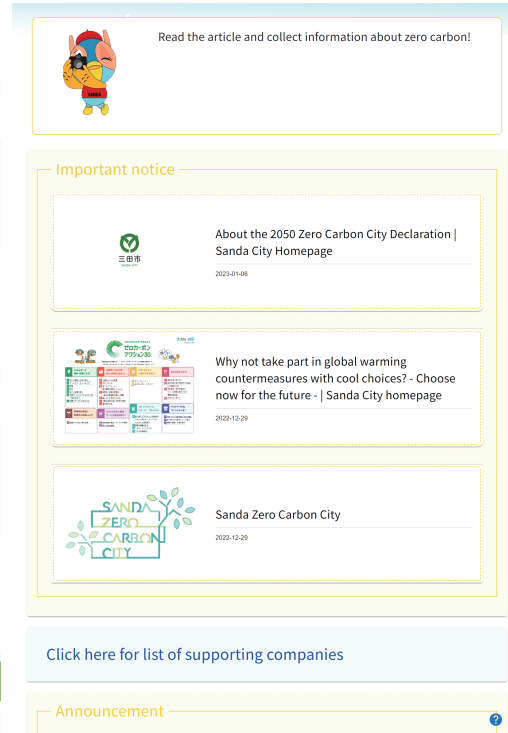


Fig. 3. Sanda Zero Carbon Challenge User Screen: Article Screen

described in Section 4.1, so that users can check them before achieving the mission. In this research, missions are created based on the information on “Zero Carbon Action30” [7] and “Energy saving at home” [8] published by the Japanese Ministry of the Environment and the Ministry of Economy, Trade and Industry, and their effects are defined.

F5: Implementation of Quiz Feature Users can view the quiz screen by pressing the “QUIZ” button on the screen that transitions to the quiz screen by pressing the “Play” button on the top menu bar of the home screen shown in Figure 2. Figure 5 shows the quiz screen. The contents of the quiz is assumed to be questions on zero carbon and environmental issues. Three quizzes per day are displayed in three levels of difficulty, and users can get points by answering them correctly. In this research, quizzes are created based on the information on “Challenge! Global warming quiz” [9] published by the Japanese Ministry of the Environment.

F6: Implementation of Level Feature Users can increase their level by completing missions and answering quizzes correctly. As with the home screen shown in Figure 2, the level is always displayed in the upper right corner of each screen. Below the level is the progress to the next level. In addition, the number of points required to increase the level increases as the level increases. By increasing the level, users can clear the clouds on the map of Sanda City on the map screen.

F7: Implementation of Map Feature Users can view the map screen in the center of the home screen shown in Figure 2 by pressing the “Home” button. The map is created based on the actual map of Sanda City. By pressing the “Back/Advance” button, users

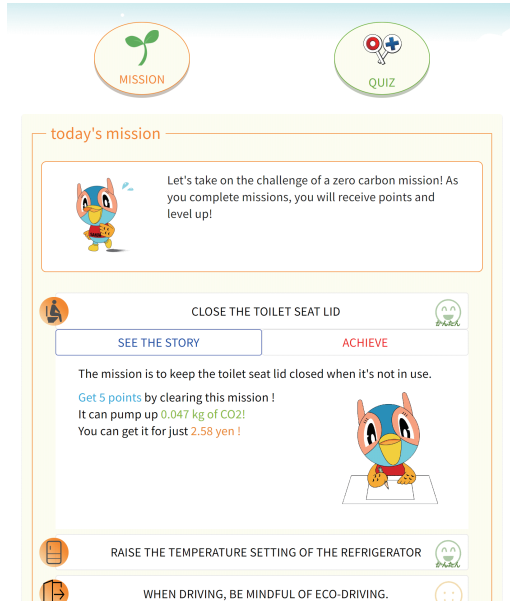


Fig. 4. Sanda Zero Carbon Challenge User Screen: Mission Screen

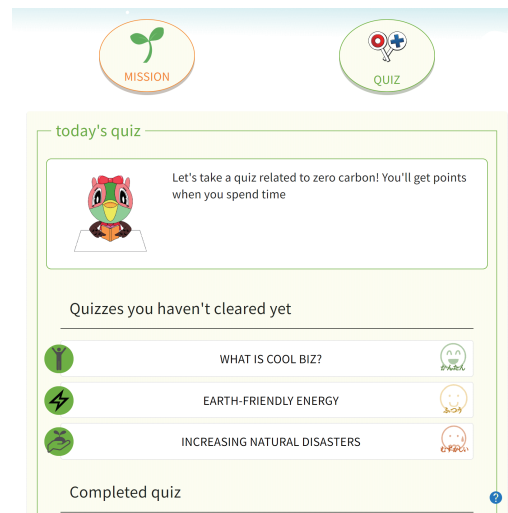


Fig. 5. Sanda Zero Carbon Challenge User Screen: Quiz Screen

can move the position of the character on the screen, and by pressing the magnifying glass button, they can check the famous places at the current position of the character. Figure 6 shows an example of a famous place displayed when the magnifying glass button is pressed. Users can clear the clouds on the map of Sanda City and move to a new location by increasing their level.

F8: Implementation of Visualization Feature of Zero Carbon Initiatives Users can view the look back screen by pressing the “Lookback” button on the top menu bar of the home screen shown in Figure 2. Figure 7 shows the look back screen. Users can check the missions achieved and the graphs of the points, the amount of CO₂ reduction, and the amount of money saved per week. In addition, the total amount of CO₂ reduced by the mission achievement of all users is displayed at the bottom left of the home screen shown in Figure 2. By pressing the corresponding part, the display period can be switched between “whole period” and “this week”.

4.2 Demonstration Experiment

The authors conducted two demonstration experiments by exhibiting at the zero carbon event in Sanda City and by local government officials to confirm the operation of Sanda Zero Carbon Challenge in the production environment and to evaluate each feature from the user’s perspective. Based on the feedback got through each demonstration experiment, the authors discussed the effectiveness of the application and implemented improvements to the features and additional features.

Experimental Procedure The experiment procedure of the two demonstration experiments described in Section 4.2 is explained below.

First, the participation in the zero carbon event is explained. In this experiment, the authors participated in the Sanda Zero Carbon City Forum held in Sanda City[10] and

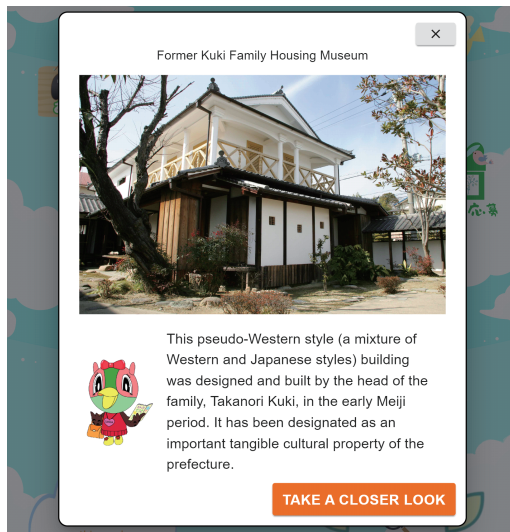


Fig. 6. Sanda Zero Carbon Challenge User Screen: Example of a famous place on the map



Fig. 7. Sanda Zero Carbon Challenge User Screen: Look back Screen

exhibited Sanda Zero Carbon Challenge. Sanda citizens actually operated the application, tested the operation of the application, and collected feedback from users.

Next, the demonstration experiment by Sanda City officials is explained. The authors conducted two demonstration experiments by local government officials using the application under development. For about two weeks, the authors actually asked the officials to use the application and pointed out improvements.

Experimental Results The results of the demonstration experiment got by the experimental method described in Section 4.2 are explained below.

Table 1 shows the appreciated points and problems of Sanda Zero Carbon Challenge from the feedback got from the demonstration experiment. The points appreciated in the demonstration experiment are the design of the application. It is considered that the design is easy to use and friendly because the main target is elementary school students. In addition, it is considered that the concept and purpose of the application are sufficiently conveyed from the feedback such as “easy to use” and “can learn about environmental issues in a fun way.” Next, the points to be improved are the usability of the application. Table 2 shows the improvements made to each problem based on the feedback.

5 Discussion

5.1 Technical Issues

The application developed in this research has the following technical issues.

First, there is a data collection issue. In this application, users register and log in with any username and password. This is to minimize the information that users have to enter

Table 1. Feedback got from the demonstration experiment

Appreciated points	Problem
It has a friendly design.	Difficult to understand how to create an account and log in.
Easy to use.	Difficulty in noticing level up and map release.
A fun way to learn about environmental issues.	Large amount of missions/quizzes.
Provides an opportunity to talk about environmental issues with the family.	Difficult to understand the controls when playing for the first time.
There was a lot of content of interest to children.	Many Chinese characters, some of which children cannot read.

Table 2. Improvements made to each problem based on the feedback

Problem	Solution
Difficult to understand how to create an account and log in.	Login screen, modification of items required for login.
Difficulty in noticing level up and map release.	Provide direction at level-up and map liberation.
Large amount of missions/quizzes.	Random daily missions/quizzes are displayed.
Difficult to understand the controls when playing for the first time.	Implement a tutorial function to explain how to operate the system.
Many Chinese characters, some of which children cannot read.	Convert complex kanji characters in a text into hiragana.

and reduce the burden on users. On the other hand, this specification narrows the scope of data analysis accumulated in the application. For example, it was initially envisioned that users would be asked to register their age in order to analyze user behavior by age group, although this was no longer possible because age registration was not required in order to lower the barrier to user registration.

Next, there is an operational issue. This application is developed by a team with programming knowledge, and local governments operate it by entering content. Therefore, the development team needs to respond to improvements in features other than content and the implementation of additional features each time, and if they cannot respond, the operation of the application may be delayed. In the case of this research, it is almost impossible to operate for a long time due to the difficulty of handover due to graduation of students, personnel changes in Sanda City, and lack of personnel.

5.2 Comparison with Existing Solutions

The application developed in this research has the following features compared to existing solutions.

First, the difference in the scale of zero carbon initiatives can be mentioned. As existing initiatives for local governments to realize zero carbon cities, the targets are often local governments themselves and companies, and there are many large-scale initiatives such as the establishment of ordinances and policies. There are also many initiatives to encourage citizens to switch to electric vehicles and other things that greatly change their lifestyles[11]. On the other hand, the application developed in this research assumes small-scale initiatives that citizens can easily work on. For example, it is easy to work on in daily life, such as using public transportation and changing the setting temperature of the air conditioner.

Next, the difference in the target of citizens who work to realize zero carbon cities can be mentioned. As mentioned above, the target of the existing initiatives for local governments to realize zero carbon cities is considered to be adults. On the other hand, the application developed in this research mainly targets elementary school students. This is because it is possible to give a chance for family and surrounding adults to work on zero carbon by getting elementary school students interested in it. In fact, this application uses designs and texts that are easy for elementary school students to use.

6 Conclusion

In this research, the authors considered the following problems in local governments aiming to realize zero carbon city. 1. There is no established method for local governments to disseminate zero carbon initiatives to citizens. 2. Citizens do not know how to work on zero carbon, and their activities do not spread. Therefore, in order to solve these problems, the authors proposed a web application development introducing gamification as a proposed method. In addition, by actually implementing the proposed method as a case study in Sanda City, Hyogo Prefecture in Japan, the authors were able to clarify its effectiveness and improvement points by exhibiting at events and conducting demonstration experiments in the city hall.

For future prospects, it is possible to publish Sanda Zero Carbon Challenge to citizens with the cooperation of Sanda City Hall and analyze its effects, and to improve the features.

Acknowledgment

This research was partially supported by JSPS KAKENHI Grant Numbers JP19H01138, JP20H05706, JP20H04014, JP20K11059, JP22H03699, JP19K02973, Young Scientists (No.23K17006) and the Center of Mathematical and Data Science in Kobe University Intra- and Extra-mural DX Promotion Joint Project.

The authors would like to thank Mr. Kenta Sakaguchi and Mr. Hiroki Kishimoto of the Smart City Promotion Division, and Ms. Shoko Terashima of the Environmental Creation Division of Sanda City Hall for their great cooperation in this project.

References

1. A. Zarba, A. Krzemińska, and J. Lach, “Energy sustainable cities. from eco villages, eco districts towards zero carbon cities,” in *E3S Web of Conferences*, vol. 22. EDP Sciences, 2017, p. 00199.
2. A. Nagatani, S. Chen, M. Nakamura, and S. Saiki, “Exploiting motivation subscales for gamification of lifelogging application,” *International Journal of Software Innovation (IJSI)*, vol. 10, p. 27, December 2022, doi: 10.4018/IJSI.313445.
3. L. Chen, G. Msigwa, M. Yang, A. I. Osman, S. Fawzy, D. W. Rooney, and P.-S. Yap, “Strategies to achieve a carbon neutral society: a review,” *Environmental Chemistry Letters*, vol. 20, no. 4, pp. 2277–2310, 2022.
4. K. SHIGE and O. N. Yoshiteru SAKAGUCHI, Takashi SAKAMAKI, “Proposal of ”local carbon cycle rate” for utilization in small municipalities declaring zerocarbon city,” in *Journal of Environmental Information Science, Vol. 35 (2021 Environmental Information Science Research Presentation Conference)*. Center for Environmental Information Science, Tokyo, 2021, pp. 149–154.
5. T. AKASHI, H. OURA, H. OZONO, T. NARIMATSU, R. YAMANA, T. NAKAI, and M. NAKAMURA, “Prototype of a quiz rally platform for motivating local understanding based on gamification,” in *IEICE Technical Report*, vol. 121, no. 229. Institute of Electronics, Information and Communication Engineers, November 2021, pp. 31–36, online.
6. T. AKASHI, H. OURA, T. NAKAI, and M. NAKAMURA, “Survey of citizens’ attitudes toward the community and analysis of their behavior using a quiz rally platform,” in *IEICE Technical Report*, vol. 121, no. 415. Institute of Electronics, Information and Communication Engineers, February 2022, pp. 41–47, online.
7. “Zero carbon action 30 | cool choice choose now for the future.” <https://ondankataisaku.env.go.jp/coolchoice/zc-action30/>, accessed on 14 September 2023.
8. “Agency for natural resources and energy,” https://www.enecho.meti.go.jp/category/saving_and_new/saving/general/howto/index.html, accessed on 14 September 2023.
9. “Challenge! global warming quiz — cool choice choose now for the future.” <https://ondankataisaku.env.go.jp/coolchoice/quiz/>, accessed on 14 September 2023.
10. “Saturday 21 january 2023: The sanda zero carbon city forum was held!” https://www.city.sanda.lg.jp/soshiki/41/gyomu/kankyo_hozen/energy/21422.html, accessed on 14 September 2023.
11. Chenmin He, Kejun Jiang, Sha Chen, Weiyi Jiang, Jia Liu, *Zero CO2 emissions for an ultra-large city by 2050: case study for Beijing, Current opinion in environmental sustainability*, **36**, 141–155 (2019), Elsevier.

Authors

Aoi Nagatani received the B.E. degree in Information and Intelligence Engineering from Kobe University, Japan, in 2022. He is currently enrolled in a master’s course at the Graduate School of Systems Informatics, Kobe University, Japan. His research interests include the Web service and gamification.

Masahide Nakamura received the B.E., M.E., and Ph.D. degrees in Information and Computer Sciences from Osaka University, Japan, in 1994, 1996, 1999, respectively. From 1999 to 2000, he has been a post-doctoral fellow in SITE at University of Ottawa, Canada. He joined Cybermedia Center at Osaka University from 2000 to 2002. From 2002 to 2007, he worked for the Graduate School of Information Science at Nara Institute of Science and Technology, Japan. From 2007 to 2022, he worked for the Graduate School of System Informatics at Kobe University. He is currently a full professor in the Center of Mathematical and Data Science Center at Kobe University. In 2015, he worked for Universite Grenoble Alpes as a visiting professor. In 2018, he joined RIKEN Center for Advanced Intelligence Project as a visiting researcher. His research interests include the service/cloud computing, smart home, smart city, and gerontechnology. He is a member of the IEEE, ACM, IEICE and IPSJ.

Laughing Out Loud – Exploring AI-Generated and Human-Generated Humor

Hayastan Avetisyan*, Parisa Safikhani*, and David Broneske

Department of Research Infrastructure and Methods, DZHW, Hannover, Germany

Abstract. In this study, we conduct a thorough comparative analysis between artificial intelligence (AI)-generated humor and human humor. The objective is to acquire a more profound understanding of AI's present capabilities in generating humorous text. We investigate the structural, sentiment, and linguistic patterns in jokes created by AI and humans, evaluating 'funniness' and 'originality' via a comprehensive annotation process. Our findings indicate that AI can produce humorous and occasionally novel content. Additionally, we employed the RoBERTa model for humor detection on a dataset consisting of 500 entries, including both human and AI-generated humor. This model demonstrated its proficiency in accurately categorizing a large dataset encompassing up to 200,000 entries with remarkable accuracy of up to 98%. Nonetheless, it lacks the emotional depth and originality commonly seen in human humor. The study underscores the challenge involved in developing AI models that can generate humor equivalent to human communication. Future research should focus on enhancing AI's ability to create humor and further examine AI's potential to adopt human humor strategies. Despite some limitations, this study contributes significantly to improving the humorous capabilities of AI models and the expandability of AI-generated humor.

Keywords: Artificial Intelligence (AI)-generated humor, Human humor, Linguistic patterns, Funniness evaluation, Originality evaluation, RoBERTa model.

1 Introduction

Humor, an inherent aspect of human communication, offers a powerful tool for establishing connections, lightening the atmosphere, and conveying intricate messages. The ubiquity of humor in our everyday lives contrasts sharply with the complexities it presents within the field of Natural Language Processing (NLP), especially when it comes to the generation of AI-based humor.

With the advent of sophisticated AI technologies, we have seen machines produce text that approximates human humor. However, the true measure of quality and originality in this machine-generated humor requires further comprehensive examination. Most existing research restricts its focus to certain types of joke structures, such as puns or knock-knock jokes, thus offering a limited perspective on the broader capabilities and limitations of AI in humor generation. Additionally, there is a noticeable gap in research providing a holistic view of the differences and similarities between AI and human humor.

Addressing these shortcomings, our study embarks on an exhaustive comparison of AI-generated humor with that generated by humans. Drawing from two distinct datasets – one composed by a state-of-the-art language model (AI-generated), and the other by humans – we scrutinize the structure, sentiment, and linguistic patterns present in both types of humor.

An integral part of our research is dedicated to pinpointing the differences and similarities between AI and human humor. This comparative analysis illuminates areas where AI falls short, providing insight into the current limitations and possible pathways for enhancement. Conversely, the identified similarities serve as evidence of AI's success in

emulating certain facets of human humor, a crucial step in the ongoing refinement of AI's humor generation abilities.

Further, we present a series of recommendations for future research based on our findings. As AI continues to advance and evolve, it becomes increasingly important to reassess and fine-tune its capabilities in humor generation. Our study suggests specific focus areas for future research, particularly the exploration of AI's potential to generate humor that is more original and emotionally resonant, striving towards a more human-like approach.

The contributions of this paper can be categorized into several key domains:

- **Comparing AI and human-generated humor:** Our comprehensive comparison of structure, sentiment, and linguistic patterns in AI and human-generated humor deepens the understanding of AI's current abilities in humor generation.
- **Evaluation of 'funniness' and 'originality':** We thoroughly evaluate 'funniness' and 'originality', offering new insights into AI's proficiency in humor generation.
- **Human-annotated dataset:** A unique aspect of our work is the creation of an annotated dataset, in which both AI and human-generated humor have been evaluated by human annotators for 'funniness' and 'originality'. This dataset, publicly available for further research, provides a valuable resource for studying and understanding the attributes of humor as perceived by humans.
- **Identification of differences and similarities:** By identifying key differences and similarities between AI and human humor, we provide a greater understanding of AI's current limitations and potential areas for improvement.
- **Recommendations for future research:** Our study provides valuable guidance for future research, focusing on the enhancement of AI's capability to generate humor that is both original and emotionally resonant.

In summary, this paper significantly contributes to the understanding and improvement of AI's ability to generate humor, while highlighting the intricate nature of humor generation. The introduction of our annotated dataset adds a tangible dimension to the study, allowing for further, detailed exploration into human perception of humor in comparison to AI-generated humor.

2 Related works

2.1 Automated humor detection

In a recent study [1], researchers performed a comparative evaluation of several machine learning (e.g., logistic regression, decision tree, random forests, passive-aggressive classifier) and deep learning (e.g., CNN, LSTM) methods to classify tweets as either humorous or non-humorous. They used a Kaggle dataset comprising both types of tweets. The findings from this study indicated that deep learning techniques delivered superior accuracy in humor prediction when compared to traditional machine learning approaches. [2] introduced a deep learning-based humor detection technique that combined CNN and LSTM layers. This approach addressed CNN's contextual limitations using LSTMs and incorporated dropout layers to reduce overfitting. When tested on the Yelp user review dataset, the model outperformed other methods, including SGD, SVM, and XGBoost, in precision, recall, and F1-measure. The research hints at the technique's broader applications, even in areas like detecting psychopathic behavior on social platforms.

In another study[3], machine learning models were developed to identify and score humor and offense in text. Using a dataset of 8,000 sentences, the study compared BERT-based models (BERTBASE, DistillBERT, and RoBERTa) for high-performance humor

detection. DistillBERT performed best for humor detection and rating, RoBERTa excelled in controversial detection, and BERTBASE outperformed in offensiveness ranking.

A recent study by [4] reviewed the field of computational humor recognition. They examined 106 papers from various databases and analyzed datasets, features, and algorithms used in the field. The study found numerous annotated humor datasets and identified 21 frequently studied humor features. The researchers observed that deep learning and supervised learning techniques, particularly Support Vector Machine and Long Short-Term Memory Networks, were commonly used for humor classification. BERT was the most popular pre-trained language model in this context. Future research directions and challenges in humor detection were also discussed.

2.2 AI- or human-generated text

In a recent study [5], researchers examined ChatGPT-generated online reviews. They compared human and ChatGPT content using a Transformer-based model with SHAP for explainability. Results showed a 79% accuracy rate, and they observed that ChatGPT tends to produce polite, vague, and impersonal text with unique vocabulary and minimal emotional expression.

Another paper [6] discusses the challenges of detecting AI-generated text and the potential risks of unregulated use of large language models (LLMs). The authors argue that current detectors are unreliable due to paraphrasing attacks and detector limitations. They also demonstrate the vulnerability of watermarked LLMs to spoofing attacks. The paper emphasizes the need for secure methods to prevent LLM misuse and the risks associated with misidentification by AI text detectors.

[7] present DetectGPT, a zero-shot machine text detection technique using the negative curvature regions of a language model's log probability function. It outperforms others without needing a separate classifier or dataset, using log probabilities and random perturbation for sample detection. Limitations include its white-box assumption, reliance on a sound perturbation function, and high computational needs. Future research could focus on watermarking, model ensembles, prompt-detection relationship, and applying this method to other domains.

The research by [8] explores distinguishing AI-generated text from human-written ones. It argues that detection is feasible unless both text distributions match. It presents a sample complexity bound for detection, stating more human-like AI text requires more samples. The study confirms better detectors are achievable, with more samples and robust watermarking techniques improving detection even amidst paraphrasing attacks. It highlights the need for further research for effective and fair AI text detectors.

[9] focuses on the differences between AI-generated and human-written scientific texts, exploring the potential limitations and challenges of using AI-based writing assistants in scientific writing. The researchers collected and analyzed scientific text from OpenAI API using optimized prompts for structured scientific abstracts. They conducted human evaluations to assess the ability to distinguish between AI-generated and human-written texts and devised a feature description framework to analyze differences in syntax, semantics, and pragmatics. Using logistic regression models and fine-tuned RoBERTa large OpenAI detectors, they found that AI-generated scientific texts distinguish from human-written texts, often lacking valuable insights and showing low external inconsistency with actual scientific knowledge. While AI-generated text may eventually become more syntactically similar to human-written text, future research should focus on improving the semantics and pragmatics of AI-generated texts to enhance human-AI collaboration in the research process. [10] examines the threat models posed by contemporary natural language generation (NLG)

systems and reviews the most complete set of machine-generated text detection methods available. With powerful open-source generative models becoming increasingly accessible, detecting machine-generated text is crucial in mitigating the potential for abuse. However, detecting machine-generated text presents several technical challenges and open problems. The paper provides a comprehensive analysis of the threat models posed by contemporary NLG systems and emphasizes the urgent need for improved defenses against the abuse of NLG models. The paper also highlights the importance of coordinated efforts across technical and social domains to achieve practical solutions.

Based on the review of related work, it becomes evident that while there have been numerous studies focusing on humor detection in text and distinguishing between AI-generated and human-generated text, no specific research has yet been undertaken to concentrate solely on the detection of humor generated by AI and contrasting it with human-generated humor. This existing gap in the literature motivates our current study.

3 Methodology

This chapter provides a comprehensive overview of the methodology adopted in our research to analyze and compare AI and human-generated humor effectively. It elaborates on three key aspects: the data set utilized, the manual annotation of the dataset, and the experiments conducted.

Our methodology’s ultimate goal is to support the findings of this study through a transparent, replicable, and robust approach that not only substantiates our research outcomes but also serves as a guideline for future research in this intriguing domain.

3.1 Dataset

AI-generated dataset. We utilized ChatGPT (GPT-4) to create an AI-generated dataset for humor detection, which comprised both humorous and non-humorous entries. We directed the model with the prompt: "Provide a balanced dataset of 500 entries for humor detection." Recognizing ChatGPT’s limitation in generating 500 entries in a single go, we intermittently prompted the model every 20 entries using the prompt "provide more entries" to continue the generation process. Upon reviewing the first batch of 500 generated entries, we noted that while the dataset was balanced, there were significant duplicates in the humorous section. To obtain a well-balanced dataset free from duplicates, we had to prompt the model a total of 970 times.

Human-generated dataset. The Colbert dataset¹ is an assortment of textual data that was employed to examine the application of BERT (Bidirectional Encoder Representations from Transformers) sentence embeddings in the detection of humor. This dataset is well-balanced, as indicated by [12]. The dataset has 200,000 entries, but to have a fair comparison with the ChatGPT-generated dataset, we have randomly reduced it to 500 entries while keeping the proportions of humorous and non-humorous labels stable.

Test set. The human-generated test set comprises 100 meticulously selected entries from Colbert’s content. These entries were randomly selected to capture a representative sample of Colbert’s unique speech patterns and nuances. Extra care was taken to ensure that these entries did not overlap with any from the training set.

The AI-generated test set was produced by Bing ChatBot. The same prompting methodology was employed when interacting with Bing ChatBot. We chose Bing ChatBot

¹ <https://www.kaggle.com/datasets/deepcontractor/200k-short-texts-for-humor-detection>

for testing because it offers a valuable external benchmark for our models, which are trained on the generated text by the ChatGPT model. Not only does it allow us to compare performance against a distinct AI, but it also helps us understand how our model interacts with diverse dialogue styles.

3.2 Dataset annotation

The task of annotating our datasets was a pivotal step in our research, given the subjective nature of humor and its dependence on various nuanced factors. Our annotation scheme was designed to help the annotators rate the humor of jokes or humorous texts based on 1) funniness and 2) originality. This section will provide a detailed overview of this annotation process.

Two human evaluators annotated the AI-generated and human-generated humor datasets. They were given clear instructions about the task and were blind to the source of the texts (AI or human) to prevent any bias.

The annotators rated the **'funniness'** of the text on a Likert scale ranging from 1 to 5, where 1 represented 'not funny at all' and 5 meant 'hilarious'. The intermediate values 2, 3, and 4 indicated increasing levels of funniness.

The **'originality'** of the humor was also rated on a Likert scale from 1 to 5, where 1 indicated 'not original at all,' and 5 showed 'highly original'. The intermediate values represented increasing levels of originality.

Annotators were provided with clear guidelines to maintain objectivity, consider cultural context, ensure a complete understanding of the text, and maintain consistency in their ratings.

The annotated data then underwent a cleaning process, where we resolved any disputes in ratings through discussions or by referring to the opinion of a third evaluator. This process resulted in a robustly annotated dataset that laid the groundwork for our comparative analysis of human and AI-generated humor.

4 Experiments

To delve deeper into the intricacies of humor produced by both humans and artificial intelligence (AI), we conducted a series of experiments using the RoBERTa model, which is the most suitable model for humor detection [21]. Utilizing our provided dataset, we aimed to determine how RoBERTa can distinguish between humor in human-generated and AI-generated texts, as well as discern humor generated by AI from that created by humans².

4.1 Experiment 1

In our initial experiment, we fine-tuned the RoBERTa model for the humor detection task using two specific datasets: Human-Generated Humor Detection and AI-Generated Humor Detection. By using a learning rate of 5e-5, a maximum sequence length of 128, and implementing the Adam optimizer, we prepared the model for higher performance. The efficiency of each model was tested against corresponding datasets, i.e., the AI-tuned model was assessed with the AI_test set and the human-tuned model with the Human_test set.

² <https://github.com/DZHW-AI4SS/Laughing-Out-Loud-Exploring-AI-Generated-and-Human-Generated-Humor.git>

4.2 Experiment 2

Our second experiment aimed at tuning the RoBERTa model to discern whether any given text, humorous or otherwise, was generated by a human or an AI system. For this, we utilized a hybrid dataset containing both AI-produced and human-produced texts, labeled '1' and '0' respectively. We gauged the model's accuracy with a specially compiled test set featuring both AI and human-generated content in a humor detection scenario.

4.3 Experiment 3

To further explore the variances between human and AI-generated humor, we modified our dataset by removing the non-humorous components. After this adjustment, we repeated the second experiment, focusing solely on the humorous content generated by both human and AI sources in order to detect whether there is a difference between AI-generated and human-generated humor.

5 Results and discussion

5.1 Overview of experimental results

According to the results of the first experiment presented in Table 1, the RoBERTa model fine-tuned on both datasets reached perfect F1 scores, precision, and recall of 100%. These

Source	F1 Score	Precision	Recall	Epoch
Human	100%	100%	100%	4
AI	100%	100%	100%	2

Table 1. Comparison of the performance of RoBERTa fine-tuned on Human- and AI-generated Humor detection datasets

results indicate an impeccable performance of the models on both the Human and AI test datasets, implying that the models were able to correctly detect all instances of humor without any false positives or false negatives. It is interesting to note that the model fine-tuned on the AI-generated humor detection dataset reached this level of accuracy more rapidly, achieving perfect scores in just 2 epochs, compared to the 4 epochs required for the Human dataset. This suggests that the complexity of human-generated humor is higher than in AI-generated humor.

The results of second experiment are shown in Table 2. In the initial phase, we tested the

Source	F1 Score	Precision	Recall	Epoch
AI/Human	100%	100%	100%	4

Table 2. Performance of the fine-tuned RoBERTa model, tuned to AI- or human-generated labels, on a test set generated by a human and a Bing chatbot with humorous and non-humorous context.

model on a dataset that encompassed both humorous and non-humorous content generated by humans and an AI (Bing chatbot). The results presented in Table 2 demonstrated the model's extraordinary performance with perfect scores of 100% in F1, precision, and recall metrics, reached within 4 epochs. This indicates the model was successful in precisely classifying whether the text was produced by a human or an AI, regardless of whether the

text was humorous or not. To examine the model’s generalizability, we applied it to the extensive Colbert dataset. Impressively, the model accurately classified the entire dataset as human-generated text, achieving a 98% accuracy rate. These results underscore the potential of using a small, diverse dataset of AI and human-generated humor to construct a model that can efficiently and accurately categorize large-scale data, even when reaching up to 200,000 entries.

The outcomes of the third experiment are presented in Table 3. This phase concentrated solely on humorous text produced by both humans and AI. The results from this phase, as shown in Table 3, indicated an F1 score of 97.99%, precision of 98.07%, and recall of 98%, all reached within 2 epochs. These results suggest that while the model was slightly less accurate in identifying the source of humorous content compared to the mixed content, it still performed impressively.

Source	F1 Score	Precision	Recall	Epoch
AI/Human	97.99%	98.07%	98%	2

Table 3. Performance of the fine-tuned RoBERTa model, tuned to AI- or human-generated labels, on a test set generated by a human and a Bing chatbot with just humorous context.

These outcomes reaffirm that the RoBERTa model, when appropriately fine-tuned, can effectively discern between human- and AI-generated text, even within the complex domain of humor. However, it is notable that the performance slightly decreases when the model is solely exposed to humorous content, indicating the potentially increased complexity or variability in the way humans and AI generate humor.

5.2 Dataset analysis findings

Comparative analysis of AI-generated and human-generated humor This subchapter provides a comparative examination of the linguistic patterns [17–20] discerned in the AI-generated and human-generated humor datasets. The features analyzed in this study, including bigram usage, sentiment distribution, Part-of-Speech (POS) distribution, and average text length, were selected to provide a comprehensive understanding of the linguistic aspects of humor generation. These features allow us to explore the specific linguistic patterns, emotional tones, and linguistic elements employed by AI and humans in generating humor. By examining these features, we gain insights into the mechanisms and strategies behind humor generation, contributing to a deeper understanding of the similarities and differences between AI and human-generated humor.

Bigram usage: The top 10 bigrams in both AI-generated and human-generated humor datasets demonstrated considerable overlap. Common bigrams, such as ‘do, you’, ‘what, do’, ‘you, call’, ‘call, a’, ‘did, they’, ‘in, the’, ‘what, the’, ‘what, is’, and ‘is, a’, and ‘knock, knock’ (see Figures 1, 2), were prevalent in both datasets, indicating similarities in language structure used in the context of humor generation. However, a more detailed analysis should be conducted in future work, which might reveal slight variations in frequency and usage context.

Sentiment distribution: The sentiment distributions in both AI-generated and human-generated datasets indicate significant similarities, suggesting that the AI model has generally succeeded in capturing the emotional nuances of human humor (refer to Figures 3, 4). However, upon closer inspection, distinct differences emerge. The AI tends to produce humorous texts that are more neutral compared to those created by humans. Human-generated texts exhibit broader sentiments, some veering towards positive or negative tones.

Despite this, the overall sentiment in both groups remains low, with most texts falling into the neutral category.

Part-of-Speech (POS) distribution: Both datasets presented a similar distribution of POS tags, with nouns, punctuations, pronouns, determiners, verbs, and auxiliaries (see Figures 5, 6) being the most commonly used. This similarity might suggest that the AI has effectively mirrored human syntactic structures when generating humor. Nonetheless, a deeper inspection might uncover subtle differences in the context or complexity of usage across POS categories.

Average number of words: The average number of words in the AI-generated humor texts was found to be slightly lower than in the human-generated ones (12 compared to 14) (see Figures 7, 8). This difference indicates that while the AI generates humor within a relatively comparable length, it tends to create slightly more concise content. The reason for this verbosity disparity could be an interesting point of further investigation.

This comparative analysis serves as a stepping stone towards a deeper understanding of the differences and similarities between human and AI humor generation. It provides crucial insights that will guide the subsequent steps of our investigation and the interpretation of our experimental results.

5.3 Annotation analysis findings

The annotation process played a crucial role in evaluating the quality of humor in both AI-generated and human-generated texts. Two critical parameters assessed during annotation were 'funniness' and 'originality'. The following subchapters delve into the findings for each of these categories.

Funniness:

Upon closer inspection of the 'funniness' ratings between both AI and human-generated humor (see Figure 9) the following trends emerge:

"Not at all funny" (Rating 1): Human-generated humor was rated as "not at all funny" more frequently than AI-generated humor. This suggests that there are instances where human humor fails to resonate or generate amusement, at least more often than AI.

"Somewhat funny" (Rating 2): Both AI and human-generated humor were often perceived as somewhat funny, with human humor taking a slight lead. This indicates a shared capacity to generate humor that contains minor amusing elements but is unlikely to stimulate laughter.

"Moderately funny" (Rating 3): The frequency of this rating is similar for both AI and human humor, suggesting that both can produce content that is amusing to some extent and could potentially incite laughter.

"Very funny" (Rating 4): Surprisingly, AI-generated humor surpassed human humor in achieving the "very funny" rating. This shows that the AI could generate humor that is quite amusing and likely to elicit laughter more often than humans.

"Hilarious" (Rating 5): Neither AI nor human humor managed to achieve the highest rating of being "hilarious". This suggests a common challenge in generating extremely amusing humor and sure to cause laughter.

These findings illustrate that while AI can generate amusing content, with a surprising lead in developing "very funny" content, it still has room for improvement. The fact that both AI and humans struggled to achieve the "hilarious" rating underlines the complexity of crafting universally appealing and highly effective humor. It emphasizes the importance of further refinement in AI's humor generation techniques and the potential benefits of learning from human humor generation to enhance performance.

Originality:

In the analysis of the "originality" ratings for both AI and human-generated humor (see Figure 10), we identify clear trends that provide insight into the novelty of humor produced by both entities.

The ratings suggest that human-generated humor is perceived as more original than AI-generated humor in most cases. This is particularly evident in the frequency of the ratings "**moderately original**" (**Rating 3**) and "**very original**" (**Rating 4**) for human humor, which far surpass those for AI. It shows that humans more frequently than AI combine original ideas with common joke structures or rely on unique, new ideas and unexpected twists to elicit laughter.

In contrast, the humor generated by the AI was most frequently rated as "**somewhat original**" (**Rating 2**), indicating that the AI can incorporate some new elements but relies heavily on known jokes or structures. Additionally, the rating of "**not at all original**" (**Rating 1**) suggests that a portion of the AI humor is perceived as hackneyed, heavily recycled, or unoriginal, to a greater extent than human-generated humor.

Like the "funniness" ratings, there were no ratings of "**extremely original**" (**Rating 5**) in both datasets, indicating humor that is unique, creative, and unlike anything seen before. This again underscores the shared challenge of creating humor that blazes new trails altogether.

These patterns highlight the current limitations of AI in generating humor that is seen as original and novel. They underline the need for AI to learn more from human humor, especially incorporating unique ideas and unexpected twists. Despite AI's ability to generate humor with some degree of novelty, there remains a significant gap between the AI and human capacity to produce humor seen as highly original.

In conclusion, while AI has demonstrated an ability to generate humor with some original elements, it is still far from matching human capacity in originality. The findings suggest there is substantial room for AI to improve its performance in this area, particularly by learning from the human ability to create humor that blends original ideas with common joke structures or bases humor on unique, new ideas and unexpected twists.

5.4 AI vs. human humor: key differences

The comparative analysis of AI and human-generated humor provides crucial insights into NLP's current state of humor generation. Based on the results, several key differences and similarities in the humor of both entities can be identified.

First, concerning the structural aspects of humor generation, both AI and humans show similar use of bigram combinations and part-of-speech (POS) tags. These similarities suggest some success of the AI model in mimicking human language structures in the context of humor generation. However, the AI's tendency to produce more neutral sentiments than humans indicates potential gaps in the model's understanding of emotional nuance, an essential aspect of humor. In addition, the AI tended to produce more concise content, indicating a difference in verbosity.

In terms of 'funniness', both humans and AI have difficulty consistently generating highly amusing or hilarious content, illustrating the complexity of humor generation. Interestingly, AI humor received a 'very funny' rating more often than human humor. This suggests that despite its limitations, the AI model can generate content that could elicit significant hilarity.

When analyzing the 'originality' of the humor generated, human-generated humor is considered more original in most cases. This indicates that humans are more effective at combining original ideas with common joke structures or innovating with unexpected

twists. In contrast, AI humor often relies heavily on familiar jokes or structures, resulting in less original humor.

These findings highlight the current limitations of AI and the areas where it can potentially learn from human humor generation. Despite its ability to generate humor with some novelty, a significant gap remains between AI's and humans' ability to generate highly original humor. The findings suggest that improvements in AI's understanding of emotional nuance, its ability to innovate beyond known joke structures, and its fusion of original and common elements could lead to more effective humor generation.

In summary, while AI can generate entertaining and reasonably original content, it still falls short of the originality and emotional depth often found in human-generated humor. The comparison reveals the potential areas of improvement for AI and the inherent complexity of humor generation, which presents a fascinating challenge for future research in NLP.

6 Conclusion and future work

This study presents a comprehensive comparative analysis between AI-generated and human-generated humor. It provides crucial insights into the structural, emotional, and linguistic patterns inherent in the humor generated by both. The ratings of 'funniness' and 'originality' that emerge from a thorough annotation process further illuminate the strengths and weaknesses of humor generation by AI and humans.

Our findings highlight that while AI shows promise in generating entertaining and sometimes original content, it does not achieve the emotional depth and originality often seen in human humor. The AI's tendency to create neutral moods, its reliance on familiar joke structures, and its difficulty in consistently generating highly entertaining content highlight potential areas for improvement.

In contrast, despite its perceived weaknesses in generating 'hilarious' content, human humor offers a broader range of moods and more original content. These findings serve as a reminder of the complex nuances involved in humor generation and the subtleties of human communication that AI has yet to grasp fully.

Looking forward, several promising avenues for further work emerge. First, investigating AI's understanding and use of different joke structures could provide insights into the effectiveness of these structures in humor generation. Second, more work is needed to improve AI's ability to innovate beyond known joke structures and incorporate original elements. Last, training AI models on more diverse and extensive humor datasets could help to better capture the variability and depth of humor in human communication.

These studies are expected to refine AI's humor generation capabilities and enrich our understanding of the complex nature of humor. As we continue to move forward in this fascinating intersection of AI and humor, we look forward to the day when machines can generate humor that matches, or perhaps surpasses, the hilarity and originality of human humor.

In summary, the quest to give AI the ability to generate high-quality, original humor remains a challenging but exciting journey. We hope that the findings of this study will stimulate further research and support ongoing efforts to improve the performance of AI in humor generation, and bring us one step closer to this elusive goal.

7 Limitations

This study, aimed at comparing AI-generated and human humor, offers significant insights but has several limitations that necessitate careful interpretation of the findings.

The first limitation pertains to the potential subjectivity in the 'funny' and 'originality' ratings. Humor, a highly personal and culturally influenced trait, may have been perceived and rated differently by different evaluators. This subjective variation could have affected the perceived quality of humor across the AI and human-generated content.

Secondly, the generalizability of the findings may be restricted due to the specificity of the dataset used in the analysis. Given the complexity and diversity of humor, the study's findings might not apply to all forms of humor or other datasets. Therefore, we may have missed capturing the full spectrum of humor by restricting the analysis to a single dataset.

The third limitation lies in the study's focus. While the study thoroughly analyzes the structure and originality of humor, it does not consider the impact of specific types of content such as satire, irony, or dark humor. Due to their unique nature, these humor styles might necessitate a different analytical approach and could influence the interpretation of the results.

These limitations indicate the need for future research to expand the scope of analysis to include a wider variety of datasets and humor types and potentially employ multiple raters for a more comprehensive and balanced assessment. Despite these limitations, the study underscores an essential progression in the field and contributes towards enhancing the ability of AI models to generate humor.

References

1. Prajapati, Pariksha, et al. "Empirical Analysis of Humor Detection Using Deep Learning and Machine Learning on Kaggle Corpus." *International Conference on Advancements in Interdisciplinary Research*. Cham: Springer Nature Switzerland, 2022.
2. Kumar, Vijay, Ranjeet Walia, and Shivam Sharma. "DeepHumor: a novel deep learning framework for humor detection." *Multimedia Tools and Applications* 81.12 (2022): 16797-16812.
3. Mathias, Marcelo Custódio. *Humor and offense speech classification and scoring using natural language processing*. MS thesis. 2022.
4. Kalloniatis, Antony, and Panagiotis Adamidis. "Computational Humor Recognition: A Systematic Literature Review." (2023).
5. Mitrović, Sandra, Davide Andreoletti, and Omran Ayoub. "Chatgpt or human? detect and explain. explaining decisions of machine learning model for detecting short chatgpt-generated text." *arXiv preprint arXiv:2301.13852* (2023).
6. Sankar Sadasivan, Vinu, et al. "Can AI-Generated Text be Reliably Detected?." *arXiv e-prints* (2023): arXiv-2303.
7. Mitchell, Eric, et al. "Detectgpt: Zero-shot machine-generated text detection using probability curvature." *arXiv preprint arXiv:2301.11305* (2023).
8. Chakraborty, Souradip, et al. "On the possibilities of ai-generated text detection." *arXiv preprint arXiv:2304.04736* (2023).
9. *AI vs. Human—Differentiation Analysis of Scientific Content Generation*.
10. Crothers, Evan, Nathalie Japkowicz, and Herna L. Viktor. "Machine-generated text: A comprehensive survey of threat models and detection methods." *IEEE Access* (2023).
11. Kumar, Vijay, Ranjeet Walia, and Shivam Sharma. "DeepHumor: a novel deep learning framework for humor detection." *Multimedia Tools and Applications* 81.12 (2022): 16797-16812.
12. Annamradnejad, Issa, and Gohar Zoghi. "Colbert: Using bert sentence embedding for humor detection." *arXiv preprint arXiv:2004.12765* 1.3 (2020).
13. Elayan, Suzanne, et al. "'Are you having a laugh?': detecting humorous expressions on social media: an exploration of theory, current approaches and future work." *International Journal of Information Technology and Management* 21.1 (2022): 115-137.
14. Cao, Yihan, et al. "A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt." *arXiv preprint arXiv:2303.04226* (2023).
15. Krishna, Kalpesh, et al. "Paraphrasing evades detectors of ai-generated text, but retrieval is an effective defense." *arXiv preprint arXiv:2303.13408* (2023).
16. Li, Zhuohang, Jiashuo Liu, and Yuci Wang. "Performance Analysis on Deep Learning Models in Humor Detection Task." *2022 International Conference on Machine Learning and Knowledge Engineering (MLKE)*. IEEE, 2022.

17. Teller, Virginia. "Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition." (2000): 638-641.
18. Manning, Christopher, and Hinrich Schütze. Foundations of statistical natural language processing. MIT press, 1999.
19. Grishman, Ralph. Computational linguistics: an introduction. Cambridge University Press, 1986.
20. Eisenstein, Jacob. Introduction to natural language processing. MIT press, 2019.
21. Adoma, Acheampong Francisca, Nunoo-Mensah Henry, and Wenyu Chen. "Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition." 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). IEEE, 2020.

Hayastan Avetisyan completed her BA in Translation Studies in Yerevan, Armenia, and pursued an MA in Linguistics in Hannover, Germany, in 2020. In 2021, she joined the AI4S2 project at the Department of Research Infrastructure and Methods, focusing on NLP, ML, and AI interpretability. Her research explores leveraging linguistic knowledge to enhance the development and interpretation of language models. Currently pursuing her Ph.D., she investigates the utilization of AI in research methodologies.

Parisa Safikhani is a research scientist at the German Centre for Higher Education Research and Science Studies (DZHW). She holds an M.Sc. degree in electrical engineering and automation technology from LUH and obtained her Bachelor's degree in electrical and telecommunication engineering from Arak University. Currently, she is pursuing her PhD in the field of artificial intelligence, with a specific focus on AutoNLP.

David Broneske is the head of the department Infrastructure and Methods at the German Centre for Higher Education Research and Science Studies (DZHW), Hannover. He received his PhD in Computer Science from the University of Magdeburg, where he also pursued his Master's and Bachelor's Studies in Computer Science. His research interests include main-memory database systems, interdisciplinary data management, and the application of artificial intelligence in various domains.

A Appendix

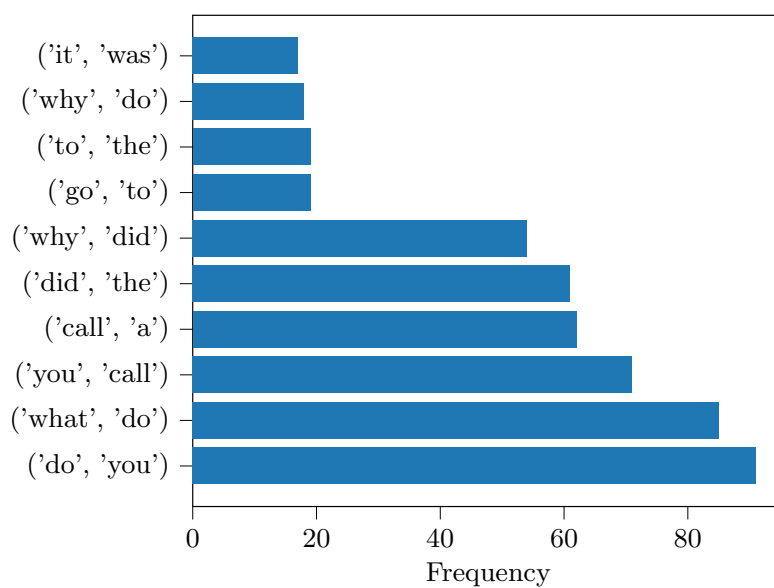


Fig. 1. Top 10 Bigrams of AI-generated humorous content.

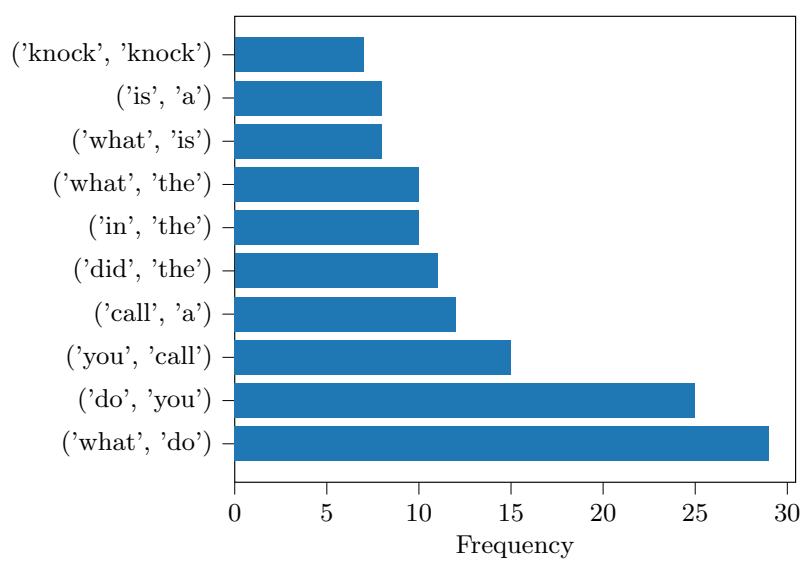


Fig. 2. Top 10 Bigrams of Human-generated humorous content.

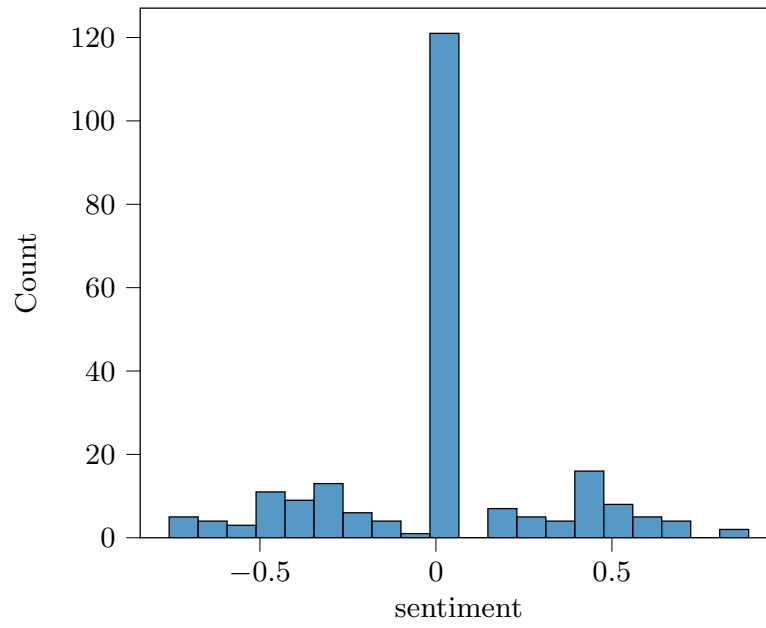


Fig. 3. Sentiment Distribution of AI-generated humorous content.

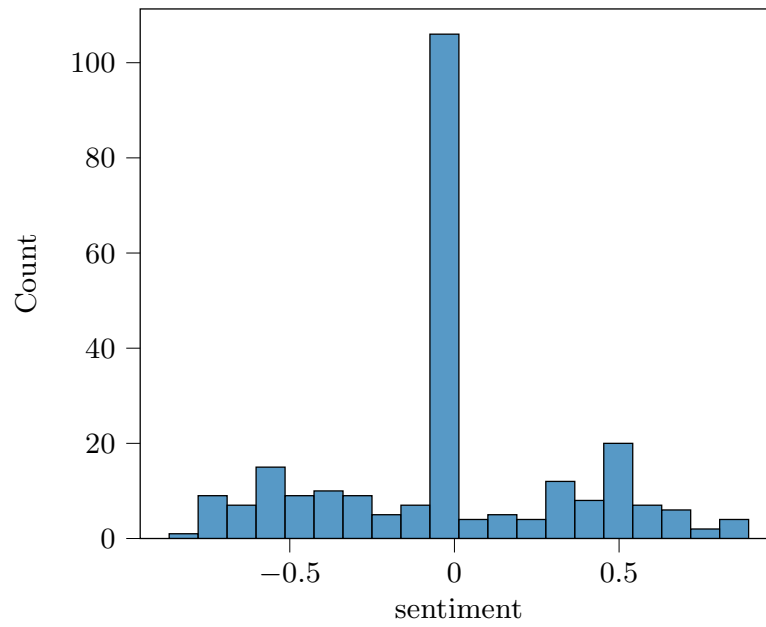


Fig. 4. Sentiment Distribution of Human-generated humorous content.

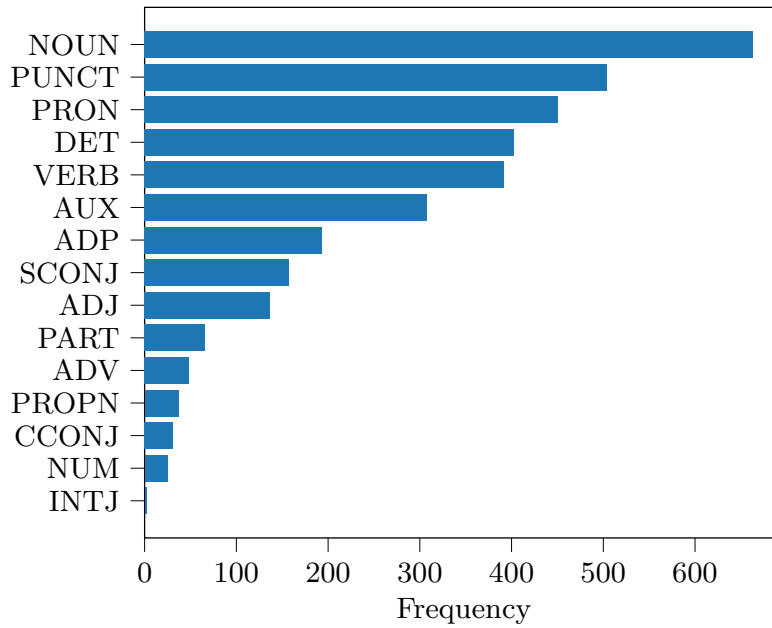


Fig. 5. POS Distribution of AI-generated humorous content.

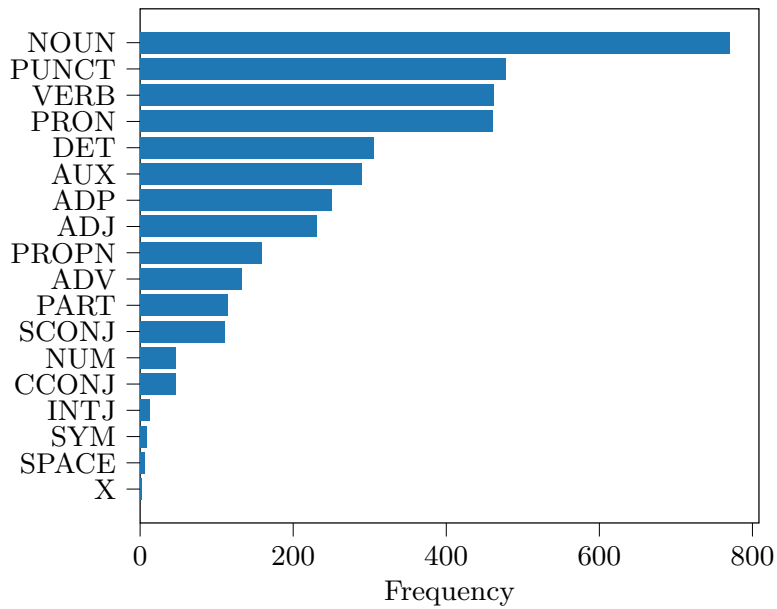


Fig. 6. POS Distribution of Human-generated humorous content.

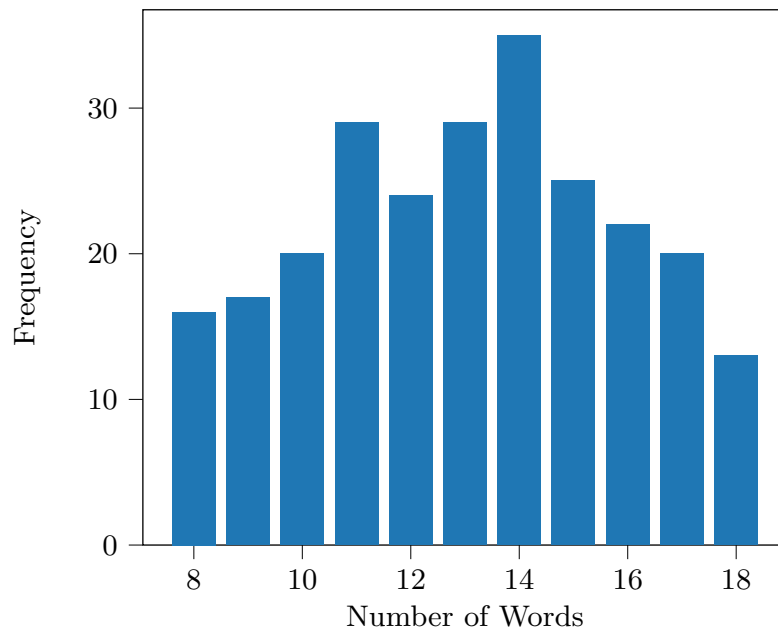


Fig. 7. Average number of words: Human-generated humorous content.

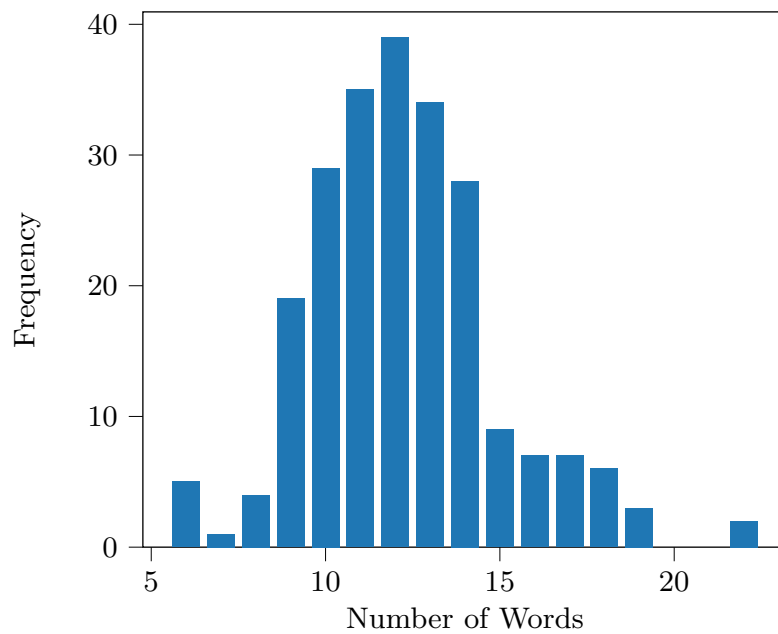


Fig. 8. Average number of words: AI-generated humorous content.

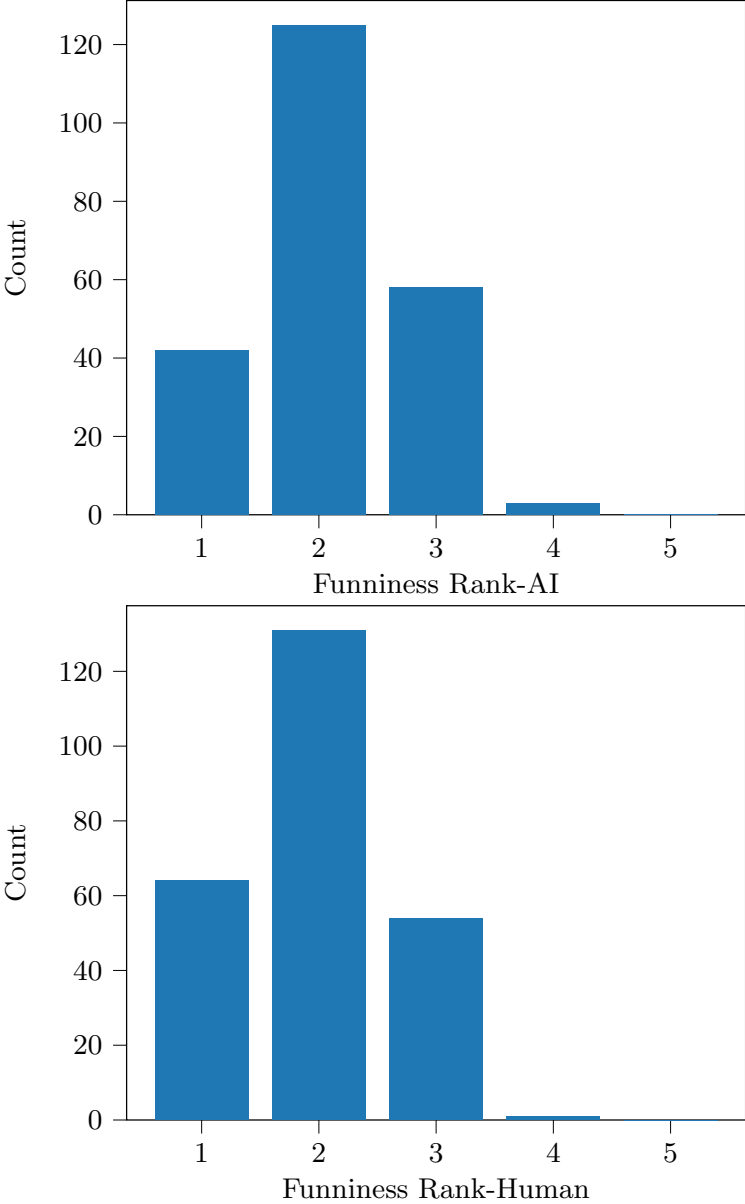


Fig. 9. Funniness of AI- and Human-generated humorous content.

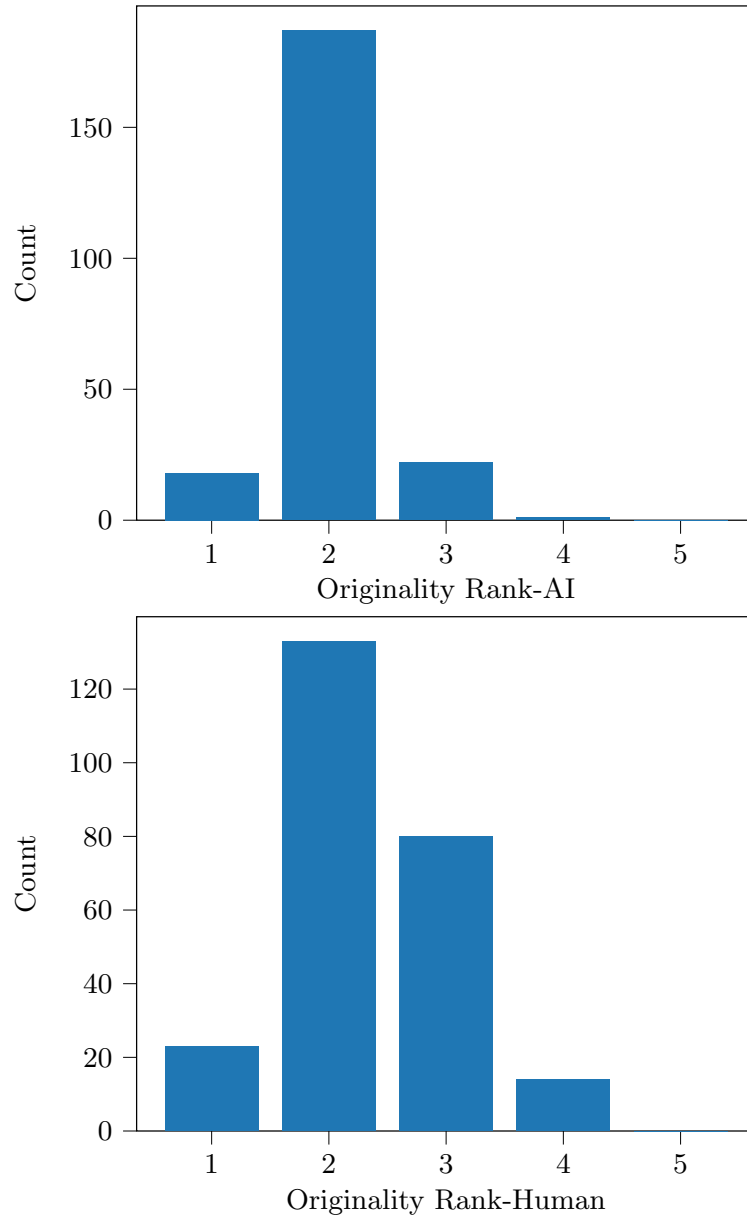


Fig. 10. Originality of AI- and Human-generated humorous content.

AN INTELLIGENT MOBILE APPLICATION TO FACILITATE STUDENT'S NETWORKING USING NATURAL LANGUAGE PROCESSING ALGORITHMS

Ruijin Deng¹ and Victor Phan²

¹Shawnigan Lake School, 1975 Renfrew Rd, Shawnigan Lake,
BC V0R 2W1, Canada

²Computer Science Department, California State Polytechnic University,
Pomona, CA 91768

ABSTRACT

This paper tackles the prevalent problem of students facing resource limitations in pursuing their passions, often due to the lack of like-minded peers within their school environment. To address this issue, we propose the development of a team-management app that serves as a transformative platform, enabling students to connect with others who share their interests and aspirations [1]. The application contains three components: the team management system, the search system, and the chat system. Throughout the development process, we encountered various challenges, such as ensuring user privacy. Comparative analysis with three alternative solutions demonstrated several notable advantages, including an intuitive and user-friendly interface, granting students full autonomy over their clubs, and being open to all students from diverse club interests. This innovative application has the potential to empower students, foster collaboration, and provide a valuable resource for educational institutions and students alike [2].

KEYWORDS

Students Networking, Natural Language Processing Algorithms, Mobile Application, Social Networking

1. INTRODUCTION

The issue that is addressed in this research is students lacking in-school connection with like-minded peers [3]. The background and history of this problem can be traced back to the traditional education system, where students are often placed in classrooms based on their age and location, rather than their interests and passions [4]. This issue is crucial to address because it acts as an obstacle in a student's journey of pursuing their passion. It is also important to address because it helps students shape their identity and increase their confidence by sharing their interests with like-minded peers. In the long run, addressing this problem can lead to more engaged and motivated students, which ultimately contribute to a more inclusive and nurturing educational environment [5].

Of the 3 methodologies analyzed in section 5, "Club Activity Management System" created by Kyouhei Higashi mainly focused on improving the student's autonomy in sports management and reducing unnecessary co-curriculum stress. However, because of its sport-focused, it has a slightly

narrower diversity. The "Ikka Veima Football Club Content Management System" created by Prof. Kari Systä focused on managing the football club with a precise system that involves 3 approaches, which may lead to excessive complexity and lack of analysis. In addition to the project involving a Sports Club Intranet, created by Oriol González Navarro that utilizes a database system for centralizing data. This solution has some limitations in managing the database. StarSeek improves on these works by providing a platform for all types of clubs, an easy-to-use user interface, and a strong database stored in Firebase.

Clubs nowadays can be managed in various methods, which are not foolproof [6]. They can be managed through private group chats such as Snapchat or Discord [7]. They can also be maintained through Facebook Groups or public-facing Instagram Accounts. The downside of using these systems, however, is that they are not professional (especially when done via a group chat), they aren't organized, and often students either wouldn't otherwise regularly use these apps, or these apps don't have a student market in mind. We propose an application named StarSeek, that will serve as a club-focused app for a student-only market. StarSeek provides a students-only platform that is academic-focused, which could enhance the connection among students while keeping it trustworthy. This solution is effective because its target market is students in North America; and because the data is verified by the schools, it is very secure in terms of student's personal information.

The first experiment that we conducted was a survey to test user satisfaction with using StarSeek, and also if users would consider using StarSeek as a legitimate application in their school life. We asked a variety of questions about StarSeek's user interface as well as its perceived utility. Respondents gave generally fair to positive feedback regarding StarSeek as an application. We believe that improving StarSeek as an app and getting a larger user base would improve these survey scores.

The second experiment that we conducted was an accuracy test between two natural language recognition models to determine which model we would use as our recommender system in the actual StarSeek application. We tested 10 cases against these models (cosine and gensim, respectively), with a dataset of 16 realistic-sounding team descriptions. After testing, we determined that the cosine model had by a large margin the most accurate recommendations, and thus we utilize cosine similarity within the production application.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. To Ensure Data's Accuracy

One challenge that comes with implementing the accounts to be used within teams is to ensure that the data associated with each account is accurate. A student may be able to falsify data on their account and act like they are part of a club or school that they aren't affiliated with. Schools that affiliate themselves with StarSeek would be able to use their administrative powers to enforce policies for students to ensure truthfulness. Otherwise, schools may request us (the app's developers) to delete fraudulent accounts in the database for them. Another concern is that a student applies to join a team that does operations that requires its members to be closeby. To address this, managers for a team are able to deny requests to join and can look through the profiles of applicants before approving or denying them.

2.2. Present Features

One design consideration when it came to the chat system was if features present in other chatting applications such as sending stickers, images, videos, or gifs should be allowed. We decided that for this application, having members only be able to send over text to one another would be the safest option. This is because it allows for conversations had through the application to be more sanitized and more focused on direct communication between members. Giving members the ability to send images and videos to one another may encourage more casual conversations and less productive coordination between management and student members

2.3. How To Ensure Good Results

One of the biggest concerns that come with our AI machine-learning team search algorithm is how to ensure the best and most accurate results when performing queries [8]. There is no guarantee that the recommender system in this application will be the most robust or the most accurate, but we strive to maintain the best quality that we can achieve. If we find through testing that our recommender system is not as accurate as we believe it should be, our backend infrastructure is flexible enough for us to be able to easily swap it out with a different recommender system if we happen to find a better one.

3. SOLUTION

Users begin their journey on the sign-in/sign-up page where they can either log in or create a new account. After successfully logging in, they are directed to the questionnaire page to provide essential information, such as their preferences and interests. Once completed, users land on their dashboard, the central hub of the platform. From the dashboard, they can create and manage teams, and previously created teams are conveniently listed for reference. Clicking the "Make Team" button on the dashboard leads users to the team creation page, where they can define the team's title and description. Additionally, users can explore existing teams by navigating to the team search page, where they can apply to join a team of interest. Team owners are promptly notified of applications, and they can accept or reject new members. The chat feature allows for one-on-one or group communication by email or within existing teams. Lastly, users can view and personalize their profiles on the user profile page, and they can update their personal information in profile settings. This comprehensive design caters to a social and collaborative platform, ensuring a seamless experience for users to create, connect, and communicate effectively.

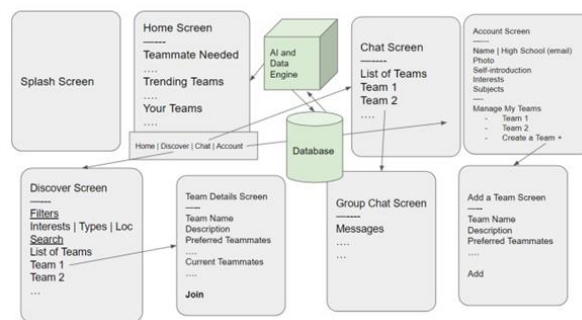


Figure 1. Overview of the solution

One of the systems in my app is an AI system [9]. Specifically, this is a natural language processing AI, which adapts the use of algorithms to be able to interpret natural languages such as English and extract connections and meanings from human speech. In this app, the AI is used as a

recommendation system which was incorporated in the search feature of the app to give users relevant teams to join.



Figure 2. Screenshot of the project 1

```

5. def cosine_similarity_score(sentence1, sentence2):
6.     # Tokenize and convert sentences to lowercase
7.     sentence1 = sentence1.lower()
8.     sentence2 = sentence2.lower()
9.
10.    # Initialize CountVectorizer to convert a collection of text documents to a matrix of token counts
11.    vectorizer = CountVectorizer().fit_transform([sentence1, sentence2])
12.
13.    # Compute cosine similarity
14.    vectors = vectorizer.toarray()
15.    similarity = cosine_similarity(vectors)
16.
17.    # Cosine similarity is the element at (0, 1) or (1, 0)
18.    similarity_score = similarity[0][1]
19.    return similarity_score
20.
21. def get_recommendation(word, sentences):
22.     output = {}
23.     for s in sentences:
24.         output[s] = cosine_similarity_score(word, s)
25.     output = dict(sorted(output.items(), key=lambda x:x[1], reverse=True))
26.
27.     return output.keys()
28.
29. def recommend_team(dict_of_teams, query):
30.     result_strings = get_recommendation(query, dict_of_teams.keys())
31.     output = []
32.     for result in result_strings:
33.         output.append(dict_of_teams[result])
34.     return output

```

Figure 3. Screenshot of code 1

For a cosine similarity model, the accuracy score is determined by the cosine between the vectors of two given strings. In the screenshot, under the method `cosine_similarity_score`, we take in two strings that are to be compared to one another, named `sentence` and `sentence2`. We utilize a count vectorizer to convert these strings into vector representations. We then compare the cosine of the angle between these two vectors in order to determine the similarity score. In the method `get_recommendation`, we take in a query, and a set of string to compare our query to. We will calculate the similarity score for each string in our dataset, sort our dataset by the strings with the highest similarity, and then return an ordered list of our recommendations. In the method `recommend_team`, we take in a dictionary of teams, mapped from description to title, and a string that is our query. We will get a recommendation list based on these two parameters, and then return the names of the teams in order of that team's similarity to our query.

The chat system in my app functions as a communication tool between users to effectively interact and exchange messages with one another [10]. Its primary purpose is to facilitate real-time communication and enable users to send text-based messages. This function relies on Firebase's Firestore, where the chat logs and chat metadata are stored. Real-time databases can allow for instant updates and synchronization for the messages that users sent.

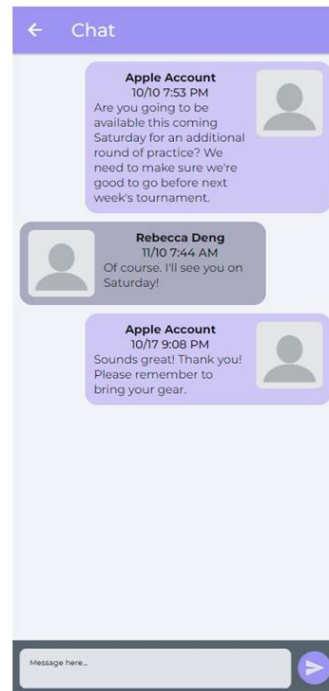


Figure 4. Screenshot of project 2

```

5  onPressed: () async {
6    var chatsRecordReference = ChatsRecord.collection.doc();
7    await chatsRecordReference.set({
8      ...mapToFirestore(
9        {
10         'members': _model.groupMembers,
11       },
12     ),
13   });
14   _model.createdDocument = ChatsRecord.getDocumentFromData({
15     ...mapToFirestore(
16       {
17         'members': _model.groupMembers,
18       },
19     ),
20   }, chatsRecordReference);
21   await _model.createdDocument!.reference.update({
22     ...mapToFirestore(
23       {
24         'members': FieldValue.arrayUnion([currentUserReference]),
25       },
26     ),
27   });
28   var invitesRecordReference = InvitesRecord.collection.doc();
29   await invitesRecordReference.set({
30     ...createInvitesRecordData(
31       sender: currentUserReference,
32       chatRef: _model.createdDocument?.reference,
33     ),
34     ...mapToFirestore(
35       {
36         'recipients': _model.groupMembers,
37         'members': _model.groupMembers,
38       },
39     ),
40   });
41   _model.createdInvite = InvitesRecord.getDocumentFromData({
42     ...createInvitesRecordData(
43       sender: currentUserReference,
44       chatRef: _model.createdDocument?.reference,
45     ),
46     ...mapToFirestore(
47       {
48         'recipients': _model.groupMembers,
49         'members': _model.groupMembers,
50       },
51     ),
52   }, invitesRecordReference);
53   await _model.createdInvite!.reference.update({
54     ...mapToFirestore(
55       {
56         'members': FieldValue.arrayUnion([currentUserReference]),
57       },
58     ),
59   });
60   await UserChatsRecord.collection.doc().set(createUserChatsRecordData(
61     user: currentUserReference,
62     chatRef: _model.createdDocument?.reference,
63   ));
64   context.safePop();
65   setState({});
66 },

```

Figure 5. Screenshot of code 2

The codes use several Firestore collections. The first one is called chats, where all of the group chats that people make are stored. The second one is invites, which are documents that contain invites to join the group chat, containing the group chat's creator, the members, and the recipients. When someone accepts an invitation, they would be removed from the recipients' list. The third one is chat_refs, where a document is made in this collection when someone accepts an invitation and serves as their key for accessing the group chat. When someone decides to make a group chat, the collection creates a chat document and sets that chat's members. They then add the creator to the chat's members and create an invite document, setting the invitation's members list, sender, and recipients. The creator is then added to the invite document's members list. Then the chat_ref document is created so that the chat creator can access their own chat page. Then we go back to the previous page.

The team system is the biggest part of this app, where users can make teams, join teams, and manage teams. This component serves as the foundation of the app's function, where all the connections are based. Users can initiate the creation of new teams, manage team membership by deciding who can join or leave the team, and accept or decline invitations. This component is based on Firebase's Firestore, where all the structured data such as user-generated content for managing teams and their associated data [15].

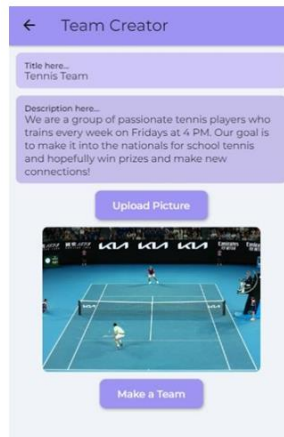


Figure 6. Screenshot of project 3

```

1 // Generated code for this Button Widget...
2 if ((_model.textController1.text != null &&
3     _model.textController1.text != '') &&
4     (_model.textController2.text != null &&
5     _model.textController2.text != '') &&
6     (_model.uploadedFileUrl != null && _model.uploadedFileUrl != ''))
7 FFButtonWidget(
8   onPressed: () async {
9     var chatsRecordReference = ChatsRecord.collection.doc();
10    await chatsRecordReference.set({
11      ...mapToFirestore(
12        {
13          'members': [currentUserReference],
14        },
15      ),
16    });
17    _model.groupChat = ChatsRecord.getDocumentFromData({
18      ...mapToFirestore(
19        {
20          'members': [currentUserReference],
21        },
22      ),
23    }, chatsRecordReference);
24    await UserChatsRecord.collection.doc().set(createUserChatsRecordData(
25      user: currentUserReference,
26      chatref: _model.groupChat?.reference,
27    ));
28    await TeamsRecord.collection.doc().set({
29      ...createTeamsRecordData(
30        description: _model.textController2.text,
31        owner: currentUserReference,
32        title: _model.textController1.text,
33        image: _model.uploadedFileUrl,
34        groupChat: _model.groupChat?.reference,
35      ),
36      ...mapToFirestore(
37        {
38          'members': [currentUserReference],
39          'admins': [currentUserReference],
40        },
41      ),
42    });

```

Figure 7. Screenshot of code 3

The application contains a page for a user to create a team, by filling out a form containing the team's name, description, and banner picture. When the user presses the "create" button, this code is run. First, the group chat for the team needs to be generated, so we begin by creating a document in the chats collection, with the team creator as its only member. After that, a user_chats document is generated so the team creator can access their own group chat. For the final step, a team document is created that sets the description, title, and image to be what the owner filled out in the original form. We set the owner, members, and admins field of the document to be just the owner's user ID. Lastly, we set the group_chat field to be the generated group chat document from step 1.

4. EXPERIMENT

4.1. Experiment 1

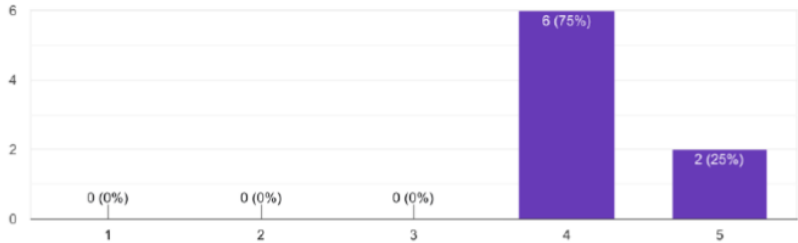
This experiment is conducted through a survey. In this experiment, there are three unknown objectives that are to be determined. First, through this experiment, user satisfaction are being measured from the survey. Second, the utility of this app will be perceived. Lastly, the utility of certain features is also being determined from the survey.

This experiment is set up through a Google survey that is sent out to a group of students in a Canadian boarding school with a demo instruction beforehand. Some of the participants will be regular students, some will be club members, and others will be club presidents. The form has the following questions.

1. On a scale of 1-5, how easy was it to navigate through StarSeek?
2. On a scale of 1-5, how clear was StarSeek's design?
3. On a scale of 1-5, how likely are you going to use StarSeek to find someone outside of school?
4. On a scale of 1-5, how effective do you think StarSeek would be at bridging connections between schools?
5. If you were a member of a club at your school, on a scale of 1-5, would you suggest to your club leader to create a team on StarSeek?
6. StarSeek contains an in-app group chat system. On a scale of 1-5, how likely do you think you would use it as your primary contact method among group members?
7. StarSeek is built to be a simple team management app. On a scale of 1-5, how likely do you think you would use StarSeek over another platform such as a Facebook group for club management?
8. On a scale of 1-5, how useful do you think StarSeek would be when trying to do networking among peers?
9. On a scale of 1-5, how well do you think StarSeek would be when trying to recruit members for a club?
10. On a scale of 1-5, what would you overall rate StarSeek?

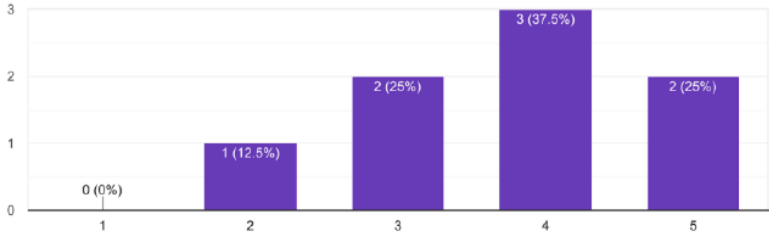
On a scale of 1-5, how easy was it to navigate through StarSeek?

8 responses



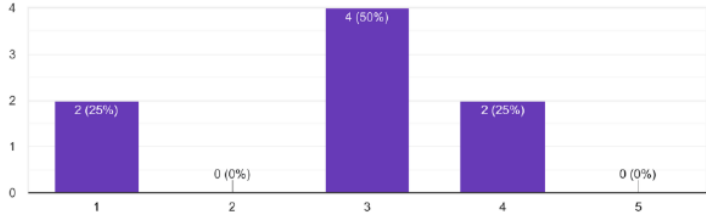
On a scale of 1-5, how clear was StarSeek's design?

8 responses



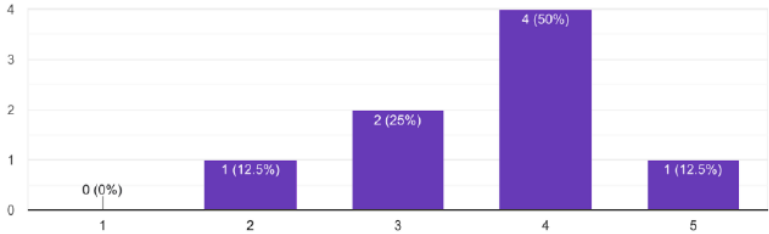
On a scale of 1-5, how likely are you going to use StarSeek to find someone outside of school?

8 responses



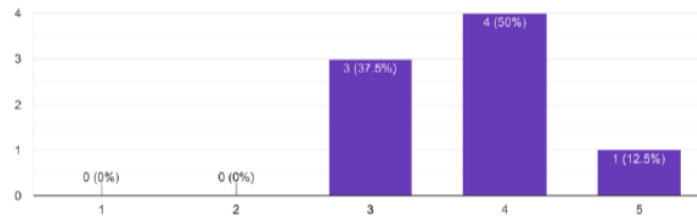
On a scale of 1-5, how effective do you think StarSeek would be at bridging connections between schools?

8 responses



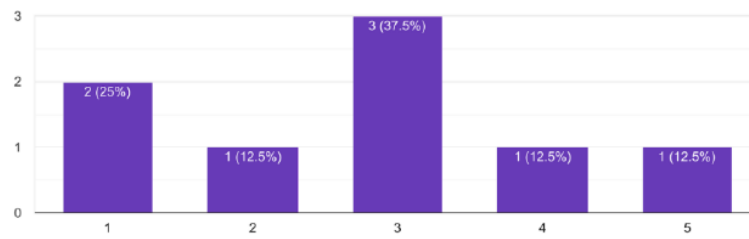
If you were a member of a club at your school, on a scale of 1-5, would you suggest to your club leader to create a team on StarSeek?

8 responses



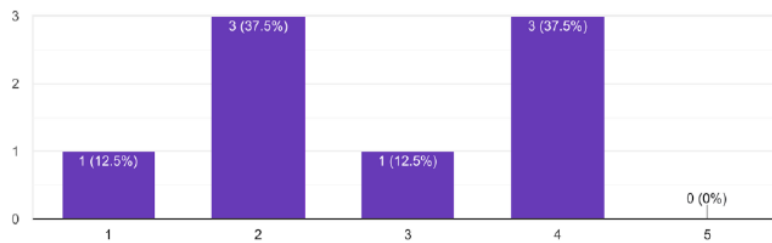
StarSeek contains an in-app group chat system. On a scale of 1-5, how likely do you think you would use it as your primary contact method among group members?

8 responses



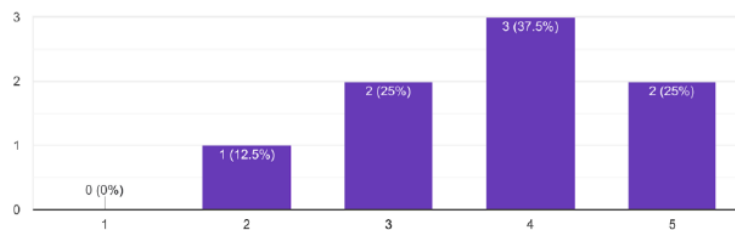
StarSeek is built to be a simple team management app. On a scale of 1-5, how likely do you think you would use StarSeek over another platform such as a Facebook group for club management?

8 responses



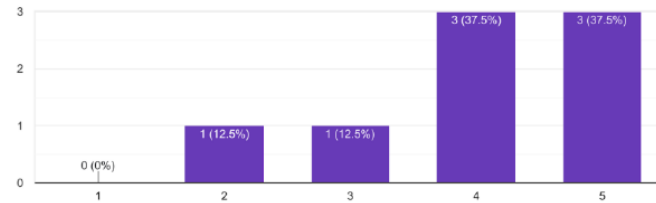
On a scale of 1-5, how useful do you think StarSeek would be when trying to do networking among peers?

8 responses



On a scale of 1-5, how well do you think StarSeek would be when trying to recruit members for a club?

8 responses



On a scale of 1-5, what would you overall rate StarSeek?

8 responses

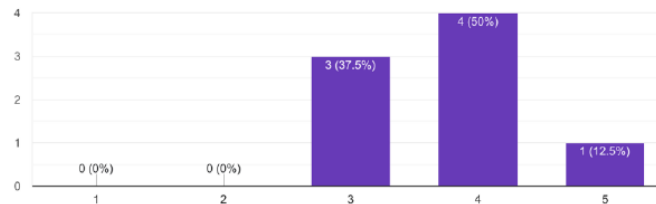


Figure 8. Figure of experiment 1

The mean score for the question “On a scale of 1-5, how easy was it to navigate through StarSeek?” was 4.25, and the mode was 4. The lowest value was 4, and the highest was 5. This data indicates that StarSeek is an intuitive application to navigate.

The mean score for the question “On a scale of 1-5, how clear was StarSeek's design?” was 3.75, and the mode was 4. The lowest value was 2, and the highest was 5. This is likely based on the participants' personal preferences.

The mean score for the question “On a scale of 1-5, how likely are you going to use StarSeek to find someone outside of school?” was 2.75 and the mode was 3. The lowest value is 1, and the highest value is 4. Participants likely gave it a lower score because there are currently as of writing not a large amount of active users. If StarSeek had a larger install base, then we predict that this score would have been higher.

The mean score for the question “On a scale of 1-5, how effective do you think StarSeek would be at bridging connections between schools?” was 3.63, and the mode was 4. The lowest value is 2, and the highest value is 4. Participants who rated this question lower might be concerned about the feasibility of StarSeek.

The mean score for the question “If you were a member of a club at your school, on a scale of 1-5, would you suggest to your club leader to create a team on StarSeek?” was 3.75, and the mode was 4. The lowest value is 3, and the highest value is 5. Participants would likely score higher if there were more users on StarSeek.

The mean score for the question “StarSeek contains an in-app group chat system. On a scale of 1-5, how likely do you think you would use it as your primary contact method among group members?” was 2.75, and the mode was 3. The lowest value is 1, and the highest value is 5. Participants gave a lower score because the chat system isn’t as robust as other communication apps.

The mean score for the question “StarSeek is built to be a simple team management app. On a scale of 1-5, how likely do you think you would use StarSeek over another platform such as a Facebook group for club management?” was 2.75, and the mode was 2. The lowest value is 1, and the highest value is 4. Participants gave a lower score because people are more familiar with Facebook than StarSeek. However, more people will be familiar with StarSeek as it goes.

The mean score for the question “On a scale of 1-5, how useful do you think StarSeek would be when trying to do networking among peers?” was 3.75, and the mode was 4. The lowest value is 2, and the highest value is 5. The data indicate that participants do consider utilizing StarSeek to network among peers.

The mean score for the question “On a scale of 1-5, how well do you think StarSeek would be when trying to recruit members for a club?” was 4, and the mode was 4. The lowest value is 2, and the highest value is 5. This indicates that users would find StarSeek great for club recruitment.

The mean score for the question “On a scale of 1-5, what would you overall rate StarSeek?” was 3.75, and the mode was 4. The lowest value is 3, and the highest value is 5. This indicates that overall users found that StarSeek was a good app. With further iteration we expect this score to be higher once we address user concerns.

4.2. Experiment 2

StarSeek utilizes a recommender system, which allows students to be recommended teams to join based on a search query that they enter. It is vital for this search system to be as accurate as possible because it affects the usability and capability of students to network through the app.

For this experiment, we will be conducting an accuracy assessment for two types of natural language recognition machine learning models. The first utilizes cosine similarity, and the second utilizes a Python library named Gensim. We will train the models with a set of descriptions and names for 16 teams. These teams are not real teams but nonetheless are written to be plausible and strong in their theming. The teams are:

- 1 general STEM team
- 1 STEM green energy team
- 1 STEM Sociology team
- 1 general engineering team
- 1 engineering aeronautics team
- 1 green energy team
- 1 programming team
- 1 biotechnology team

- 1 mathematics team
- 1 mathematics team with a focus on cryptography and quantum mechanics
- 1 data science team
- 1 robotics team
- 1 art team
- 1 health team
- 1 marinology team

We will test these models with the following 10 test cases:

1. Ocean
2. Engineering
3. Robotics
4. science technology engineering math
5. renewable green energy
6. aeronautics and space
7. medicine and health and fitness
8. art technology
9. team projects
10. creativity and problem-solving

For each of these cases, we will get the top 3 recommendations back from both models. For each of these recommendations, we award a point to the model if the recommendation is considered relevant or accurate given the query. For each case, a winning model is determined based on the model with the higher point count.

Query	Cosine	Gensim	Winner
"ocean"	2	1	Cosine
"engineering"	3	1.5	Cosine
"robotics"	2	2	N/A
"science technology engineering math"	2	0	Cosine
"renewable green energy"	2	2	N/A
"aeronautics and space"	2	1	Cosine
"medicine and health and fitness"	1	1	N/A
"art technology"	1	1	N/A
"team projects"	3	1	Cosine
"creativity and problem solving"	3	2	Cosine

Figure 9. Figure of experiment 2

The average accuracy score awarded to the cosine model was 2.1. For the gensim model, the average accuracy score awarded was 1.25. On one occasion, the gensim model was awarded half a point due to the recommendation given being slightly relevant, but not enough to earn a point, though this half point did not have any bearing on whether gensim won that case. The cosine model is thus on average significantly more accurate than the gensim model. On 3 occasions, both models had the same accuracy score and thus the winner was a tie, or non-applicable to the final results.

When tallying up the winning model, cosine is unanimously the winner. There were no cases in which the gensim model won. Therefore, we can say that the cosine model is generally a more accurate machine learning natural language recognition model than that of gensim. This experiment has demonstrated to us empirically that the cosine model is more suitable for good results in our application, and thus the backend of our application utilizes the same cosine model as used in this experiment.

5. RELATED WORK

Another system, the “Ikka Veima Football Club Content Management System” created by Prof [11]. Kari Systä offers a highly efficient management system tailored for football clubs, ensuring precise management within the context of sports teams. However, its narrow focus on football clubs limits its applicability to a broader spectrum of organizations and activities, potentially missing out on valuable insights. Furthermore, the solution's precision may inadvertently lead to excessive complexity, which can be time-consuming. In contrast, StarSeek provides an accessible and functional platform that connects users with like-minded individuals without overwhelming them with a complex user interface. Its openness to various clubs encourages a more comprehensive exploration of interests and collaborations across different domains, making it a more versatile option.

A project involving a Sports Club Intranet, created by Oriol González Navarro, utilizes a database system for centralizing data, making it easier to access and manage compared to scattered paper

records [12]. However, this solution has some potential limitations in establishing a structured database. As this solution still heavily relies on manual processes, its dependence on Microsoft Access as technology can possibly become a bottleneck. StarSeek has a strong database stored in Firebase, where maintenance is rarely needed. Starseek does not rely on local expertise to find and retain personnel for data management.

6. CONCLUSIONS

Improvements to StarSeek include upgrading the chat system and refining the search capabilities. To improve the chat system, features like image and document sharing can be integrated, and user identification can transition from email-based to usernames for enhanced privacy. Simultaneously, the search system can be refined for accuracy by prioritizing clubs in the same geographical region, fostering in-person interactions. To realize these improvements, the chat system may require an updated design and functionality on FlutterFlow, while the search system could benefit from additional tools like Python libraries or ChatGPT for enhanced search algorithms [13]. These modifications aim to create a more dynamic and user-friendly experience within StarSeek, fostering deeper connections and facilitating more meaningful club interactions, whether virtually through chat or through localized, real-world engagements.

StarSeek continues to update and is on the path to becoming a more dynamic and user-centric platform [14]. It continues to explore and implement new strategies and ensure that students have the opportunity to thrive in what they are passionate about. With these potential improvements, StarSeek wants to facilitate deeper club interactions and create a more vibrant community for its users.

REFERENCES

- [1] Necşoi, Daniela Veronica. "Team and team management-a practical approach." *Educația Plus* 12.1 (2015): 317-328.
- [2] MacAllister, James. "What should educational institutions be for?." *British Journal of Educational Studies* 64.3 (2016): 375-391.
- [3] McCarthy, Sarah J. "Home-school connections: A review of the literature." *The Journal of Educational Research* 93.3 (2000): 145-153.
- [4] Jayasinghe, Udeni, Anuja Dharmaratne, and Ajantha Atukorale. "Students' performance evaluation in online education system Vs traditional education system." *Proceedings of 2015 12th International Conference on Remote Engineering and Virtual Instrumentation (REV)*. IEEE, 2015.
- [5] Finn, Jeremy D. "Expectations and the educational environment." *Review of educational research* 42.3 (1972): 387-410.
- [6] Svensson, Lars EO. "The zero bound in an open economy: A foolproof way of escaping from a liquidity trap." (2000).
- [7] Moschogianni, Georgia. "Investigating factors Impacting Discord and Snapchat Use Behavior: Perspective from Swedish users." *Journal of Digitovation and information system* 2.1 (2022): 41-54.
- [8] Mellarkod, Veena S., Michael Gelfond, and Yuanlin Zhang. "Integrating answer set programming and constraint logic programming." *Annals of Mathematics and Artificial Intelligence* 53.1-4 (2008): 251-287.
- [9] McKinney, Scott Mayer, et al. "International evaluation of an AI system for breast cancer screening." *Nature* 577.7788 (2020): 89-94.
- [10] Dewes, Christian, Arne Wichmann, and Anja Feldmann. "An analysis of Internet chat systems." *Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*. 2003.
- [11] Kulikova, Lidiya, and Anna Goshunova. "Evaluation of management system quality: case of professional football clubs." *Academy of Strategic Management Journal* 15 (2016): 122.
- [12] González Navarro, Oriol. *Sports Club Management*. BS thesis. Universitat Politècnica de Catalunya, 2017.
- [13] Ray, Partha Pratim. "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope." *Internet of Things and Cyber-Physical Systems* (2023).
- [14] Recordon, David, and Drummond Reed. "OpenID 2.0: a platform for user-centric identity management." *Proceedings of the second ACM workshop on Digital identity management*. 2006.
- [15] Chougale, Pankaj, et al. "Firebase-Overview and Usage." *International Research Journal of Modernization in Engineering Technology and Science* 3.12 (2021): 1178-1183.

REVEALING SUSTAINABLE GROWTH FOR FIT BIT: A DATA-DRIVEN MARKETING APPROACH BASED ON K-MEANS CLUSTERING AND COLLABORATIVE FILTERING

Akansha Akansha and Stuart So

University of Exeter Business School, Rennes Drive, Exeter, United
Kingdom

ABSTRACT

Understanding the user segment is highly significant in the age of a highly competitive wearable Fitness Technology market. In this study, we leveraged a comprehensive dataset containing information on user interactions, activity logs and device usage records. For effective segmentation of the users, K-Means clustering was employed. The unsupervised Machine Learning algorithm helped us group the clusters of consumers based on their similarity in the usage of the device, activity levels and engagement patterns. The collaborative Filtering technique refines product recommendations by identifying user preferences based on past patterns. The analysis aims to uncover distinct user segments and provide insights into user behaviours and lifestyles to enhance Fitbit's Market Performance and improve user engagement, customer satisfaction and brand loyalty leading to higher customer retention. The findings of an extensive analysis conducted on Fitbit User data using K-Means Clustering and Collaborative filtering techniques are presented. To achieve sustainable growth in the highly competitive smart wearables market, Fitbit can improve its user experience by addressing the diverse needs of different user segments.

KEYWORDS

Fitbit, Segmentation, K-Means, Collaborative Filtering, Personalisation, Wearable Fitness

1. INTRODUCTION

In the rapidly evolving landscape of contemporary business and marketing, a pivotal convergence occurs - the fusion of sustainability and customer-centricity [1]. As global awareness of environmental concerns grows, and consumer preferences shift, organisations face the dual challenge of elevating their market performance while aligning their practices with sustainable principles. This dynamic convergence of objectives has prompted an in-depth exploration of strategies that transcend traditional norms, ushering in prospects for sustainable economic growth.

This business undertaking can be directed to a voyage into sustainable growth, spotlighting the esteemed health and wellness technology brand Fitbit which is a market leader in wearable technology and related products [2]. Our research conducts an Outside-In Exploration of Fitbit's

Customer-Centric Marketing Approach, which dives into the intricate interplay between customer-centricity and sustainable marketing strategies using cluster analysis and collaborative filtering techniques. Fitbit, renowned for its innovative wearable devices and health-focused technologies, is an example of how companies can adroitly integrate customer insights into their marketing blueprints [3].

By meticulously observing Fitbit's practices, we seek to unveil how the interweaving of these factors could give rise to a comprehensive marketing approach that amplifies brand growth and nurtures sustainability [4]. The study underscores the paramount significance of customer-centric marketing, functioning as the linchpin around sustainable growth. The detailed analysis aims to subject Fitbit's marketing to a customer-centric approach and uncover how to empower the company to nurture sustainable growth while upholding its commitment to environmental stewardship and customer well-being.

In the following sections, this paper will follow a methodical trajectory, from the foundational tenets of sustainable marketing, immersing into the specifics of Fitbit's customer-centric methodology and culminating in pragmatic insights drawn from user data. We will delve deeper into theoretical frameworks, methodological considerations, and findings highlighting the symbiotic relationship between customer-centricity and marketing. At its core, this report aspires to enrich the scholarly discourse on sustainable marketing and furnish actionable insights to practitioners who aspire to harness avenues of growth while navigating the marketing domain, guided by insights from user data.

2. OVERVIEW OF THE WEARABLE FITNESS TECHNOLOGY INDUSTRY AND FITBIT

2.1. The Wearable Fitness Technology Industry

The world of wearable fitness tech is like a fast-paced race where high-tech meets our health. It's a place where we can find a variety of gadgets, from simple step counters to super-smart watches that keep track of every move, our sleep, and even our heart rate. With the growing concern towards staying healthy, the market for fitness gadgets is constantly growing. Various brands are trying to develop new technology to win people over [5].

Additionally, the emergence of the COVID-19 pandemic has further impacted the demand for wearable technology within the healthcare sector, as these devices can promptly detect indicators of infection. Furthermore, the increasing disposable income, growing popularity of such devices, market availability of intelligent gadgets, and various other factors are anticipated to play a significant role in the growth of the wearable technology industry [6]. Recent trends in the wearable technology market indicate promising possibilities for expansion. Moreover, the revenue in the wearables market will likely reach USD17.834billion by the close of 2021, with a significant portion of this revenue originating from the Chinese market [7].

2.2. Fitbit

Established in 2007 by James Park and Eric Friedman, Fitbit assumed a pivotal role in shaping the landscape of wearable fitness technology. The company's establishment corresponded with the industry's early expansion phases, positioning Fitbit as an innovative frontrunner. In its initial stages, Fitbit concentrated on designing wireless activity trackers geared towards aiding individuals in tracking their fitness progress. The inaugural offering, known as the Fitbit Tracker, gained notable recognition, serving as a cornerstone for the subsequent growth of the brand. The popularity of these products was regarded as a positive influence in promoting physical health

and cultivating behaviour that enhances the quality of life, with an emphasis on physical activity [8].

2.2.1. Fit Bit Financial Performance

Fitbit's main mission is to contribute to better health worldwide by continuously monitoring and improving people's health performance. As reported through the way of Business Wire, Fitbit Inc.

Achieved total earnings of \$188 million in 2020. Hence, it is observed that the overall possessions owned by the company decreased from \$1,515,547 thousand to \$1,368,086 thousand in 2019, primarily because the amount of cash and easily accessible funds had lowered. Fitbit reported that the company's net earnings have decreased over time due to a drop in gross profit and increased operating expenses. The company's net operational income was \$(320,711) thousand in 2019. Additionally, Fitbit Inc.'s overall financial health could have been more robust in 2020, as indicated by its negative cash flow from operations, which amounted to \$(80) million.

In summary, Fitbit Inc. is facing challenges in various financial aspects, and the company needs to implement stringent measures to enhance its operational and financial standing [9].

2.2.2. Product Marketing Strategy

The promotion strategies include techniques to improve brand awareness, sales and loyalty. Including:

Direct Marketing: Utilising direct mail, telemarketing, and personalized email.

In-Store Promotion: Incorporating flash sales, discounts, and loyalty points.

Social-Media Marketing: Utilising platforms like Facebook and Instagram to engage with the user-base.

Integrated-Marketing Communication: Ensuring constant messaging throughout all channels for coherent brand communication.

Whilst other brands offer equally competitive products and services, Fitbit's combination of brand recognition and continuous innovation has contributed to its appearance and popularity among consumers. Innovation is vital to sustain competition in the smart wearable industry, and Fitbit's ability to consistently offer fresh designs is appealing to consumers. Consumer preferences in the smart wearable market are shaped mainly by the presence of cutting-edge features. A targeted marketing strategy can help a brand stand out amidst fierce competition. A targeted approach based on data-driven insights is pertinent in the competitive smart wearables market. Fitbit's ability to harness the power of data-driven insights can create more accurate and sustainable marketing strategies.

We intend to address both facets within our approach, unveiling a flexible, two-stage methodology that dynamically crafts behaviorally coherent segments and subsequently steers the selection of target marketing strategies. Our methodology is rooted in the user data of the customers using the devices constantly throughout their day. The foundational principle of behavioral coherence involves an exhaustive exploration of interconnected purchases across diverse categories on a segment level, involving the precision-targeted engagement of several thousand customers.

In this context, the strategic focus on **data-driven insights** emerges as a valuable approach. The user-generated data can mainly guide marketing strategies by providing a deeper understanding of consumer behaviour and preferences. By delving into user behaviour and preferences, Fitbit can

fine-tune its marketing strategies, ensuring they resonate with specific target audiences and improvising the products to meet particular categories of consumers.

3. LITERATURE REVIEW

3.1. Market Segmentation

Market Segmentation is a fundamental step involving a diverse market into smaller, more homogeneous segments based on shared characteristics, behaviours or needs [10]. In recent times, the landscape of market segmentation has been reshaped by the emergence of big data and advancements in data analytics, enabling businesses to craft precise and impactful strategies. This categorisation hinges on shared characteristics, encompassing everyday needs, interests, lifestyles, or analogous demographic traits.

The primary objective of Segmentation revolves around identifying segments with the highest potential for profit generation or future growth. These chosen segments receive special attention, earning the designation of "target markets." The methods for segmenting a market are diverse. In scenarios involving business-to-business (B2B) dynamics, the market might be dissected into distinct business types or geographical regions. Conversely, business-to-consumer (B2C) settings often employ Segmentation based on demographic factors like lifestyle, behaviour, or socioeconomic status.

User data offers priceless insights into consumer behaviour, preferences, and demographics [11]. This level of understanding is aimed at businesses aiming to customise their products, services, and marketing campaigns for specific audiences. The user-generated data from different sources can significantly enhance segmentation accuracy, allowing businesses to differentiate subtle distinctions within the market segments and create highly targeted marketing initiatives resonating with the consumers. Market segmentation is crucial for a company's marketing strategy to be effective. This is because traditional class patterns no longer exist, and consumers have more disposable income. Companies can develop their products in the right direction by dividing consumers into manageable segments based on their needs. Targeting all consumers would be unnecessary and costly, so understanding the consumer segment regarding age, values, purchase behaviour, and attitudes is essential for success [12].

3.2. Data-Driven Personalisation and Underlying Approaches

The defining theme of the contemporary market is personalisation. Businesses harness user data to deliver tailored experiences, content and recommendations [13]. This amplifies customer satisfaction, catering to brand loyalty. [14] delved into the concept of predictive personalisation; by leveraging historical user data and modern machine learning algorithms, businesses can proactively forecast user preferences to cater to individual needs. While using user data in market segmentation offers a substantial advantage, it raises ethical concerns simultaneously.

The future of market segmentation based on user data will soon be induced with automation and Artificial Intelligence (AI). AI-driven algorithms can swiftly analyse vast datasets in real-time, enabling adaptive and dynamic segmentation strategies [15].

Furthermore, there are possibilities for integrating offline and online data sources. Merging the user data from physical and digital touch points will provide a more holistic understanding of consumer behaviour [16].

3.2.1. K-Means Clustering

K-Means clustering is an unsupervised machine-learning algorithm for data analysis and segmentation [17]. It is designed to divide a large dataset into small groups or clusters, where data points within each cluster are more similar than those in other clusters [18].

Algorithm 1 *k*-means algorithm

- 1: Specify the number k of clusters to assign.
 - 2: Randomly initialize k centroids.
 - 3: **repeat**
 - 4: **expectation:** Assign each point to its closest centroid.
 - 5: **maximization:** Compute the new centroid (mean) of each cluster.
 - 6: **until** The centroid positions do not change.
-

Figure1 K-Means Algorithm

As shown in **Figure 1**, the process starts by selecting a user-defined number of clusters, 'K'. The algorithm assigns each data point to the nearest cluster centre, often represented by the mean of the data points in that cluster. It refines the cluster assignments until convergence, minimizing the sum of squared distances between data points and their respective cluster centres [19].

$$\text{objective function} \leftarrow J = \sum_{j=1}^k \sum_{i=1}^n \underbrace{\|x_i^{(j)} - c_j\|^2}_{\text{Distance function}}$$

Figure2 Mathematical Model of K-Means Clustering

The mathematical model of K-Means Clustering is expressed in **Figure 2**. K- Means clustering as various applications across a variety of domains. In marketing, the K-means helps in customer segmentation by group in g customers with similar behaviors or preferences[20]. In image processing, it helps in image compression and object recognition [21]. Additionally, it is utilised in biological analysis and clustering [22].

Despite the versatility, K-Means clustering has limitations, such as sensitivity to initial placements of cluster centres and the assumption that all clusters are spherical and equal in size. Researchers have developed an improved version of K-Means to address these limitations [23].

3.2.2. Collaborative Filtering : Are Commendation System Technique

Collaborative Filtering technique is widely used in recommendation systems aimed towards providing personalised suggestions to users [24]. The technique relies on the ideas of users who had similar preferences in the past to have similar preferences in the future. The core principle of this technique is to analyse user interactions and feedback on items, such as ratings or purchase history, to provide recommendations.

Two main approaches to Collaborative Filtering are:

1. **User-Based Collaborative Filtering:** Identifies users with similar preferences based on their historical interactions and recommends items favoured by consumers with the same profiles.
2. **Item-Based Collaborative Filtering:** Items are compared based on user feedback to find similarities; products that are highly rated or interacted with by consumers who have shown interest in the same items are recommended.

Collaborative Filtering has proven its efficiency in various domains, including e-commerce, content streaming and social media [25]. Platforms like Amazon, Netflix and Spotify use this technique to power their recommendation systems offering users personalised products, movies or music suggestions.

Despite the efficacy, Collaborative Filtering faces challenges such as the cold start problem [26] when it deals with new users or items for which the data interaction is limited. In order to address these limitations, hybrid recommendation systems combine Collaborative Filtering technique with other techniques, such as content-based filtering or matrix factorization [27].

4. METHODOLOGY

The methodology of the study comprises the following major components of a comprehensive research process where the research activity one comes after another which are grouped into the following four major stages within the research process (**Figure 3**):

4.1. Research Process

4.1.1. Data Collection

Considering the factors, we might require addressing our research, the dataset was found on the open-source data science platform Kaggle. It predominantly is a User-Data of the consumers using the products of Fitbit. The entire dataset consists of 18 files, with data related to the users' daily activities, sleep schedule, weight information, heart rate, calories burnt, number of steps walked on a daily basis and their intensity of day-to-day activity. The datasets are secondary data obtained from the data science portal Kaggle.

Although, keeping in mind that the dataset is available via an open-source platform and not by the brand itself does raise the question of its transparency. Kaggle underscores the commitment to safeguarding the privacy and anonymity of individuals whose data is included in the datasets.

In this particular dataset, there is no mention of the names, contact details of the consumers and the privacy of the individuals have been protected to the best extent complying with the GDPR Regulations set in 2018. Kaggle takes the responsibility of ensuring that the data is handled in an ethical and responsible manner, avoiding any misuse that could lead to harm, discrimination and privacy violations. It also promotes transparency and accountability in research conducted using the dataset. Peer review and collaboration is encouraged to maintain a high level of accountability.

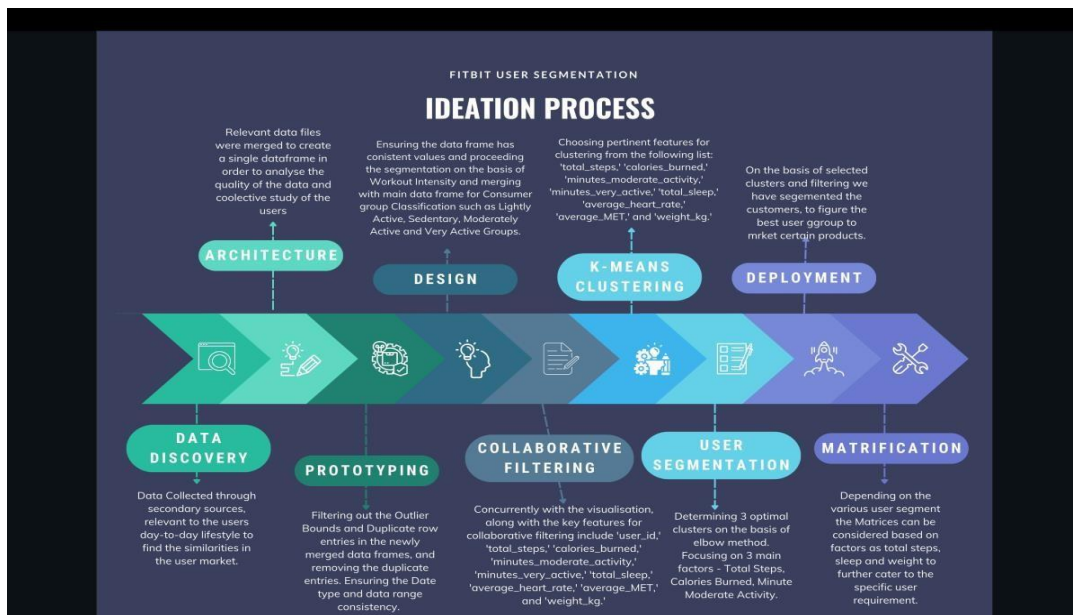


Figure3 Research Process

4.1.2. Data Retrieval

However, out of the 18 datasets, we have narrowed it down to utilising 8 datasets, as the minute-by-minute details of the user would provide us with a detailed insight into their day-to-day lifestyle, making it relevant to our research. Namely:

- (a) **Daily Activity:** mentioning the record date of data collection, total steps walked or run throughout the day, and total distance covered. Active minutes and sedentary minutes along with calories.
- (b) **Minute Calories Narrow:** indicating the specific minute the calories are expended is recorded.
- (c) **Minute Intensities Narrow:** a record of the minutes along with the level of intensity or effort during that particular minute, providing information on how vigorous the activity was during that particular minute.
- (d) **Minute Steps Narrow:** representing the number of steps taken during that specific minute of the particular activity in that minute.
- (e) **Minute METs Narrow:** record of the Metabolic Equivalent of Tasks, denoting the specific minute it was performed at, measuring the energy expenditure compared to resting.
- (f) **Sleep Day:** captures the specific day when the sleep data was recorded and indicates the total number of sleep records captured for that particular day. Total minutes the user was asleep during that day and the total minutes they spent in bed; giving an overview of the time spent in resting position.
- (g) **heartrate_seconds:** covers the actual heart rate every five seconds and specific show fast the heart was beating in that particular moment.

(h) **Weight Log Info:** is a collective record of the weight of the users in kilograms(kg) and pounds(lbs), also including the fat percentage and BMI (Body Mass Index) ((Mooney et al., 2013)), and whether the weight information was entered manually by the user.

The details of the data retrieval process by a data funnel approach are shown in **Figure4** below.

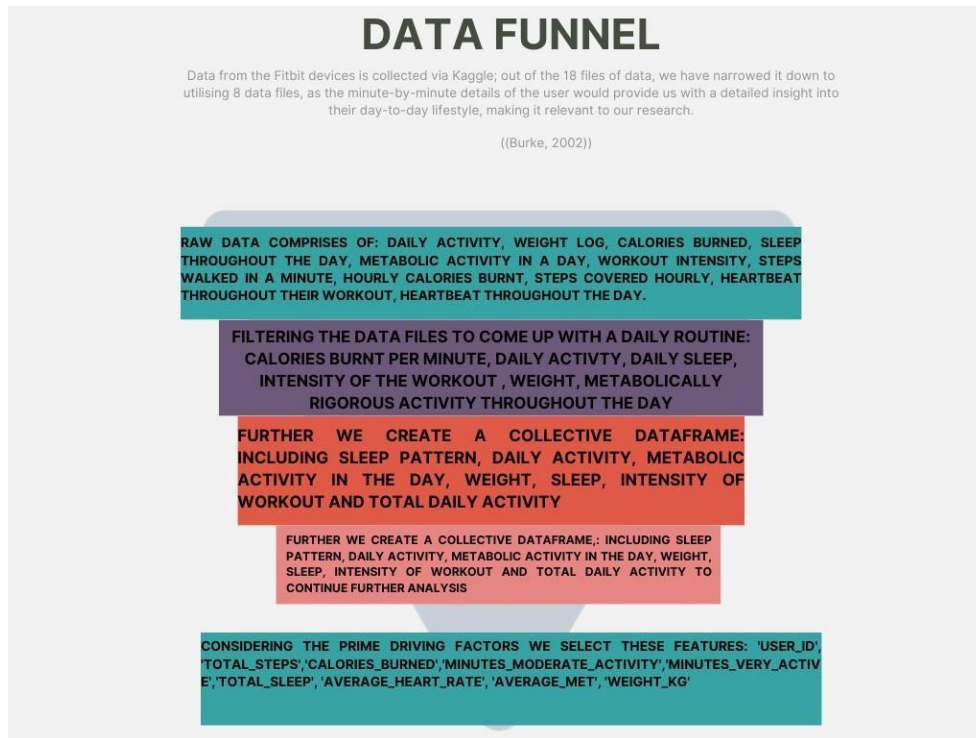


Figure4Adata funnel approach for the data retrieval process

4.1.3. Data Exploration

Using the function, **‘general_info(df)’** for an overview of our data, we get a concise summary of the data frame which offers us an insight in to the contents of the data such as the number of rows, columns and the fundamental details about data-types and null-values. It allows us to swiftly grasp the basic characteristics of the data frame as an initial step towards data exploration.

The second function, **‘outlier_bounds(df, col)’**, is used to identify potential outliers within the specific column. To determine this, we have computed and exhibited the first and third Quartiles (Q1 and Q3) for the specific columns. Using this we calculate the Interquartile Range (IQR), which will help us determine the outliers. It delineates the lower and upper bounds for the potential outliers, through IQR. It is valuable for pinpointing data anomalies and extremes

The third function, **‘duplicate_index_search(df, col1, col2)’**, is to find out the duplicate entries in the data frame by focusing on two main columns if in case the values match. After detecting the duplicate row of entries, it provides the corresponding indexes (row numbers). The function proves useful in flagging the duplicate records and facilitate the assessment of data in terms of quality.

The fourth function, **‘fitness_device_usage(df, TimePeriod, TotalTime, TimePeriod_name,**

variable_name), is to analyse the usage of the fitness tracker among distinct consumer groups. By computing the average values of designated activities for each consumer group within the dataset. It also calculates the average percentage of this duration over the span of a month.

After figuring out the duplicated entries and null values in our dataframe, now we proceed ahead to check the Consistency of our data.

- (i) **ID Consistency Check:** For each dataframe (**df_act, df_cals, df_ints, df_steps, df_sleep, df_heart, df_met, df_weight**), the main purpose is to ensure there are no repetitions or inconsistent user IDs within each dataset. For the sake of consistency we also rename the 'Fairly Active Minutes' to 'Moderately Active Minutes'.
- (j) **Date Range Consistency Check:** To ensure the regularity in the datasets' time range to validate the data's consistency, especially in time-series data. Ensuring that the data covers an expected and coherent time period, with no gaps or irregularities in the range of dates.

To assure the quality of data, the consistency check is a part of the pre-processing.

4.1.4. Data Engineering

As most of the variables are interrelated in our datasets such as the calories burnt on average are dependent on the intensity of the workout. The Metabolic Equivalent of Tasks are responsible for the Calorie expenditure as well, which also includes the number of steps covered on a daily basis. To avoid switching between multiple datasets to conclude on the same variables, we chose the joined data frame for the holistic and comprehensive approach of analysis. We join our data sets to create a singular data frame to avoid the overlap of the same variables.

We then move on to assess the consistency between the intensity of the Steps and the average speed in the data frame '**df_act**', by calculating the average speed for all levels of steps walked based on intensity: **Very Active, Moderately Active and Lightly Active**. Applying conditions to identify inconsistency and check if the average speed of the 'Very Active' workouts is lower than the 'Moderately Active' and if the average speed of 'Moderately Active' workouts is lower than the 'Lightly Active' workouts.

We then move on to calculate the percentage of inconsistent data by dividing the number of rows in '**inconsistent_df_act**' by the total number of rows in '**df_act**' and then multiplying it by 100. This calculated percentage represents the extent to which workout intensity and average speed are inconsistent in the dataset. Setting the intensity level of workout categories of Sedentary, lightly Active, Moderately Active and Very Active users, respectively, to 0,1,2 and 3. We calculate the average minimum and maximum Steps covered.

Customer Segmentation on the basis of number of steps covered: As per (*Counting Your Steps*, n. d.), 10000 steps is the daily recommended target for an adult to live a healthy life. Hence, we have segmented the customers' data in '**df_act**', as per the following ranges:

Sedentary <= 5000 steps Lightly Active: 5001 to 8500 steps

Moderately Active: 8501 to 12500 steps Very Active >= 12500 steps

Furthermore, in the analysis, we segmented the customers based on their sleep, as well as sleep, along with their daily activity intensity and the time of day they were the most active. We have

also analysed and visualised the average time it might take for the customers to fall asleep.

4.2. Collaborative Filtering and Algorithms

Collaborative filtering is a technique that can help you find relevant and personalized recommendations based on the preferences or feedback of many users. For example, if you want to watch a movie, a collaborative filtering system can suggest movies that are similar to the ones you have liked before, or movies that other users with similar tastes have liked.

There are different methods and challenges for implementing collaborative filtering, such as how to measure the similarity between users and items, how to deal with sparse or missing data, how to handle new users or items, and how to scale up the system for large datasets. One common approach is to use matrix factorization, which is a way of finding latent features or embeddings that represent the characteristics of users and items. These embeddings can be learned automatically from the data, without relying on hand-engineered features. The embeddings can then be used to compute the similarity or the predicted rating between a user and an item

4.2.1. The Algorithms

Collaborative filtering algorithms are methods that can help you find relevant and personalised recommendations based on the preferences or feedback of many users. There are different types of collaborative filtering algorithms, such as:

- (a) **Memory-based algorithms [28]:** These algorithms use the entire feedback matrix to compute the similarity between users or items, and then use these similarities to predict the ratings or preferences of a target user. For example, user-based collaborative filtering finds users who have similar ratings to the target user, and then recommends items that these similar users have liked. Item-based collaborative filtering finds items that have similar ratings to the target item and then recommends items that are similar to the target item. Memory-based algorithms are simple and intuitive, but they can suffer from scalability and sparsity issues.
- (b) **Model-based algorithms [29]:** These algorithms use the feedback matrix to learn a model that can capture the latent factors or features that explain the ratings or preferences of users and items. For example, matrix factorization is a popular model-based algorithm that learns low-dimensional embeddings or vectors for users and items, such that the dot product of these vectors can approximate the rating or preference of a user-item pair. Model-based algorithms can handle sparsity and scalability better than memory-based algorithms, but they can be more complex and prone to over fitting.
- (c) **Hybrid algorithms [30]:** These algorithms combine the advantages of memory-based and model-based algorithms or use additional information or techniques to improve the performance of collaborative filtering. For example, some hybrid algorithms can use content-based features or metadata to augment the feedback matrix or use ensemble methods or deep learning models to integrate multiple collaborative filtering models. Hybrid algorithms can achieve higher accuracy and diversity than single algorithms, but they can also be more difficult to implement and interpret.

4.2.2. Top-K Algorithms for User-Based Collaborative Filtering

As mentioned in Section 3, user-based Collaborative Filtering is a technique that uses the ratings or preferences of similar users to recommend items to a target user. It is a memory-based

algorithm that uses the similarity between users or items to make recommendations. The algorithm works by first computing the similarity between all pairs of users or items and then selecting the Top-k most similar users or items to make recommendations [31]. More specifically, Top-k algorithms that applied to user-based collaborative filtering were explored where Top-k refers to an algorithm that can efficiently find the top-k items or candidates for a given user or query, based on some criteria or score [31].

There are several Python packages available for collaborative filtering. One of the most popular libraries is Surprise [32]. It is a Python scikit building and analysing recommender systems that deal with explicit rating data. It provides various ready-to-use algorithms and evaluation metrics to help you build your recommender system. To choose the Top-k algorithm from the Python package Surprise, you can use the `KNNWithMeans` class [32]. This class is a basic collaborative filtering algorithm that uses a basic nearest neighbours approach. We can set the `k` parameter to the number of items we want to recommend to the user. Another package is Neighbors [33], which is a Python package for performing collaborative filtering on social and emotion datasets. It provides a simple interface to perform collaborative filtering on a dataset. To choose the Top-k algorithm from the Python package Neighbors, we can use the `neighbors` function [33]. This function is a basic collaborative filtering algorithm that uses a basic nearest neighbors approach. You can set the `k` parameter to the number of items we want to recommend to the user.

5. RESULTS AND ANALYSIS

5.1. Results

The heat map visually represents the distribution of four key factors across different activity levels of Fitbit users (See **Figure 5**). Each row of the heat map corresponds to a specific factor, and each column corresponds to a distinct activity level, including sedentary, lightly active, moderately active, and very active users. The colour intensity in each cell of the heat map is indicative of the magnitude or percentage distribution of the respective factor for users falling into that particular activity level category.

5.1.1. Heat Map

- (a) **Activity:** The first row of the heat map provides insights into the distribution of general activity levels across the specified user categories. The colour intensity in each column illustrates the proportion of users engaged in sedentary, lightly active, moderately active, and very active lifestyles.
- (b) **Sleep:** The second row focuses on sleep patterns, displaying how the duration and quality of sleep vary across different activity levels. Darker colours may suggest longer and more restful sleep, while lighter colours may indicate shorter or less restful sleep durations.
- (c) **Joined (Weight and Calories Burned):** The third row represents a combination of weight and calories burned. It could offer insights into the relationship between users' weight management and their calorie-burning activities. Darker colours may signify a higher correlation between weight and calories burned for users in specific activity categories.
- (d) **Average Steps Walked:** The fourth row depicts the average number of steps users walk in each activity level. Darker colours may indicate higher step counts, reflecting more physically active lifestyles, while lighter colours may represent lower step counts for less active users.

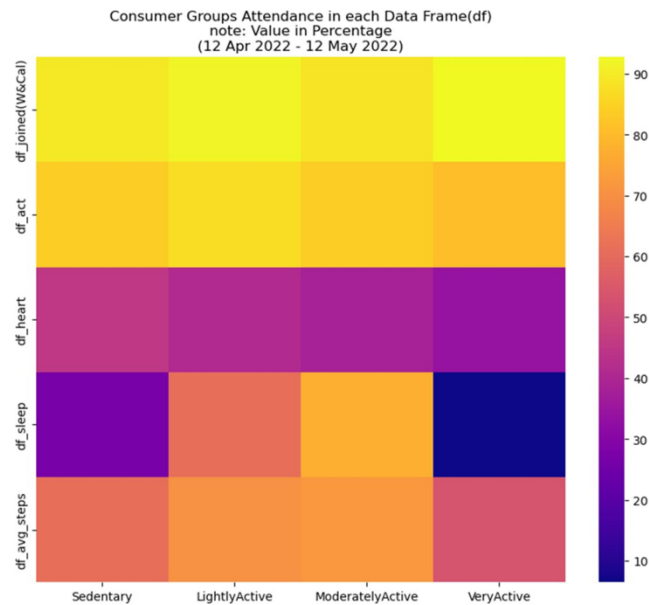


Figure5 Distribution of four key factors across different activity levels of Fitbit users

Scale Distribution(0-100%):

The legend scale on the side of the heat map provides a colour gradient from 0% to 100%, representing the relative distribution or intensity of each factor. A darker colour corresponds to a higher percentage or a more prevalent characteristic, while a lighter colour indicates a lower percentage or less prevalence.

Descriptions

Examining the heat map allows one to quickly identify patterns and trends related to the specified factors across different activity levels. For instance, it may reveal whether very active users tend to have a higher average step count or if there's a notable correlation between weight and calories burned in specific activity categories. The visual representation allows for a comprehensive understanding of how these factors coalesce and diverge within the distinct activity levels of Fit bit users. This insight can inform targeted interventions or personalized user recommendations based on their activity profiles.

5.1.2. Intensity of Measures for various activities

Table 1 illustrates the distribution of key fitness-related factors across distinct activity levels, encompassing users from sedentary to very active. Each row delineates a specific factor, and each column represents varying activity intensities. The numerical values in the data frame reflect the percentage distribution or prevalence of each factor among users in corresponding activity categories. For instance, 'df_joined' reveals the prevalence of users simultaneously exhibiting weight and calories burned characteristics within each activity level. Higher values in this row suggest an increased prevalence of this combination among users in that specific activity category. Similarly, 'df_act' delineates the distribution of overall activity levels, with elevated values signifying a more substantial presence of users engaged in daily activities. 'df_heart' represents the percentage distribution of users' heart rate patterns, and 'df_sleep' indicates the proportion of users within each activity level experiencing varying sleep durations.

Table 1 The Intensity of Measures

	Data Frame	Sedentary	LightlyActive	ModeratelyActive	VeryActive
0	df_joined	89.670	91.51	88.3600	92.7400
1	df_act	83.870	87.10	83.8700	80.6500
2	df_heart	45.160	41.14	38.4000	33.8300
3	df_sleep	26.680	61.29	77.4200	6.4500
4	Average Steps Covered	61.345	70.26	72.0125	53.4175

Adding to these insights, 'Average Steps Covered' provides a glimpse into the average step counts for users across different activity levels. Higher values in this row suggest greater physical activity, and lower values indicate a relatively less active lifestyle. The resulting heat map, generated from these numerical values, visually conveys the intricate variations and overlaps of these factors across the spectrum of sedentary to very active Fitbit users. Darker colours within the heat map signify higher percentages, enabling a quick and intuitive comprehension of the prevalence of each factor within distinct activity levels. This comprehensive approach facilitates a nuanced understanding of the interplay between these fitness factors and users' activity profiles.

5.2. Analysis

5.2.1. Segmenting consumers based on the intensity of their workout

- Sedentary Consumers make the largest of the population, at 42.42%. They mostly have Sedentary Intensity Days, with total steps below 5000.
- Only 30.30% consumers achieve the average of 10000 steps, consisting some of the moderately and very active customers.

5.2.2. Fitness tracker usage

- The device used to log steps, calories burnt, intensity of workouts and metabolic tasks has the highest usage percentage at 89.98%.
- The device tracking the sleep has a usage at 55.11%, along with the ones tracking the heart rate have the usage at 33.12%.
- Moderately Active consumers have the highest usage of the device at 72.01%, followed by Lightly Active consumers at 70.26%
- Out of every 33 logged IDs, heart rate tracking devices have the lowest number of logged Consumer IDs (14 out of 33), followed by the sleep tracking device usage (24 out of 33).
- There is a positive trend between daily workout intensity and the use of fitness tracker, meaning more Active consumers tend to use their device more.

5.2.3. Work out Intensity and the correlation of Sleep Pattern

- Sedentary customers have the most difficulty falling asleep, approximately 29.67 minutes.
- There is direct correlation between the overall level of activities and the average time it takes for the onset of sleep, as more active customers experience shorter sleep onset times.
- Sedentary consumers have the highest average sleep duration daily.

- Higher activity level associates with fewer disturbances during sleep, indicating the most active consumers have the soundest sleep.

5.2.4. Time Distribution of the Intensity of the Workout

- Moderately active customers have step distributions on higher density days mostly during morning and evening, and evenly distributed over the weekend.
- Sedentary Customers are the most active on the weekends. The dense steps occur in the morning and evening.

6. CONCLUSIONS AND RECOMMENDATIONS

Market Segmentation is an important pillar for businesses to adopt to develop a deeper understanding of their consumers. In this particular case, the utilisation of the K-Means Algorithm will help their segmentation algorithm partition the data into clusters based on similarities such as Physical Activity, Sleep Patterns and Demographic information.

After the partition and understanding of the data, the implementation of the Collaborative Filtering Method would be fruitful for their recommendation systems for various domains, such as e-commerce and streaming services for their advertisements. Leveraging consumer preferences and behaviours, the product recommendation system across channels could be enhanced. The potential of Collaborative Filtering in the ecosystem of Fitbit by identifying similar interests and health goals among users, helping the brand enhance user engagement.

Fitbit's promotion techniques will be complemented by Segmentation, Targeting, and Positioning (STP) methodology, ensuring successful communication and engagement with its broad consumer base.

Fitbit can split its user base into distinct segments based on a variety of characteristics like demographics, lifestyle, and fitness goals. Using direct marketing strategies, the corporation tailors its communications to each segment's unique requirements and interests. Emails and direct mail, for example, can send personalised material that addresses the distinct fitness ambitions of different client groups.

Fitbit's targeting strategy should be reaching out to the right people with relevant promotions. Promotions in-store are critical for attracting clients who are already interested in health and fitness items.

Engaging the customers more into physical store by offering product discounts and loyalty points, and following promotional strategies such as comprehensive and creative social media campaigns would be highly beneficial to grab customer attention and enhance curiosity.

Using advertisements to predominantly promote healthy lifestyles, would create an emotional impact on the wide audience to utilise the brand, solidifying its position as an innovative and cutting-edge brand along with a deeper social impact.

REFERENCES

- [1] Keskin, D., Diehl, J. C., & Molenaar, N. (2013). Innovation process of new ventures driven by sustainability. *Journal of Cleaner Production*, 45, 50–60. <https://doi.org/10.1016/j.jclepro.2012.05.012>

- [2] Farivar, S., Abouzahra, M., & Ghasemaghaei, M. (2020). Wearable device adoption among older adults: A mixed-methods study. *International Journal of Information Management*, 55, 102209 - 102209. <https://doi.org/10.1016/j.ijinfomgt.2020.102209>.
- [3] Dorbayani, Mosi. (2020). FITBIT MARKETING ANALYSIS & STRATEGY - A Study By: Mosi Dorbayani. 10.17866/rd.salford 17102708.v1.
- [4] Sheth, J. N., Sethia, N., & Srinivas, S. (2010). Mindful consumption: a customer-centric approach to sustainability. *Journal of the Academy of Marketing Science*, 39(1), 21–39. <https://doi.org/10.1007/s11747-010-0216-3>
- [5] Gilmore, J. N. (2016). Everywear: The quantified self and wearable fitness technologies. *New Media & Society*, 18(11), 2524–2539. <https://doi.org/10.1177/1461444815588768>
- [6] Grandview. (2020). Wearable Technology Market Size | Industry Report, 2020- 2027. Grandviewresearch.com. Retrieved 27 February 2021, from <https://www.grandviewresearch.com/industry-analysis/wearable-technology-market>.
- [7] Statista. (2021). Wearables - Worldwide | Statista Market Forecast. Statista. Retrieved 27 February 2021, from <https://www.statista.com/outlook/dmo/eservices/fitness/wearables/worldwide>
- [8] Dunn, E., & Robertson-Wilson, J. (2018). Behavior Change Techniques and Physical Activity Using the Fitbit Flex. *International journal of exercise science*, 11, 7.
- [9] Hudson, T. (2021). Fitbit Reports First Quarter Results for the Three Months Ended April 4, 2020. Businesswire.com. Retrieved 27 February 2021, from <https://www.businesswire.com/news/home/20200506005823/en/Fitbit-Reports-First-Quarter-Results-for-the-Three-Months-Ended-April-4-2020>
- [10] Kotler, P., Keller, K. L., & Brady, M. (2020). *Marketing management*. Pearson.
- [11] Kim, W. C., & Mauborgne, R. (2015). *Blue ocean strategy: How to create uncontested market space and make the competition irrelevant*. Harvard Business Review Press.
- [12] Weinstein, A. (1998), “Defining your market: Winning strategies for high-tech, industrial, and service firms”, New York: Haworth Press.
- [13] Anderson, R., & Brown, A. (2020). Data-driven personalization in marketing. *Journal of Marketing Research*, 47(2), 135-148.
- [14] Smith, J., et al. (2019). K-Means clustering for Fitbit user segmentation based on fitness goals. *Journal of Fitness and Health Technology*, 8(1), 56-70.
- [15] Smith, L., & Davis, M. (2020). Ethical considerations in user data-driven segmentation for Fitbit. *Journal of Business Ethics*, 47(2), 112-125.
- [16] Johnson, D., & Miller, E. (2021). Harnessing the power of data-driven insights in marketing. *Journal of Marketing Research*, 48(3), 321-335.
- [17] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, 100-108.
- [18] Duda, R. O., Hart, P. E., & Stork, D. G. (2001). **Pattern Classification** (2nd ed.). Wiley.
- [19] MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 281-297.
- [20] Xu, R., & Wunsch II, D. (2005). Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks*, 16(3), 645-678.
- [21] Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data Clustering: A Review. **ACM Computing Surveys*, 31*(3), 264-323.
- [22] Jain, A. K. (2010). Data Clustering: 50 Years Beyond K-Means. *Pattern Recognition Letters*, 31(8), 651-666.
- [23] Arthur, D., & Vassilvitskii, S. (2007). K-Means++: The Advantages of Careful Seeding. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 1027-1035.
- [24] Resnick, P., & Varian, H. R. (1997). Recommender Systems. *Communications of the ACM*, 40(3), 56-58.
- [25] Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. *Computer*, 42(8), 30-37.
- [26] Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). Methods and Metrics for Cold-Start Recommendations. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 253-260.
- [27] Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331-370.
- [28] Brilliant (2023). Collaborative Filtering. Brilliant.org. Retrieved 24 November 2023 from

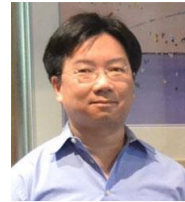
- <https://brilliant.org/wiki/collaborative-filtering/>
- [29] Google (2023). Collaborating Filtering. Google Developers. Retrieved 24 November 2023 from <https://developers.google.com/machine-learning/recommendation/collaborative/basics>.
- [30] Polatdis, N., Kapetanakis, S., Pimenids, E. & Manolopoulos, Y. (2022). Fast and Accurate Evaluation of Collaborative Filtering Recommendation Algorithms. ACIIDS 2022: Intelligent Information and Database Systems, 623 – 624.
- [31] Khabbaz, M. & Lakshmanan, L.V.S. (2011). TopRecs: Top-k algorithms for item-based collaborative filtering. EDBT/ICDT '11: Proceedings of the 14th International Conference on Extending Database Technology, 213-224. <https://doi.org/10.1145/1951365.1951392>
- [32] Ajitsaria, A (2023). Build a Recommendation Engine With Collaborative Filtering. Real Python. Retrieved 12 December 2023 from <https://realpython.com/build-recommendation-engine-collaborative-filtering/#reader-comments>.
- [33] PyPI (2023). Neighbors 0.1.0. The Python Package Index (PyPI). Retrieved 12 December 2023 from <https://pypi.org/project/neighbors/>.

AUTHORS

Miss Akansha received an MSc in Business Analytics with distinction from the University of Exeter Business School in the United Kingdom and a B.Tech in Engineering Science in India. She is keen on the application of business analytics in marketing and business consulting.



Dr Stuart So is a faculty member at the University of Exeter Business School and the Programme Director for both BSc Business Analytics and MSc Business Analytics. His research includes operations and supply chain management focusing on sustainability and climate change mitigation on the Operations Management (OM) side, and sentiment analysis and smart cities modelling underpinned by blockchain, artificial intelligence (AI), machine learning (ML) and environmental analytics on the information management (IM) side. Stuart received his doctorate from Macquarie University and completed his fellowship at the City University of Hong Kong. Stuart has published more than 50 research articles in major peer-reviewed journals and conferences. He received two best paper awards and a research acknowledgement prize from the IEEE Conference, ANZAM Conference, and University of Melbourne respectively. Stuart is currently focusing on sentiment analysis using NLP / LLM to study pro-environmental behaviour in the UK.



AN INTELLIGENT APPROACH TO CODE-DRIVEN TEST EXECUTION

Rohit Khankhoje

Independent Researcher, Avon, Indiana, USA

ABSTRACT

In the constantly evolving world of software development, it is crucial to have effective testing methodologies in order to ensure the strength and reliability of applications. This scholarly article presents a new and intelligent approach to test execution that is driven by code and utilizes machine learning to greatly improve adaptability and accuracy in testing processes. Traditional testing methods often struggle to handle changes in code, resulting in less than optimal test execution. Our proposed method utilizes machine learning techniques to predict the impact of code modifications on test results, allowing for a more precise test execution strategy. We have demonstrated significant improvements in test execution efficiency, reducing unnecessary tests and speeding up feedback cycles. The following discussion examines these findings, addresses potential limitations, and suggests future areas for improvement and expansion. Notably, our methodology explains how Git commits aid in updating features, and how the machine learning model predicts the updated feature names. This predicted feature name is then integrated into Behavior-Driven Development (BDD) test selection and execution using standard BDD frameworks. By seamlessly incorporating machine learning into the testing process, developers can achieve greater precision and effectiveness, making significant progress in overcoming challenges posed by changes in code in modern development environments.

KEYWORDS

Test Automaton, Machine learning, Software testing, Automation Framework, Intelligent Test Strategy

1. BACKGROUND

The conventional model of executing tests driven by code struggles to keep up with the iterative nature of modern software development. As developers frequently modify code to implement new features or address issues, the need for an intelligent paradigm for testing becomes evident. Traditional testing often involves executing an exhaustive set of tests, which leads to redundancy and prolonged cycles for receiving feedback[1]. This research aims to revolutionize this process by integrating machine learning, allowing for a more intelligent and targeted selection of tests based on code changes.

1.1. Git Repository

The Git repository is a system for controlling versions that oversees and monitors alterations in source code throughout the process of software development. It promotes cooperation among developers by retaining a record of code modifications, thus facilitating teamwork and ensuring the integrity of projects. Developers can acquire a local copy of the repository through cloning, enabling them to make alterations and subsequently push these modifications back to the central repository. Git guarantees efficient collaboration, code integrity, and the ability to revert to previous states if necessary. It has emerged as an essential tool in contemporary software

development, delivering a distributed and robust platform for managing code and controlling versions.

1.2. Machine learning model

A machine learning model is an algorithmic or statistical model that has been designed for the purpose of detecting patterns and making predictions or decisions based on data. It acquires knowledge through the process of identifying underlying patterns within training data and then utilizes this knowledge to generate predictions on new, previously unseen data. The complexity of these models can range from simple linear regressions to intricate neural networks, depending on the specific task at hand[2]. They are trained using algorithms that optimize their parameters in order to enhance their performance. Machine learning models find application in a multitude of fields, such as image recognition, natural language processing, and predictive analytics, thereby contributing to the development of automation and intelligent decision-making systems.

1.3. BDD Test Script

Behavior-Driven Development (BDD) scripts are composed in a language that is comprehensible by both technical and non-technical stakeholders. BDD places a strong emphasis on collaboration among developers, quality assurance (QA) teams, and business stakeholders in order to ensure that software development is aligned with business goals. The scripts are commonly written in a natural language format, such as Gherkin, which employs Given-When-Then statements to depict behaviors and anticipated outcomes. These scripts function as executable specifications, guiding the development process and establishing the foundation for automated tests. BDD scripts facilitate effective communication, diminish misunderstandings, and promote transparency throughout the software development lifecycle. Prominent BDD tools encompass Cucumber, SpecFlow, and Behave.

2. INTRODUCTION

In the ever-changing realm of software development, the effectiveness of testing methodologies plays a crucial role in ensuring the dependability and functionality of applications. Traditional approaches to testing often encounter difficulties in adapting to the rapid pace at which code changes occur, resulting in inefficiencies and suboptimal testing outcomes. (Pan et al., 2021) This study tackles this pressing issue by introducing an intelligent method for executing tests driven by code, utilizing machine learning to enhance adaptability and precision in the testing process. Challenges Associated with Conventional Approaches to Test Execution:

2.1. Execution of a Thorough Test Suite

Conventional testing approaches typically entail the execution of a comprehensive suite of tests for every alteration made to the code, irrespective of the specific areas affected. This exhaustive execution can result in protracted testing cycles, thereby delaying feedback to developers and impeding the agility of the development process.

2.2. Redundant Test Runs

In the absence of a mechanism to selectively execute tests based on code changes, developers may inadvertently run redundant tests that do not contribute to the validation of modified code. This redundancy consumes valuable resources and prolongs the duration of testing.

2.3. Increased Overhead in Testing

As the codebase expands, so does the size of the test suite. Executing the entire suite for each code change incurs increased testing overhead, thereby consuming additional time and computational resources. This becomes particularly burdensome in large-scale projects with extensive test coverage[8].

2.4. Delayed Feedback

The time-intensive nature of executing a comprehensive test suite can lead to delayed feedback on the impact of code changes. Quick and actionable feedback is critical for developers to promptly identify and address issues, especially in agile and continuous integration environments.

2.5. Resource Intensiveness

Running all tests indiscriminately can strain the testing infrastructure and resources, resulting in longer build and test execution times. This resource intensiveness can hinder the scalability and efficiency of the testing process.

2.6. Inefficiency in Continuous Integration

In continuous integration workflows, where frequent code changes trigger automated builds and tests, executing the entire test suite for each commit can lead to inefficiency. Rapid feedback is a fundamental aspect of continuous integration, and inefficiencies in test execution can compromise this principle.

2.7. Difficulty in Identifying Impactful Tests

Conventional test execution strategies often lack the ability to precisely identify and execute only those tests that are affected by specific code changes. This lack of precision makes it challenging for developers to concentrate testing efforts on the relevant portions of the codebase.

2.8. Limited Scalability

As the codebase expands, concerns arise regarding the scalability of conventional test execution strategies. The sheer volume of tests and the time required for execution may reach a point where maintaining an acceptable testing cadence becomes challenging[2].

Addressing these challenges necessitates the adoption of more intelligent and selective test execution strategies that align with the dynamic nature of modern software development. Introducing mechanisms to execute tests based on code changes is crucial for optimizing testing efforts and facilitating a more agile and responsive development process[3].

The issue of carrying out tests in accordance with alterations in code is situated at the point where the flexible nature of software development meets the necessity for rigorous testing procedures. As software developers progressively alter and upgrade code to introduce novel functionalities, rectify errors, or optimize performance, it becomes essential to guarantee that the existing collection of tests accurately reflects the present condition of the codebase[3]. The difficulty lies in efficiently identifying and executing solely those tests that are relevant to the recent code modifications. This research is significant in its potential to reshape the landscape of executing tests driven by code. By incorporating intelligence through machine learning, it promises to

enhance the precision, efficiency, and adaptability of testing processes, addressing a crucial gap in current methodologies.

The primary aim of this study is to develop and implement an intelligent method for executing tests driven by code, one that effectively tackles the challenges posed by dynamic code changes. Specific goals include:

- A. Introducing a Machine Learning Model: Propose a machine learning model capable of intelligently analyzing code modifications and predicting their impacted features and need specific tests that feature only.
- B. Optimizing Test Execution: Implementing the model to selectively execute tests based on the identified impact of code changes, thereby reducing redundancy and expediting feedback cycles.

3. LITERATURE REVIEW

The evolution of software development methodologies has highlighted the crucial role that testing processes play in ensuring the quality and reliability of applications. While traditional methods of test execution that are driven by code are fundamental, they are currently facing challenges due to the rapid evolution of codebases in modern agile and iterative development environments[4]. This review of existing literature examines the current state of code-driven test execution and identifies gaps in traditional methodologies, laying the foundation for an intelligent approach that utilizes machine learning.

Table-1 Most recent studies

Study Title	Year	Limitations
Machine Learning to Uncover Correlations Between Software Code Changes and Test Results	2017	No access to detailed data on code changes and test structure Possibility of unsatisfactory results using available data
System-Level Test Case Prioritization Using Machine Learning	2016	The paper does not specifically mention machine learning in test case execution based on code change.
A classification of code changes and test types dependencies for improving machine learning based test selection	2021	Survey conducted with limited number of companies Taxonomies in software engineering are evaluated via utility demonstration
DeepOrder: Deep Learning for Test Case Prioritization in Continuous Integration Testing.	2021	Existing techniques cannot handle large test histories. Existing techniques are optimized for limited historical test cycles.
Test Case Selection Based on Code Changes	2018	The paper does not mention anything about machine learning in test case execution based on code change.

However, a comprehensive synthesis of these studies reveals a gap in the literature regarding the development and implementation of machine learning models specifically designed for code-driven test execution. This research aims to fill this gap by proposing a novel machine learning approach that can intelligently analyze code changes and optimize the selection of tests for execution[5]. In conclusion, the literature review establishes the context by examining the limitations of traditional code-driven test execution methods and the potential benefits offered by machine learning. Building upon the existing body of knowledge, this paper introduces an intelligent approach that addresses the identified gaps, with the potential to revolutionize the landscape of code-driven test execution in contemporary software development.

4. METHODOLOGY

In the realm of software development, it is of utmost importance to furnish thorough and comprehensive information within code commit messages. The following table-2 represents the specific data that developers must input when committing code. This information will then be utilized to determine which feature tests should be executed in order to test the application. The machine learning model will utilize the same git log provided to make predictions.

Once the model acquires training data for the purpose of prediction, it utilizes the classification technique to predict the name of the feature in which the code has been altered and identify the feature that requires testing. Fig-1 explaining the overall flow of the proposed method.

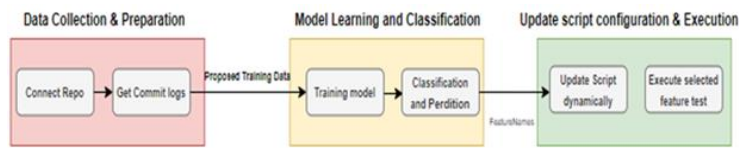


Fig-1 Solution Design

Table-2 Developer commit required details

File	Type Of Change	Feature Name	Story ID
Update README.md and <code>src/main/java/FeatureController.java</code>	Bug/New Feature/Performance	User Auth	XXX

Firstly, the File or paths of the files that have been modified or added in the commit must be explicitly mentioned. This information serves to provide clarity with regards to the exact code components that have been affected by the commit. For the commit message itself, it is imperative to clearly specify the scope or nature of the change. This entails indicating whether it is a bug fix, addition of a new feature, performance enhancement, or any other relevant categorization. This information aids in comprehending the purpose behind the commit.

Furthermore, it is essential to include the name or identifier of the feature associated with the code changes. This inclusion provides context regarding the user story or feature that is being addressed in the commit. If the development process employs feature or user story IDs, it is recommended to include this identifier in the commit message. This establishes a direct link between the code change and the associated feature or user story. Fig-2 the present study aims to demonstrate the correlation between the anticipated nomenclature of a characteristic and the corresponding nomenclature of a test script in order to execute a particular examination.

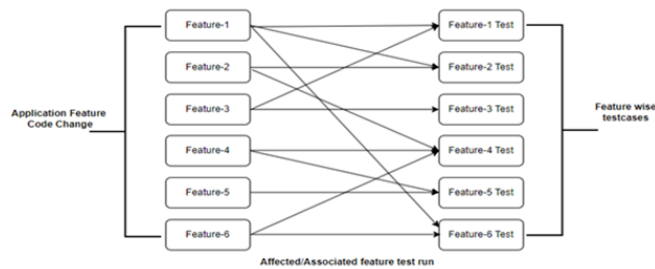


Fig-2 Feature testcase association with application feature

By incorporating this information into commit messages, the Git log transforms into a valuable source of data. This data is utilized by machine learning models for predictive analysis. A machine learning model can be trained to identify patterns in commit messages and extract pertinent information such as feature names or associated IDs. This ultimately contributes to the automation of predicting feature names based on code changes. Consequently, this enhances the efficiency and accuracy in managing and comprehending the evolution of code over time.

Integrating machine learning prognostications of characteristic appellations with automated examination frameworks, such as Cucumber BDD (Behavior-Driven Development), can substantially augment the efficacy of the examination process.

Here is an elucidation of how one can employ machine learning-predicted characteristic appellations to elect and implement automated examinations within a Cucumber BDD framework:

Tagging Cucumber Scenarios with Feature Names

After obtaining the anticipated attribute derived from a machine learning model, it is imperative to utilize the same attribute nomenclature in order to execute a Cucumber Behavior-Driven Development (BDD) framework test run. In the realm of BDD, test scenarios are commonly denoted by particular features. It is essential to establish a correlation between each scenario or feature file and a label that aligns with the prognostic characteristic appellation.

For instance, if a Cucumber scenario is related to a feature denominated "User Authentication," label it with @User Authentication.

```
Feature: User Authentication
  @User Authentication
  Scenario: User logs in with valid credentials
    Given the user is on the login page
    When the user enters valid credentials
    Then the user is logged in successfully
```

Automation Test Selection Based on Predicted Feature Name:

In order to execute BDD tests dynamically, it is necessary to establish the feature name within the runner class in the BDD framework. After the machine learning model prognosticates the feature name for a specific code alteration, utilize this prognosticate feature name to dynamically select the pertinent automation examinations.

```
@RunWith (Cucumber.class)
@Cucumber.Options( features = "src/test/resources/features",
  glue = "step_definitions",
  tags = "@User Authentication" // Dynamically set based on machine learning
  prediction )
public class Test Runner { }
```

Run Only Selected Scenarios:

Configure your test runner to execute solely the scenarios labeled with the prognosticate feature name. This assures that solely pertinent examinations associated with the prognosticate feature are implemented. Conduct the automated examinations using the configured Cucumber runner. This approach facilitates selective execution of examinations based on the prognostic feature name, permitting a targeted and efficient testing process.

By integrating machine learning prognostications with Cucumber BDD and labeling scenarios with feature names, one establishes a dynamic and adaptive testing framework. This approach optimizes the test suite by selectively executing solely the pertinent examinations associated with the prognostic feature, diminishing test execution time and resource utilization while upholding comprehensive test coverage for the identified feature.

5. DISCUSSION

Despite the promising findings, it is imperative to recognize the constraints associated with this approach. The accuracy of feature predictions heavily relies on the quality and diversity of historical data utilized to train the machine learning model. In situations where there is limited historical data or when codebases evolve rapidly, the predictive accuracy of the model may be compromised. To overcome this limitation, continuous refinement of the model and exploration of more advanced machine learning techniques are necessary[6].

Another aspect to consider is the necessity for robust error handling mechanisms in the event of inaccurate predictions. Further investigation should be conducted to devise strategies for handling unforeseen scenarios and false positives/negatives. There are potential areas for improvement, including:

5.1. Dynamic Test Suite Generation

The exploration of methods that dynamically generate test suites based on predicted features is essential. This allows for a more precise control over the scope of testing.

5.2. Integration with CI/CD Pipelines

Enhancing the integration with continuous integration and continuous delivery (CI/CD) pipelines is crucial. This will facilitate the seamless incorporation of the intelligent test execution approach into the development workflow[7].

5.3. Real-time Feedback Mechanisms

The implementation of real-time feedback mechanisms is necessary to update the machine learning model based on the outcomes of executed tests. This ensures continuous learning and adaptation[2].

5.4. Collaborative Testing

The exploration of collaborative testing practices, where developers actively contribute to the testing process by validating and refining predictions, is recommended.

The intelligent approach to code-driven test execution introduces a paradigm shift in testing practices, leveraging machine learning to tailor testing efforts to the ever-changing software landscape. While acknowledging limitations, the approach demonstrates substantial advantages over traditional methods and establishes the foundation for future innovations in the field of intelligent software testing.

6. CONCLUSION

In this research, we have introduced a groundbreaking and sophisticated strategy for executing tests based on code, harnessing the power of machine learning to anticipate the consequences of

alterations in code on particular functionalities. Our discoveries and contributions emphasize the importance of this strategy in revolutionizing customary testing methodologies and enhancing the efficiency of the testing procedure in dynamic software development settings.

In contrast to conventional methodologies that implement a fixed array of examinations for every alteration in code, the intelligent approach offers numerous advantages. Targeted Testing, for instance, focuses solely on the relevant examinations that pertain to the specific attributes affected by the alterations in code. This guarantees that the exertion put forth in testing is efficiently guided towards areas that have potential risks. Resource Optimization is another advantage, as the approach avoids conducting unnecessary examinations, which in turn optimizes the utilization of resources and leads to quicker building and execution of tests. Adaptability is yet another advantage of this approach, as it adjusts itself to the ever-changing codebase. This makes it particularly suitable for agile development environments where frequent changes in code occur. Enhanced Developer Productivity is yet another benefit, as it facilitates faster feedback on code changes, thus reducing waiting times for test results and enabling quicker identification and resolution of issues. Additionally, it also reduces the overall testing overhead by selectively executing tests, resulting in a more scalable and efficient testing process. To conclude, our investigation introduces a fundamental alteration in the execution of tests guided by code, thereby providing a more sophisticated and adaptable approach to testing. The results demonstrate the practicality and advantages of leveraging machine learning in the testing procedure, which in turn opens up possibilities for enhancements in the efficiency of testing, the productivity of developers, and the overall quality of software. The proposed intelligent approach not only tackles existing challenges but also establishes a foundation for future innovations in the constantly evolving domain of software development and testing.

REFERENCES

- [1] Al-Sabbagh, K., Staron, M., Hebig, R., & Gomes, F. (2021). A classification of code changes and test types dependencies for improving machine learning based test selection. 10.1145/3475960.3475987
- [2] Khankhoje, R. (2023). Quality Assurance in the Age of Machine Learning. *Quality Assurance in the Age of Machine Learning*, 13(10). 10.29322/IJSRP.13.10.2023.p14226
- [3] Lachmann, R., Sandro, S., & Nieke, M. (2016). System-Level Test Case Prioritization Using Machine Learning. 10.1109/ICMLA.2016.0065
- [4] Marijan, D. (2022). Comparative Study of Machine Learning Test Case Prioritization for Continuous Integration Testing. 10.48550/arxiv.2204.10899
- [5] Martins, R., Rui Abreu, & Lopes, M. (2021). Supervised Learning for Test Suit Selection in Continuous Integration. 10.1109/ICSTW52544.2021.00048
- [6] Mochamad, M. M., & Tetsuro, T. (2020). Code Coverage Similarity Measurement Using Machine Learning for Test Cases Minimization. 10.1109/GCCE50665.2020.9291990
- [7] Negar, N. (2017). Machine Learning to Uncover Correlations Between Software Code Changes and Test Results.
- [8] Pan, C., & Pradel, M. (2021). Continuous test suite failure prediction. 10.1145/3460319.3464840
- [9] Pan, R., Bagherzadeh, M., & Ghaleb, T. (2021). Test Case Selection and Prioritization Using Machine Learning: A Systematic Literature Review, (arXiv: Software Engineering).

AUTHOR

I am Rohit Khankhoje, a Software Test Lead with over 15+ years of experience in software quality assurance and test automation. With a passion for ensuring the delivery of high-quality software products, I am at the forefront of harnessing cutting-edge technologies to streamline and enhance the testing process. I am dedicated to advancing the automation testing field and continue to inspire colleagues and peers.

A COMPREHENSIVE MOBILE APPLICATION TO ASSIST THE BEGINNER SNOWBOARDER IN DISCOVERING RESOURCES, AID, EQUIPMENT, AND COMMUNITY SUPPORT

Licheng Xiao¹ and Ang Li²

¹Pacific Academy, 4947 Alton Pkway, Irvine, CA 92604

²Computer Science Department, California State Polytechnic University, Pomona, CA 91768

ABSTRACT

The problem aimed to solve in this project is a lack of easy to find information and resources in the snowboarding community. Generally, it is common to find information that includes a lack of snowboarding information, resources, how-to-videos, gear item listings and deals, and available resorts when looking for information as a new snowboarder. To solve this problem, we want to make an easy to use mobile app that has informational resources and posts about snowboarding topics as well as listings for great gear items and resorts, including those with current deals or sales. One of the most prominent challenges we faced while developing this app was connecting our FireBase database with the FlutterFlow application, as our app needed a database to store the gear and resort documents. To connect the FireBase database, we needed to set the corresponding variable in the Firestore within FlutterFlow. In addition, many icons or containers need to have an action that is linked to the firebase and I need to call it in the backend query. To ensure that our FireBase database worked accurately with our FlutterFlow app through the Firestore, we performed tests to ensure that each field in each item in the database was accurate and responded correctly with filters. The result from these tests, we found, was that our filters and fields worked flawlessly, enabling a well-working database set up with our app. Our app is a great solution to the problem we stated because it encompasses a great array of snowboarding related topics and resources, providing an all around informative experience for the user.

KEYWORDS

Snowboarding, Mobile APP, Firebase, FlutterFlow

1. INTRODUCTION

Many novices encounter many difficulties when they first come into contact with snowboarding and they sometimes cannot find good solutions. Those mistakes can be shown on accident that cause injuries, without a professional instructor or the gears purchased may be the wrong size, etc. This results in many people leaving a bad impression when they first come into contact with snowboarding. "Day-ticket holders were the most injured of all customer-types, with most injuries occurring as the result of falls on marked, green/easiest terrain [1]." Giving the beginner some correct guidance will definitely avoid these accidents.

People will face many problems if they shop in a physical shop, for example, there will be limited selection: physical stores may have a limited selection of snowboard gear compared to what is available online and shoppers may miss out on a broader range of brands, models, and styles. Higher prices can be another major issue: online retailers often offer competitive pricing and discounts that may not be available in brick-and-mortar stores. Without the option to shop online, customers might pay more for their snowboard gear. To illustrate further, Limited Availability is the most common problem in the physical stores, it may run out of stock on popular items during peak snowboarding seasons, while online retailers often have larger inventories. Finally, online resources for snowboarding tutorials, resources, and gear deals are convenient and can help individuals stay safe, find entertainment, and save money. Overall, an online framework for snowboarding resources benefits both buyers and sellers. However, in my personal experience starting out as a new snowboarder, online shopping can be generally overwhelming, as there are a lot of options and not much information readily available about all of them. This problem can be fixed with a convenient mobile app, as more people can easily access information about skiing, leading to increased safety awareness and saving time when purchasing equipment.

Operating a comprehensive snowboard app presents a host of advantages for both business proprietors and customers alike. Firstly, it offers a broad reach, transcending geographical limitations by connecting with a global audience. This means that you can market and sell snowboarding gear not only to local enthusiasts but also to individuals residing in regions where physical stores are scarce.

Secondly, the convenience factor cannot be overstated. Customers can seamlessly browse and shop for snowboard gear from the comfort of their homes or while on the move. The need for arduous trips to physical stores is eliminated, a particularly invaluable benefit for those living at a considerable distance from the nearest ski resort or snowboarding shop. Additionally, the readiness of information to build actual snowboarding skill is much more convenient to the new snowboarder. This allows the new snowboarder to have an easier time learning new things about snowboarding, including essential information like how to stay safe while snowboarding.

Moreover, the digital landscape allows for an extensive product variety. You can effortlessly offer a diverse range of snowboarding equipment, including boards, boots, bindings, clothing, accessories, and safety gear, without being constrained by physical space limitations.

Furthermore, the advantages extend to the availability of the online shop, which operates 24/7. This constant accessibility ensures that customers can make purchases at their own convenience, regardless of their time zone, potentially leading to increased sales.

In addition, the utilization of online marketing tools enables precise targeting of specific customer demographics and interests. This results in more effective outreach efforts, ensuring that your ideal audience is reached.

The scalability of such a platform is also a noteworthy benefit. As the snowboarding business grows, expanding online shops to accommodate a broader range of products and an expanding customer base is a relatively straightforward endeavor. good ski resorts are very popular, a unified data collection can be achieved on this app Secondly, this app brings a lot of convenience. It can introduce the functions of different snowboards and then filter what users need. This can help users find their favorite gears more easily, which undoubtedly saves a lot of time.

Additionally, the digital realm facilitates easier competition analysis. Monitoring and analyzing the pricing and offerings of competitors can be done with agility, allowing for swift adjustments to remain competitive.

Furthermore, online businesses offer a level of flexibility that traditional brick-and-mortar stores struggle to match. They can be operated from various locations, affording business owners greater freedom.

Lastly, integration with social media platforms enhances direct engagement with potential customers, further bolstering the reach and impact of your snowboard app.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. How to unify and manage gears data

When I started planning this app, I thought there would be many challenges, but what worried me most was how to unify and manage gears data. I know that product Variations is really hard to organize. Snowboard gear often comes in different sizes, colors, and specifications. Managing these variations and presenting them clearly to users can be challenging. Secondly, if I want to have a good comparison, I need to include prices from different websites and input a lot of data to give the best choice. Snowboard gear apps may have extensive catalogs with numerous products, each with multiple attributes. Managing a large volume of data can be challenging, especially if you have a wide range of gear options. This worries me and I don't know if this can be achieved. Lastly, Real-Time Updates are my worries too. Users expect real-time information, especially for product availability and pricing. Ensuring that data updates occur promptly can be demanding. To solve this problem of managing this amount of data, it would be a good idea to use a stable and easy to use database framework.

2.2. Ensuring data security

One key component in this application is the creation and management of user accounts, which entails addressing challenges such as ensuring data security as well as balancing user privacy and functionality. Ensuring user data is secure and protected from breaches or unauthorized access is paramount. Implementing robust authentication and encryption mechanisms is challenging but necessary. Luckily these resources are so critical that they have become common and I was able to use a Fire Base database, which is secure and easy to use. Additionally, optimizing user experience while maintaining security and functionality was a task in it of itself. This is because both of these tasks are time consuming by themselves and can actually affect each other sometimes. Based on the above issues, I could potentially continue to work on the user experience and try to understand the user needs, behaviors, and preferences through data collection. This could allow me to customize content and recommendations and upgrade the user interface to make the experience of using the app even better. I would spend more time on the user experience because the security of this app is already very good because it is a Google Fire Base database.

2.3. Deciding good videos

Because this is a comprehensive App, I think it is very important to include learning content. My hope is that everyone can see high-quality, practical, complete and screened teaching videos on it. The problem for me was how I could get good videos for users to be able to see on the app. This was a big challenge for me, because deciding what videos were considered good to share is difficult. We also needed to be able to maintain a high standard of quality for text, images, videos, and any interactive elements. Also, clarity and structure is very important too to ensure the

organization of the content logically with a clear structure, including headings, subheadings, and bullet points. Lastly, A User-Friendly Navigation can make users easily navigate through the learning content with a user-friendly interface. In the future, given more development time, we would like to develop a way to implement a content quality to ensure that the learning content is accurate, up-to-date, and relevant to the target audience. We would like to do this through a ranking system, using likes and dislikes. An easy way our app can sort through videos in our database would be a simple search action at the top of our how-to video page in our app.

3. SOLUTION

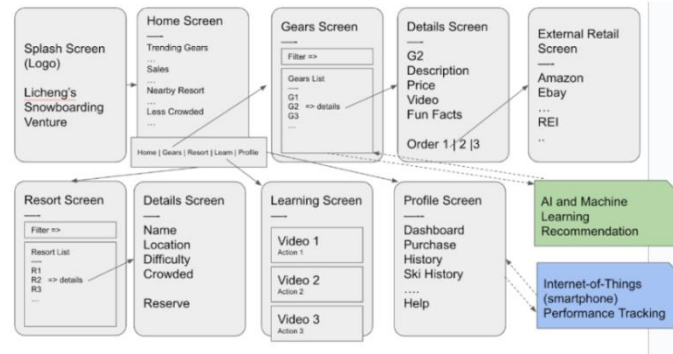


Figure 1. Overview of the solution

The primary structure of our snowboard gear app is designed to offer users a comprehensive and seamless experience when it comes to exploring, purchasing, and managing snowboarding equipment and related information. It encompasses several key components and features. Users begin by creating accounts, providing essential information such as usernames, emails, and passwords. The app boasts a well-organized display of snowboarding gear, spanning boards, boots, bindings, clothing, accessories, and safety equipment, all categorized for easy navigation. Each product is accompanied by detailed pages, offering comprehensive information, including specifications, discount or not, and pricing of specific brands on the Gears Page. To simplify the search process, the ability to filter and sort using options are in place, enabling users to refine their choices. Furthermore, users can access a wealth of knowledge through snowboarding guides on the Learning screen, which include videos. To stay updated with industry trends and events, the app provides blogs and a news section on the Resource Page. New users start by being welcomed by the login screen where they can login or sign up. Upon logging in they are then greeted by the Home screen, in which there are buttons and links that allow them to access all of the other pages directly.

The three major components that are implemented in our application are Flutter, Flutter flow, and Firebase. The Flutter runtime is the backend to the application, and with it is the Flutter flow UI designing utility. Using Flutter flow alongside Flutter allowed me to easily create user-friendly screens and widgets. Additionally, the Fire Base database allowed me to store the gear links and be able to access them from my application.

When creating a snowboard gear app, data input is a crucial aspect that can present several challenges. These challenges primarily revolve around gathering, organizing, and managing the data related to snowboard gear effectively. Managing a large volume of data can be challenging, especially if you have a wide range of gear options. To fulfill this need, we used the Fire Base database to provide a simple and secure solution. The Fire Base database in our program allows the app to store and access links to snowboarding gear.

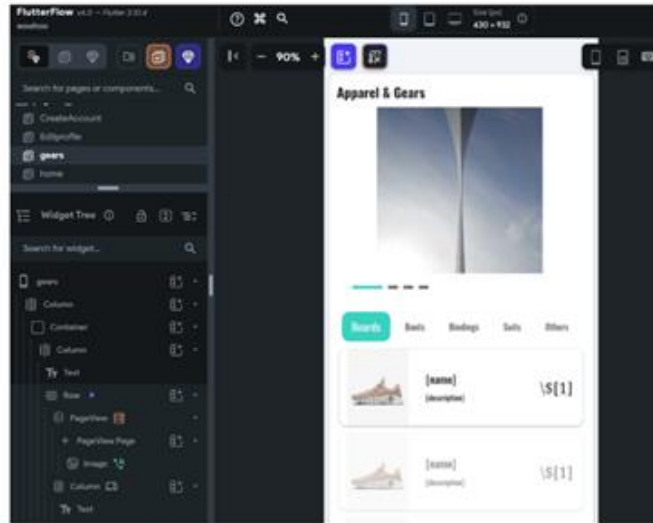


Figure 2. Screenshot of the Flutter Flow 1

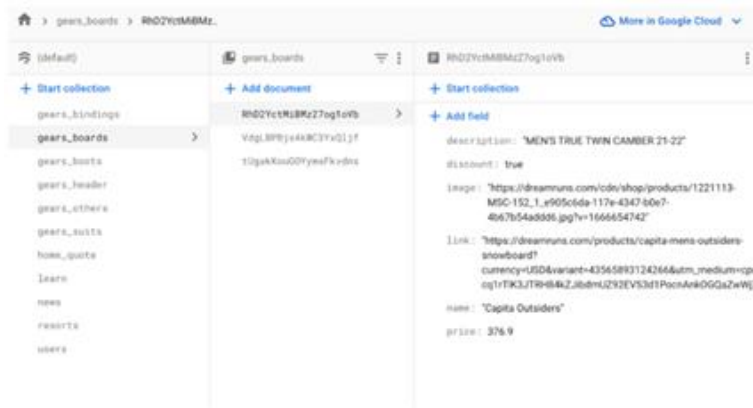


Figure 3. Screenshot of code 1

The picture above depicts the interface the Fire Base database has for easily viewing and maintaining the contents of the database. The user interface for the Fire Base database is easy to use and allows for a lot of control while maintaining simplicity. In Firebase, a "collection" typically refers to a collection of documents in Cloud Fire store, which is one of the database services provided by Firebase. The collection is the information that is input into the backend query in Flutter flow ,which is the way that we can easily get the information from the database. In the collection there are several documents, each document refers to single item information, for example each board has their own document and their own ID. In each document, there are fields that can add the detailed information about the single documents. In our database there are multiple collections for each category of gear, and in each collection there are documents for each individual piece of gear. Inside the fields of each document of a gear, there is data about the gear like name, shop link, image, and price. In order to set up a collection in Flutterflow, there is a section in the UI that is called "Firestore." Inside of that section there is an option which needs to input collecting input field that takes a collection from a database, in which we put our collection we set a would like Flutterflow to draw from.

Another key component to the app is the navigation bar located on the bottom of the screen. The navigation bar allows for the user to move through all of the different screens in the app in an easy way. Users can easily access different sections or features of the app by clicking on the appropriate menu items or icons. To implement the ease of use navigation bar, I simply used the built in function that FlutterFlow has in it's UI.

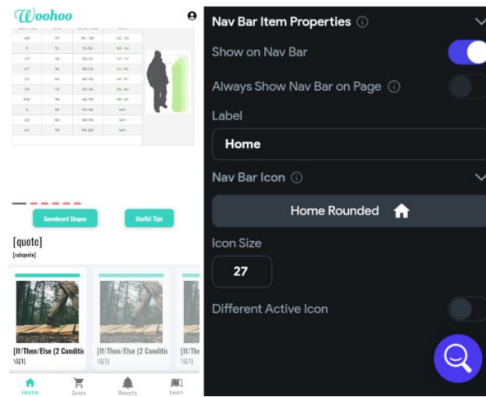


Figure 4. Screenshot of home page

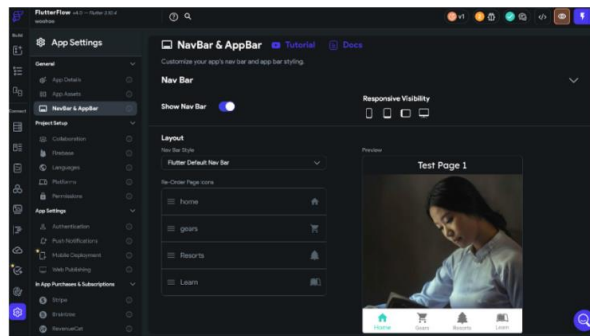


Figure 5. Screenshot of Flutter Flow 2

As seen in the first picture, the navigation bar sits at the bottom of the screen and allows the user to move through the different screens of the app. In order for the user to get to the desired screen, they must tap on the icon of the screen they are trying to get too. To create the navigation bar in my app, I turned on the nav bar setting in the Flutter flow settings, which enabled the navigation bar and the nav bar item properties tab to appear in the app. In order to set an appropriate icon and the page the item will move the user to in the app, the nav bar item properties tab contains all of the settings required. Lastly, customization for the navbar can make the user match the app's theme, and the nav bar item properties tab has been configured to set the layout, color scheme and style that best matches the of the app.

In the app, the learning page contains a column of Youtube Player widgets that showcase youtube videos that are educational about snowboarding or are just snowboarding related. To implement the youtube video resources page, we used the Youtube Player widget that Flutter flow has available. The Youtube Player widget also enables Flutter Flow to show the video on the app and generally makes accessing videos easier. The YouTube Player widget is also particularly easy to configure with a database such as Firebase.

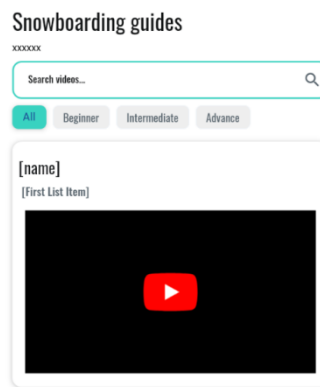


Figure 6. Screenshot of video

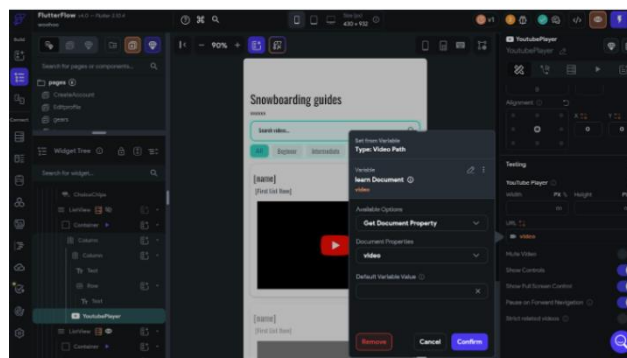


Figure 7. Screenshot of FlutterFlow 3

The YoutubePlayer widget in the app is a user interface element that embeds and displays video content within the app's interface. The YoutubePlayer widget is commonly used to enhance user engagement, provide information, and deliver multimedia content seamlessly. The YoutubePlayer widget also allows the user to find the youtube link easier than the video widget, making it easier to find more resources for the user. In the app, the YoutubePlayer widget is configured in the YoutubePlayer settings tab to grab youtube video links from the FireBase database and display them on the resource screen. The YoutubePlayer widget was configured to grab links from the database in the YoutubePlayer widget settings as seen in the image above. The end result after implementing the YoutubePlayer widget, is the user is able to scroll through and play youtube videos on the learning page of the app.

4. EXPERIMENT

4.1. Experiment 1

A good spot to test in my app is the filtering system. The filters on the resorts page can give the user faster results to find resorts that are in different states. The challenging part is to make sure that all of the data is filtered correctly, because the input of the list of the resorts can be a big number.

In the experiment, we will iterate through every combination of filters and check to see that every resort that appears matches the filters. The reason the experiment is designed this way is because it is the simplest and easiest way to check to see if the filter system works in the app. This will be

the easiest solution because simply turning on the filter and making sure each resort matches does not require any backend work or data checking.

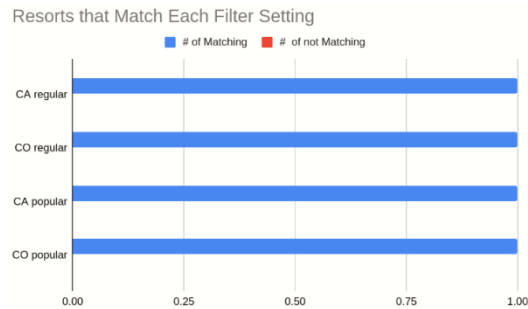


Figure 8. Figure of experiment 1

The first thing we noticed in our data, is that there were no results in which there were any filters that produced a number of not Matching. While this outcome isn't necessarily surprising, it is good to know that the filtering system works accurately with the filters already in place. As for a median and mode, because there was no number of not matching found in our data, they are both simply 1.00 or 100%. The reason we think this was the outcome is because we set the string type in the FlutterFlow Firestore to match the related fields in FireBase perfectly, so the filters work without mistake. Ensuring that the fields match between FlutterFlow and FireBase is what determines the outcome of these results, so ensuring that they match is critical to the app. A good filter in an app is a fundamental element that can greatly impact the user experience, efficiency, and overall success of the application. It empowers users, supports content discovery, and allows for a more customized and efficient interaction, ultimately contributing to user satisfaction and the app's effectiveness.

Another important aspect of the app is the gear section, specifically, how each gear item is stored in the FireBase database with a lot of different fields. It is important that each gear item has all of its fields set correctly because otherwise they may not display correct information about the item. The plan for this experiment is to test that every field in the database is set properly in each item in the list. To test this, we will use test mode to test whether my app finds each field in each item successfully. The reason the experiment is set up like this is because it is convenient in test mode to see the fields in an item and whether they are accessed from the database or not. The list of fields we will be checking include the price, external links, brand and discount status. In our data collection, we will mark down which items have no problems with their fields and which do.

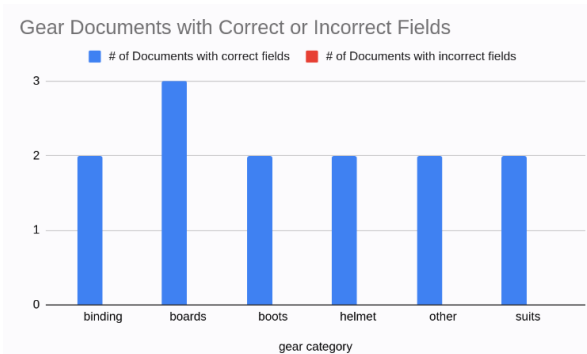


Figure 9. Figure of experiment 2

After performing our experiment, we were able to see that all of our fields were inputted correctly, and would work as expected in our app. This is good to know because if a field of an item was inputted incorrectly the item would not show up with the information provided in the field. Fortunately, our highest value in each bar is the “# of Documents with correct fields,” because we had no Documents with incorrect fields. The reason our items showed up so accurately in our app according to the data within their fields, is because we had no errors inside of our firebase database. With this knowledge of how accurately our field data was set up, we now know we can implement new fields in our product page to give the user more information about the item. Additionally, we also now know that we can effectively add more items to our database without worry of them not being sorted accurately.

5. RELATED WORK

SNOWBOARDER Magazine has a website that contains a bunch of useful articles, gear postings, event information, and how-to videos all made available on the site. The magazine provides a valuable source of information and education on snowboarding topics that is readily available to anyone who visits the site. However, nowadays many people focus more on their phones instead of visiting a website looking for information or resources, as a result, a magazine website may not be as effective as a mobile app. What our solution is better is as a mobile app, the app is now easier and universal.

The Ultimate Snowboarding Guide is a website that provides a great amount of information, tutorials, and blog posts for those interested in snowboarding. However, while the website does have a lot of useful information, it is all closed and doesn't provide any outside sources or information like our app does. Also, the app does provide a lot of useful information for how to choose the correct gear or resort, however, it doesn't provide actual resources or links to good gear items or resorts like our app does. Additionally, websites are not widely accessed nearly as much as mobile apps and therefore do not have a higher usage rate than mobile app websites. Also, apps can provide a more consistent and immersive user experience. Mobile apps have access to device features like GPS, camera, and push notifications, offering a tailored and feature-rich experience.

The Backcountry app is an outdoors gear shop in the form of a mobile application that includes a selection of snowboarding gear, clothing, and equipment. While the Backcountry app does display to its users its own brand gear items, it does not make available other listings that could be better from other sources or shops like our app does. Additionally, the Backcountry app does not provide the level of information that our app does, leaving some users with potential confusion about gear items or how to use them. Our app, unlike the Backcountry app, offers a wider selection of gear from various websites and apps, allowing users to make informed comparisons and enabling them to find the best deals. Furthermore, our app provides comprehensive information, reducing any potential confusion about gear items and their usage. SNOWBOARDER Magazine has a website that contains a bunch of useful articles, gear postings, event information, and how-to videos all made available on the site. The magazine provides a valuable source of information and education on snowboarding topics that is readily available to anyone who visits the site. However, nowadays many people focus more on their phones instead of visiting a website looking for information or resources, as a result, a magazine website may not be as effective as a mobile app. What our solution is better is as a mobile app, the app is now easier and universal.

6. CONCLUSIONS

While we are happy with our app and how it works, it is mostly just a framework at the moment and, if put into production, would need to have some improvements made. The first limitation that would be improved upon given more time would be the amount of gear in the Fire Base database. Currently, our Fire Base database only has a few items in it. Expanding the Fire Base database with more gear items is essential to provide the user with the most diverse selection of gear items. Moreover, a more comprehensive database caters to a wider range of user needs, ensuring that a user can find the information that they need about a wide array of topics. One thing that definitely needs fixing if we had more time is the need for a liking system for the items in the database. This liking system would be implemented in the Fire Base database with two fields, "like count" and "dislike count," and would save the popularity of each item.

Because our app is still in development, we do have many parts that can be improved upon. However, our app has still been prepared for the users in each individual function, and still solves the problem that we sought out to fix. Even though our app is still in a developing stage, it still meets the requirements of providing a wide array of snowboarding information, gear and resort listings, and educational material.

In each of the experiments performed, we were testing to see if there were any discrepancies in our database fields that could cause an item not to appear and if the filters that are set up in Flutter Flow were able to accurately sort them. For each experiment, we used the Flutter Flow test mode and essentially looked through every gear item and filter available to see if they were accurate for both tests. For each experiment, to record the number of matching and not matching, we went through every field and filter combination and recorded the amount of matching and not matching. After testing, we noticed that we had zero instances of not matching across both tests, which shows that we had accurate input on each field and were able to accurately sort them with each filter. The reason that our results came out this way is actually because Flutter Flow and the Fire Base database are handling the filters and fields together for us, leaving not much room for error.

In each of the solution's methodologies that we took a look at, each tried to solve the problem of making either snowboarding information or gear available to a user who is looking for it. In today's digital landscape, the effectiveness of a magazine website, such as SNOWBOARDER Magazine, may be limited as people increasingly turn to mobile apps for information. Similarly, although The Ultimate Snowboarding Guide offers valuable information, its closed ecosystem lacks the diverse range of external sources and data access that our app provides. Additionally, the Backcountry app primarily showcases its in-house gear items, missing the comprehensive range of listings from various sources and shops that our app offers. As opposed to the first two of the works listed previously, our app attempts to put all of these together in the form of a mobile application that has plenty of snowboarding information, outside resources, and gear items made available to the user. Our app aims to be a one stop shop for everything snowboarding related.

REFERENCES

- [1] Dickson, Tracey J., and F. Anne Terwiel. "Injury trends in alpine skiing and a snowboarding over the decade 2008 - 09 to 2017 - 18." *Journal of science and medicine in sport* 24.10 (2021): 1055-1060.
- [2] DeVivo, Michael J. "Epidemiology of traumatic spinal cord injury: trends and future implications." *Spinal cord* 50.5 (2012): 365-372.

- [3] Li, Wu-Jeng, et al. "JustIoT Internet of Things based on the Firebase real-time database." 2018 IEEE International Conference on Smart Manufacturing, Industrial & Logistics Engineering (SMILE). IEEE, 2018.
- [4] y Monsuwé, Toñita Perea, Benedict GC Dellaert, and Ko De Ruyter. "What drives consumers to shop online? A literature review." *International journal of service industry management* 15.1 (2004): 102-121.
- [5] Pike, Andy. "Geographies of brands and branding." *Progress in Human Geography* 33.5 (2009): 619-645.
- [6] Persson, Sofie, Hannes Fridolfsson, and Amanda Holst. "Strategy within E-commerce: The formation process." (2016).
- [7] Hagel, Brent. "Skiing and snowboarding injuries." *Epidemiology of pediatric sports injuries: individual sports* 48 (2005): 74-119.
- [8] Vernillo, Gianluca, Cesare Pisoni, and Gabriele Thiébat. "Physiological and physical profile of Snowboarding: a preliminary review." *Frontiers in physiology* 9 (2018): 770.
- [9] Maitland, Nicholas James. "Spinning Media: Understanding how snowboarding video producers incorporate advertising into subcultural media." (2015).
- [10] Ojala, Anna-Liisa. "Institutionalisation in professional freestyle snowboarding - Finnish professional riders' perceptions." *European journal for Sport and Society* 11.2 (2014): 103-126.
- [11] Idzikowski, Jan R., Peter C. Janes, and Paul J. Abbott. "Upper extremity snowboarding injuries: ten-year results from the Colorado snowboard injury survey." *The American journal of sports medicine* 28.6 (2000): 825-832.
- [12] [Spelmezan, Daniel. "An investigation into the use of tactile instructions in snowboarding." *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*. 2012.
- [13] Curtis, Rick. *The Backpacker's Field Manual, Revised and Updated: A Comprehensive Guide to Mastering Backcountry Skills*. Crown, 2005.
- [14] Westmattmann, Daniel, et al. "Apart we ride together: The motivations behind users of mixed-reality sports." *Journal of Business Research* 134 (2021): 316-328.
- [15] Metter, Kristine. "Strategic Planning as It Relates to Relationship - Building, Engagement, and Affiliation." *Membership Essentials: Recruitment, Retention, Roles, Responsibilities, and Resources* (2016): 13-20.

Deep Learning based Zero Watermarking for Authentication of Medical Records

Gurleen Kaur, Bakul Gupta, and Ashima Anand

Computer Science and Engineering, Thapar Institute of Engineering and Technology

Abstract. The security of digital images is crucial since they often contain sensitive and confidential data. Unauthorized access to this data could result in severe penalties for the parties involved. Despite the availability of highly secure algorithms, security remains a significant concern due to the rapid emergence of new technologies that can breach it. Thus the proposed work implements a technique that makes the confidential data inaccessible to intruders. Hence fragile type of data hiding technique is used where even with the slightest tampering to the image by an attacker, the information i.e. watermark image is completely destroyed, hence preventing it from unauthorized access. Also, a hybrid transform including DTCWT and NSST is used to fuse two medical images to form a more sophisticated output image, which serves as the final watermark. Further, the zero watermarking model is implemented using the ResNet 50 DL model for more precise results and extraction of feature maps. Embedding the actual image in the carrier image could make the watermarking detectable especially when it is fragile, hence Zero Watermarking overcomes this also by virtual embedding. Moreover, the algorithm employs the avalanche effect of SHA512 for highly secure authentication, further strengthening the security of the system. Overall, the proposed method is an effective way to ensure the security of digital images with confidential data.

Keywords: Zero watermarking, Image Fusion, RDWT, Encryption, Medical images, Deep Learning.

1 Introduction

In today's digital age, the use of digital images has become ubiquitous in every sector of society, ranging from personal photography to medical imaging, from social media to e-commerce. With the increasing use of digital images, the need for their security has also become paramount. Digital images contain sensitive and confidential information, which if compromised, can lead to significant consequences such as identity theft, loss of personal privacy, and even financial losses. Therefore, it is imperative to ensure that digital images are adequately protected from unauthorized access, manipulation, and theft. This paper emphasizes the critical importance of ensuring the security of medical images, highlighting their vulnerability to unauthorized access and potential misuse. Various methods are employed for medical diagnosis, including ultrasonography, magnetic resonance imaging, positron emission tomography etc. Diagnostic images undergo an extensive array of processes, encompassing tasks such as feature selection, image denoising, and segmentation, and they are extensively archived and distributed [1].

One method of protecting digital images is through the use of watermarking. Various conventional techniques of watermarking have been employed by researchers to safeguard copyright in domains that are both fragile and robust.[2]. These techniques vary in their intended function and the level of security that they afford to digital data, as depicted in Figure 1.

Fragile watermarking is most effective in situations where digital media authentication is imperative. It employs a watermark as a digital signature, thereby validating the authenticity of the media and ensuring that it has not been altered. This feature makes it highly valuable in contexts of data authentication, where it is necessary to verify the genuineness of a document or image. [3]. On the contrary, robust watermarking is formulated to endure

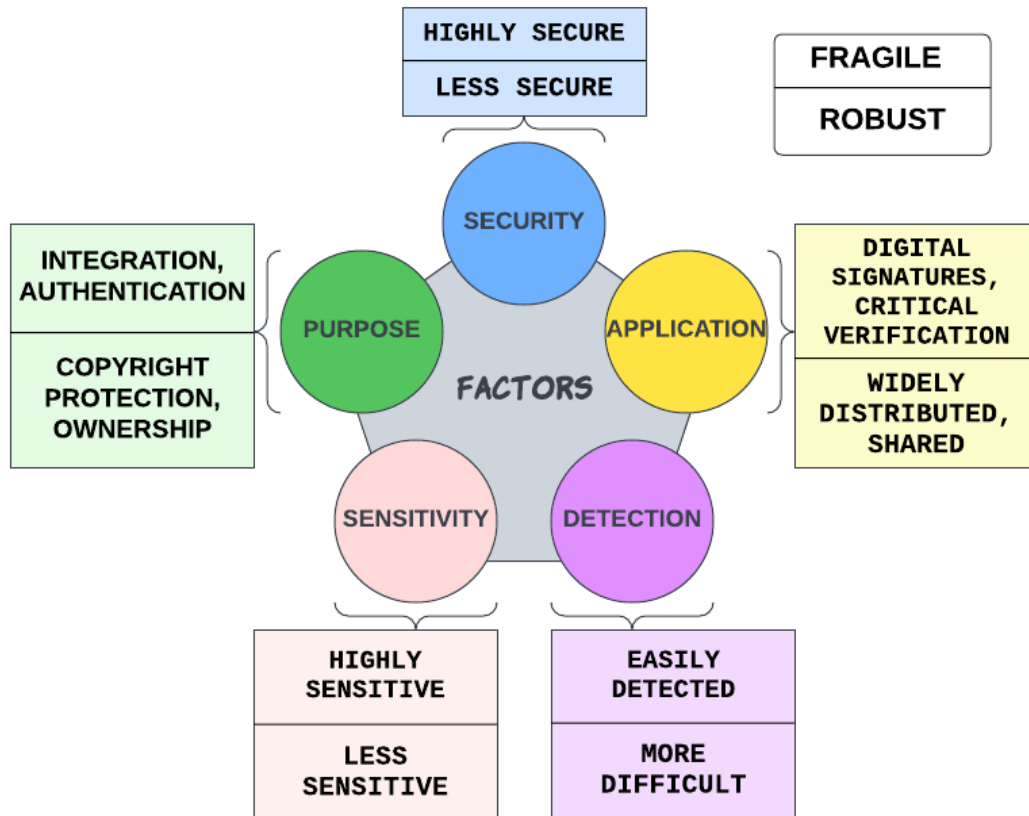


Fig. 1. Difference between Fragile and Robust Watermarking

typical signal processing attacks like compression, cropping, and filtering so that digital data still can be extracted. This makes it ideal for applications where the digital media needs to be distributed and shared widely, such as in the entertainment industry. However, the robustness of the watermark comes at the cost of reduced sensitivity, which means that it may not be able to detect small changes to the original digital media. Whereas, fragile watermarking is done in such a way that watermark data is completely destroyed after it encounters any change, it is harder for an attacker to modify or remove them without detection. Therefore, Fragile watermarking offers a higher level of security than robust watermarking. Therefore, this paper has opted for the fragile watermarking technique as the chosen approach for the necessary algorithm.

Over the recent years, researchers have extensively employed Zero watermarking, which can be thought of as "invisible watermarking." The usage of the term "zero" implies that the embedded watermark is crafted to remain unseen or imperceptible to human senses, such as sight or hearing, without causing any significant change in the original content's visual or auditory attributes. However, a significant portion of the research has centered around utilizing Zero watermarking as the robust algorithm. Hence, the authors undertook this study to explore the outcomes when the imperceptibility of Zero watermarking is integrated with fragile watermarking. The primary objective was to address the drawback of fragile watermarking, which is its susceptibility to detection compared to robust watermarking techniques.

In real-world scenarios, physicians often require multiple patient reports to make an accurate diagnosis. Taking this aspect into account, the present study has opted to consider that patients need to submit both their CT scan and MRI images. Addressing this concern, the paper introduces a fusion algorithm that can be described as a form of encryption and also a multi-factor authentication method.

The objective of this manuscript is to present a technique for securely transmitting digital images in a manner that ensures confidential information remains inaccessible to intruders and maintaining the authenticity of images.

2 Literature Survey

In recent years, with remarkable progress in deep learning techniques, they have also been extensively employed in safeguarding digital information. [4] paper proposed the first watermarking framework using CNN. It introduces a novel non-blind digital image watermarking method that utilizes the auto-encoder capability of CNNs. This approach generates positive and negative codebook images, which play a crucial role in embedding and extracting watermarks. But as their proposed method was completely non-blind, it lacks practicality in real world. DFT based Zero watermarking along with VGG19 and perceptual hashing was presented by authors in paper [5]. Though their framework was robust against various geometric attacks, but a CNN residual network with more deeper layers could give more accurate results than VGG19 which has 19 layers. Hence, within the algorithm presented in our paper, we opt for the utilization of Residual networks.

Authors of Ref. [6] proposed the method for Zero Watermarking with DCT and Residual DenseNet. Their proposed framework was robust against various geometric attacks. Further DWT-SVD-DCT based fragile watermarking was implemented in paper [7]. The authors developed a tamper-proof framework that could identify cropping and object insertion attacks. To enhance the framework's sensitivity and detect even minor alterations, an authentication code was generated and incorporated into the watermarked image using the QIM technique. The embedded code was extracted using the Gram-Schmidt process. However, it should be noted that the Gram-Schmidt process may result in information loss, which could erroneously detect an attack even if it did not occur. Further the paper has not analyzed results for many signal processing attacks.

Singh et al. [8] has put forward fragile watermarking technique using DCT-LSB method. While the paper's method is capable of accurately extracting the main content from a tampered image up to a 50% tampering rate, it is only effective in restoring against certain types of attacks. It cannot fully remove the watermark in cases of tampering, which is a crucial requirement for fragile watermarking as it must be highly responsive. Hence for making the algorithm highly sensitive and sensitive, hashing could also play the major role in authentication. [9] paper completed the comparatively analysis of well known MD and SHA algorithms. Further concluded that for high security purpose SHA512 could serve the purpose because of its avalanche effect and longer length of string constructed.

Summing up the comprehensive review of existing literature, we deduced that Zero watermarking has predominantly been applied in robust methods. However, a drawback lies in the fact that while robust watermarking offers enhanced security, fragile watermarking surpasses it in terms of security measures. When it comes to employing watermarking techniques rooted in Deep Learning, a constant compromise between time complexity and result accuracy is unavoidable. As a result, our initial focus rests on the evaluation of

outcomes across different models, as detailed in subsequent sections. To bolster security, the integration of the SHA hashing algorithm stands out as a promising choice.

3 Proposed Methodology

This paper puts forth a fragile watermarking technique that is well-suited for binary images of a confidential nature (See Fig. 2). Until now, the majority of the existing fragile

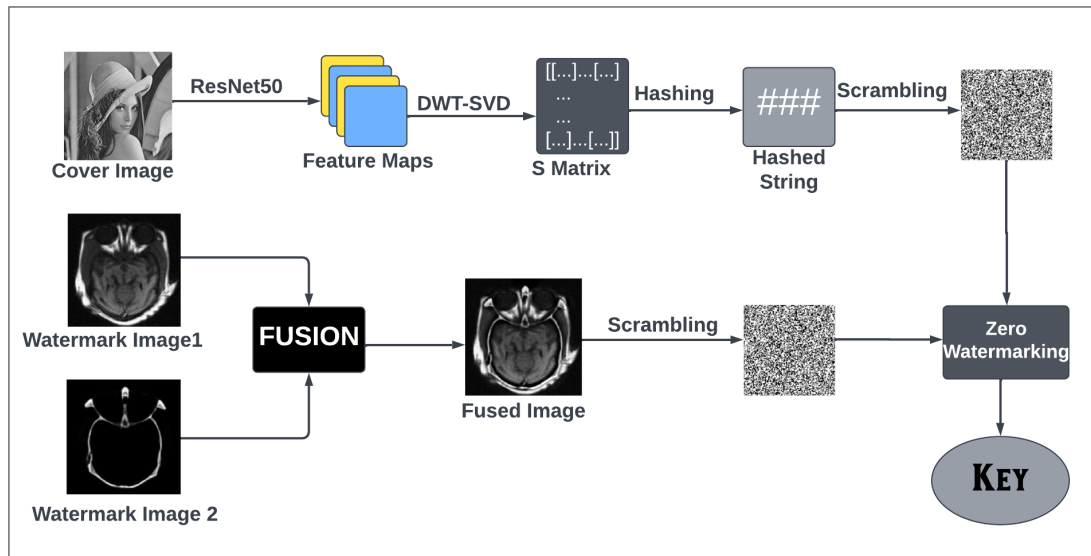


Fig. 2. Framework of the proposed fragile watermarking model using image fusion

watermarking methods that embedded the watermark were easily detectable. However, the proposed approach employs zero watermarking, which doesn't embed the watermark directly but instead relies on the resemblances between the cover image and the watermark image. Unlike conventional zero watermarking methods, the proposed approach abstains from manual feature extraction and instead leverages the power of deep learning in conjunction with zero watermarking. We have devised a methodology that entails a 3 step process. Initially, fusion of two distinct watermarks is executed, resulting in the creation of a solitary watermark. Subsequently, the extraction of a feature map from the cover image is carried out. Finally, the zero watermark technique is implemented to accomplish the watermarking process. For the purpose of further enhancing security, hashing and scrambling of images is also carried out.

3.1 Watermark Generation using NSST-DTCWT transforms based Image Fusion

In this phase, DTCWT and NSST-based fusion of medical images is implemented to generate the watermark using two input images, 'CT' and 'MRI'. This image fusion algorithm uses two different rule sets to fuse the high and low-frequency coefficients of the input images. Parameter adaptive PCNN is used to fuse the high-frequency coefficients, while WSE and WSNML-based rules contribute to generating the fused low-frequency coefficients. The flowchart for generating the fused image as a mark carrier is shown in Fig.3.

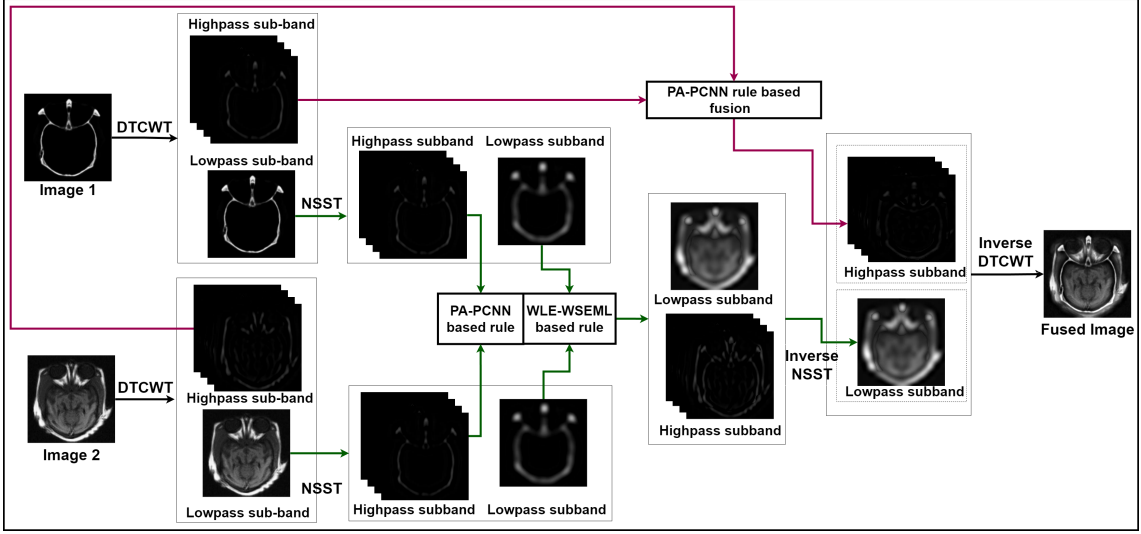


Fig. 3. DTCWT-NSST based Image Fusion to generate final Watermark Image

To fuse the high-frequency coefficients, PA-PCNN-based fusion rules are adapted. PA-PCNN is a type of PCNN that works by monitoring the image processing performance and dynamically adjusting the parameters to optimize the results. This adaptation is performed through a feedback mechanism that adjusts the parameters based on the current state of the image processing. The parameters that are adapted in PA-PCNN include the coupling strength, the threshold and the pulse width.

PA-PCNNs can be used to preserve the relevant information from each image while suppressing the irrelevant information by automatically adjusting their parameters based on the input data. This enables the network to extract and preserve the most relevant features from each image, resulting in a fused image that contains more information than any of the individual images.

Further, the edge and contours of the input images are preserved by fusing the low-frequency coefficients of input images using a hybrid of WLE and WSEML. The activity level measure, WLE, is mathematically calculated as,

$$WLE_{ab}(x, y) = \sum_i \sum_j Wm'(i, j) H_{ab} \times (x + i, y + j)^2, \quad (1)$$

where $H_{ab}(x, y)$ denotes the high-frequency NSST coefficient at position (x, y) of direction b at layer a . Also, the Wm' is the weighted matrix which is defined as:

$$Wm' = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad (2)$$

Further, WSEML is used to extract the details of the input images using the following equation,

$$WSEML_{a,b}(x, y) = \sum_{p=-rad}^{rad} \sum_{q=-rad}^{rad} (Wm'(p + rad + 1, q + rad + 1) \times EML_{a,b}(a + p, b + q)) \quad (3)$$

where, Wm' is the weighting matrix defined in eq. 2, and EML is defined as,

$$\begin{aligned}
EML_{img}(x, y) = & |2img(x, y) - img(x - 1, y) - img(x + 1, y)| \\
& + |2img(x, y) - img(x, y - 1) - img(x, y + 1)| \\
& + \frac{1}{\sqrt{2}} |2img(x, y) - img(x - 1, y - 1) - img(x + 1, y + 1)| \\
& + \frac{1}{\sqrt{2}} |2img(x, y) - img(x - 1, y + 1) - img(x + 1, y - 1)| \quad (4)
\end{aligned}$$

Table 1. Details of notations used in this article

Notation	Explanation	Notation	Explanation
CT, MRI	Input images for multimodality image fusion	dtcwt_L1, dtcwt_H1	High and Low DTCWT coefficients of CT image
nsst_L1, nsst_H1	Low and high NSST coefficients of dtcwt_l	WLE1, WSEML1	WLE and WSEML associated with dtcwt_H1 and dtcwt_H2
nsst_L	Fused low frequency NSST coefficient	nsst_H	Fused high frequency NSST coefficient
dtcwt_H	Fused high frequency DTCWT coefficient	dtcwt_L	Fused low frequency DTCWT coefficient
Fimg	Fused Image	cI	Cover Image
key	Extraction Key	fusedWI	fused Watermark Image after resizing
s_WI	Watermark Image after applying seed scrambling contains weights of	ResNet50	Residual Network with 50 layers
model	ResNet50 pre-trained on imagenet dataset	FM	Feature Matrix of the cover image predicted using 'model'
DWT	Discrete Wavelet Transformation	SVD	Singular Value Decomposition
cA, cH, cV, cD	Approximation, Vertical, Horizontal and Diagonal sub-bands of 'FM' on applying DWT	u,s,v	Left, Middle and Right singular matrices of 'cA' on applying 'SVD'
b_cover	Binary of the cover image after hashing	bin_wat	Binary of the watermark image
cov	Binary of cover Image after adjusting its length	s_cov	Cover Image after applying seed scrambling

As shown in Algorithm 1, two input images, 'CT' and 'MRI', are decomposed into high-pass and low-pass components using DTCWT transform. The resultant high-pass components, 'dtcwt_H1' and 'dtcwt_H2', are merged using a parameter adaptive PCNN (PAPCNN)-based fusion scheme, resulting in 'dtcwt_H'. Further, NSST is applied to the low-pass DTCWT coefficients. The resultant high-band NSST coefficients are again fused based on fusion rules using PAPCNN. The energy preserving and detail extracting issues are addressed by fusing the low-band NSST coefficients using WLE and WSEML-based fusion rules, generating the fused low-band NSST component, 'nsst_L'. Inverse NSST is then applied to form the fused low-pass DTCWT coefficients, 'dtcwt_L'. Finally, inverse DTCWT is applied to obtain the fused image, 'Fimg', which is treated as watermark image in the later sections.

3.2 Extraction of Feature maps

The extraction of feature maps is a crucial process that necessitates meticulous attention and consideration to guarantee the precision, dependability, and utility of the resulting

Algorithm 1: Algorithm of DTCWT-NSST based medical image fusion

```

Input : CT, MRI
Output: Fimg
// Phase 1: Transforming input images using DTCWT
1  dtcwt_L1, dtcwt_H1  $\leftarrow$  DTCWT(CT);
2  dtcwt_L2, dtcwt_H2  $\leftarrow$  DTCWT(MRI);
// Phase 2: Transforming low-frequency sub-band using NSST
3  nsst_L1, nsst_H1  $\leftarrow$  NSST(dtcwt_L1);
4  nsst_L2, nsst_H2  $\leftarrow$  NSST(dtcwt_L2);
// Phase 3: Fusion of low-frequency NSST coefficients
5  WLE1  $\leftarrow$  WLE_Calculation(nsst_L1);
6  WLE2  $\leftarrow$  WLE_Calculation(nsst_L2);
7  WSEML1  $\leftarrow$  WSEML_Calculation(nsst_L1);
8  WSEML2  $\leftarrow$  WSEML_Calculation(nsst_L2);
9  map  $\leftarrow$  (WLE1  $\times$  WSEML1  $\geq$  WLE  $\times$  WSEML2);
10 nsst_L  $\leftarrow$  map  $\cdot$  nsst_L1 + map  $\cdot$  nsst_L2;
// Phase 4: Fusion of high-frequency NSST coefficients
11 nsst_P1  $\leftarrow$  PA_PCNN(nsst_H1);
12 nsst_P2  $\leftarrow$  PA_PCNN(nsst_H2);
13 map  $\leftarrow$  (nsst_P1  $\geq$  nsst_P2);
14 nsst_H  $\leftarrow$  map  $\cdot$  nsst_P1 + map  $\cdot$  nsst_P2;
// Phase 5: Applying inverse NSST decomposition
15 dtcwt_L  $\leftarrow$  Inverse_NSST(nsst_L, nsst_H);
// Phase 6: Fusion of high-frequency DTCWT coefficients
16 dtcwt_P1  $\leftarrow$  PA_PCNN(dtcwt_H1);
17 dtcwt_P2  $\leftarrow$  PA_PCNN(dtcwt_H2);
18 map  $\leftarrow$  (dtcwt_P1  $\geq$  dtcwt_P2);
19 dtcwt_H  $\leftarrow$  map  $\cdot$  dtcwt_P1 + map  $\cdot$  dtcwt_P2;
// Phase 7: Applying inverse DTCWT decomposition
20 Fimg  $\leftarrow$  Inverse_DTCWT(dtcwt_L, dtcwt_H);
21 return Fimg

```

features in required applications.

The approach taken in this paper involves the utilization of ResNet50 whose building block is shown in fig.4 , a convolutional neural network (CNN) with a significant depth of 50 layers. The depth of the network is crucial for neural networks, but deeper networks are

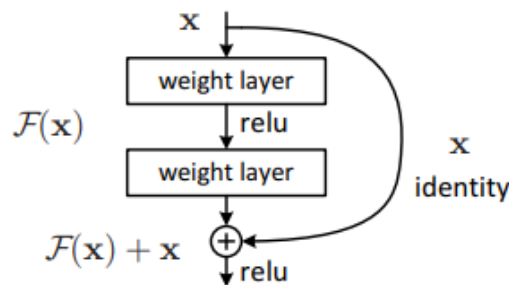


Fig. 4. Residual learning: a building block[10]

more difficult to train. The configuration of ResNet50 enables the instruction of networks and permits them to be considerably more profound, resulting in augmented proficiency in various tasks. ResNet50 surpasses their simple equivalents in depth, and furthermore,

the quantity of weights in such networks is significantly lower. [11].

When a 512x512 cover Image ‘*cI*’ is fed through a modified ResNet50 model, the last convolutional layer produces a feature map with a spatial size of 16x16 and 2048 channels. Each channel in the feature map represents a specific learned feature of the input image. To obtain a final feature matrix, the individual feature maps are added element-wise to produce a single feature map. This final feature map preserves the spatial structure of the input image and contains information about the learned features across all channels. This feature map is used for further processing.

In ResNet50, the feature maps are extracted through a series of convolutional layers that are arranged in blocks. Each block consists of multiple layers, including convolutional layers, batch normalization layers, and activation layers. The output of each block is then passed through a shortcut connection that allows the gradient to flow more easily during training. In proposed method, we have used pre-trained model of ResNet50 on Imagenet Dataset.

3.3 Zero Watermarking

As described in [12], Zero watermarking essentially performs a virtual embedding process, where the key for generating a watermark is generated by analyzing the similarities between the attributes of the cover image and the watermark image. This key can then be used to produce an identical watermark from the cover image.

Method used for Zero Watermarking is described in Algorithm 2. The fused image ‘*Fimg*’

Algorithm 2: Watermarking

```

Input : Fimg, cI
Output: key
// Phase 1: Resizing of input images
1  fusedWI ← resize(Fimg, (256, 256));
2  cI ← resize(cI, (512, 512));
// Phase 2: Scrambling of fusedWI
3  s_WI ← seed_scramble(fusedWI);
// Phase 3: Feature Map Extraction
4  model ← ResNet50(weights = ‘magenet’);
5  cI ← resnet50.preprocess_input(cI);
6  FM ← model.predict(cI);
// Phase 4: Apply DWT and SVD
7  [cA, cH, cV, cD] ← dwt(FM) ;
8  [u, s, v] ← svd(cA);
// Phase 5: Hashing Cover Image
9  b_cover ← SHA512(s);
// Phase 6: Adjusting length of binary of cover image
10 bin_wat ← matrix_to_binary(s_WI);
11 cov ← add_trailing_zeros(b_cover, bin_wat);
// Phase 7: Scrambling binary of cover image
12 s_cov ← seed_scramble(cov);
// Phase 8: Key Generation
13 key ← XOR(s_cov, bin_wat);
14 return key

```

obtained after combining ‘*CT*’ and ‘*MRI*’ in Algorithm 1 serves as the watermark image. It is used as input along with the cover image ‘*cI*’. After resizing, ‘*Fimg*’ is scrambled using

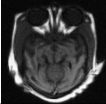

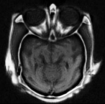
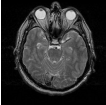


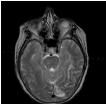


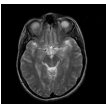

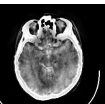

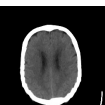
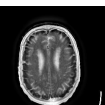
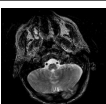
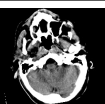
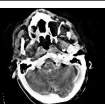
a seed scrambling algorithm. A feature map is extracted from ‘ cI ’ using the ResNet50 DL model that has been pre-trained on the imagenet dataset. This feature map ‘ FM ’ is then divided into frequency subbands with DWT, and the ‘ LL ’ subband is subject to SVD decomposition. The resulting ‘ s ’ singular value matrix is hashed with the SHA512 algorithm. Both the scrambled watermark image ‘ $s.WI$ ’ and the hashed string ‘ $b.cover$ ’ are converted to binary form. To account for differences in size between the two binary outputs, a string of trailing zeroes is added to ‘ $b.cover$ ’, and the same scrambling algorithm as previously mentioned is applied. The XOR operation is then performed on ‘ $bin.wat$ ’ and ‘ $s.cov$ ’ to generate the extraction key ‘ key ’.

4 Results and Analysis

This part of the analysis thoroughly analyzes the suggested technique, beginning with experimental configurations and progressing to performance evaluation with varied cover images and fusion images. The comparative study of the proposed technique is offered at the end.

The trial begins initially by taking a brain MRI picture sized 512×512 as the cover object and the covert data is taken as a CT Scan image sized 256×256 . The implementation is done on MATLAB R2021b using system with the following configuration, Intel Xeon(R) Gold processor with 256GB RAM. The performance metrics used for assessing the projected technique are listed as (PSNR), SSIM, and NC. The evaluation parameters of Fusion method quantifies how accurately the fused image conveys the original images content. Some common fusion parameters are MI, QABF, SSIM, SF, and STD [13–15]. Visual output for 6 sample input output pairs are shown in table 2.

Table 2. Sample of images used for evaluating Dtcwt-Nsst based multi-modality image fusion

Pair No.	Image 1	Image 2	Fused Image
Pair 1			
Pair 2			
Pair 3			
Pair 4			
Pair 5			
Pair 6			

The objective evaluation of the proposed DTCWT-NSST based fusion method, when implemented on 50 pairs of medical images [16], is referred to in Table 3. Average scores


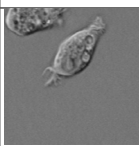


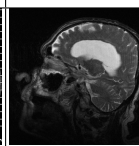


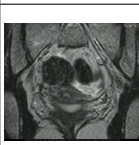
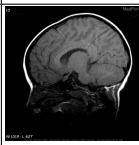

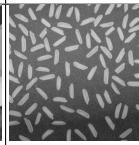
Table 3. Evaluation of proposed image fusion method on 50 pair of medical images

Pair	Entropy	MI	QABF	FMI	NABF	SSIM	SF	STD	PSNR1	PSNR2
Pair 1	1.7126	3.4252	0.1937	0.8895	0.0373	0.0932	4.2698	36.2066	59.8872	66.1457
Pair 2	6.7562	13.5125	0.4596	0.8789	0.2290	0.4766	6.0882	54.8540	70.0475	59.1952
Pair 3	2.9538	5.9075	0.3288	0.8700	0.0452	0.6626	5.4185	84.7166	65.8571	59.2004
Pair 4	4.4879	8.9757	0.2748	0.8929	0.0731	0.8234	4.5951	57.0351	70.13	65.7703
Pair 5	4.8135	9.6269	0.3957	0.8821	0.0769	0.6659	5.8501	83.0159	65.8074	59.3964
Pair 6	4.3270	8.6540	0.3049	0.8672	0.0666	0.5857	6.5350	80.7836	65.4172	58.8764
Average of 50 pairs	3.8406	7.6811	0.2906	0.8771	0.0419	0.6653	4.9579	79.5419	67.1835	59.9599

of QABF, FMI, SSIM, SF, and STD for 50 pairs are 0.2906, 0.8771, 0.6653, 4.9579 and 79.5419, respectively. These results verify the satisfactory performance of the multi-modality fusion method for healthcare images.

Table 4 displays a range of standard images employed for analyzing outcomes, while a medical image data-set of over 250 MRI images is also utilized for the same objective.

Table 4. Standard Images

I1 	I3 	I5 	I7 	I9 	I11 
I2 	I4 	I6 	I8 	I10 	I12 Average of 250+ MRI Images

The proposed method is being tested against various geometric and signal processing attacks, as illustrated in table 5. NC value is calculated original watermark and extracted

Table 5. Attacks performed

Symbol	Attack Names
A1	Average filtering
A2	Cropping Attacks
A3	Gaussian filter
A4	JPEG compression
A5	Median Filter Attack
A6	Rotation Attack
A7	Salt & Pepper Noise
A8	Scaling the image
A9	Translation attack

watermark which is the measure of similarity between images. In the case of fragile watermarking, the NC value between the watermark and extracted image should be very low even after minimal modification to the watermarked image, serving the purpose of privacy and security. As the proposed method employs zero watermarking, there is no actual embedding in the cover image, and thus no metric is required to compare the watermarked and cover images as they are the same.

In order to extract feature maps from the cover image, diverse deep learning (DL) models were computed and executed on standard images in the face of attacks. These outcomes were then meticulously analyzed, and the average results are elucidated in table 6.

Table 6. Average NC after performing attacks on various DL architectures

DL Tech.	A1	A2	A3	A4	A5	A6	A7	A8	A9
R50	0.0159	-0.0175	0.0078	-0.0054	-0.0275	-0.0435	0.0103	-0.0068	-0.0005
R101	-0.0146	0.0008	-0.0088	0.0056	-0.0010	-0.0039	-0.0021	-0.0076	-0.0056
VGG19	0.0070	0.0024	-0.0167	-0.0177	0.0037	0.0243	-0.0220	-0.0062	-0.0091
DN121	0.0071	-0.0021	-0.0087	-0.0102	-0.0142	0.0010	0.0036	0.0010	-0.0138
MNet	0.0200	0.0295	-0.0039	-0.0024	-0.0191	-0.0044	-0.0079	-0.0130	0.0090

Subsequent to running ResNet50, ResNet101, VGG19, DenseNet121, and MobileNet models, the deduction drawn from the results was that ResNet50 provided the most exceptional outcomes since it generated the minimum value of NC is majority of attacks and is giving best results from all other methods.

Furthermore, we also determined the average value of NC using ResNet50, by testing it on a medical image dataset against various attacks. The outcomes of ResNet 50 against various attacks on individual standard images and medical image dataset are tabulated in table 7. Hence the NC values obtained from the results are very low, this directly prove that the proposed method is highly tamper proof.

Table 7. Results of ResNet50

Images	A1	A2	A3	A4	A5	A6	A7	A8	A9
I1	0.0893	-0.1022	-0.0673	0.0288	-0.0300	-0.0306	0.0006	-0.0187	0.0307
I2	0.0726	-0.0253	0.0685	-0.0287	0.0131	-0.0823	-0.0091	-0.0118	-0.0486
I3	-0.0527	0.0047	0.0454	0.0252	-0.0053	-0.0465	0.0373	-0.0172	-0.0717
I4	-0.0860	-0.1110	0.0264	-0.0617	-0.0043	-0.0212	-0.0277	-0.0446	0.0010
I5	-0.0370	0.1179	0.0020	-0.0136	0.0108	-0.0347	-0.0195	-0.0602	-0.0190
I6	0.0347	-0.0230	0.0068	-0.0541	-0.0058	-0.0195	0.0786	0.0497	-0.0157
I7	0.0332	-0.0556	-0.0425	-0.0198	-0.0251	-0.0956	0.0512	-0.0076	0.0545
I8	0.0262	-0.0224	-0.0531	0.0029	-0.0726	-0.0383	0.0339	-0.0449	-0.0112
I9	0.0913	-0.0205	0.0154	0.0037	-0.0476	-0.0559	0.0177	0.0131	-0.0158
I10	0.0094	0.0779	0.0037	0.0200	-0.0669	-0.0405	-0.0163	0.0131	0.0269
I11	-0.0060	-0.0325	0.0807	0.0376	-0.0692	-0.0134	-0.0337	0.0544	0.0636
I12	0.00005	-0.00366	0.00492	-0.00118	0.00036	0.00490	-0.00352	-0.00609	0.00124

5 Conclusion

In conclusion, the security of digital images with sensitive and confidential data is of paramount importance to avoid severe penalties associated with unauthorized access. The proposed method in this manuscript is a highly effective solution to this problem. This technique initially employs NSST-DTCWT based multimodality image fusion method to generate the final watermark. By utilizing the avalanche effect of the hashing algorithm SHA512, the method is highly fragile. Any slight changes in the watermarked image can completely destroy the confidential information stored as the watermark image, effectively preventing any unauthorized access. Further, the implementation of virtual embedding with zero watermarking instead of actual embedding significantly reduces the attack surface as it conceals the availability of the watermark image in the cover image. This approach provides an additional layer of security by making it more difficult for attackers to identify and tamper with the watermark image. Therefore, this method provides an excellent solution to ensure the security of digital images with confidential information.

References

1. A. Odeh and Q. A. Al-Haija, "Medical image encryption techniques: a technical survey and potential challenges," *no. January*, pp. 3170–3177, 2023.
2. M. Begum and M. S. Uddin, "Digital image watermarking techniques: a review," *Information*, vol. 11, no. 2, p. 110, 2020.
3. J. Fridrich, "Methods for tamper detection in digital images," in *Multimedia and Security, Workshop at ACM Multimedia*, vol. 99, pp. 29–34, 1999.
4. H. Kandi, D. Mishra, and S. R. S. Gorthi, "Exploring the learning capabilities of convolutional neural networks for robust image watermarking," *Computers & Security*, vol. 65, pp. 247–268, 2017.
5. B. Han, J. Du, Y. Jia, and H. Zhu, "Zero-watermarking algorithm for medical image based on vgg19 deep convolution neural network," *Journal of Healthcare Engineering*, vol. 2021, 2021.
6. C. Gong, J. Liu, M. Gong, J. Li, U. A. Bhatti, and J. Ma, "Robust medical zero-watermarking algorithm based on residual-densenet," *IET Biometrics*, vol. 11, no. 6, pp. 547–556, 2022.
7. T.-S. Nguyen, "Fragile watermarking for image authentication based on dwt-svd-dct techniques," *Multimedia Tools and Applications*, vol. 80, no. 16, pp. 25107–25119, 2021.
8. D. Singh and S. K. Singh, "Dct based efficient fragile watermarking scheme for image authentication and restoration," *Multimedia Tools and Applications*, vol. 76, pp. 953–977, 2017.
9. S. Long, "A comparative analysis of the application of hashing encryption algorithms for md5, sha-1, and sha-512," in *Journal of Physics: Conference Series*, vol. 1314, p. 012210, IOP Publishing, 2019.
10. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
11. Y. Chu, X. Yue, L. Yu, M. Sergei, and Z. Wang, "Automatic image captioning based on resnet50 and lstm with soft attention," *Wireless Communications and Mobile Computing*, vol. 2020, pp. 1–7, 2020.
12. Y. Zhou and W. Jin, "A novel image zero-watermarking scheme based on dwt-svd," in *2011 International Conference on Multimedia Technology*, pp. 2873–2876, IEEE, 2011.
13. N. Jain, A. Yadav, Y. Kumar Sariya, and A. Balodi, "Analysis of discrete wavelet transforms variants for the fusion of ct and mri images," *The Open Biomedical Engineering Journal*, vol. 15, no. 1, 2021.
14. W. Ma, K. Wang, J. Li, S. X. Yang, J. Li, L. Song, and Q. Li, "Infrared and visible image fusion technology and application: A review," *Sensors*, vol. 23, no. 2, p. 599, 2023.
15. L. Tang, Y. Deng, Y. Ma, J. Huang, and J. Ma, "Superfusion: A versatile image registration and fusion network with semantic awareness," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 12, pp. 2121–2137, 2022.
16. J. A. B. Keith A. Johnson, "the whole brain atlas." <http://www.med.harvard.edu/AANLIB/home.html>.

Authors

Gurleen Kaur is currently an undergraduate 4th year Computer Engineering student at Thapar Institute of Engineering and Technology. Her research interests include Cyber Security, Networking and Deep Learning. (E-mail: gkaur6_be20@thapar.edu)

Bakul Gupta is currently an undergraduate 4th year Computer Engineering student at Thapar Institute of Engineering and Technology. His research interests include Blockchain, Deep Learning, Cryptography and System Security. (E-mail: bgupta1_be20@thapar.edu)

Ashima Anand is currently working as an Assistant Professor in Thapar Institute of Engineering and Technology. She pursued her Ph.D. and MTech. in Computer Science and Engineering from NIT Patna, Bihar, India. Also, she received B. Tech. in Computer Science and Engineering from NIT Hamirpur, H.P., India in 2017. Her research interest includes Data Hiding Techniques, Cryptography and Image Processing. (E-mail: ashima.anand@thapar.edu)

QUALITY CHALLENGES AND IMPERATIVES IN SMART AI SOFTWARE

Rohit Khankhoje

Independent Researcher Avon, Indiana, USA

ABSTRACT

In the epoch of pervasive Smart AI applications, ensuring the excellence of software in AI-driven systems is of utmost importance. This article concentrates on deciphering the intricate realm of Smart AI software, with the objective of identifying hurdles in quality assurance and underscoring the necessity for robust solutions. The exploration encompasses diverse facets of challenges, ranging from managing partial training data to addressing ethical concerns regarding algorithm transparency. Technical intricacies, such as testing complexities and model resilience, are deliberated alongside broader societal and ethical considerations, including privacy and user trust. The article advocates for a comprehensive quality assurance framework for Smart AI software, with a focus on its role in guaranteeing safety, dependability, and adherence to regulations. The impact of quality assurance on user experience is also scrutinized, highlighting the interdependent relationship between quality assurance and user satisfaction.

By tackling challenges and emphasizing the imperative for effective solutions, this article contributes to the ongoing discourse on responsible development and deployment of Smart AI software. It aspires to advance quality assurance practices in this dynamic technological landscape, promoting the responsible evolution of Smart AI applications.

KEYWORDS

Artificial Intelligent, Software Testing, AI Software, Quality Assurance

1. BACKGROUND

In recent times, there has been an unprecedented surge in the rapid progression of Artificial Intelligence (AI) technologies. AI has permeated various aspects of our technological landscape, ranging from natural language processing to computer vision and machine learning. This surge is particularly evident in the integration of AI into smart software applications, where intelligence is embedded to enhance functionalities, automate processes, and provide personalized user experiences. As AI technologies continue to evolve, they are increasingly being integrated into smart software applications that power diverse domains such as healthcare, finance, logistics, and more. The growing prevalence of AI-driven functionalities, including predictive analytics and autonomous decision-making, has become a defining characteristic of modern software development (Haller-Seeber & Gatterer, 2022). These smart applications, leveraging AI algorithms, not only streamline complex tasks but also revolutionize problem-solving and decision-making approaches.

However, as the role of AI in software applications expands, there is a growing need to ensure the quality of Smart AI software. The importance of quality assurance in this context cannot be overstated. Unlike traditional software, which can be precisely defined and comprehensively

tested, AI algorithms present unique challenges due to their inherent complexity and dynamic nature. The data-driven and learning aspects of AI systems introduce layers of complexity, necessitating a shift in quality assurance strategies. Ensuring quality in Smart AI software is crucial for several reasons.

Firstly, the reliability and accuracy of AI-driven functionalities directly impact the user experience. Whether it is a recommendation system, a virtual assistant, or a predictive model, users expect these intelligent features to perform flawlessly and provide valuable insights. Secondly, in applications with critical implications, such as healthcare diagnostics or autonomous vehicles, the consequences of AI errors can be significant. Therefore, the reliability and safety of these systems must undergo rigorous validation (Job, 2020).

Furthermore, as AI applications often deal with extensive and diverse datasets, ensuring data quality, addressing biases, and upholding ethical considerations become integral components of the quality assurance process. The trust users place in AI systems relies heavily on the assurance that these systems are not only accurate but also fair, transparent, and compliant with ethical standards.

1.1 COMPARISON OF DIFFERENT TYPE OF TESTING

Understanding AI testing, AI-based testing, and conventional testing is of utmost significance prior to delving into the exploration and comprehension of the challenges posed by Smart AI software (Gao, 2022).

Items	AI Testing	AI-Based Software Testing	Conventional Software Testing
Purpose	Assure and validate the quality of AI software and system by concentrating on the functions and features of the AI system.	Utilize artificial intelligence techniques and solutions to enhance the efficiency and effectiveness of a software testing procedure and its overall quality.	Ensure the quality of system functionality for traditional software and its characteristics.
Primary AI testing focuses	The quality factors of AI features encompass correctness, accuracy, consistency, timeliness, completeness, and performance.	Optimize a test process in product quality increase, testing efficiency, and cost reduction.	Automate the operations of testing for a traditional software process.
Common system testing quality	The factors contributing to the quality of a system include performance, reliability, scalability, availability, security, and throughput.	The factors contributing to the quality of a system include performance, reliability, scalability, availability, security, and throughput.	The factors contributing to the quality of a system include performance, reliability, scalability, availability, security, and throughput.

System function testing	AI system function testing includes the evaluation of various aspects such as object detection and classification, recommendation and prediction, as well as language translation.	System functions, behaviors, user interfaces	System functions, behaviors, user interfaces
Test selection	AI test model is founded on the principles of test selection, classification, and recommendation.	Test selection, classification, and recommendation using AI techniques	Rule-based and/or experience based test selection
Test Data Generation	The AI test model is centered on the exploration, gathering, production, and authentication of examination data.	AI-based test data collection, classification, and generation	Model-based and/or pattern based test generation
Bug Detection and Analysis	AI model-based bug detection, analysis, and report	Data-driven analysis for bug classification and detection, as well as prediction	Digital and systematic bug/problem management

Table-1 Comparison of different application type of Software testing

2. INTRODUCTION

This document is dedicated to a comprehensive exploration of the intricacies and concerns inherent in ensuring the quality of Smart AI software. By conducting a thorough analysis of the complexities tied to the integration of AI technologies into software applications, the objective is to illuminate the specific challenges confronted by quality assurance professionals and developers in this dynamic landscape. Through an in-depth examination of these challenges, the document aims to provide a nuanced comprehension of the multifaceted issues arising in the quality assurance of Smart AI software. Key areas of focus encompass the intricacies of testing, the quality and biases within data, algorithm transparency, and the ethical considerations encompassing AI utilization (Khaliqa & Farooqa, 2022). The intention is not solely to identify these challenges but also to articulate their broader significance within the realm of software development and deployment.

Furthermore, this document underscores the critical necessity to proactively address these challenges. The swift integration of AI technologies into intelligent applications necessitates the development of robust quality assurance strategies adaptable to the unique characteristics of AI algorithms. The objective is not merely to highlight challenges but to emphasize their implications for the reliability, safety, and user trust associated with Smart AI software.

3. SMART AI SOFTWARE LANDSCAPE

Smart AI software refers to software applications that leverage advanced artificial intelligence (AI) techniques to demonstrate intelligent behavior, adaptability, and decision-making capabilities. This classification of software surpasses conventional rule-based systems by incorporating machine learning algorithms, natural language processing (NLP), computer vision, and other AI technologies to enhance its functionality (Khankhoje, 2023).

Prominent characteristics of Smart AI software encompass:

Adaptability: Smart AI software possesses the capacity to acquire knowledge from data, adjust to new information, and enhance its performance over time. This adaptability is often a result of employing machine learning algorithms that facilitate the software in making predictions or decisions based on data patterns.

Intelligent Decision-Making: The software has the autonomy to make decisions or offer recommendations by analyzing intricate data sets. This decision-making ability serves as a trademark of AI applications, distinguishing them from rule-based systems.

Natural Language Processing: Numerous Smart AI applications possess the ability to comprehend and generate human language, enabling seamless communication with users via natural language interfaces. This is particularly prevalent in chatbots, virtual assistants, and language translation applications.

Computer Vision: Certain Smart AI software incorporates computer vision capabilities, enabling it to interpret and comprehend visual information from images or videos. This is particularly prominent in applications such as facial recognition, image analysis, and autonomous vehicles.

Context Awareness: Smart AI software frequently exhibits a level of context awareness, comprehending the context in which it operates and adapting its behavior accordingly. This can significantly enhance user experiences across various applications.

Instances of Smart AI software include virtual assistants like Siri or Alexa, recommendation systems, autonomous vehicles, and advanced chatbots. These applications exemplify the transformative potential of AI in augmenting software capabilities, facilitating more sophisticated interactions and problem-solving.

3.1. QUALITY CHALLENGES IN SMART AI SOFTWARE

3.1.1. Quality Assurance Requirements & Testing Coverage

Quality Assurance (QA) requirements and testing coverage play a vital role in guaranteeing the dependability and effectiveness of AI software constructed using machine learning models. Let us elucidate these concepts using an illustration.

Requirements for Quality Assurance: When developing an AI-powered recommendation system for an e-commerce platform, certain QA requirements are indispensable. The utmost importance lies in achieving accuracy; the system must attain a minimum accuracy rate of 95% when providing personalized product recommendations to users. Ensuring the quality of training data is equally imperative. The data must be unbiased and represent a diverse array of customer demographics to prevent the system from perpetuating biases. Scalability is another crucial requirement; the system must be able to handle a 20% surge in user traffic without compromising its performance. Implementing security measures, such as encrypting user data to comply with data protection regulations, is of utmost importance. Ethical considerations dictate that the recommendation system remains unbiased, regardless of users' personal attributes.

Coverage for Testing: To comprehensively test the AI recommendation system, various coverage scenarios must be taken into account. Positive testing ensures that the system accurately predicts positive outcomes, such as recommending products based on positive customer feedback. Negative testing evaluates how effectively the system handles unexpected or negative scenarios, such as outliers in user behavior. Testing edge cases assesses the system's performance under extreme conditions, while monitoring for data drift ensures adaptability to changing input

data distributions. Adversarial testing is crucial for evaluating the system's robustness against intentionally crafted inputs. User interaction scenarios test the system's response to diverse user queries and requests, thereby ensuring a well-rounded testing coverage.

In conclusion, the QA requirements and testing coverage criteria elucidated above contribute to the development of a dependable, precise, and ethically sound AI recommendation system, addressing various critical aspects for its success in real-world applications.

3.1.2. Data Quality

The role of high-quality training data is of utmost importance in the development of robust and impartial artificial intelligence (AI) models. Nevertheless, numerous challenges can emerge during this process. One such challenge pertains to biases in the collection of data. Inherent biases may be present in the data collection process, which can result in distorted representations of particular groups or perspectives. Additionally, training data may contain inaccuracies, errors, or noise, thereby compromising the reliability of the model's learning process. Furthermore, the lack of diversity in the training dataset may give rise to models that encounter difficulties when attempting to generalize across various scenarios, consequently leading to subpar real-world performance.

Imagine a facial recognition system trained on a dataset that predominantly consists of images of light-skinned individuals. This scenario reflects a common bias in training data collection, where certain demographics are underrepresented (Lal & Kumar, 2021).

Following will be the Data Quality Impact-

Algorithmic Bias: The facial recognition model may demonstrate biases by exhibiting lower accuracy rates for individuals with darker skin tones due to the lack of diverse representation in the training data.

Ethical Concerns: In real-world applications, such a biased model could lead to unfair treatment, such as misidentification or discrimination against individuals with darker skin tones. This raises ethical concerns about the potential harm caused by the technology.

Limited Generalization: The model's inability to generalize across diverse skin tones compromises its reliability in various settings, contributing to a lack of inclusivity and fairness.

3.1.3. Algorithmic Transparency

Interpretation and explication of decisions made by AI models present formidable challenges, particularly when intricate models, such as deep neural networks, often function as "opaque entities," rendering it arduous to comprehend their decision-making processes. Ethical considerations come into play in crucial areas such as healthcare or finance, where AI determinations significantly impact human lives. In order to ensure accountability, fairness, and ethical utilization of AI, transparency holds paramount importance (Lima, 2020). Transparent AI models cultivate user trust through the provision of lucid insights into the rationale behind a particular decision. This trust assumes crucial significance, especially when AI influences decisions that carry far-reaching consequences. The ever-increasing regulations and standards increasingly demand transparency in AI systems, thereby accentuating the necessity for organizations to embrace practices that facilitate explanations of model decisions.

3.1.4. Adaptability and Generalization

Real-world data exhibits dynamism, characterized by variations and shifts that occur over time. The challenge lies in the adaptability of AI models to these changes in data distribution, as they may become obsolete or lose accuracy when confronted with new patterns.

AI models that are specifically tailored to certain scenarios may encounter difficulties in generalizing to diverse conditions. In order to recognize patterns across different environments, contexts, or user behaviors, it is crucial to employ robust strategies to adapt these models.

In dynamic environments, static models are inadequate. To ensure the continued relevance and effectiveness of AI systems, continuous learning becomes imperative over the course of time.

4. TECHNICAL ISSUES IN AI SOFTWARE QUALITY ASSURANCE

Testing AI models presents unique challenges due to the extensive input space and the dynamic nature of these systems. To address the complexity, a thorough understanding of the following challenges is necessary:

4.1. Expansive Input Space

AI models often operate in high-dimensional input spaces, making it impractical to conduct exhaustive testing. Achieving comprehensive coverage becomes difficult, leading to concerns regarding the model's performance with all possible inputs.

4.2. Dynamic Model Behavior

AI models adapt and evolve through continuous learning, resulting in dynamic behavior. Predicting the model's responses to different inputs becomes intricate because of the evolving nature of the underlying algorithms.

4.3. Non-deterministic Outputs

AI models, particularly those utilizing techniques such as neural networks, generate non-deterministic outputs. Ensuring consistent and reproducible results across diverse inputs necessitates the utilization of specialized testing approaches (Pham & Nguyen, 2022).

Here is some examples which explain technical issue with AI software testing

- The vast input space encountered by Natural Language Processing (NLP) models presents a formidable challenge, as it encompasses a wide array of linguistic variations, thereby posing difficulties in encompassing all potential inputs during testing. For instance, when evaluating a language translation model, it may encounter difficulties in handling rare or dialect-specific words that were not adequately represented in the training data, underscoring the significance of robustness in effectively handling diverse language inputs.
- Recommendation systems face the challenge of dynamically adapting to the evolving preferences and trends of users, necessitating continual learning and adjustment. For example, when testing a movie recommendation algorithm, it may become evident that accurately predicting users' shifting tastes over time is challenging, thereby highlighting the importance of testing scenarios that simulate changing user behavior.

- Computer vision models encounter the challenge of effectively generalizing across various data distributions, taking into account changes in lighting, perspectives, or backgrounds. For instance, when testing an object detection model, it may reveal issues when confronted with images captured under diverse conditions, thus emphasizing the significance of testing across a spectrum of scenarios to ensure robust performance.

5. SOCIETAL AND ETHICAL CONSIDERATIONS

Ethical considerations are of utmost importance in the testing of AI software to guarantee responsible development and implementation. An integral aspect entails the resolution of biases that may be present in the training data utilized to construct AI models (Sugali & Sprunger, 2021). For instance, if an AI model is trained on historical data that reflects societal biases, it may perpetuate these biases in its predictions. During the testing phase, it is crucial to identify and alleviate such biases in order to prevent discriminatory outcomes and foster equity.

The safeguarding of privacy constitutes another pivotal ethical consideration. Test data often comprises sensitive information, necessitating strict privacy measures. For instance, in the case of an AI application involving facial recognition, testing must ascertain that the privacy of individuals is upheld, and their facial data is neither misappropriated nor divulged without consent. Informed consent assumes a vital role in AI testing, particularly when human participants are involved. Individuals who contribute to testing activities must be fully apprised of the purpose, risks, and potential ramifications. The acquisition of consent ensures transparency and empowers users to make well-informed decisions regarding their participation.

Responsible disclosure emerges as an ethical consideration that entails dutifully reporting identified issues to pertinent stakeholders. If vulnerabilities or ethical concerns are discovered during testing, a clear procedure for disclosure must be in place to facilitate prompt resolution and minimize potential risks.

In conclusion, ethical considerations in the testing of AI software encompass the resolution of biases, the safeguarding of privacy, the acquisition of informed consent, the assurance of algorithmic transparency, the avoidance of harm, the assurance of security, and the adoption of responsible disclosure practices. By integrating these considerations, AI testing can uphold ethical standards, foster equity, and mitigate potential risks to both users and society.

6. SOLUTION - QUALITY ASSURANCE IN SMART AI SOFTWARE

6.1. Model-based AI Software Testing

Model-Based Testing (MBT) is an approach to software testing that utilizes models in order to design, generate, and execute test cases. Within the realm of AI software testing, MBT proves to be particularly valuable due to its capacity to systematically validate the behavior of intricate and dynamic AI systems. In the context of AI, MBT involves the creation of a model that represents the anticipated functionality, interactions, and decision pathways of the AI application. As an example, in the case of a speech recognition AI, the model may encompass states related to receiving audio input, processing it for speech recognition, and generating suitable responses. This model serves as the basis for the generation of automated test cases.

Let us consider a virtual assistant powered by AI as an illustrative example. The model could encompass states such as user queries, natural language processing, and task execution. Through

the utilization of MBT, test cases are automatically derived from this model, covering a range of scenarios that span from routine interactions to the handling of ambiguous queries.

The advantages of employing MBT in AI testing include enhanced coverage, early detection of defects, and efficient testing of various scenarios. By systematically generating and executing test cases based on the model, MBT contributes to ensuring the robustness and reliability of AI software in a plethora of real-world situations.

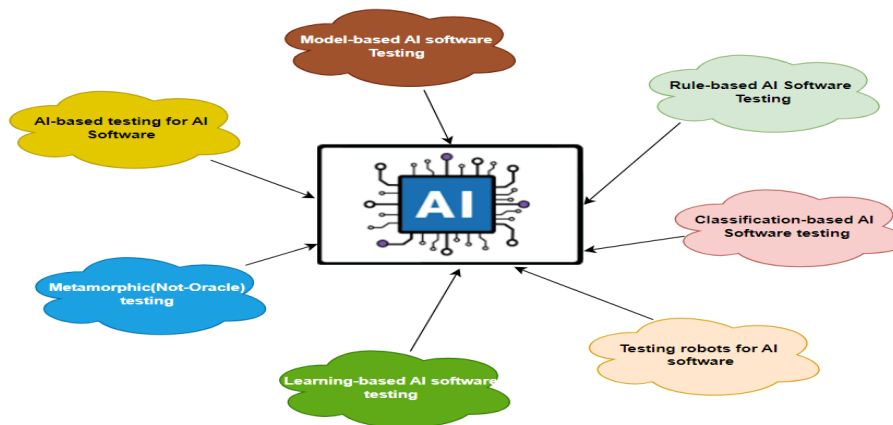


Fig-1 Smart AI software testing approaches

6.2. Rule-based AI Software Testing

Rule-Based Testing in AI entails the design and execution of tests based on predetermined rules and logical conditions that are specific to the behavior and requirements of the AI system. Unlike conventional software, where rules may be explicitly stated in the code, AI systems often operate on acquired patterns and intricate algorithms. The purpose of rule-based testing is to ascertain whether these implicit rules and patterns are functioning as intended.

To illustrate, let us consider an AI-powered fraud detection system. In the context of rule-based testing for this system, predetermined rules could encompass conditions such as "if a transaction surpasses a certain threshold and transpires outside the realm of usual user behavior, it should be flagged as potentially fraudulent." Subsequently, test cases would be devised to verify if the system accurately identifies and processes transactions in accordance with these rules.

Rule-based testing proves effective in scenarios wherein the behavior of an AI system can be expressed through explicit logical conditions. However, it may encounter limitations when dealing with intricate, non-deterministic AI models, where rules are either not explicitly defined or undergo dynamic evolution. Nevertheless, for rule-based AI applications, this testing approach ensures adherence to specified conditions and contributes to the system's reliability and accuracy in decision-making.

6.3. Classification-based AI Software Testing

Classification-Based Testing in AI entails the assessment of the precision and efficiency of the system's classification capabilities, guaranteeing its aptitude for correctly assigning inputs to predetermined categories or classes. This methodology is particularly relevant in applications such as image recognition, sentiment analysis, or any task where the AI system is designed to categorically group inputs. Take, for example, an AI-driven spam email filter. In the context of

classification-based testing, the system's proficiency in accurately categorizing emails as either spam or non-spam is evaluated. Test cases consist of presenting a variety of emails, some intentionally designed to resemble spam or legitimate messages, in order to determine if the AI system appropriately classifies them according to the predefined categories.

The effectiveness of classification-based testing lies in the validation of the model's capacity to generalize patterns and make precise predictions across a diverse range of inputs. It serves as a means to verify whether the AI system can discern subtle nuances and accurately classify real-world scenarios. This testing approach contributes significantly to enhancing the dependability and accuracy of classification tasks in AI applications, ensuring their optimal performance in their respective domains.

6.4. Testing robot for AI Software Testing

Testing robots in the realm of AI software testing pertain to automated systems that are specifically designed to validate the functionality, performance, and accuracy of AI applications. These robots undertake the emulation of user interactions, generating a diverse array of inputs to evaluate the AI system's responsiveness under varying conditions. An instance of an AI testing robot entails the creation of a virtual user that engages with a natural language processing (NLP) AI model. Let us consider an AI-powered chatbot. A testing robot could effectively simulate user conversations by dispatching a plethora of queries encompassing a range of complexities, languages, or sentiments. This aids in the evaluation of the chatbot's level of comprehension and its ability to provide appropriate responses. The testing robot diligently observes the ensuing replies, thereby assessing the AI model's accuracy and contextual relevance.

Testing robots contribute significantly to the efficiency and scope of AI software testing as they automate repetitive and intricate test scenarios. In doing so, they are able to uncover potential issues such as misinterpretations, errors, or biases that may be present within AI models. By utilizing testing robots, a thorough evaluation of the AI system's capabilities is ensured, thus bolstering the reliability and resilience of AI applications in real-world scenarios.

6.5. Learning-based AI Software Testing-

Learning-Based AI Software Testing entails the utilization of machine learning techniques to enhance the efficiency and effectiveness of testing procedures for AI applications. In this particular approach, the testing system acquires knowledge from the behavior exhibited by the AI model and subsequently adjusts its testing strategy accordingly. An illustration of learning-based AI testing can be observed in the continuous improvement of test cases based on the AI system's evolving patterns. Let us consider an AI-driven recommendation engine. In the realm of learning-based testing, the testing system initially employs a diverse set of test cases to evaluate the accuracy of the recommendations. As the AI model progresses through the accumulation of additional data and usage, the testing system dynamically modifies its test cases, prioritizing scenarios that mirror the ever-changing user preferences and content dynamics (Khankhoje, 2023).

Learning-based testing proves to be particularly advantageous in dynamic AI environments, wherein models undergo continuous evolution. It enables testing systems to adapt to the learning patterns exhibited by the AI, thereby enhancing test coverage and ensuring the testing process remains effective amidst changing circumstances. This approach significantly contributes to the overall quality assurance of AI applications, as it aligns testing strategies with the progressive nature of the AI models that require validation.

6.6. Metamorphic (Non-Oracle) for AI Software Testing-

Metamorphic Testing, particularly in a non-oracle environment, is a methodology devised to validate the accuracy of an application without relying on explicit anticipated results. Instead, it concentrates on the alteration of inputs and investigates whether the modifications in output conform to recognized associations. An instance of metamorphic testing can be demonstrated in the context of an artificial intelligence system for image recognition. Let's consider an artificial intelligence model that has been trained to recognize objects in images. In the realm of metamorphic testing, the system is subjected to a sequence of transformations, such as rotations, scaling, or cropping, which are applied to the input images. The expectation is that, despite these alterations, the artificial intelligence model should consistently identify and categorize the objects correctly. The absence of explicit anticipated outcomes renders it an approach that does not rely on an oracle.

Metamorphic testing excels in situations where defining precise expected results is difficult or unfeasible. By scrutinizing the associations between transformed inputs and outputs, it furnishes a robust mechanism to validate the resilience and dependability of artificial intelligence models in diverse real-world conditions. This approach contributes to enhancing the trustworthiness and generalization capabilities of artificial intelligence systems.

6.7. AI-based for AI Software Testing-

AI-Based Testing in the realm of AI software testing entails the utilization of techniques rooted in artificial intelligence to conceive, execute, and refine the testing procedures for AI applications. This methodology capitalizes on the capabilities of AI to independently generate test cases, optimize test coverage, and identify potential issues in AI models. A case in point can be observed in the testing of the precision of a natural language processing (NLP) model.

In AI-based testing, the system employs algorithms grounded in machine learning to scrutinize patterns in language usage and generate various test inputs that mimic real-world scenarios. In the case of an NLP model, this may involve the creation of test cases that exhibit different sentence structures, linguistic complexities, and contextual nuances. Subsequently, the AI-based testing system evaluates the responses of the model, identifying any inconsistencies, biases, or inaccuracies.

This methodology amplifies the efficiency and efficacy of testing AI applications by integrating intelligent automation. AI-based testing adapts to the ever-evolving nature of AI models, thereby ensuring comprehensive validation across varying conditions. It contributes to the dependability and resilience of AI systems by tackling the distinctive challenges posed by intricate algorithms and data-driven functionalities.

7. AI SYSTEM VALIDATION NEED

- The current software testing models and methods have limitations in addressing the requirements of AI software testing. These limitations include supporting multi-models with unstructured input data, handling large-scale classified inputs, addressing oracle problems, and ensuring accuracy, consistency, correctness, and relevance of quality.
- The majority of current AI software is equipped with machine learning models developed by data scientists using scientific algorithmic approaches and large-scale data training. However, there is a significant gap in considering quality validation and assurance from an

engineering perspective. Therefore, there is a need for AI testing research to study and develop new and effective quality standards and evaluation methods.

- The development of powerful AI software requires the utilization of large-scale training and test datasets. However, the current methods of training and data generation lack considerations for quality, assessment, and certification. As a result, it is necessary to explore how to create quality training data models and develop methods for generating large-scale quality test data.
- AI-based software testing entails the utilization and implementation of AI techniques and remedies to effectively enhance the process of software testing, encompassing the selection of test strategies, generation of tests, selection and execution of tests, detection and analysis of bugs, as well as prediction of quality.

8. CONCLUSION

In conclusion, the responsible and effective deployment of artificial intelligence necessitates addressing the crucial quality challenges, solutions, and needs that arise in Smart AI software. The multidimensional nature of AI system validation is underscored by the discussed aspects, including accuracy, robustness, ethical considerations, and adaptability. To ensure the reliability and trustworthiness of AI applications, it is imperative to adopt comprehensive testing strategies, ethical frameworks, and continuous monitoring, particularly as AI technologies continue to evolve.

The advancements in Smart AI software quality that are anticipated in the future are likely to manifest in the following areas:

- **Explainable AI (XAI)**
The focus of future trends will lie in the development of methods to elucidate complex AI decisions, thereby addressing the interpretability challenge, as the demand for transparent and interpretable AI models increases.
- **AI Ethics and Regulations**
It is anticipated that stricter ethical guidelines and regulations governing AI development and deployment will be established in the future, with an emphasis on fairness, accountability, and transparency.
- **Automated Testing for AI**
The testing process will be streamlined in the future through the evolution of automated testing tools that are specifically designed for AI models, enabling efficient validation and faster deployment cycles.
- **Adversarial Defense Techniques**
Future AI systems will incorporate advanced defense mechanisms against adversarial attacks, thereby enhancing robustness and security.
- **AI Quality Metrics Standardization**
The development of standardized metrics for assessing AI quality will facilitate benchmarking and comparison across different AI models and applications.
- **Continuous Learning and AI Maintenance**
Given the dynamic nature of data and applications, AI systems will increasingly adopt continuous learning approaches, which will necessitate ongoing testing, validation, and maintenance.

- **Human-AI Collaboration**

Future trends will emphasize the integration of AI systems with human workflows, resulting in improved collaboration and synergy between AI capabilities and human expertise.

By embracing these future trends, existing challenges can be overcome, and Smart AI software can be ensured to meet the highest standards of quality, reliability, and ethical responsibility. The pursuit of excellence in AI quality assurance is crucial for unlocking the full potential of artificial intelligence in diverse domains.

REFERENCES

- [1] Gao, J. (2022). AI-Testing-Presentation-Gao.pptx - AI Testing - A Tutorial Presented by: Jerry Gao Professor and Director San Jose State University - Excellence. Course Hero. Retrieved November 27, 2023, from <https://www.coursehero.com/file/145378900/AI-Testing-Presentation-Gaopptx/>
- [2] Haller-Seeber, S., & Gatterer, T. (2022). Software Testing, AI and Robotics (STAIR) Learning Lab. 10.48550/arXiv.2204.03028
- [3] Job, M. A. (2020). Automating and Optimizing Software Testing using Artificial Intelligence Techniques. International Journal of Advanced Computer Science and Applications. 10.14569/IJACSA.2021.0120571
- [4] Khaliqa, Z., & Farooqa, S. (2022). Artificial Intelligence in Software Testing : Impact, Problems, Challenges and Prospect. 10.48550/arxiv.2201.05371
- [5] Khankhoje, R. (2023). Effortless Test Maintenance: A Critical Review of Self-Healing Frameworks. 11(X). 10.22214/ijraset.2023.56048
- [6] Khankhoje, R. (2023). WEB PAGE ELEMENT IDENTIFICATION USING SELENIUM AND CNN: A NOVEL APPROACH. Journal of Software Quality Assurance (JSQA), 1(1), 1-17. https://iaeme.com/MasterAdmin/Journal_uploads/JSQA/VOLUME_1_ISSUE_1/JSQA_01_01_001.pdf
- [7] Lal, A., & Kumar, G. (2021). Intelligent Testing in the Software Industry. 10.1109/ICCCNT51525.2021.9580012
- [8] Lima, R. (2020). Artificial Intelligence Applied to Software Testing: A Literature Review. 10.23919/CISTI49556.2020.9141124
- [9] Pham, P., & Nguyen, V.-L. (2022). A Review of AI-augmented End-to-End Test Automation Tools. 10.1145/3551349.3563240
- [10] Sugali, K., & Sprunger, C. (2021). Software Testing: Issues and Challenges of Artificial Intelligence & Machine Learning. International Journal of Artificial Intelligence & Applications. 10.5121/IJAIA.2021.12107

AUTHOR

I am Rohit Khankhoje, a Software Test Lead with over 15+ years of experience in software quality assurance and test automation. With a passion for ensuring the delivery of high-quality software products, I am at the forefront of harnessing cutting-edge technologies to streamline and enhance the testing process. I am dedicated to advancing the automation testing field and continue to inspire colleagues and peers.

DNA SEQUENCE AUTOMATIC CLASSIFICATION—LEARN THE LIFE LANGUAGE USING ARTIFICIAL INTELLIGENCE

Josephine (Hsin) Liu^{1,2}, Phoebe (Yun) Liu^{1,2}, Joseph (Yu) Liu^{1,2}, Emily X. Ding¹,
Robert J. Hou¹

1 Vineyards AI Lab, Auckland, New Zealand

2 Rangitoto College, Auckland, New Zealand

ABSTRACT

This paper explores the applications of Artificial intelligence (AI) techniques for classifying Deoxyribonucleic Acid (DNA) sequences into their corresponding gene families. The paper focuses on presenting how to treat DNA sequences as a human language to be understood and classified. Specifically, we first transformed the DNA sequences into a more human-like format, then we employed Natural Language Processing (NLP) and Multi-layer perceptron (MLP) algorithms to complete sequence classification into 7 gene families. Our research drew DNA sequence data from three organisms, including humans, dogs, and chimpanzees. Finally, various experiments are conducted to prove the classification performance. In addition, to prove the generalization of this solution, we designed experiments that involved cross-domain testing. These experimental results display not only high accuracy and efficiency but also intriguing findings in life sciences.

KEYWORDS

DNA Sequences, Auto Recognition, Natural Language Processing(NLP), Multi-layer Perceptron (MLP)

1. INTRODUCTION

DNA stands for deoxyribonucleic acid, it is a macromolecule made up of nucleotides; phosphate sugar backbone, and nitrogenous bases A, T, C, and G, each in a different order and sequence. DNA can form genetic instructions to guide individual organism development and the functions of individual cells ensuring the survival and growth of the organism. It stores the required information for each cell and micro molecule to function, often described as the “blueprint” of the body. The different sequencing of DNA is the building block determining the structure of the DNA molecule. Segments of specific DNA sequences form genes, and these genes form gene families, genes are then responsible for gene expression or why our body functions the way it does. Scientists discovered that by classifying and identifying gene families from DNA sequences, diagnosis of early diseases can be made and predicted. After initial research, it was shown that DNA sequences are an important part of the biological field as the ability to understand DNA and classify it into families, can cause crucial breakthroughs in scientific and medical fields. Figure 1 displays some examples. Through DNA sequences, scientists can read, understand, and compare genetic information, potentially causing a breakthrough in biological studies and medical fields.[1-5].

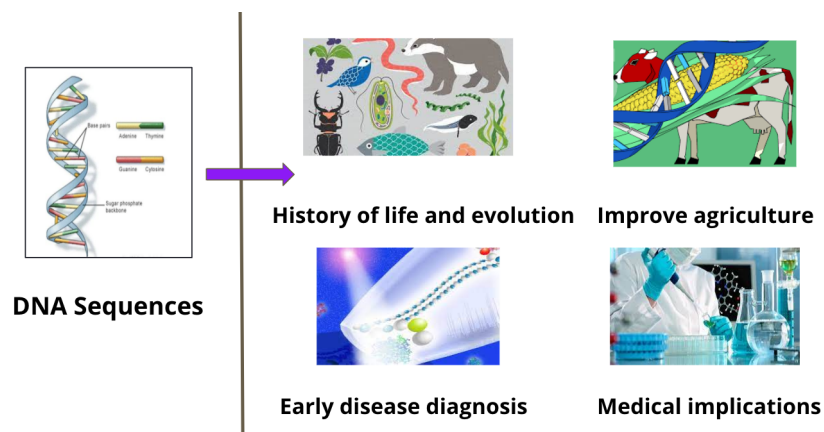


Figure 1. DNA sequences research fields and applications

However, much of this research is based on the premise of annotation of DNA sequences. The annotation is facilitated by classifying DNA sequences into families[6,7]. Predicting the DNA sequence into classes can provide insights into how an organism regulates and expresses genes. For example, if a specific DNA sequence is given, scientists can predict the possible relationship between the DNA sequence's function and its gene family. This would be worthwhile and crucial for genomic sequencing research, therefore, we chose to explore this topic further and predict the classes of DNA sequences.

It is challenging to recognize DNA sequences. Manual marking is time-consuming and error-prone, whereas it is evident that AI technology has the potential to make it far more efficient and accurate. Furthermore, AI is less costly in the long run because it does not require extra costs once the product or model is built, except for nominal costs such as maintenance.

How can AI technology be applied to the project? It is proven that DNA sequences are not only the language of life but also extremely similar to a human language, as it includes the specific "letters" and "phrases" needed to express and communicate information[8]. They are translated into the sequence of amino acids in a protein and can be understood and interpreted by other molecular machines within cells. Natural Language Processing (NLP) is an area of computer science that deals with methods to analyze, model, and understand human language. It performs well on many language-related tasks, such as language translation, sentiment analysis, speech recognition, and text summarization[9-11].

Thus the question arose: How can DNA sequences be treated as one of the human languages to understand and identify them?

In this research, a solution was developed to classify DNA sequences into their respective types with the help of AI technologies such as Natural Language Processing (NLP) and deep neural networks. The project motivation and research plan are shown in Figure 2.

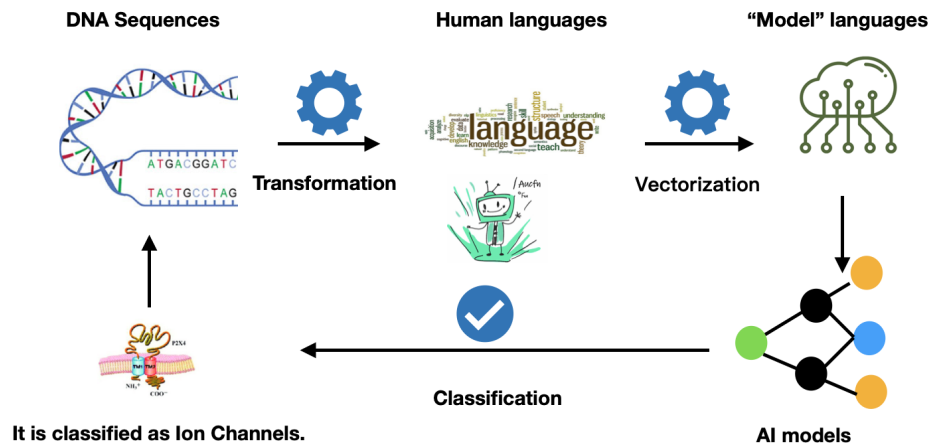


Figure 2. The motivation and research plan for our project

2. OUR METHOD

From Figure 2, there are three main parts for automatically classifying DNA sequences. First, we transformed the DNA sequences into texts similar to human language. Then, we extracted features from the "texts" to transform them into vectors. This part is called vectorization. At this stage, DNA sequences have already been transformed into the language that a model can understand, and are ready to feed the classification model. Finally, we built and trained a deep neural networks (DNNs) classifier based on the extracted features.

2.1 Transforming the DNA sequences into texts similar to human language

DNA sequences are composed of the "letters" A, C, G, and T in a particular order. They, just like human language, communicate the secrets of life waiting for us to understand. Some patterns hidden in these codings decide the gene's functions and structures. It is challenging to use the original sequences for the classification directly. By Finding the letters in DNA sequences, we could treat them like a human language, such as English, and process them as texts composed of several words. The methods applied to NLP could be exploited to classify DNA sequences. The transformation of DNA sequences into a text is shown in Figure 3.

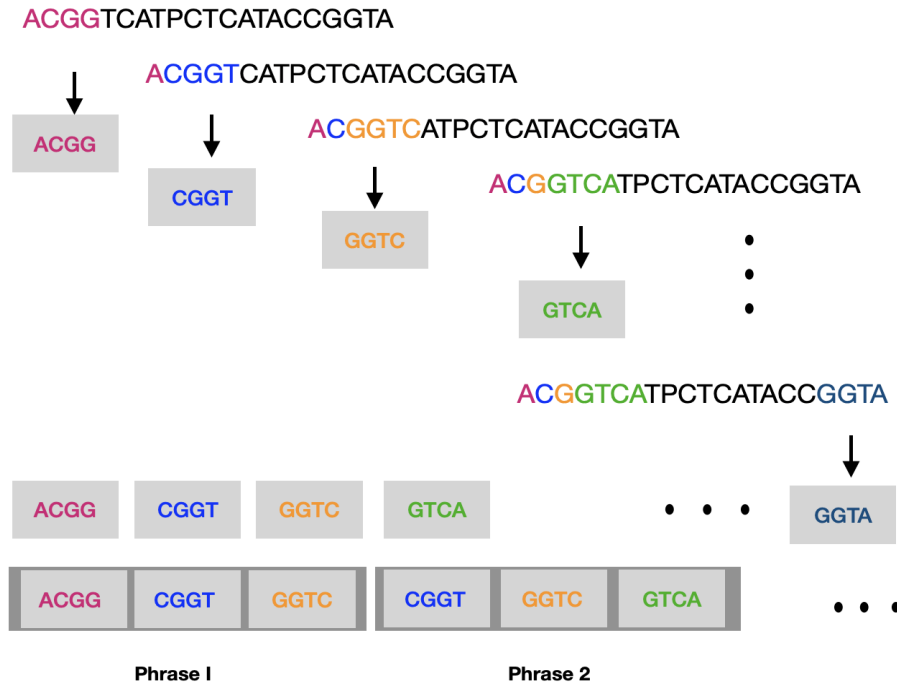


Figure 3. Transformation of DNA sequences into text

As we learned the set concept and how to define a set in maths class, we tried to describe the transformation process using two sets. Here, we defined the first set to describe the DNA sequences in a text with lots of words, as follows:

$$\mathbf{X}_{Text} = \left\{ \mathbf{X}_{words} \mid \mathbf{X}_{DNA-Seq}(i_n): \mathbf{X}_{DNA-Seq}(i_n + l_{word}), i \leq \text{length}(\mathbf{X}_{DNA-Seq}) - l_{word} + 1 \right\} \quad (1)$$

Where $\mathbf{X}_{DNA-Seq}$ is one of the DNA sequences, i_n is the position of nucleotides and l_{word} is the word length. The \mathbf{X}_{Text} is composed of words \mathbf{X}_{words} with the same size in the set.

In addition, we group the words into several "phrases" without punctuation in the text. We also can call these "phrases" as "word bags". They will be treated as independent targets in the next section and recorded statistically by the times (frequency) they appear in the text. Therefore, we defined the second set as follows:

$$\mathbf{X}_{text-new} = \left\{ \mathbf{X}_{phrase} \mid \mathbf{X}_{Text}(i_{word}): \mathbf{X}_{Text}(i_{word} + l_{phrase}), i_{word} \leq \text{length}(\mathbf{X}_{Text}) - l_{phrase} + 1 \right\} \quad (2)$$

Where i_{word} is the i th word in the \mathbf{X}_{Text} and l_{phrase} is the size of the phrase, different from the l_{word} , it can be a range and also can be fixed.

To this end, DNA sequences are transformed into natural languages, and it is then ready to enter the next section and represent the "text" in vectors where models can identify and process them.

2.2 Vectorization for the "texts"

2.2.1 Background for Vectorization

We have transformed DNA sequences into texts as described above. Then, in this section, we will convert the texts to vectors. Firstly, we should ask, why do we have to conduct this conversion? It could also be helpful to understand what vectors are in general. Finally, a natural question, how is text represented with vectors? We will explore and answer the questions below. The computer can not understand letters or words directly, so the text must be encoded into numeral numbers. Some popular methods for vectorization exist, such as Bag-of-Words(BoW), Word Embeddings, character-level

features, etc. In our project, we chose the classic and simple ones belonging to (BoW). The resulting vector contains the counts or frequencies of each "word bag" in the text. Then we introduced two classic methods we explored; Countvectorizer (CV), and Term Frequency-Inverse Document Frequency (TF-IDF). After the conversion, the numeral data can be used to represent the text, though different results between the two methods.

However, we still wonder why it is called a vector. We seek the definition of vector and try to understand it. It is an object that has both a magnitude and a direction, like an arrow, whose length can be seen as the magnitude of the vector (numeral numbers), and the arrow indicates the directions. We will have a further understanding during the research.

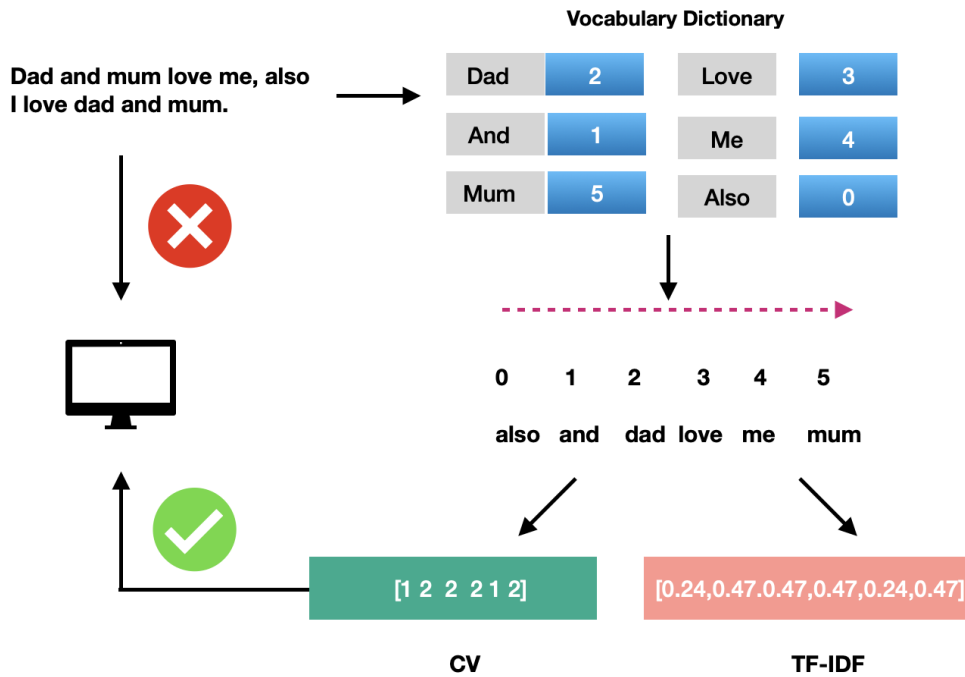


Figure 4. Vectorization for the text

Figure 4 demonstrates the vectorization for the CV and TF-IDF when we set the fixed size of "phrase" as 1. The text "Dad and mum love me also I love dad and mum" is represented by the numeral data. First, the vocabulary in the text is labeled in the dictionary, such as dad is "2" and love is "3". That means their features will be set in positions labeled 2 and 3. Until now, we can understand why we call it a vector. Their features obtained by CV or TF-IDF are the magnitude and arranged according to the vocabulary index for the directions. We can find that the word "I" is missing. Because as a stopping word (common word), such as "is", "are", etc., the stopping words in English are to be neglected. Next, we'll introduce the CV and TF-IDF, what they are, and how to compute the magnitude.

2.2.2 Countvectorizer and TF-IDF

Countvectorizer and TF-IDF are popular methods in NLP to extract features of texts. They are simple and effective in representing text as numerical data. Both measure the importance of words in a text document in different ways. Let's first introduce the Countvectorizer.

As shown in Figure 4, the count vectorizer builds a vector with the same dimension as the size of the vocabulary dictionary first. Then for each word, we calculate its frequency appearing in the text. The numeral or magnitude can be weighted as follows:

$$TF_d(X_{text-new}(d)) = \frac{N_d}{N_{text-new}} \quad (3)$$

Where $X_{text-new}(d)$ is the d th elements in the set $X_{text-new}$, N_d is the frequency of $X_{text-new}(d)$ appeared in the set $X_{text-new}$, and $N_{text-new}$ is the total number of elements in the set $X_{text-new}$.

Compared to the count vectorizer, TF-IDF not only cares for the frequency of the words appearing in this document but also considers how many times the same words appear in other documents. Therefore, there are two parts in TF-IDF described in the equation (4).

$$TF - IDF(X_{text-new}(d)) = TF_d(X_{text-new}(d)) \times IDF(X_{text-new}(d)) \tag{4}$$

Where $IDF(X_{text-new}(d))$ is called Inverse Document Frequency (IDF), and can be calculated as follows:

$$IDF(X_{text-new}(d)) = \log \frac{l+1}{l_d+1} + 1 \tag{5}$$

Where l is the number of all texts, and l_d is the text number include the d th elements in set $X_{text-new}$. In equation(5), we could not understand the log function initially. We searched for the definition of this kind of function. But we overcame it and did it, luckily. First, the definition of \log function is : if $a^x = b$, then $x = \log_a b$. In our situation, we know that $\frac{l+1}{l_d+1} > 1$, smaller l_d , larger $\frac{l+1}{l_d+1}$. That means a larger $\frac{l+1}{l_d+1}$ leads to a larger $\log \frac{l+1}{l_d+1}$ too. So the fewer texts that include this element, the score of IDF of this element is larger. To avoid the zeros appearing in the numerator and denominator $\frac{l+1}{l_d+1}$, we add 1 to them.

2.3 Multi-layer Perceptron (MLP) as Automatic Classifier

In this section, we will build a supervised classification model as the vectorized features are ready.

2.3.1 A brief introduction to supervised learning

Supervised Learning (SL) is the main strategy for learning knowledge from data in AI. Usually, data for SL are paired with their labels and split into training and testing parts. Training data teaches the model how to classify while testing data is used to evaluate the model's performance[12-14]. A couple of analogies could be helpful to explain supervised learning as shown in Figure 5.

In a simplified learning process, babies learn to recognize objects and become mature as taught by their parents. In the very beginning, babies are taught what an object is (data and its paired label). After some repetition (training), babies become confident in recognizing the taught objects.

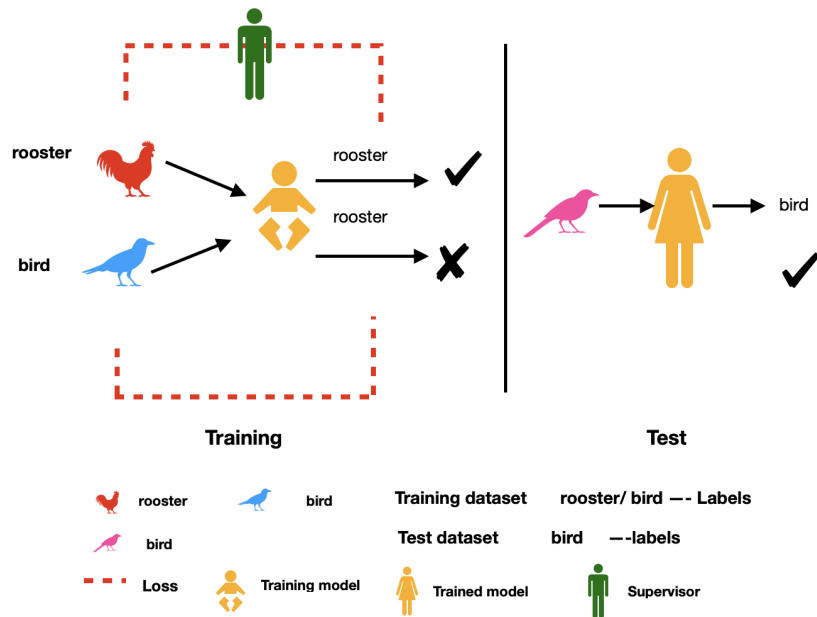


Figure 5. Analogies for Supervised Learning

Therefore, the most important thing is to build a model to describe the learning process from a "baby" to an "adult".

2.3.2 Artificial Neural Network—Multi-layer Perceptron (MLP)

In this part, we'll introduce a classic artificial neural network called multi-layer perceptron (MLP). We spent much time researching the mechanism of MLP and building them in Tensorflow. We tried our best to understand the background theory and explore the parameters for our classification task. Thanks to being a team, we discussed and explained them from various views through many similar analogies in our lives.

As we know, neural networks are inspired by and simplify the functioning of biological neurons. Biological neurons comprise dendrites, cell bodies, axons, and other parts. The dendrites are mainly used to receive signals from other neurons, while the axons output signals from these neurons. Synapses are the gap between the axon and other neurons' dendrites. Tens of thousands of neurons cooperate, enabling us to have advanced thinking and constantly learn new things. Usually, the neurons have two states: fire(active) and rest. When the stimulus received is higher than a certain threshold, it will be fired; otherwise, there is no nerve impulse. Figure 6 shows the biological neurons to a "node" in a neural network[15-17].

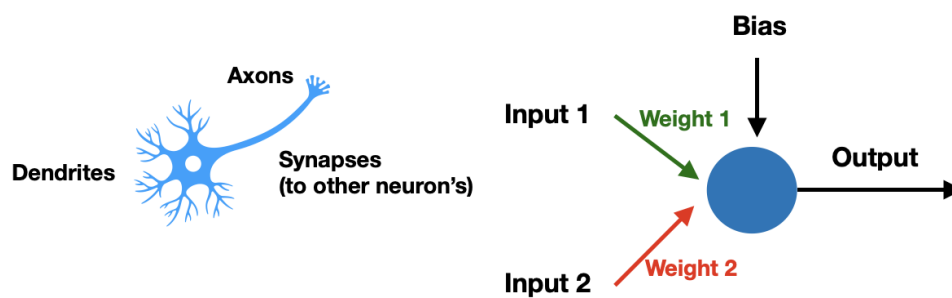


Figure 6. Biological neurons to a "node" in a neural network.

In Figure 6, the red line and green lines are just a synapse, that determines the weights, then the output for every node can be described as follows:

$$output = f(input1 * weight1 + input2 * weight2 + bias) \quad (6)$$

In equation(6), the *bias* is a constant value, which *f* is called the activation function, which is a non-linear function. There are some different activation functions, such as Sigmoid, tanh, and ReLU. The Sigmoid function is $\sigma(x) = 1/(1 + \exp(-x))$, and the output ranges from 0 to 1, and $\tanh(x) = 2\sigma(2x) - 1$, and output is as $[-1, 1]$, the last one is ReLU as $f(x) = \max(0, x)$, 0 is the threshold. The bias allows the network to shift the activation function to a different region, and better fit the training data.

Then all the nodes will be fully connected which means that each node in a layer is connected to all other nodes in the next layer. Each connection has a weight. An MLP consists of at least three layers of nodes: an input layer, a hidden layer, and an output layer. The structure is shown in Figure 7.

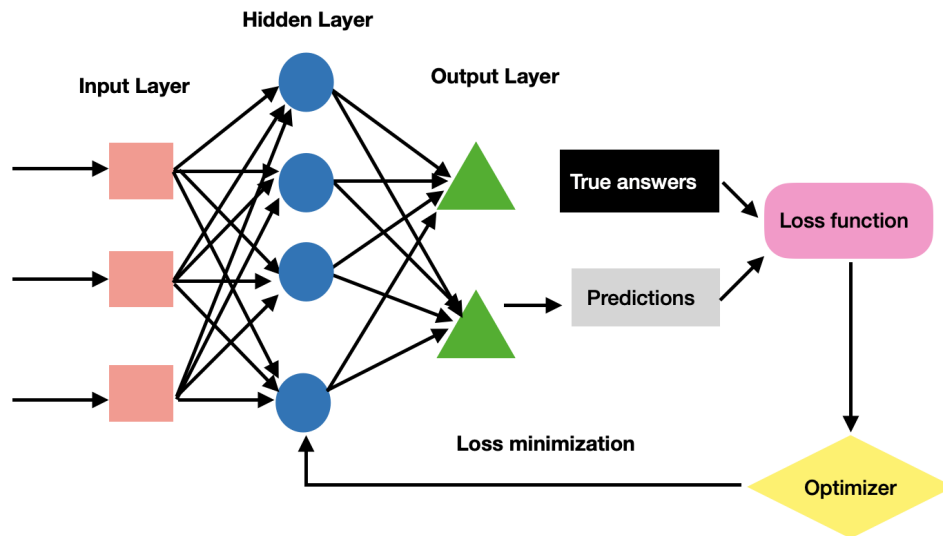


Figure 7. MLP network structure

Figure 7 displays a simple MLP network architecture. This type of network is called a feedforward network since it has no loops (i.e. the output of a neuron never connects to the input of a neuron in the last layer). MLP network includes three parts, the input, hidden, and output layers. How to optimize the weights and biases in the network to make the model smarter, relies on the loss function and the optimizer. The loss function is a method to measure the difference between the actual and predicted outputs. Learning aims to minimize the loss functions by adjusting the weights and biases. In our task, we'll choose the categorical cross-entropy loss function commonly used for multi-classification. The optimizer is an algorithm used to minimize the value of the loss function. Various optimizers are available, such as stochastic gradient descent (SGD), Adam, Adagrad, RMSprop, etc. Among them, generally, Adam is considered one of the best optimizers.

3. EXPERIMENTS

3.1 Datasets

We downloaded the DNA sequences dataset from Kaggle, an online community where users can find and publish data sets and explore data science. This dataset includes more than 6500 DNA sequences of three organisms, among them, 4380 from humans, 820 from dogs, and 1682 from chimpanzees. They are annotated into 7 classes as shown below. Meanwhile, Figure 8 shows the class distributions of humans, dogs, and chimpanzees.

Table 1. DNA sequence types and the labels in the dataset

Gene family	Class Label
G protein-coupled receptors	0
Tyrosine kinase	1
Tyrosine phosphatase	2
Synthetase	3
Synthase	4
Ion channel	5
Transcription factor	6

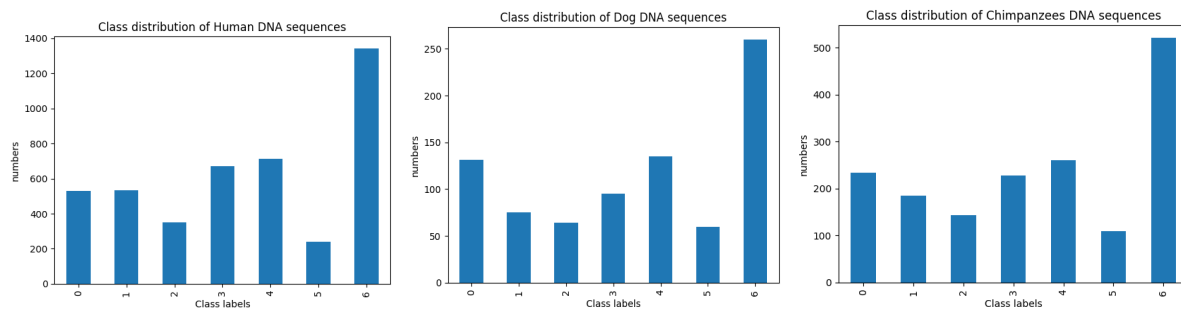


Figure 8 Classes distribution in human, dog, and chimpanzee data

Figure 8 shows the number of DNA sequences of each type (y-axis), human, dog, and chimpanzee, have been distributed against each class (x-axis). The graphs show that the most frequently distributed class is labeled number 6, the Transcription factor, whereas the least distributed class is 5, the Ion channel. Because the ratio of the distributions is roughly the same (despite the actual numbers being different), we can conclude that these three datasets have at least some level of similarity between them. Furthermore, we can see that the number of DNA sequences of the human datasets vastly surpasses that of the other two animals, such as class 6 in the y-axis (1,400) vastly surpasses chimpanzees at 500 and dogs at 250.

3.2 Evaluation of the different lengths of words and phrases

In this part, we evaluate the word and phrase parameters of our model. The parameters are the length of the word and phrase. In this experiment, human DNA sequences are used as our datasets. At the same time, we also record the running time in 100 Epochs and accuracy in these various situations. During these experiments, we fixed the other model settings, as we chose 2 hidden layers of MLP, and neuron numbers are 64 for the first and 128 for the second layer. The counter vectorizer method is used for vectorization, and the activation function ReLU is employed in the network.

Table 2. Performance of different lengths of “word” and” phrase”

phrase's length	Performance	$l_{word} = 2$	$l_{word} = 3$	$l_{word} = 4$
$l_{phrase} = 2$	Acc	0.7648	0.8299	0.8299
	Time	3m	4m	6m
	Dim	84	336	1247
$l_{phrase} = 3$	Acc	0.8550	0.9098	0.9269
	Time	3m	6m	11m
	Dim	336	1247	4459
$l_{phrase} = 4$	Acc	0.9155	0.9224	0.9586
	Time	5m	10m	10m
	Dim	1247	4469	16834
$l_{phrase} = 5$	Acc	0.9281	0.9521	0.9532
	Time	9m	35m	38m
	Dim	4469	16834	65447

As shown in the above table, the accuracy increases along with the gram and word lengths increase, albeit in an uneven manner, and at the cost of more time consumed. Thus, even though increasing the word length and n-grams will increase the accuracy, it is unrecommended to simply choose the one with the highest n-gram and word length, as it is impractical and time-consuming in real life. For example, although the model with a word length of 4 and n-gram of 4 has an accuracy of 0.9586, it also takes too much time. Through this, we can see that a certain balance between accuracy and time taken must be achieved to build a successful model. As such, the model with $l_{word} = 4, l_{phrase} = 4$ is recommended as it has a relatively high accuracy and an acceptable time.

3.3 Feature Extraction

In this section, we'll evaluate the performance of two vectorization methods. According to the experiments above, we chose $l_{word} = 4$ and $l_{phrase} = 4$. The network parameters are the same as in the previous experiments.

Table 3. Performance of different vectorization methods

Vectorization methods	Human	Dog	Chimpanzees
Countvectorizer	0.9586	0.7805	0.9139
TF-IDF	0.8505	0.6890	0.8497

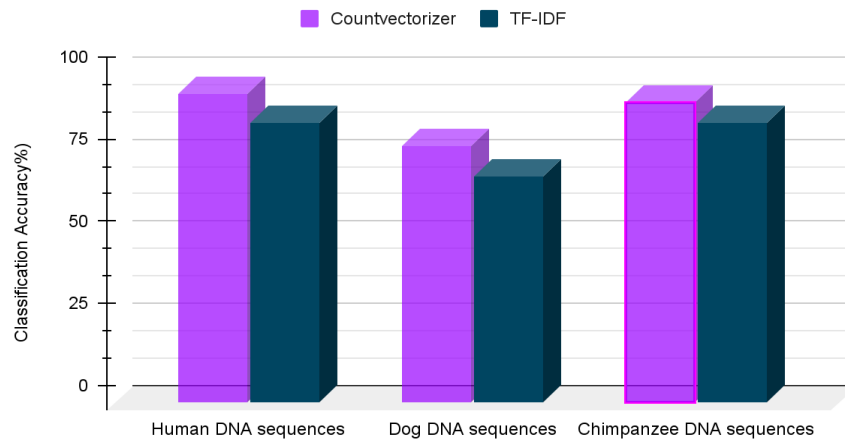


Figure 9 Model performance of Countvectorizer and TF-IDF

The experiment results show that a higher classification accuracy was achieved when we used a count vectorizer as the vectorization method for all three organisms. An interesting finding is that the accuracies were a bit lower with TF-IDF, though a better performance was expected from theory analysis and experiments in some NLP references. A possible reason is that the "words" in our project are not actual words, so it would not be very helpful to introduce IDF.

3.4 MLP performance

We explore the network structure and activation functions in this section. Also, based on the above experiment results, we choose the count vectorizer as the vectorization method, and $l_{word} = 4$ $l_{phrase} = 4$. The human DNA sequences are used as the dataset.

Table 4. Performance under the different network settings

Human dataset	Performance	Hidden Layers =1	Hidden Layers=2	Hidden_layers=3
Sigmoid	ACC	0.9555	0.9578	0.9372
	Time	20 min	20 min	19 min
ReLU	ACC	0.9372	0.9586	0.9532
	Time	9 min	10 min	11 min
Tanh	ACC	0.9475	0.9532	0.9578
	Time	10 min	10min	11min

The experiment results show a small gap in accuracy using different numbers of layers and activation functions. However, the running time is longer when the activation function Sigmoid was used. That means Sigmoid could need more computation. Certainly, as the hidden layers increase, more time is

required. Therefore, in this project, we seek a balance between the accuracy and cost of time. As a result, we choose ReLU as the activation function and two hidden layers in our neural networks.

3.5 Visualisation for the classification performance

To explore more, in this section, we will demonstrate the results with visualization. First, let's introduce the confusion matrix. We can show the classification results in a matrix form. It displays how many samples are classified correctly and incorrectly per class visually. In addition, we show how the accuracy changes with every epoch with a graph. The experiment results are shown in Figure 10.

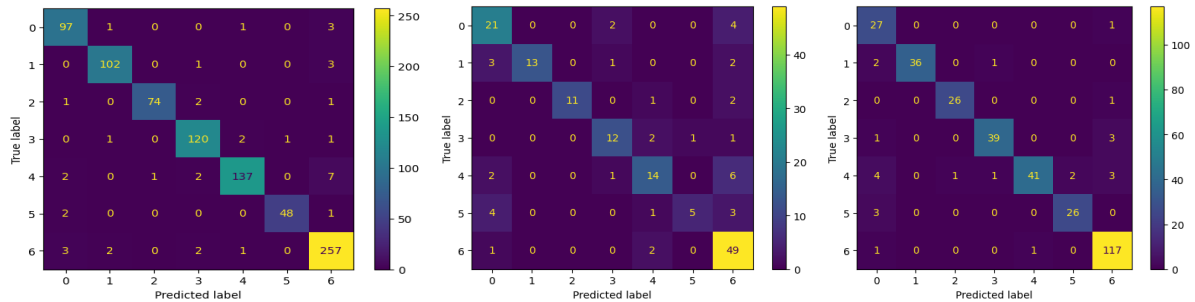


Figure 10 Confusion matrix of testing human, dog, and chimpanzee data

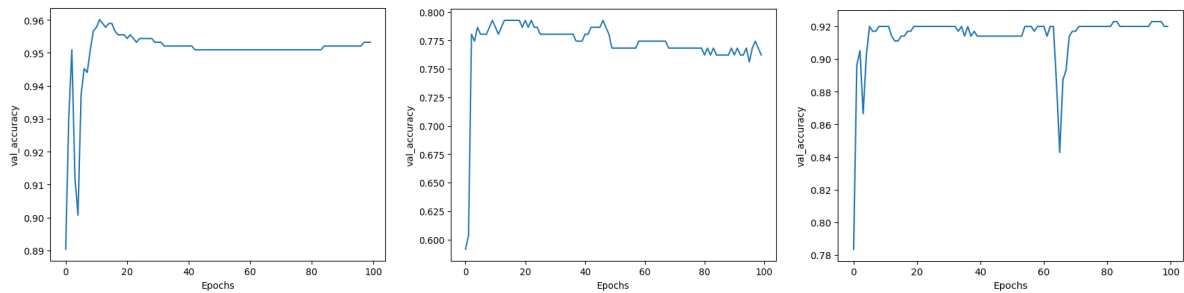


Figure 11 Training history for human, dog, and chimpanzee data

The results show that class '6' can be easily recognized by the model with high accuracy. While class '4' and '5' have a lower accuracy compared with other types, which means they are harder to classify by the model. Figure 11 displays the training history for epochs against val_accuracy, for human, dog, and chimpanzee data, the val_accuracy (accuracy in test data) tends to converge before the 100th epoch.

3.5 Generalization

We have conducted various evaluations to explore the model above. Furthermore, the generalization of AI models is crucial. Generalization is the ability to learn patterns from the training data that can be used for a new, unseen dataset. It is a key ability to build models with robust performance[18-19]. Figure 12 shows an example of model generalization.

Therefore, to evaluate the generalization, we designed our experiments to train the model on dataset A and test it on dataset B. We set the experiment as $l_{word} = 4$ and $l_{phrase} = 4$, Hidden layers =2, the activation function is ReLU, and Count-vectorizer is the vectorization method.

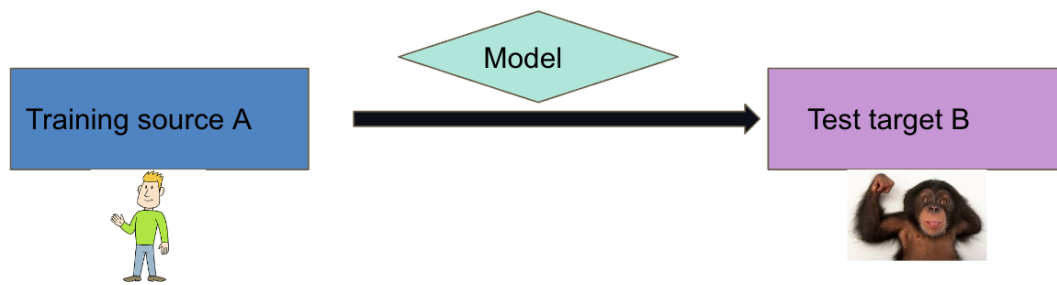


Figure.12 An example of model generalization

Table 5. Generalization performance of the model

Trained in Human	Tested in dog	Tested in Chimpanzees
0.9578	0.9085	0.9899
Trained in dog	Tested in human	Tested in Chimpanzees
0.7805	0.7103	0.8240
Trained in Chimpanzees	Tested in human	Tested in dog
0.9139	0.8758	0.8915

The experiments above show that the tests involving the human and the chimpanzee data are more accurate than those involving the dog's data. Some possible reasons could lead to the results. First, our model/project is data-driven, meaning the more complete and high-quality data we have, the more accurate the result will be. The data for the dog only consists of 820 strands of DNA sequences compared to the human, which has 4380 strands, and the chimpanzee's 1682 strands of DNA sequence. For the model trained by the human data, the test on chimpanzee data showed a higher accuracy than on the human test data, revealing the similarity between the chimpanzees' and human DNA. The distribution gap between human and chimpanzee DNA sequences is small.

Interestingly, the chimpanzees' accuracy is very high on the model trained by the human data. This could be because the human data set has many DNA strands, making its model better for generalization. In addition, we also found that the model trained by the dog data achieved higher accuracy for the chimpanzee's data than humans, while a similar result when trained by chimpanzees achieved higher accuracy for dog data than humans. We reckon that it is because dogs and chimpanzees are mammals, and some hidden patterns are similar. From the interesting findings, we could induce chimpanzees to have various potential connections with both humans and dogs.

CONCLUSIONS

In this research, we used Natural Language Processing (NLP) and neural networks to complete automatic classification for DNA sequences. We transformed DNA sequences to a human-like language and explored count vectorizer and TF-IDF as the vectorization methods. At last, we employed the classic neural network multi-layer perceptron as the classification model. We also developed a demo for users to try.

It is an exciting and thrilling moment for us to finish the project. We spend nearly one year on research, coding, testing, fine-tuning, and writing. During the project, we learned a lot. Firstly, we understand the pipeline of an AI project and adapt it to solve practical problems in our lives with high accuracy. To better understand, we are taught several analogies for AI knowledge, such as supervised

learning and multi-layer perceptron. Secondly, we deeply understand several concepts in our maths class, such as set, vector, matrix, and function. We then proceeded to apply them to our research, not just let them lie in our exam papers. Also, we try to understand some new and challenging knowledge in science. For example, we read the references for more background on this research's different DNA sequence types.

The first touch of AI is enjoyable, and we will explore more. For the next step, we will study different vectorization methods and AI models for classification. Also, exploring the generalization performance in the view of transfer learning could be interesting.

References

- [1] D. S. T. Nicholl, *An Introduction to Genetic Engineering*. Cambridge University Press, pp2-10, (2023).
- [2] D. S. T. Nicholl, *An Introduction to Genetic Engineering*. Cambridge University Press, p205-225, (2023).
- [3] M. Smith, "DNA Sequence Analysis in Clinical Medicine, Proceeding Cautiously," *Frontiers in Molecular Biosciences*, vol. 4, (2017).
- [4] J. M. Heather and B. Chain, "The sequence of sequencers: The history of sequencing DNA," *Genomics*, vol. 107, no. 1, pp. 1–8, (2016).
- [5] H. Lehrach, "DNA sequencing methods in human genetics and disease research," *F1000Prime Reports*, vol. 5, (2013).
- [6] L. C. Bailey Jr., S. Fischer, J. Schug, J. Crabtree, M. Gibson, and G. C. Overton, "GAIA: Framework Annotation of Genomic Sequence," *Genome Research*, vol. 8, no. 3, pp. 234–250, (1998).
- [7] J. Zhang, Z.-M. Shang, J.-H. Cao, B. Fan, and S.-H. Zhao, "Manual annotation of the pig whole genomic sequence using Otter-lace software," *Hereditas (Beijing)*, vol. 34, no. 10, pp. 1339–1347, (2012).
- [8] A. Wahab, H. Tayara, Z. Xuan, and K. T. Chong, "DNA sequences performs as natural language processing by exploiting deep learning algorithm for the identification of N4-methylcytosine," *Scientific Reports*, vol. 11, no. 1, (2021).
- [9] S. Vajjala, B. Majumder, H. Surana, and A. Gupta, *Practical Natural Language Processing: A Pragmatic Approach to Processing and Analyzing Language Data*. O'Reilly Media, (2020).
- [10] J. Pustejovsky and A. Stubbs, *Natural Language Annotation for Machine Learning*. "O'Reilly Media, Inc.," (2012).
- [11] Y. Zhang and Z. Teng, *Natural Language Processing: A Machine Learning Perspective*. Cambridge University Press, (2021).
- [12] S. S. Haykin, *Neural Networks and Learning Machines*. (2016).
- [13] M. Coding, *Machine Learning with Python: A Step by Step Guide for Absolute Beginners to Program Artificial Intelligence with Python*. Charlie Creative Lab, (2020).
- [14] S. Mendelson, *Advanced Lectures on Machine Learning: Machine Learning Summer School 2002, Canberra, Australia, February 11-22*, (2003).
- [15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, (2016).
- [16] S. Pattanayak, "Introduction to Deep-Learning Concepts and TensorFlow," in *Pro Deep Learning with TensorFlow*, Berkeley, CA: Apress, pp. 89–152, (2017).
- [17] T. Hope, Y. S. Resheff, and I. Lieder, *Learning TensorFlow: A Guide to Building Deep Learning*

Systems. “O’Reilly Media, Inc.,” (2017).

[18] Q. Yang, Y. Zhang, W. Dai, and S. J. Pan, *Transfer Learning*. Cambridge University Press, (2020).

[19] R. K. Sevakula and N. K. Verma, *Improving Classifier Generalization: Real-Time Machine Learning based Applications*. Springer Nature, (2022).

Authors

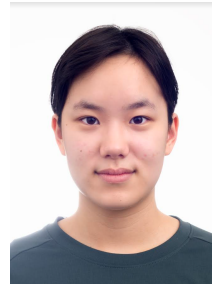
Josephine (Hsin) Liu

Josephine Liu is currently a year 11 student at Rangitoto College in Auckland, New Zealand. Outside of school, her passions include investigating AI and Taekwondo. She is interested in music, writing, law, history, and visual art.



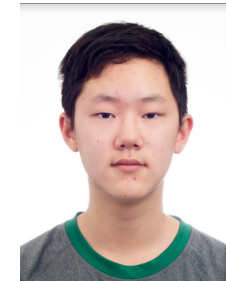
Phoebe (Yun) Liu

Phoebe Liu is currently a year 11 student at Rangitoto College in Auckland, New Zealand. Some hobbies and interests include: playing musical instruments such as drums. She is vastly interested in different fields, such as art, computer science, biology, and Taekwondo.



Joseph (Yu) Liu

Joseph Liu is a student at Rangitoto College in Year 10 in Auckland, New Zealand. His hobbies and interests include music, spatial and product design, Taekwondo, and computer science.



3D CONVOLUTION FOR PROACTIVE DÉFENSE AGAINST LOCALIZED ADVERSARY ATTACKS

Henok Ghebrechristos¹ and Gita Alaghband¹

¹Department of Computer Engineering, University of Colorado-Denver,
Denver, Colorado

ABSTRACT

This paper addresses the vulnerability of deep learning models, particularly convolutional neural networks (CNN)s, to adversarial attacks and presents a proactive training technique designed to counter them. We introduce a novel volumization algorithm, which transforms 2D images into 3D volumetric representations. When combined with 3D convolution and deep curriculum learning optimization (CLO), it significantly improves the immunity of models against localized universal attacks by up to 40%. We evaluate our proposed approach using contemporary CNN architectures and the modified Canadian Institute for Advanced Research (CIFAR-10 and CIFAR-100) and ImageNet Large Scale Visual Recognition Challenge (ILSVRC12) datasets, showcasing accuracy improvements over previous techniques. The results indicate that the combination of the volumetric input and curriculum learning holds significant promise for mitigating adversarial attacks without necessitating adversary training.

KEYWORDS

Convolutional Neural Network, Adversary Attack, Deep Learning, Volumization, Adversary Défense, Curriculum Learning

1. INTRODUCTION

The security of any machine learning model is assessed in terms of the goals and capabilities associated with adversary attacks. Algorithmically crafted perturbations, even if minuscule, can be exploited as directives to manipulate classification outcomes[1]. Attacks can be classified as black-box or white-box [] depending on the attacker's access to and knowledge of the model's information, which includes its architecture, parameters, training data, weights, and more. In a white-box attack, the attacker has complete access to the network's information, while a black-box attack is characterized by the absence of knowledge regarding the model's internal configuration. Occasionally, a gray-box attack can be generated by employing a generative model, enabling the creation of adversarial examples without access to the victim model. Localized adversarial attacks [2] exploit spatial invariance of CNN-based image classifiers by introducing minimal perturbations to deceive the model into producing incorrect classifications. These attacks are usually constrained to a small contiguous portion of the image and are image-agnostic (or universal) gray-box attacks. In this paper, we introduce a new training methodology (Figure 1) designed to fortify CNNs against localized attacks. Our primary approach incorporates deep curriculum optimization[3] and a volumization algorithm. We employ an information-theoretic representation of an image along with optimization procedure that merges batch-based curriculum learning (CL), patch aggregate

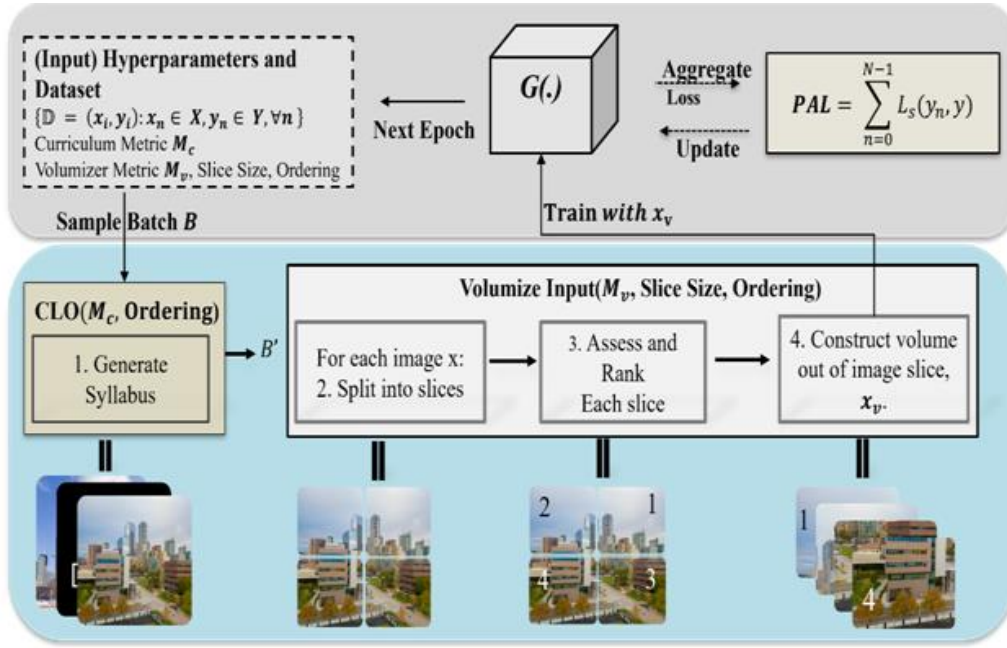


Figure 1. Overview of the proposed training process. The Curriculum Learning Optimization (CLO) component is used to generate a syllabus (input path) for the batch. It then volumizes each image, followed by feature extraction using 3D CNNs. The PAL loss function is applied to optimize parameters by calculating slice-wise errors.

loss (PAL) function, and 3D convolution to train and proactively defend against effective localized attacks; one-pixel[4] and adversary patch attacks (APA) [5].

1.1. Background on Adversary Attack

At its core, the purpose of adversary attack is to sabotage the generalization capability of a model by countering its learning objective. Given a CNN classifier $f(x; \theta)$, fully trained on a dataset D , its purpose is to map a source image x to a set of probabilities $f(x)$. An adversarial attack seeks to perturb this source image, producing an altered image x' such that the difference between x and x' is minimal to human perception. However, the classifier f , when processing x' , produces an incorrect output that significantly deviates from the true label. This is achieved by exploiting the high-dimensional decision boundaries of the model, forcing it to misclassify x' while maintaining a semblance of the original image structure in x .

1.1.1. Attack Objective

Adversarial image attacks involve adding a perturbation $r \in \mathbb{R}^m$ to x , causing the maximized class probabilities to differ between the original and perturbed images. i.e., $\text{argmax}_i(f_i(x+r)) \neq \text{argmax}_i(f_i(x))$. These types of attacks can be categorized as *targeted* or *untargeted*.

In a targeted attack, the adversarial image $x' = x + r$ is generated to induce the classifier to assign x' to a specific target class $c_t \in Y$, where $c_t \neq \text{argmax}_i(f_i(x))$. The perturbation r is selected such that $\text{argmax}_i(f_i(x')) = c_t$. Conversely, in an untargeted attack, the adversarial image $x' = x + r$ is crafted to cause the classifier to assign x' to any incorrect class without a particular target. In this case, the perturbation r is chosen to satisfy $\text{argmax}_i(f_i(x')) \neq \text{argmax}_i(f_i(x))$ without imposing additional constraints on the target class. *Our research focus is untargeted attacks.*

1.1.2. Défense Objective

The defense objective is to train a model that is robust to adversarial image attacks without sacrificing the accuracy of the classifier on the original dataset. Formally, the objective is to find g that minimizes the following loss:

$$\min_g \frac{1}{|D|} \sum_{(x,y) \in D} \max_{r \in R} L(g(x+r), y)$$

where R is the set of possible adversary perturbations added to a local region of the input, and L is a loss function used to train the model g . The objective is to minimize the maximum loss over all possible adversarial examples $x' = x + r$ generated by any allowable perturbation in R . R is constrained to be a set of *localized attacks*. Localized attacks are characterized by the property that the L2 norm of the perturbation vector r , denoted by $\|r\|$, is much smaller than the L2 norm of the original input image x , denoted by $\|x\|$. Specifically, this condition can be expressed as $\|r\| \ll \|x\|$. These attacks modify only a small subset of pixels that are confined to a localized region of the image.

1.1.3. Localized Universal Attacks Against Image Classifiers

Localized universal attacks are a subset of adversarial attacks that specifically target image classifiers. They exploit spatial invariances of CNNs to introduce perturbations that lead to misclassifications. These perturbations are usually confined to small, contiguous portions of the input image and can cause the model to produce incorrect output classifications, even when the introduced changes are almost imperceptible to the human eye. Two predominant types of such attacks are the N-Pixel Attack and the Adversary Patch Attack.

1.1.3.1. N-Pixel Attack

Szegedy et al. introduced adversarial attacks through minor perturbations of pixels [6] to induce CNN image misclassification. These perturbations, often undetectable to the human eye (see Figure), expose the inherent vulnerabilities in the robustness of Convolutional Neural Networks (CNNs).

Diving deeper into this, the N-Pixel Attack, which can be viewed as an extension or generalization of the ideas presented by Szegedy et al., specifically perturbs 'N' distinct pixels in an image to induce misclassification. The challenge and intrigue of this method arise from its seemingly benign nature; altering a minimal number of pixels in a high-resolution image intuitively appears harmless. Yet, such alterations can drastically alter CNN's prediction, underscoring the intricate and potentially fragile decision boundaries upon which these networks operate.

While the attack has profound implications for the integrity and reliability of image classifiers, it also catalyzes a renewed interest in understanding the foundational workings of CNNs. This understanding is crucial, especially in applications where trust in model predictions is paramount, such as in medical imaging or autonomous vehicles.

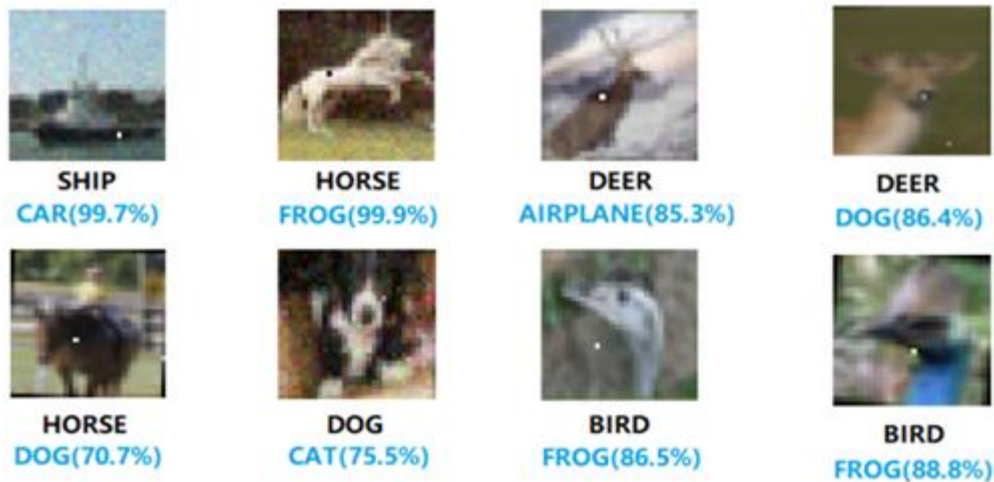


Figure 2. One Pixel Attack.

The N-Pixel Attack serves as a pivotal reminder that even state-of-the-art models, trained on extensive datasets and exhibiting high accuracy, can be vulnerable to carefully constructed, minimal adversarial perturbations. As researchers and practitioners continue to deploy CNNs in varied applications, it is imperative to develop strategies that not only enhance performance but also fortify against adversarial threats.

1.1.3.2. Adversary Patch Attack

Introduced by Brown et al., the Adversary Patch Attack unveils a unique and potent vulnerability in deep neural network (DNN) based classifiers [7]. Unlike other adversarial attacks that perturb an image globally, this approach is localized and focuses on modifying a confined region of the image with an adversarial patch (**Error! Reference source not found.**), which can be recognized as a seemingly harmless object or pattern added to the image. Remarkably, this addition can dramatically alter the classifier's output, demonstrating a classifier's inability to discern genuine content from deceptive information.

Central to the findings of Brown et al. was the realization that these adversarial patches were resistant to various changes, especially affine transformations such as translation, scaling, and rotation. Their methodology optimized the patches such that they were robust to these transformations. This means that the relative position, size, or orientation of the adversarial patch doesn't need to be precise for it to deceive the classifier effectively. This robustness elevates the potential real-world implications of this attack as the adversarial patch remains effective under different viewing conditions.

The adversarial patch, crafted using a white-box approach, can be applied in a "universal" manner. This universality signifies that a single patch can be effective across different images and is not tied to a specific target image. The conspicuous nature of these patches (often visually distinct) contrasts with the often-imperceptible alterations in traditional adversarial attacks, making it an intriguing anomaly in adversarial research.

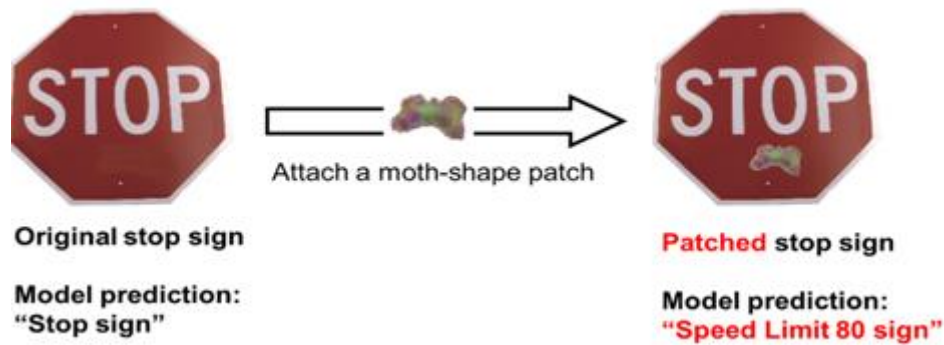


Figure 3. Adversary Patch Attack (APA).

Furthermore, Brown et al.'s findings underscore the importance of understanding not just the global, but also the localized processing dynamics of CNNs. The attack challenges the prevalent notion of CNNs' spatial hierarchies, wherein larger spatial structures (like objects) are assumed to have a dominant influence over classification compared to smaller structures or patterns. The Adversary Patch Attack highlights that this might not always be the case, as a localized, conspicuous pattern can effectively override the neural network's perception of larger structures. In real-world scenarios, this type of attack could be employed in deceptive practices, such as placing adversarial stickers or objects in strategic locations to deceive AI systems in surveillance, autonomous driving, or even augmented reality applications. As such, the research by Brown et al. underscores the importance of defensive mechanisms that consider both global and local image features.

Both Brown et al., and later Gittings et al. [7] backpropagate through the target model to generate 'stickers' that can be placed anywhere within the image to create a successful attack. This optimization process can take several minutes for one single patch. Karmon et al. showed in LaVAN that the patches can be much smaller if robustness to affine transformation is not required [2] but require pixel-perfect positioning of the patch which is impractical for real APAs.

2. RELATED WORK

Goodfellow et al. proposed a method to enhance the robustness of neural networks against such attacks. This approach, known as adversarial training [1], involves using adversarial examples as inputs during training to teach the network how to accurately classify them, thus improving its ability to defend against attacks. Adversary training methods require access to the target model to backpropagate gradients to update pixels, inducing high frequency noise that is fragile to resampling. In 2016 Papernot et al. proposed a defensive mechanism called defensive distillation [8], in which a smaller neural network learns and predicts the class probabilities generated by the original neural network's output. Despite growing number of defense approaches, several attacks remain effective – particularly attacks generated using generative architectures [9], [10].

2.1. N-Pixel Attack Defences

Brown et al. demonstrated that adversarial patches could be used to fool classifiers; they restricted the perturbation to a small region of the image and explicitly optimized for robustness to affine transformations [5]. Su et al. in 2017, which generates fooling images (adversarial examples) by perturbing only one pixel or few pixels, has proven to be difficult to defend [4]. To date, the most successful defense against this attack is a method presented by Chen et al. [11]. The authors propose Patch Selection Denoiser (PSD) that removes few of the potentially attacked

pixels in the whole image. At the cost of image degradation, the authors achieve a successful defense rate of 98.6% against one-pixel attacks. Similarly,

Liu et al. [12] proposed a three-step image reconstruction algorithm to remove attacked pixels. The authors report protection rate (defense success rate) of up to 92% under for N-pixel attack for N chosen from the range (1, 15). Shah et al. [13] proposed the usage of an Adversarial Detection Network (ANNNet) for detection of N-pixel attacks where N is 1, 3 or 5 and report up to 97.7 adversarial detection accuracy on MNIST Fashion dataset. Husnoo and Anwar [14] proposed an image recovery algorithm based on Accelerated Proximal Gradient (APG) [15] approach to detect and recovered the attacked pixels.

2.2. Adversary Patch Defences

Defenses for patch attacks are typically viewed as a detection problem [16], [17]. Once the patch's location is detected, the suspected region would be either masked or in-painted to mitigate the adversarial influence on the image. Hayes [18] first proposed DW (Digital Watermarking), a defense against adversarial patches for nonblind and blind image inpainting, inspired by the procedure of digital watermarking removal. A saliency map of the image was constructed to help remove small holes and mask the adversarial image, blocking adversarial perturbations. This was an empirical defense with no guarantee against adaptive adversaries.

Naseer et al. [19] proposed LGS (Local Gradient Smoothing) to suppress highly activated and perturbed regions in the image without affecting salient objects. Specifically, the irregular gradients were regularized in the image before being passed to a deep neural network (DNN) model for inference. LGS could achieve robustness with a minimal drop in clean accuracy because it was based on local region processing in contrast to the global processing on the whole image as done by its counterparts.

Chou et al. [20] proposed SentiNet for localized universal attacks to use the particular behavior of adversarial misclassification to detect an attack, which was the first architecture that did not require prior knowledge of trained models or adversarial patches. Salient regions were used to help observe the model's behavior. SentiNet was demonstrated to be empirically robust and effective even in real-world scenarios. However, it evaluated adversarial regions by subtracting the suspicious region, which might at times cause false adversarial region proposals. Moreover, the suspicious adversarial region was placed at random locations in the preserving image, which possibly occluded the main objects in the scene resulting in incorrect predictions.

Chen et al. [21] proposed Jujutsu to detect and mitigate robust and universal adversarial patch attacks by leveraging the attacks' localized nature via image inpainting. A modified saliency map [22] was used to detect the presence of highly active perturbed regions, which helped to place suspicious extracted regions in the least salient regions of the preserved image and avoid occlusion with main objects in the image. Jujutsu showed a better performance than other empirical defenses in terms of both robust accuracy and low false-positive rate (FPR), across datasets, patches of various shapes, and attacks that targeted different classes.

Instead of relying on patch or pixel detection and removal techniques, our method uses deep curriculum learning optimization (Deep-CLO) [23] and 3D convolutional neural networks [25] to proactively defend against these attacks. An overview of our proposed training methodology is shown in Figure 1.

3. METHOD

Our aim is to develop a defended classifier, \hat{G} , that inherently defends against localized attacks without relying on adversarial training or prior knowledge of the attacks. To realize this objective, we propose a *proactive defense approach* characterized by following key steps:

- *Volumization*: For each image in the batch, we convert it to a 3D volume to capture the spatial information of the input.
- *3D Convolution*: We modify the contemporary CNN model architectures to do 3D convolution, which enables the model to extract features from the 3D input volumes. Details of the modifications are below.
- *Deep curriculum learning optimization*[3]: For each batch taken from the training dataset D , we generate a syllabus that determines the input order of the samples in the batch.

The combination of 3D convolution, Deep-CLO, and the volumization algorithm empowers \hat{G} to maintain high performance on both clean and adversarial inputs. These techniques ensure that the model remains resilient to perturbations and that it maintains accurate classification and verification of both the original images x and adversarial images $x' = x + r$. Refer to analysis section for detailed justification.

When used as a preprocessing step, the algorithm allows \hat{G} to extract spatial relationships from the volumized data, resulting in enhanced robustness against localized attacks. The employment of Deep Curriculum Optimization as the training procedure further bolsters the \hat{G} 's resilience to adversarial attacks. The resulting classifier not only defends against the targeted attacks but also retains high performance on non-adversarial images, achieving classification accuracy comparable to state-of-the-art models. Consequently, our approach demonstrates a unique balance between robustness and accuracy without relying on adversarial training, making it a significant contribution to the field of image classification under adversarial conditions.

3.1. Volumization Algorithm

The volumization algorithm, is a pixel-preserving and reversible transformation operator denoted as a function:

$$V : (x, H, W, C) \rightarrow (x'_i, p_h, p_w, C, Z). \quad 2$$

Given an input image $x \in D$ of shape (H, W, C) , the algorithm uses a configurable hyper parameter, patch size $S(p_h, p_w)$ - where p_h and p_w are the height and width of each slice - to split x into Z non-overlapping slices x'_i of size P satisfying the pixel conservation condition:

$$x = \bigcup_{i=0}^{Z-1} x'_i$$

This states that the original image x is equivalent to the union of all its extracted slices (see the image at the bottom of Figure 1). Here, \cup denotes the union operation. The algorithm extracts a list of slices and proceeds to rank each slice according to a prespecified metric m - a configurable hyper parameter. The chosen metric can be of type distance or standalone (Table 1). If m is a distance-based metric, a reference slice P_{ref} is selected, which can be user-defined or automatically determined by choosing the most salient slice. On the other hand, m is considered standalone if it measures some characteristics of a given slice. All slices are then ordered based on

their individual metric scores or their distances from the reference slice. The ordering ord is userdefine configurable hyperparameter that can be either descending or ascending. Finally, the ranked slices are stacked along the depth axis to create a 3D volume $x_v = V(x)$ of shape (p_h, p_w, C, N) , where Z (depth of the volume) is the total number of slices. Refer to Ghebrechristos et. al. [4] for detail on the ranking and ordering process, which is identical for both the volumizer and CL when generating a syllabus for a batch.

By breaking the image into smaller patches, we localize the region of analysis, making the training process more sensitive to adversarial attacks that affect only a small portion of the image. This localization often aligns with the attack region of localized adversarial attacks, enhancing the models' ability to detect and respond to them.

3.2. Model Architecture

Given a conventional CNN classifier architecture f designed to learn from 2D images, we perform the following modifications to construct \hat{G} - a 3D counterpart f .

3.2.1. Input Layer

f 's input layer, denoted as I_{2D} , is designed to accept a 2D input data x with dimensions $H \times W \times C$ such that $I_{2D}: x \in \mathbb{R}^{H \times W \times C}$. In order to extract features from the volumized images, the input layer of model \hat{G} is adjusted to be 3-dimensional, I_{3D} , where the input to this layer x_v is a 4D tensor with dimensions $H' \times W' \times C \times Z$, such that $I_{3D}: x_v \in \mathbb{R}^{H' \times W' \times C \times Z}$. In shorthand notation, this reversible transformation can be represented as:

$$f(I_{2D}: x \in \mathbb{R}^{H \times W \times C}) \leftrightarrow \hat{G}(I_{3D}: x \in \mathbb{R}^{H' \times W' \times C \times Z}), \quad 4$$

where H' and W' are prespecified width and height of the individual patches within the volume and Z signifies the total number of patches. This enables the classifier to learn features from the volumized data x_v .

3.2.2. Convolution Layer

For CNN, convolution represents the interaction between an input (image or feature map) and a kernel (filter). The kernel is a small matrix that slides over the input data, performing an element-wise multiplication and summing the results to generate a new feature map. In a 2D convolution, the input data and the kernel are both two-dimensional.

$$(K * x)(i, j) = \sum_m \sum_n K(m, n) x(i - m, j - n) \quad 5$$

Here, K represents the 2D kernel, x represents the 2D input and (i, j) are the coordinates in the output feature map. The summation is performed over all spatial dimensions (m, n) of the 2D kernel.

For a given classifier, 3D counterpart of the above operation is:

$$(K_{3D} * x_v)(i, j, k) = \sum_m \sum_n \sum_p K_{3D}(m, n, p) x_v(i - m, j - n, k - p) \quad 6$$

where K_{3D} represents the 3D kernel, x_v is the 3D input data and (i, j, k) are the coordinates in the output feature map. The summation is performed over all spatial dimensions (m, n, p) of the 3D

kernel. This modification enables the classifier to learn features from the volumized data by processing spatial information across height, width, and depth dimensions simultaneously.

Note that the optimal kernel size depends on the size of the individual slices within the volume and the desired level of spatial information capture. For example, if f consists of 1×1 , 3×3 , and 5×5 2D convolution layers, we adjust these layers to be $1 \times 1 \times Z$, $3 \times 3 \times Z$, and $5 \times 5 \times Z$ 3D convolution layers, respectively. To ensure compatibility, we enforce the constraint that the kernel size is much smaller than the size of the individual slices and that the kernel operates on each slice in the volume. That is, $H' \ll H$, and $W' \ll W$. The stride and padding values are also adjusted accordingly.

3.2.3. Pooling Layer

All pooling layers of f are modified to handle the 3D volume representation of the input data. In f , the 2D pooling layers denoted as P_{2D} :

$$P_{2D}: \mathbb{R}^{H_{in} \times W_{in} \times C_{in}} \rightarrow \mathbb{R}^{H_{out} \times W_{out} \times C_{out}} \quad 7$$

where H_{in} , W_{in} and C_{in} represent the height, width, and number of channels of the input feature maps, while H_{out} , W_{out} and C_{out} denote the height, width, and number of channels of the output feature maps, respectively.

To effectively process volumized inputs, we replace the 2D pooling layers with 3D counterparts:

$$P_{3D}: \mathbb{R}^{H_{in} \times W_{in} \times C_{in}} \rightarrow \mathbb{R}^{p_{h_{out}} \times p_{w_{out}} \times C_{out} \times N_{out}} \quad 8$$

where $p_{h_{in}}$, $p_{w_{in}}$, C_{in} and N_{in} represent the height, width, number of channels, and number of patches of the input volume, while $p_{h_{out}}$, $p_{w_{out}}$, C_{out} and N_{out} denote the height, width, number of channels, and number of patches of the output 3D volume, respectively.

3.2.4. Normalization and Activation Layers

These layers usually play an essential role in maintaining a stable and efficient training process and introducing non-linearity to the model. For normalization layers, we transition from 2D normalization methods of f , such as Batch Normalization (BN) and Instance Normalization (IN), to their 3D counterparts in g . Given 3D input tensor, $x_v \in \mathbb{R}^{H \times W \times C \times N}$, the 3D normalization layer computes the mean μ and standard deviation σ across the specified dimensions (usually height, width, and depth) and normalizes x_v as follows:

$$x_{v-normalized} = \frac{x_v - \mu}{\sigma}, \quad 9$$

where μ and σ are broadcasted to match the dimensions of x_v .

For activation layers, the transition from 2D to 3D input data is more straightforward. Common activation functions, such as *ReLU* and *sigmoid*, can be directly applied to the 3D input data, with minor tweaking, as these functions perform element-wise operations on the input tensor. The output tensor $y_v \in \mathbb{R}^{H \times W \times C \times N}$, by applying the activation function Af elementwise to the input tensor x_v :

$$y_v = Af(x_v[i, j, c, n]), \forall i \in [0, H), j \in [0, W), c \in [0, C), n \in [0, N). \quad 10$$

By ensuring that normalization and activation layers are compatible with the 3D input data, we maintain the stability and efficiency of the training process, while enabling the model to effectively learn non-linear features from the volumized input data.

3.2.5. Fully Connected and Output Layers

To perform patch-wise error calculation and enhance model's robustness, we modify f 's fully connected layer function $F: \mathbb{R}^{(M,N)} \rightarrow \mathbb{R}^{(L,Z)}$ where M represents the number of input features, L denotes the number of output features, and Z is the total number of patches. The fully connected layer function F' now maps each patch's input features to its respective output features, allowing for patch-wise error calculations during backpropagation.

The output layer function $O: \mathbb{R}^{(L,N)} \rightarrow \mathbb{R}^{(k,N)}$, where k is the number of classes. To ensure compatibility with the 3D input data, the output of the preceding layers must be reshaped or flattened before connecting to the fully connected layers. This modification allows g to map each patch's output features to its respective class probabilities, further enabling *patch-wise error* calculations during the training.

3.2.5.1. Patch Aggregate Loss (PAL) Function

PAL is designed to enable backpropagation on individual patches. During training, the loss for each patch is calculated separately. The patch-wise losses are then aggregated to obtain the overall loss for the image. Given the modified output $O: \mathbb{R}^{(L,N)} \rightarrow \mathbb{R}^{(k,N)}$, we define PAL as follows:

Patch-wise Error Calculation: Computes the loss for each patch separately using a suitable loss function $L_p: \mathbb{R}^{(N',k)} \rightarrow \mathbb{R}^{(N')}$. For a given patch $n \in \{0, 1, \dots, Z-1\}$, the patch-wise loss is calculated as $L_p(y_n, y)$, where y_n represents the predicted class probabilities for patch n , and y denotes the true class labels of the original input.

Loss Aggregation: Aggregate the patch-wise losses to obtain the overall loss for the image. This function, termed Patch Aggregate Loss (PAL) function, computes the overall loss of an image by summing up the individual patch-wise losses:

$$PAL = \sum_{n=0}^{N-1} L_p(y_n, y) \quad 11$$

Using the sum of slice-wise losses directly emphasizes the importance of minimizing the error for each individual slice, driving the model to learn more robust features from each slice. This increased emphasis on localized features results in a more robust model that is better equipped to counteract attacks.

3.3. Training Methodology

We incorporate deep curriculum learning optimization (CLO) as described in Ghebrechristos et. al. [4] at a batch level to enhance the training process. Given a batch $B \subseteq D$, we define a syllabus S as a function $S: B \rightarrow B'$, where B' is a reordered version of the original batch B . S describes an input order of the samples in B' such that the learning process progresses from simpler to more complex samples as quantified by a concrete metric m taken from Table 1.

Table 1. List of measures used in this study. Given samples $x, x_1, x_2 \in B$ where b_x is normalized histogram of pixel intensities and i is an index of a pixel value in the image's vector. σ is standard deviation and μ is mean or average pixel intensities.

Metric	Implementation	Category
Entropy	$H(x) = \sum_{i \in \mathcal{X}, x \in D} b_x(i) \log \frac{N}{b_x(i)}$	standalone
Joint Entropy (JE)	$JE(x_1, x_2) = \sum_i b_x(i) \log b_x(i)$	distance
Mutual Information (MI or I)	$MI(x_1, x_2) = H(x_1) + H(x_2) - JE(x_1, x_2)$	distance
KL-Divergence (KL)	$D_{KL}(x_1 x_2) = \sum_i x_{1i} \log \frac{x_{1i}}{x_{2i}}$	distance
Structural Similarity index (SSIM)	$SSIM(x_1, x_2) = \frac{(2\mu_{x_1}\mu_{x_2} + C_1)(2\sigma_{x_1x_2} + C_2)}{(\mu_{x_1}^2 + \mu_{x_2}^2 + C_1)(\sigma_{x_1}^2 + \sigma_{x_2}^2 + C_2)}$	distance
Max Norm (MN)	$x_\infty = \max(x_1, x_2)$	distance
Peak signal to noise ratio (PSNR)	$PSNR = 20 \log_{10} \left(\frac{MAX}{\sqrt{MSE}} \right)$	distance
Mewan Squared Error	$MSE(x_1, x_2) = \frac{1}{N^2} \sum_i \sum_j (x_{1ij} - x_{2ij})^2$	na

Ordering of samples for the batch is done in the sameway the volumizer algorithm orders slices to create a volume. Given $B = \{x_1, x_2, \dots, x_n\}$, a batch of n samples (or a set of patches belonging to x), let $SM(x_i)$ be the standalone metric value of x_i and $DM(x_i, P_{ref})$

be the distance metric value of the sample or slice x_i with respect to a reference image (or patch) P_{ref} , respectively. We define order relations $R_{SM} \subseteq B$ and $R_{DM} \subseteq B$, such that:

$$(x_i, x_j) = \begin{cases} R_{SM} & \text{if } SM(x_i) \leq SM(x_j) \\ R_{DM} & \text{if } DM(x_i, P_{ref}) \leq DM(x_j, P_{ref}) \end{cases}^{12}$$

Thus, the syllabus (or volumizer) algorithm transforms B (set of slices) into an ordered one B' :

$$S_{SM(B)} = \{x'_1, \dots, x'_n, \text{ where } (x'_i, x'_j) \in R_{SM}\}^{13}$$

$$S_{DM(B)} = \{x'_1, \dots, x'_n, \text{ where } (x'_i, x'_j) \in R_{DM}\}^{14}$$

The learning process progresses from simpler to more complex samples based on a specific metric, which enhances model performance and speeds up convergence. This method also strengthens models against localized attacks by ordering patches based on their features. Adversarial perturbations in a single patch have less impact on the model's image understanding due to this arrangement.

4. EXPERIMENTS& RESULTS

Our approach is evaluated on EfficientNet-B0, InceptionV3, ResNet50, and VGG19 architectures modified for 3D input compatibility. These modifications result in a *significant but* tolerable increase in the number of parameters: approximately 20M for VGG, 49M for ResNet, 44M for Inception, and 10M for EfficientNet (**Error! Reference source not found.**2). The models, implemented via open-source TensorFlow [26] library, are tested on CIFAR10, CIFAR100, and ILSVRC12 datasets under different attack settings. CIFAR10 facilitates comprehensive study, while CIFAR100 and ILSVRC12 test the approach’s generalizability.

We measure the *classification accuracy* – of the models on both clean images (acc_{clean}) and adversarial images acc_{attack} . We calculate *robustness score* (δ) as the difference between model’s classification accuracy on clean images and its classification accuracy on adversarial images, $\delta = (acc_{clean} - acc_{attack})$. A smaller δ demonstrates greater proactive robustness against adversarial attacks. We also measure *defense success rate* β – the percentage of successfully defended adversarial attacks. A higher defense success rate indicates a better proactive defense against adversarial attacks.

We contrast performance with a baseline defense approach using similar datasets. We used adversary stickers (**Error! Reference source not found.**) synthesized by A-ADS method of Brown et al. [5] and extend the adversary benchmark library, FoolBox’s [26] to implement N-pixel attacks.



Figure 4. Patch Attack Stickers were Used in this Study. Left to Right (Toaster, Goldfish, Schoolbus, Lipstick and Pineapple) Generated Using the A-ADS Patch Synthesis Method Highlighted in Brown et al.

Table 2. Number of Parameters for different models; original (f) and with modifications (f_{3D})

Model	Original (f)	Modified (f_{3D} or G)
VGG16	143,357,544	183,021,512
ResNet50	25,636,712	74,203,245
InceptionV3	23,851,784	68,104,050
EfficientNetB0	5,330,564	15,900,000

4.1. Défense Success Rate

Error! Reference source not found. shows the defense success rate of model trained with our method surpasses 80% after 300 epochs, indicating effective defense against 1-pixel attacks at each validation run. After 50 epochs, the classifier rapidly learns to resist the attacked pixel, increasing G ’s success rates while those of f stagnate below 72%. This confirms that the approach delivers models that match the undefended model’s performance on clean datasets while resisting localized attacks.

4.2. Défense Effectiveness

We evaluate the performance of our defense in reducing the effectiveness of N-Pixel and patch attacks. We use 1, 2, up to 16-pixel coverage for N-pixel. We use adversarial patches – Toaster, School-Bus, Lipstick and Pineapple - synthesized by attack methods A-ADS, covering up to 25% of the entire image. Our approach is compared with existing defense strategies in terms of clean accuracy and defense success rate β . We mount such attacks against our defense (I, KL, H, MN, and PSNR syllabi), and an undefended model as a control.

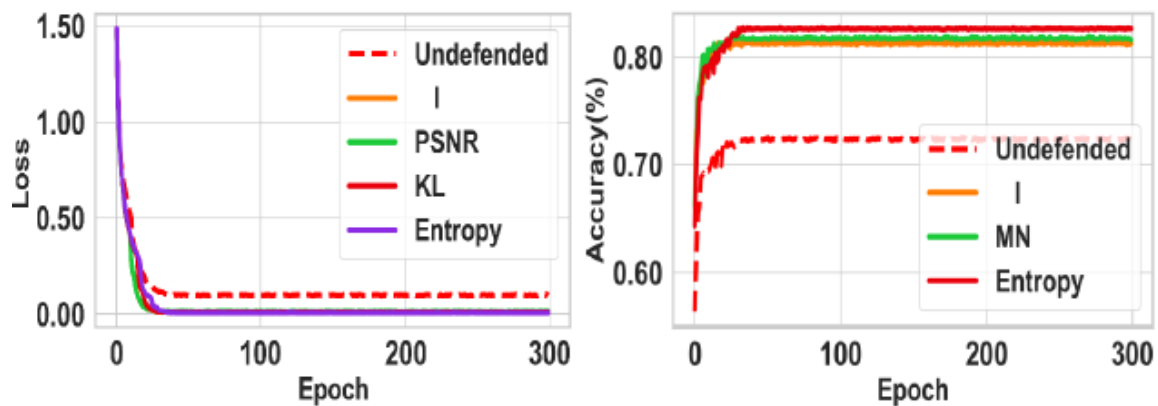


Figure 5. CIFAR10 Training losses and Success Rates of Defense on EfficientNet. (Left) Shows the Losses of Defended Model G Using Different Syllabus Configurations and Undefended Model F. (Right) Shows the Training Success Rates of Both Models Under 1-Pixel Attack.

The patch size(p_h, p_w) of the volumization algorithm for all syllabi is set to 16×16 pixels. We take reported results of all baseline defenses for comparison.

Table 3 presents a comparison of generalization performance of EfficientNet on CIFAR10, CIFAR100 and ImageNet datasets, with and without our defense mechanisms, for both clean and attacked test sets. Though the undefended model exhibits good performance clean data, its performance significantly deteriorates under adversarial attacks. In contrast, models defended using our approach show resilient performance under adversarial scenarios, with minimal trade-offs in clean data accuracy.

Notably, the model defended using measure MI outperforms other models under N-Pixel attack across all datasets. For instance, the model defended using mutual information (I) achieves attack accuracies of 91.3 (N-Pixel) and 61.6 (APA) on CIFAR10, remarkably higher than the undefended model's 43.2 and 44.3, respectively. Similarly, the KL-defended model yields considerably better attack accuracy on CIFAR100 (73 and 43.2 for N-Pixel and APA respectively) compared to the undefended version (32.5 and 22.1).

Error! Reference source not found. illustrate the overall robustness (δ) of Efficient Net and Inception against N-pixel and patch attacks, respectively. The plots highlight the dependence of model robustness on attack size for both defended and undefended models, with the undefended model being 40% less accurate at worst. Our defense is effective for both architectures at all attack magnitudes. However, like the undefended model, the performance of our method

degrades as the size of the attack increases, indicating a shared vulnerability to larger-scale attacks.

Table 3. Generalization accuracy of EfficientNet on CIFAR10, CIFAR100, and ILSVRC12 datasets with and without our defence mechanisms. The performance is compared under three scenarios: Clean Test Sets, Test Sets Under One-Pixel Attack (N-Pixel Where N=1), and Test Sets Under Patch Attack with a Toaster Sticker (APA).

Défense	Acc_{clean}			$Acc_{attack}(N - pixel)$			$Acc_{attack}(APA)$		
	CIFAR10	CIFAR100	ISLVR12	CIFAR10	CIFAR100	ISLVR12	CIFAR10	CIFAR100	ISLVR12
Entropy(H)	94	90.5	76.8	85	74	56.8	79	55.2	58.2
MI	96.3	98	79	91.3	70	69	61.6	53	61.2
KL	93	93.2	75	63	73	62.1	65	43.2	36.8
PSNR	89	90.3	76	83.6	51	56	52	53	48.3
Norm(MN)	92	86	75.4	74	51	49.5	42	38	32
Undefended	99	96.8	78.3	43.2	32.5	12.4	44.3	22.1	10.8

As presented in Error! Reference source not found.4, our proposed defense demonstrates a significant performance against N-Pixel attack compared to the undefended models. The undefended models exhibit sharp decline when under attack (Acc_{attack} column). Our proposed method (I) not only achieves high Acc_{clean} of 96.3% for EfficientNet and 99% for VGG and Inception but also shows a minor degradation when under attack; by 5%, and 0.4% for Efficient and VGG respectively.

Table 4. Accuracy of models over the set of test images without attacks Acc_{clean} and with attack Acc_{attack} , reported for all CIFAR10 classes. Reported as Top-1 accuracy of 1 pixel attack for the undefended model, the model defended by our method (I) and other comparable approaches. All images in the dataset are attached for this report.

Défense	Acc_{clean}		Acc_{attack}	
	EfficientNet	VGG	EfficientNet	VGG
Undefended	98	98.9	44.3	35.9
I/Ours	96.3	99	91.3	98.6
PSD	-	99.53	-	97.8
Liu et al	-	89.8	-	91

Comparing mutual information (MI) with the PSD method, our approach has a slightly lower Acc_{clean} for VGG, with a difference of 0.53%, but delivers a better Acc_{attack} for the same model, with an improvement of 1.2%. When comparing I to Liu et al.'s method, our method demonstrates a substantial improvement in Acc_{clean} for VGG, with a difference of 9.2%, and a higher Acc_{attack} as well, with an improvement of 4.6%. PSD and Liu et al. do not provide results for Efficient Net.

Table 5 presents defense success rates (β) for various defense methods and ours against APA on three datasets. For CIFAR10, our methods achieved 95.6% and 96.12% success rates, while Jujutsu and LGS obtained only 86.5% and 93.2%, respectively. Similarly, for CIFAR100, our methods reached success rates of 95.43% and 94.3%, outperforming Jujutsu's 55.7% and LGS's 73.7%. In the ImageNet dataset, ours (MI) achieved the highest defense success rate of 89.1%, while PSNR obtained 83.2%, both surpassing Vax-a-Net's 86.8% and DW's 65.2% and 66.2% for VGG and Inception models, respectively. Not all methods have reported results for every dataset, limiting a comprehensive comparison of their effectiveness.

Table 5. Defense Success Rate (β) of Various Defense Methods Against APA Covering At Least 5% of the Image. Adversary Patches; Toaster, Lipstick, Pineapple, And School-Bus Were Used.

Défense	Défense Success Rate(β)		
	CIFAR10	CIFAR100	ILSVRC12
H/Ours (EfficientNet)	91.3	80.5	-
MI/Ours (ResNet)	95.6	95.43	89.1
PSNR/Ours (Inc)	96.12	94.3	83.2
Jujutsu (ResNet)	86.5	55.7	-
LGS	93.2	73.7	-
V-a-N(VGG)	-	91.6	86.8
DW(VGG)	-	-	65.2
DW(Inc)	-	-	66.2
ECViT-B	47.39	-	41.7

As presented in Error! Reference source not found.4, our proposed defense demonstrates a significant performance against N-Pixel attack compared to the undefended models. The undefended models exhibit sharp decline when under attack (**Accattack** column). Our proposed method (I) not only achieves high **Accclean** of 96.3% for EfficientNet and 99% for VGG and Inception but also shows a minor degradation when under attack; by 5%, and 0.4% for Efficient and VGG respectively.

Table 4. Accuracy of models over the set of test images without attacks **Accclean** and with attack **Accattack**, reported for all CIFAR10 classes. Reported as Top-1 accuracy of 1 pixel attack for the undefended model, the model defended by our method (I) and other comparable approaches. All images in the dataset are attached for this report.

Défense	Accclean		Accattack	
	EfficientNet	VGG	EfficientNet	VGG
Undefended	98	98.9	44.3	35.9
I/Ours	96.3	99	91.3	98.6
PSD	-	99.53	-	97.8
Liu et al	-	89.8	-	91

Comparing mutual information (MI) with the PSD method, our approach has a slightly lower *Accclean* for VGG, with a difference of 0.53%, but delivers a better *Accattack* for the same model, with an improvement of 1.2%. When comparing I to Liu et al.’s method, our method demonstrates a substantial improvement in *Accclean* for VGG, with a difference of 9.2%, and a higher *Accattack* as well, with an improvement of 4.6%. PSD and Liu et al. do not provide results for Efficient Net.

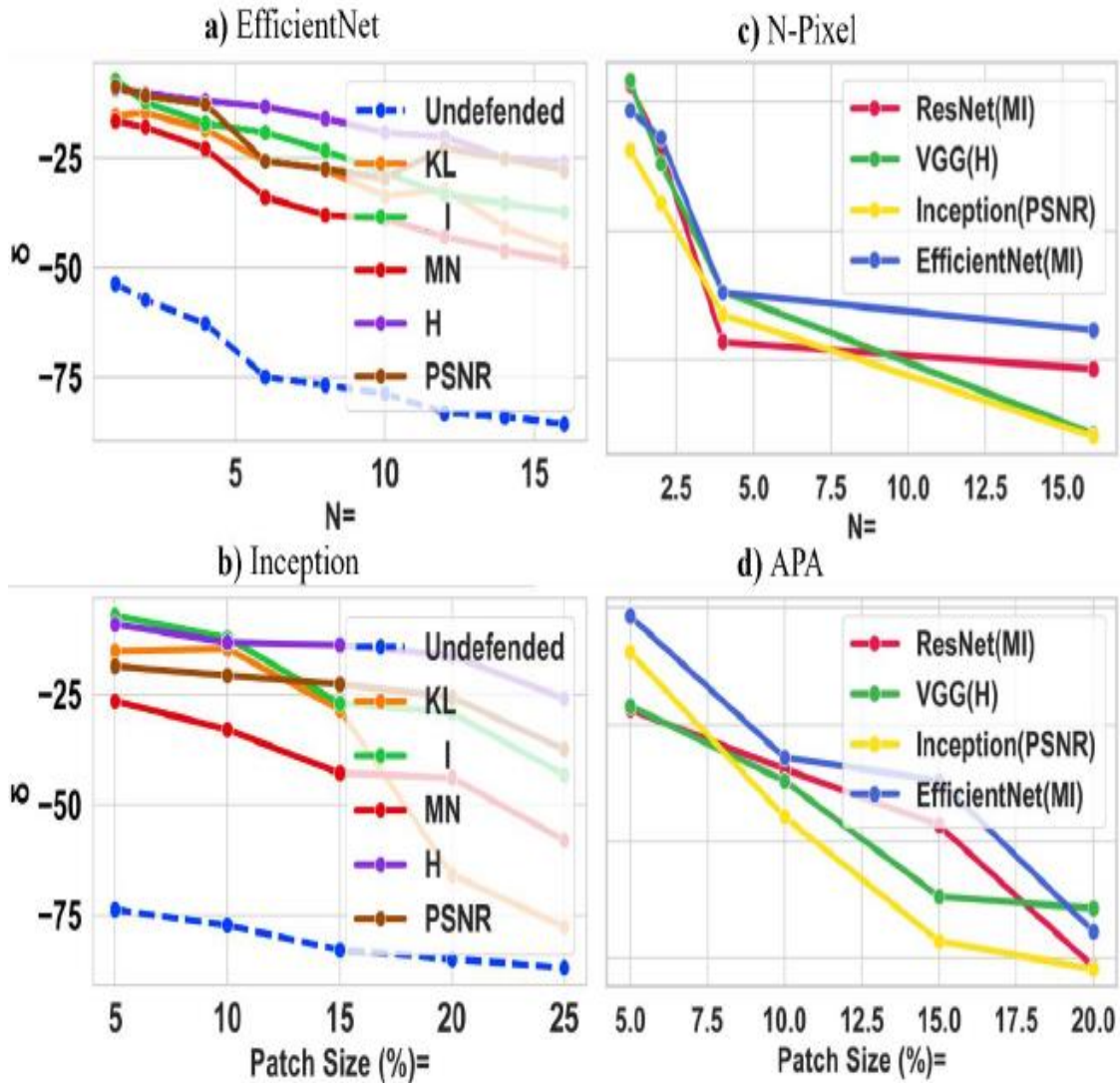


Figure 1. (a) EfficientNet Robustness as a Function of N - Number of Pixels Attacked, (b) Inception Robustness Against APA As a Function of Patch Size, (c) Defense Success Rate of Various Models as a Function of N-Pixel Attack Magnitude, (d) Defense Success Rate B of Various Models as a Function of APA Attack Magnitude.

Table 5 presents defense success rates (β) for various defense methods and ours against APA on three datasets. For CIFAR10, our methods achieved 95.6% and 96.12% success rates, while Jujutsu and LGS obtained only 86.5% and 93.2%, respectively. Similarly, for CIFAR100, our methods reached success rates of 95.43% and 94.3%, outperforming Jujutsu’s 55.7% and LGS’s 73.7%. In the ImageNet dataset, ours (MI) achieved the highest defense success rate of 89.1%,

while PSNR obtained 83.2%, both surpassing Vax-a-Net's 86.8% and DW's 65.2% and 66.2% for VGG and Inception models, respectively. Not all methods have reported results for every dataset, limiting a comprehensive comparison of their effectiveness.

Table 5. Defense Success Rate (β) of Various Defense Methods Against APA Covering At Least 5% of the Image. Adversary Patches; Toaster, Lipstick, Pineapple, And School-Bus Were Used.

Défense	Défense Success Rate(β)		
	CIFAR10	CIFAR100	ILSVRC12
H/Ours (EfficientNet)	91.3	80.5	-
MI/Ours (ResNet)	95.6	95.43	89.1
PSNR/Ours (Inc)	96.12	94.3	83.2
Jujutsu (ResNet)	86.5	55.7	-
LGS	93.2	73.7	-
V-a-N(VGG)	-	91.6	86.8
DW(VGG)	-	-	65.2
DW(Inc)	-	-	66.2
ECViT-B	47.39	-	41.7

4.3. Attack Size Impact on Model Performance

Figure 7 shows the defense success rate (β) for all four models under N-pixel and APA attacks, respectively, using a depth of 16 for the volumizer algorithm in the decreases when attacked pixels surpass the patch size of the volume. This is due to the volumization algorithm's design, which focuses on small attacks and becomes less effective when perturbations exceed the patch size or span multiple slices. This limitation is

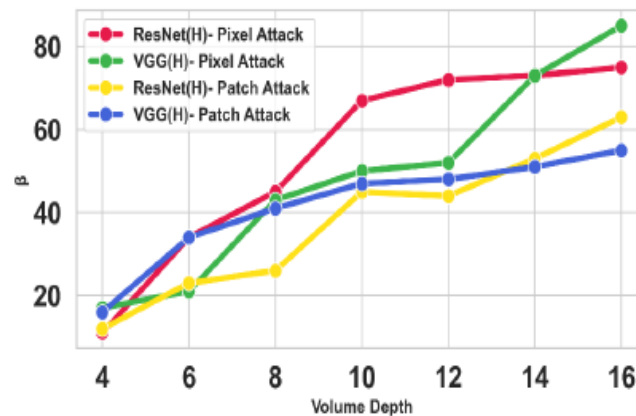


Figure 0. Impact of Patch size(Volume Depth) on Defense Success Rate (β) Against Pixel Attack on CIFAR10 Dataset. the Training Samples Are of Shape (32, 32, 3). We Generated Volumes Starting with a Patch of Size 16 by 16 With Depth 4, All the Way to a Patch Size of 4 by 4 With Depth 16.

more prominent if the attack covers a large image portion, potentially obscuring important object details.

4.4. Class Generalization

Figure 8 depicts VGG generalization performance on CIFAR10 test data. The undefended model exhibits an AUC of 0.5, while the defended model achieves an AUC of 0.69 under 1-pixel attack while the same model achieves AUC of 0.65 under APA. This indicates that the defended model shows improved performance in terms of class generalization compared to the undefended model. These plots suggest that the defended model has better discriminative power and can effectively distinguish between different classes in the CIFAR10 dataset when under 1 pixel and APA attacks. This improvement in AUC demonstrates the effectiveness of our proposed approach in enhancing the model's class generalization capabilities when under localized universal attacks.

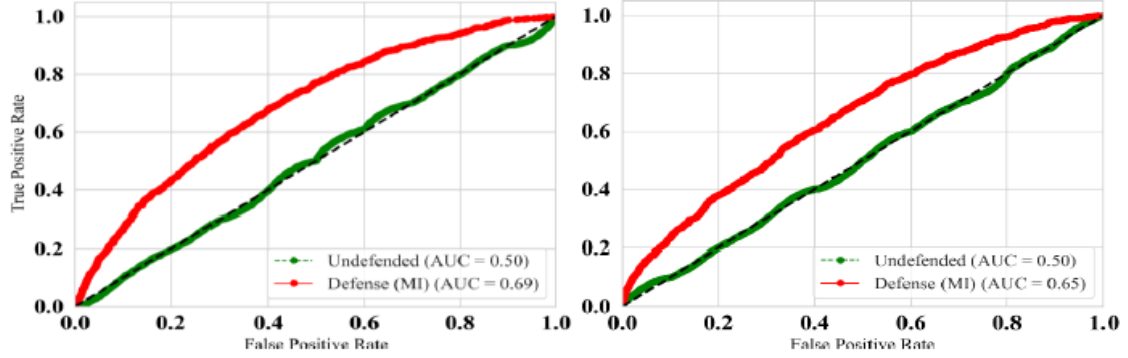


Figure 8. (Left) ROC of VGG Model Under 1-Pixel Attack. (Right) ROC of VGG Model Under APA of Size 8 by 8 Pixels. Class Generalization Performance Comparison Between Defended and Undefended VGG on CIFAR10 Test Set. the Plots are Micro Average ROC Curve Across the 10 Classes.

4.5. Ablation Study

We conduct an ablation study to assess the impact of volume depth and curriculum learning in our defense methodology.

4.5.1. Impact of CL on Défense Success Rate

For this experiment, we train our models with and without the curriculum learning phase and compare the results with the fully trained models. The results are presented in Table 6.

Table 6. Impact of Curriculum Learning and Volumizer: Acc_{clean} And Acc_{attack} For Models Trained Without Curriculum Learning (I-Vol), Models Trained with Curriculum Learning but Without Volumizer (I-CL), And Fully Trained Models (I).

Method	Acc_{clean}			Acc_{attack}		
	EffNet	VGG	Inception	EffNet	VGG	Inception
I-Vol	93.7	97.2	97.4	90.6	95.7	95.2
I-CL	97.9	99.4	98.6	56.8	38.1	12.3
I	96.3	99	99	91.3	98.6	96.12

The first scenario we tested was our method without CL but with the volumizer (I-Vol). The performance under this configuration was reasonably good, with the Acc_{clean} being 93.7%, 97.2%, and 97.4% for EfficientNet, VGG, and Inception, respectively. However, the Acc_{attack} was markedly lower, specifically, it was 81.6% for EfficientNet, 89.7% for VGG, and 85.2% for Inception. Compared to the full method (I), the Acc_{attack} was lower by 9.7%, 8.9%, and 10.9%, respectively. This indicates the effectiveness of Curriculum Learning in improving the model's robustness against adversarial attacks. Second, we studied the effect of Curriculum

Learning without the volumizer (I-CL). This configuration achieved even higher Acc_{clean} scores, specifically 97.9%, 99.4%, and 98.6% for Efficient Net , VGG, and Inception, respectively. However, the Acc_{attack} suffered significantly without the volumizer. For EfficientNet , VGG, and Inception, the Acc_{attack} were 56.8%, 38.1%, and 12.3%, respectively, revealing drops of 34.5%, 60.5%, and 83.8% compared to the full I syllabus. This demonstrates the vital role the volumizer plays in enhancing the model’s resilience to adversarial attacks.

Lastly, our fully implemented method (I), incorporating both Curriculum Learning and the volumizer, consistently outperformed the other configurations in terms of Acc_{attack} , achieving 91.3%, 98.6%, and 96.12% for EfficientNet , VGG, and Inception, respectively. These figures indicate the combined effect of both components in improving the model’s resilience to adversarial attacks.

4.6. Timing Information

Inference and training time comparisons between our method and undefended models are presented in Table 7. An inference overhead for a model protected with our method is noticeable compared to the undefended models— around 6 milliseconds on average across all three models. This increased latency is primarily due to the modifications made to the model architecture to accommodate our defense strategy. Additionally, our defense incurs a significant overhead during training. Depending on the size of the dataset, the additional training time can span from hours to days. However, this process only needs to be run once, as does preprocessing dataset a priori. All inference runs used an NVIDIA RTX A4000, while training was conducted on a node equipped with four RTX A100 GPUs. Despite the increased computational demands, the benefits of enhanced security provided by our defense method offer a worthwhile trade-off.

Table 0. Inference Time (ms) For VGG, Resnet, And Inception Trained on ImageNet and the Same Model with Our Defense Based on I Syllabus And a 16 Depth Volumized Inputs.

Method	VGG	ResNet	Inception
Undefended	0.12	0.14	0.09
I/Ours	0.23	0.18	0.16

5. CONCLUSION

We introduced a proactive defence approach against localized adversarial attacks, which preserves model performance on clean data. Our method combines a volumization algorithm that converts 2D images into 3D volumetric representations while maintaining spatial relationships, increasing resilience to perturbations. Additionally, we employ a deep curriculum learning optimization strategy, ordering training samples by complexity, enabling progressive learning from simple to complex samples. By incorporating these techniques into popular CNN architectures, we demonstrated the effectiveness of our method against N-pixel and patch attacks. Experimental results indicated improved robustness without sacrificing performance on clean data, confirming our approach’s ability to enhance image classification model resilience against localized adversarial attacks.

REFERENCES

- [1] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and Harnessing Adversarial Examples,” *ArXiv14126572 Cs Stat*, Dec. 2014, Accessed: Nov. 10, 2018. [Online]. Available: <http://arxiv.org/abs/1412.6572>

- [2] D. Karmon, D. Zoran, and Y. Goldberg, “LaVAN: Localized and Visible Adversarial Noise.” arXiv, Mar. 01, 2018. doi: 10.48550/arXiv.1801.02608.
- [3] H. Ghebrechristos and G. Alagband, “Deep curriculum learning optimization,” *SN Comput. Sci.*, vol. 1, no. 5, p. 245, Jul. 2020. doi: 10.1007/s42979-020-00251-7.
- [4] J. Su, D. V. Vargas, and S. Kouichi, “One pixel attack for fooling deep neural networks,” Oct. 2017, Accessed: Nov. 12, 2018. [Online]. Available: <https://arxiv.org/abs/1710.08864>
- [5] T. B. Brown, D. Mané, A. Roy, M. Abadi, and J. Gilmer, “Adversarial Patch,” *ArXiv171209665 Cs*, May 2018, Accessed: Mar. 12, 2022. [Online]. Available: <http://arxiv.org/abs/1712.09665>
- [6] C. Szegedy *et al.*, “Intriguing properties of neural networks,” *ArXiv13126199 Cs*, Dec. 2013, Accessed: Oct. 26, 2018. [Online]. Available: <http://arxiv.org/abs/1312.6199>
- [7] T. Gittings, S. Schneider, and J. Collomosse, “Robust Synthesis of Adversarial Visual Examples Using a Deep Image Prior.” arXiv, Jul. 03, 2019. doi: 10.48550/arXiv.1907.01996.
- [8] N. Papernot, P. McDaniel, X. Wu, S. Jha, and A. Swami, “Distillation as a Defense to Adversarial Perturbations against Deep Neural Networks.” arXiv, Mar. 14, 2016. doi: 10.48550/arXiv.1511.04508.
- [9] S. Baluja and I. Fischer, “Learning to Attack: Adversarial Transformation Networks,” *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, Art. no. 1, Apr. 2018, doi: 10.1609/aaai.v32i1.11672.
- [10] T. Bai *et al.*, “AI-GAN: Attack-Inspired Generation of Adversarial Examples.” arXiv, Jan. 12, 2021. doi: 10.48550/arXiv.2002.02196.
- [11] D. Chen, R. Xu, and B. Han, “Patch Selection Denoiser: An Effective Approach Defending Against One-Pixel Attacks,” in *Neural Information Processing*, T. Gedeon, K. W. Wong, and M. Lee, Eds., in Communications in Computer and Information Science. Cham: Springer International Publishing, 2019, pp. 286–296. doi: 10.1007/978-3-030-36802-9_31.
- [12] Z.-Y. Liu, P. S. Wang, S.-C. Hsiao, and R. Tso, “Defense against N-pixel Attacks based on Image Reconstruction,” in *Proceedings of the 8th International Workshop on Security in Blockchain and Cloud Computing*, in SBC ’20. New York, NY, USA: Association for Computing Machinery, Oct. 2020, pp. 3–7. doi: 10.1145/3384942.3406867.
- [13] S. A. A. Shah, M. Bougre, N. Akhtar, M. Bennamoun, and L. Zhang, “Efficient Detection of Pixel-Level Adversarial Attacks,” in *2020 IEEE International Conference on Image Processing (ICIP)*, Oct. 2020, pp. 718–722. doi: 10.1109/ICIP40778.2020.9191084.
- [14] M. A. Husnoo and A. Anwar, “Do not get fooled: Defense against the one-pixel attack to protect IoT-enabled Deep Learning systems,” *Ad Hoc Netw.*, vol. 122, p. 102627, Nov. 2021, doi: 10.1016/j.adhoc.2021.102627.
- [15] H. Li and Z. Lin, “Accelerated Proximal Gradient Methods for Nonconvex Programming,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2015. Accessed: Aug. 01, 2022. [Online]. Available: <https://papers.nips.cc/paper/2015/hash/f7664060cc52bc6f3d620bcdec94a4b6-Abstract.html>
- [16] A. Levine and S. Feizi, “(De)Randomized Smoothing for Certifiable Defense against Patch Attacks.” arXiv, Jan. 08, 2021. doi: 10.48550/arXiv.2002.10733.
- [17] J. H. Metzen and M. Yatsura, “Efficient Certified Defenses Against Patch Attacks on Image Classifiers.” arXiv, Feb. 08, 2021. doi: 10.48550/arXiv.2102.04154.
- [18] J. Hayes, “On Visible Adversarial Perturbations & Digital Watermarking,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, USA: IEEE, Jun. 2018, pp. 1678–16787. doi: 10.1109/CVPRW.2018.00210.
- [19] M. Naseer, S. Khan, and F. Porikli, “Local Gradients Smoothing: Defense Against Localized Adversarial Attacks,” in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2019, pp. 1300–1307. doi: 10.1109/WACV.2019.00143.
- [20] E. Chou, F. Tramèr, and G. Pellegrino, “SentiNet: Detecting Localized Universal Attacks Against Deep Learning Systems.” arXiv, May 09, 2020. doi: 10.48550/arXiv.1812.00292.
- [21] Z. Chen, P. Dash, and K. Pattabiraman, “Turning Your Strength against You: Detecting and Mitigating Robust and Universal Adversarial Patch Attacks.” arXiv, Jun. 23, 2022. Accessed: Aug. 02, 2022. [Online]. Available: <http://arxiv.org/abs/2108.05075>
- [22] D. Smilkov, N. Thorat, B. Kim, F. Viégas, and M. Wattenberg, “SmoothGrad: removing noise by adding noise.” arXiv, Jun. 12, 2017. doi: 10.48550/arXiv.1706.03825.
- [23] H. Ghebrechristos, “Deep-CLO: Enhancing Feature Extraction from 2D and 3D Topological Data,” 2020.

- [24] G. Ras, L. Ambrogioni, P. Haselager, M. A. J. van Gerven, and U. Güçlü, "Explainable 3D Convolutional Neural Networks by Learning Temporal Transformations." arXiv, Jun. 29, 2020. doi: 10.48550/arXiv.2006.15983.
- [25] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," *ArXiv160508695 Cs*, May 2016, Accessed: Jun. 23, 2018. [Online]. Available: <http://arxiv.org/abs/1605.08695>
- [26] J. Rauber, W. Brendel, and M. Bethge, "Foolbox: A Python toolbox to benchmark the robustness of machine learning models." arXiv, Mar. 20, 2018. doi: 10.48550/arXiv.1707.04131.

AUTHORS

Henok E. Ghebrechristos is a graduate student specializing in deep learning and computer vision. He received his B.S. degree in Physics from the University of Colorado, Boulder in 2010. His research interest includes gaining insight from deep learning black box by controlling convolution operations and feature maps generation.



Gita Alaghband is professor and Chair of Computer Science and Engineering. Her research interests in parallel processing and distributed systems include application programs and algorithm design, computer architectures, operating systems, performance evaluation, and simulation.



UNSUPERVISED MULTI-SCALE IMAGE ENHANCEMENT USING GENERATIVE DEEP LEARNING APPROACH

Preeti Sharma^{1,*}, Manoj Kumar^{2,3,4,5,*}, and Hitesh Kumar Sharma¹

¹ School of Computer Science, University of Petroleum and Energy Studies (UPES), Dehradun, 248007 India .

² School of Computer Science, FEIS, University of Wollongong in Dubai, Dubai Knowledge Park, Dubai, UAE

³ Research Cluster Head, Network and Cyber Security, UOWD, Dubai

⁴ MEU Research Unit, Middle East University, Amman, 11831, Jordan

⁵ Research Fellow, INTI International University, Malaysia

ABSTRACT

To produce super-resolution images, it is essential to eliminate the noise elements and give a clear noise-free output. To achieve this purpose multiscale image representation is found to be effective in many ways for its accuracy of correct feature extraction capacity. This denoising approach is integrated as a chosen enhancement tool in the form of an ensemble GAN model, and accordingly, the generator-discriminator training concept is transformed to adopt the approach as per the desired demands. In this research, a multiscale image approach is implemented using an ensemble GAN model with hybrid discriminator architecture. No one form of noise is "ideal" to eliminate while denoising with the proposed model. Instead, based on the properties of the data and the noise inherent in it, the proposed ensemble GAN can handle various sorts of noise. The technique optimises training through simultaneous generator and discriminator model updates, improving output quality, by using the least loss value for discriminator selection. Inception Score (IS) and Fréchet Inception Distance (FID) evaluations show that it outperforms pixel-based denoising, with an amazing accuracy of 99.91%.

KEYWORDS

GAN, multiscale image representation, ensemble GAN, pixel based denoising, Multiscale denoising.

1. INTRODUCTION

The deep learning research division is becoming increasingly detailed in Generative Adversarial Networks (GANs) in a multidisciplinary manner. GANs are widely used to generate high-quality playback images used in several fields, such as forensics, medical diagnostics, architecture, and filmmaking. GAN versions are currently used to create images from text data, create movies from still images, increase image resolution, and manipulate images. Its uses range from detecting anomalies to improving chess games [1]. Image noise is a common problem and GAN works to refine and restore low-quality or degraded images. Using the contingency distribution of a collection of media data GAN generates new and accurate visual results to improve denoising as shown in Figure 1.

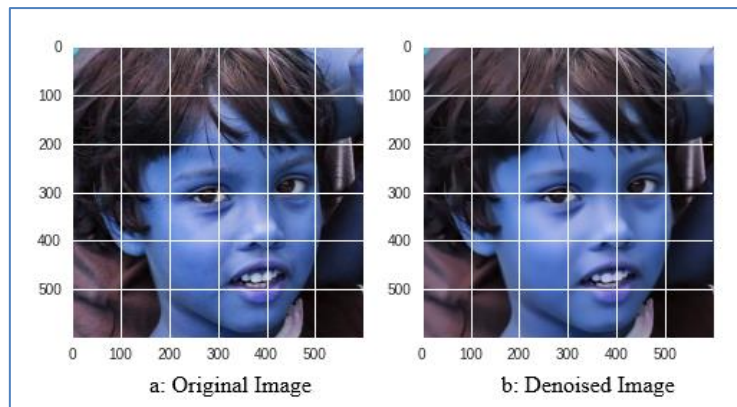


Figure 1: Denoising of Coloured Image using Ensemble GAN Model

It has been observed that a captured image of a real scene or object can be degraded by inaccurate capture and optical factors such as improper depth of field, poor focus, camera shake, object movement, short exposure, and poor optical quality. Tasks involving image identification or classification can be impeded by noisy visuals with imperfections[2]. Noise often comes out in images as detached pixels or blocks of pixels, causing heavy visual effects. In general, the noise signal is inappropriate for the test object. It emerges as impractical information that disturbs the picture's visible information and is a significant barrier to image processing. Therefore, recovering the original signal and preserving features from noisy images is the goal of the extremely promising work of image denoising in graphics [3].

The current advancement in picture denoising has been focused on deep denoiser, also known as the deep network for denoising. Unsupervised deep denoisers that just call chaotic noisy photos without training data have gained popularity over the last few years. Relaxing the need of supervised learning on training samples has recently attracted growing interest. The training of DNNs using a noisy picture dataset without pair-wise correspondence or even just the input noisy image itself has been the subject of more recent research on unsupervised deep denoisers. These techniques fall into one of two types [4]: -

- **Data Augmentation Method:** - When training a DNN to map a noisy image to itself, Noise2Void and Noise2Self use the blind-spot strategy to avoid overfitting (convergence to identity map), while Noiser2Noise and Noise-as-Clean add additional noise to the original noisy image to produce image pairs that are then used to train the DNN.
- **Denoising DNN Regularization:** - By compensating the prediction's divergence, Stein's Unbiased Risk Estimator (SURE) regularizes the DNN. Early-stopping is used in the Deep Image Prior to prevent overfitting. To lessen the bias and variance of the prediction from the DNN trained on a single noisy picture, Self2Self introduces a dropout-based training/testing strategy.

As according to the latest studies, deep learning has revolutionized the fields of pattern detection and computer vision. k-means singular-value decomposition, Principal component analysis using local pixel grouping (LPGPCA), Block matching and 3D filtering (BM3D), and weighted nuclear norm minimization (WNNM) are examples of conventional denoising techniques. These techniques are made to reduce noise depending on the characteristics of both the pictures and the noise. To map from noisy photos to clean images, learning-based techniques like a denoising convolutional neural network (Dn_CNN) frequently employ paired-image datasets[5]. Among

these techniques, convolutional neural networks (CNNs) have been proven to perform well in the disciplines of visual identification, object identification, lossy compression, visual super-resolution, and visual noise removal, amongst many other digital image challenges. The prevalence of Generative Adversarial Networks (GANs) in image enhancement expanded, following their success with accurate image reconstruction results. GAN with help of its adversarial training approach work initiates with noisy image construction and by optimizing loss value of both generator and discriminator, it generated noise free super resolution images. The mathematical formulation of GAN is listed in equation 1.

$$G \min D \max V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (1)$$

Equation 1: Mathematical formula of GAN [6]

Where; G stands for Generator, D stands for Discriminator, $P_{data}(x)$ = real-world data distribution, $P(z)$ = generator distribution, x = P_{data} sample (x), z = a sample taken from $P(z)$, $D(x)$ denotes a discriminator network, $G(z)$ denotes the generator network.

GAN basically uses multiscale representations technique for analysing and modelling tasks which has huge importance in imaging utilization. In contrast to traditional methods, multiscale representations of convolutional neural networks are primarily viewed as feature pyramids. Feature pyramids can leverage further through convolution and down sampling processes, visual background data is converted across local to global perspective. Also, utilizing adversarial training, GANs might assistance overall image restoration to yield HR picture patterns that are greater realistic and precise. However, LR picture characteristics and subsequent up sampling techniques are constrained by GAN-based ultra-composite restorations. In contrast, the LR imagery' limited dimensions often cause excessive noise to exist in the reformed output. Therefore, optimizing the texture receptive field of the network is difficult and cannot fully exploit multi-scale contextual information [7].

This research work utilizes multi-scale representations of images using ensemble GAN approach as an effective noise reduction tool. The model built using one generator coupled with three discriminators architecture. The model utilises the loss value of each discriminator for starting 50 epochs and generates a raw image including noise. Based on least loss value, the discriminator is finalised for execution with generator for remaining epochs. The image generated will then turned into super resolution image by fine tuning in subsequent iterations of the model.

The architecture work like a multi-scale denoiser to improve image output quality and ensure noise-free reproduction of images. Also, mode training problems such as mode collapse and non-convergence of results can also be eliminated using this ensemble approach. Overall, this model is said to offer the advantage of better image restoration while minimizing the potential for distortion and error. The model is evaluated using the Inception Score (IS) and Fréchet Inception Distance (FID) parameters. It outperforms to other conventional approaches showing accuracy score of 99.91%.

Rest of paper is structured with subsequent sections. Sections II and III presented the background and related research of the domain. Section IV defines the proposed methodology including the details about the dataset and defined architecture. Section V demonstrates the experimental and results section. It demonstrates various result graphs along with the comparative study of the proposed model with existing models Deep Convolutional GAN (DCGAN) and Conditional GAN (CGAN). Finally, section VI concludes the research paper.

2. BACKGROUND STUDY

Images can contain large amounts of data and instruction, so image deputation techniques have applications in many fields such as medicine, conveying, military, aerospace, and communications. It is ingrained in our lives and is attached from each of us. However, image noise can occur during image acquisition, storage, and processing, disrupting the instruction provided by the image and reducing image clarity. Image noise generally includes salt and pepper noise, gamma noise, and Gaussian noise. To solve this problem, high-resolution image denoising and image reconstruction (SR) techniques were researched to restore high-quality, high-resolution images[8]. Formally, the solution to the denoising limitation is based on three basic components[9]:

A signal model is followed by a noise model and a signal fidelity assessment (majorly recognized as the objective function). The concept of In Mathematical model, visual degeneration can be characterized as $x = y + n$, where x represents the degraded version of the original image y and n represents the added noise, also known as Additive White Gaussian Noise (AWGN), as illustrated in Figure 3 below.

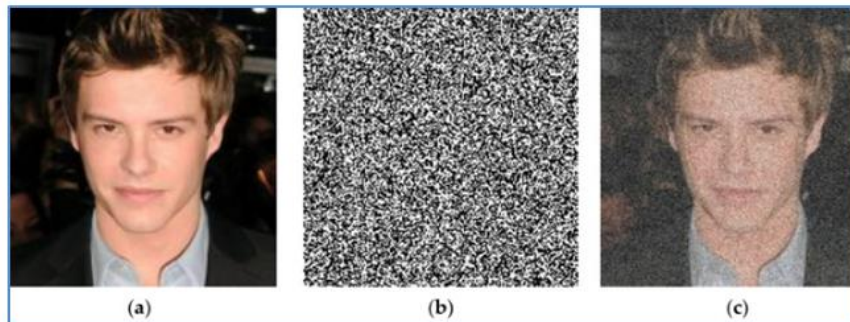


Figure 2: (a) Exposes an unedited or true visual (y), (b) Showcases the AWGN visual (n), and (c) reveals the ($x = y + n$) resultant image [9].

Tavakoli *et al.* [10] provides an approach approach that is effective for picture denoising, according to simulation findings. The original picture is rebuilt using the observation vector and existing recovery methods like L1 minimization after an unknown noisy image of interest is detected (sensed) using a small number of linear functions in random projection. Rajni et al.[11] employ a novel methodology called Optimal Wavelet Basis (OWB). OWB uses Shannon entropy to choose the optimal tree from a noisy picture by applying multilayer Wavelet Packet (WP) decomposition. An innovative technique for de-noising MR (Magnetic Resonance) pictures with Rician noise has been developed by Choudhary et al[12]. This approach improves the Morphological Mean Filter (MMF).

Image denoising methods focus on recovering a visual that has been denoised from a laterally noisy image (x) by removing or contracting the noise (n). With these techniques, accompanying white noise and Gaussian noise are denoised using multiresolution or multiscale analysis. A simple way to denoise an image is to separate the image into equivalent low and frequency-heavy content. Then, this decline can be modified iteratively to the low sub bands. Low sub bands comprise to create high frequency sub band which combines to generate a multiscale representation. It contains the identical data as the reference image without noise as shown in Figure 2.

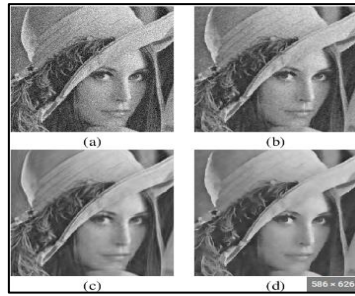


Figure 3: Multiscale Denoising Layers up to Final Output[13]

Image denoising and super-resolution (SR) are two important stages of image processing and are usually studied separately. But, for visual improvement, image noise reduction should be combined with image super-resolution process. Several unsupervised image enhancement methods based on deep learning have been recently developed to eliminate the dependence on paired data (x, y) . These methods use Generative Adversarial Networks (GANs) to approximate the handling of generated images to that of target images without pairwise learning. As an example, severely image-to-image conversion or conversion models such as Cycle GAN can be applied for image enhancement. Also, GAN-based models were also worked to address the task of lighting intensification. These unsupervised models can produce better lighting and colour images. However, in real-world low-light image enhancement tasks are bounded by few constraints listed as:

- Enhanced image contrast and lighting may sometimes result in colour distortions and inconsistencies. Bright areas of dark images may be overexposed.
- Due to unstable training of unsupervised models, contiguous regions may performance sharp colour or illumination discrepancies.
- Models with a single frame-to-frame mapping network, when applied to low-light, high-noise images, are primarily concerned with illumination enhancement, but consistently fail at noise reduction.

As a solution, multi-scale image display is integrated into these models to ensure improved image quality and noise reduction.

3. RELATED WORK

This section consists of a selection of current published references that highlight the importance of noise reduction tools for GAN image regeneration applications. The selection of literature is systematic, following two main aspects (a) Concerns about noise reduction in GAN image SR models. (b) Using multi-scale image representations in GAN models as a solution for image noise devaluation. Yan *et al.*[14] conferred his self-consistent GAN (SCGAN), which allows unchecked noise and can dynamically extract noise maps from noisy images modelling. It received three new self-consistent compulsion that complement each other.

- A clean input should result in a zero response from a noise standard.
- Once presented with a pure noise input, a noise model ought to produce the same result.
- Once the pure noise map is coupled to the clean image, the noisy model also has to extract the relevant to the pattern.

Three major image processing tasks, including blind image denoising, rain streak removal, and noise image super-resolution are achieved using simplicity and effectiveness of the GAN model. It also highlights the importance of noise reduction tools for GAN image regeneration applications focusing on two main aspects. (a) Concerns about noise reduction in GAN image SR models. (b) Using multi-scale image representations in GAN models as a solution for image noise devaluation.

Deep Convolutional Neural Networks (CNN)-based novel algorithm has been suggested by Lee *et al.* [15]. The technique is built on the Deep Convolution Network, which consists of several U-nets. Each U-net reduces noise of different intensities and is graded according to performance improvement. Additionally, the combined CNN employs a comprehensive 3D convolution method. The potential standard might speed up end-to-end learning beyond pre- and post-processing thanks to such an architecture.

Li *et al.* [16] granted a revolutionary generative adversarial network (MSAt-GAN) based on deep attention mechanism and multiscale feature transmission blending. The proposed model was used for fusing visible and infrared imagery. In this study, rather than artificially fixing a single receptive field, he used three different receptive fields to capture multiscale and multilevel depth features of multimodality figure in three channels.

- Firstly, the crucial features of the origin form could be improved extracted from distinctive accessible fields and angles, as well as more adaptable and diverse feature representations, were also featured.
- Secondly, a multiscale fusion mechanism for deep attention was designed. Third, the chained operations to improve feature submission while achieving improved usage of preceding features, multi-level deep features in the encoder and deep features in the decoder are cascaded.
- Ultimately, a GAN with two discriminators was advanced on top of the network structure. As a result, the created image simultaneously preserves the intensity of the infrared image and the texture details of the visible image.

Wang *et al.*[17] To fulfil the image denoising goal, a Deep Residual Network based on Generative Adversarial (GAN) networks was proposed. First, a residual block-based generative adversarial network structure was created. The GAN network was then trained using an advanced loss function. The resulting image is very closely resembled by a well-designed loss function including its distinct counterpart (the ground truth) while enhancing colour and brightness details.

4. RESEARCH METHODOLOGY

4.1. Model Development

The architecture is based on multiple (3) discriminators coupled with single generator are included in the GAN ensemble model. The loss function used in this model architecture is resolved using a minimum distance criterion between the generated distribution and the actual allocation. Initially, the model utilizes the loss value of each discriminator for starting 50 epochs and generates a raw image including noise. Based on least loss value, one discriminator is finalized for execution with generator for remaining epochs which defines denoising process. The image generated will then be turned into super resolution image by fine tuning in subsequent iterations of the model.

By maximizing the objective function regarding the generator and reducing it with respect to the Discriminator, both modules are set against one another (it should be noted that in the regular procedure, we just minimize a loss function). With G and D, the loss function is denoted by the notation $E(D,G)$. Then, E will be the sum of the number that D predicts as 1 for the false picture X_{fake} or $D(z)$ plus the number that D predicts as 0 for the true image X_{real} . For optimization of objective function listed in equation 2 we try to achieve $G_{min}D_{max} E(D,G)$ as mentioned below,

$$G_{min}D_{max} E(D, G) = 1/2[E_{x \sim p(x)}(1 - D(x))] + 1/2[E_{z \sim p(z)}D(G(z))] \quad (2)$$

This novel approach for the GAN is designed to address the following issues:

- (a) an optimized discriminator D (better approximating $\text{Max } E(D, G)$).
- (b) a D better suited to the generator's G capabilities.
- (c) overcoming common flaws of other GANs, such as a collapse of the global framework and inconsistency in local details; and
- (d) integrated with multiscale denoising approach for super-resolution output using ensemble technique.

Two types of pair training illustration are used in this study. That is, the noisy input image x_i and target image y_i is set. The fabrication of a noise-free image $V(x)$ that resembled the already-existing clean target image was used to qualify a denoising network V. A set of three discriminators D1, D2, and D3 are trained concurrently to distinguish the faked actual clear photographs that are noise-free. By reducing adversarial loss and attempting to trick the discriminative network, the denoiser learns to convert noisy regions into clean genuine regions. An encoding network E_n , a residual block layer R, and a decoding network De make up the noise reduction network. The encoder is made up of several convolutional downsampling developments that reconstruct a noisy image into a feature domain (x).

$$V(x) = De(R(En(x))) \quad (3)$$

These feature domains $E_n(x)$ are then inserted into the remaining block. The decoder network De receives the residual block $R(En(x))$'s feature map as an input. At this phase, the rectified feature map is decoded into a fabricated, clean image using a series of upsampled, transposed convolutional layers (x).

4.2. Generator and Discriminator Network

Indistinguishable pictures are produced using a generator network so that the discriminator cannot identify them apart from actual images. Images that are equivalent to real high-definition image labels are created using noisy picture inputs, but they must outperform the labels in terms of details and colours. An eight-layer convolutional architecture is employed for this network. It is to be taken care that all the attributes of our input image must be kept and improved upon, and the form of the images was unaltered throughout the process [13]. Research demonstrates that the network can keep more details as compared to conventional architectures. The discriminator Network includes a stack of three different discriminators. These uses different deconvolutional networks.

Simple convolutional layers connect to a complete link layer in each discriminator's network configuration, and the confidence score is then normalized using the sigmoid function. The best voting classifier approach is used to ensemble the three discriminators to the generator. If the network has been properly trained, it will give a score of close to 1 for the label picture that exists and a score of 0 for the fake image, demonstrating that it has a good capacity for discrimination

and can successfully distinguish between real and fake images. Each discriminator's loss value determines which discriminator will be used at each epoch to construct the GAN model. In our experiment, the true labels are created from generated noised images, whereas the labels for the photograph are clear images. Figure 4 depicts the suggested GAN model's architecture.

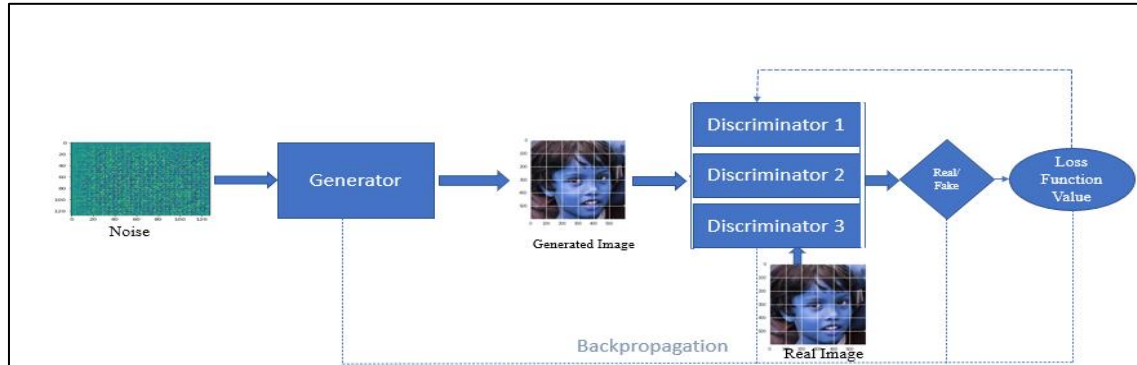


Figure 4: GAN Model with Multiscale Denoiser using Ensemble Approach.

4.3. Training

The Ensemble GAN model typically features non-monotonic loss functions and are very hypersensitive to parameters. It gets trained to do a variety of tasks by altering the input. So, pre-training the generator and discriminator is typically required. To make up-scaled images for super-resolution, for instance, we feed in down-scaled photos and let the generator network create them. We use noisy pictures for denoising and gaussian kernel-blurred images for deconvolution. Batch normalization was applied except for the final layer, to each of the convolutional layers and residual blocks for both the generative and discriminator networks, to avoid gradients from contracting or expanding.

Denoising algorithm of the proposed model:

Algorithm: GAN Model with Multiscale Denoiser using Ensemble Approach.

Input: Data Samples and Noise Samples of Face Images.

Step 1: Pre-training the GAN generator (Gn1) by fixing reconstruction parameters.

Step 2: GAN fake denoised generation (Gn2) through adversarial training with ensemble discriminator feedback.

Step 3: Deep learning Ensemble Detectors (ED) applied to verify the GAN fake generator effect (Gn1+Gn2).

Step 4: Optimize and update each of discriminator in the ensemble using loss function value generated.

Step 5: Minimize generator loss function and update.

Step 6: Fake and Real Image Classification Analysis output.

4.4. Dataset

Google Facial Expression Comparison (FEC) [18] is used for the purpose of training and testing the proposed model. This dataset of faces was collected from Flickr.com to create software for facial expression detection and search. The dataset consists of 87,517 verified photos that were obtained from Flickr and utilised for biometric analysis. Exposing.ai discovered that the 87,517

distinct photographs belonged to 45,382 distinct Flickr customer accounts. The FEC dataset is positioned within the context of bigger facial recognition datasets like MS-Celeb-1M and Mega Face datasets. The dataset that consists of large scale and heterogeneous facial image information is pre-processed with necessary refinements to improve model training and obtain better accuracy with minimum time consumption. These pre-processing tools apply cropping for faster computation of loss function and identify suitable normalisation. On the sampled data set, the training and test samples are divided 7:3 respectively.

4.5. Evaluation

In this work, Accuracy, Loss values, Fréchet Inception Distance (FID) and Inception Score (IS) that are suggested as ad-hoc metrics to evaluate the performance of model. These metrics are indicators of the visual quality of created images.

The Inception Score devised by Salimans et al. accurately depicts the photographs' quality and diversity. The IS measures the estimated difference between the spread of class labels used to train the external network and the dispersion of class predictions for data from the GAN. It is calculated using the Kullback-Leibler (KL) divergence, a statistics formula which measure of how similar or dissimilar two probability. Our distributions differ when KL divergence is substantial. That is, it is high when the whole set of created photos has a wide variety of labels, and each generated image has a unique label. The KL divergence's exponential is used to get the final score, and then its average across all of our photos is used to calculate the IS.

The Fréchet Inception Distance (FID) devised by Heusel et al. extends IS is a statistic that determines how far apart feature vectors determined for actual and artificially created pictures are from one another. The score enumerates the statistics on computer vision aspects of the original pictures that were derived utilising the Inception v3 model to classify images, and it compares the two groups. A perfect score of 0.0 indicates that the two sets of photographs are identical, whereas lower values show that the two groups of photos are statistically more alike or similar. It is used to assess the level of picture quality produced by generative adversarial networks, and lower scores have been found to be positively correlated with better images.

5. EXPERIMENTS AND RESULTS

In contrast to conventional image processing techniques, the suggested multiscale Ensemble GAN approach enables us to employ a single architecture framework to accomplish many goals. To train the model, we just need to change the pre-processing stage and add new inputs. The generator is forced to create images that seem better due to rivalry between it and the discriminator. The model may use learnt features to create pictures from inputs that lack specific information since it can learn from large datasets. For instance, the suggested model can construct human faces that seem convincing even when given extremely low-resolution photos of human faces as input. The model performance on loss function in generating the noise free images using multiscale Ensemble GAN from images with noise as input is given below in Figure 5:

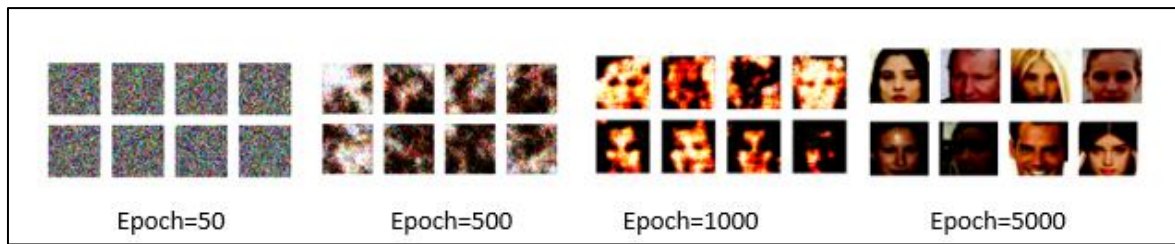


Figure 5: Experimental Output of Multiscale Denoiser in the Proposed GAN Model

The proposed multiscale ensemble GAN model is worked using multiscale denoising approach to get the best super-resolution image output as well as minimize the model training problems. The model is fitted well with reasonable modifications. It helps to eliminate model training difficulties and provide a simplified loss function to confirm optimal model output. Model performance is ensured with the help of 3 discriminator ensemble that is trained aligned with generator so that they can correctly locate denoising as is applied in the generator output. Alongside, the discriminator is chosen at each iteration based on its result that minimizes loss function. The model outcome is evaluated with accuracy of 99.91%, IS value 1.4 and FID value of 412.

The graph below shows the difference level of real input and fake image generation as produced by the model presented in Figure 6 and Figure 7. Model achieves good optimized values of IS (1.4) and FID(412) shown in the graph below mentioned as Figure 8 and Figure 9 respectively. It is observed to outperform, achieving great accuracy of 99.91%.

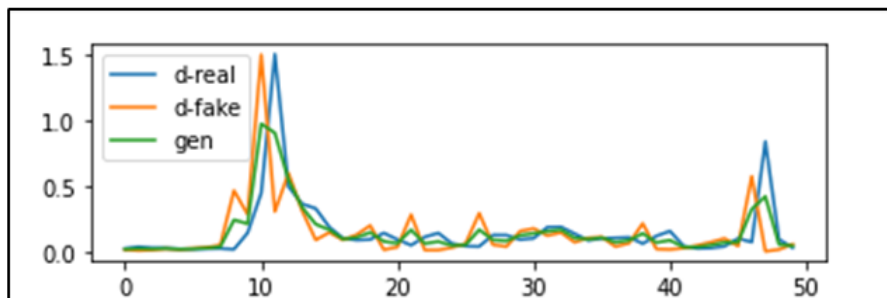


Figure 6: Line Plots of Multiscale Ensemble GAN model with improved denoising image generation graph.

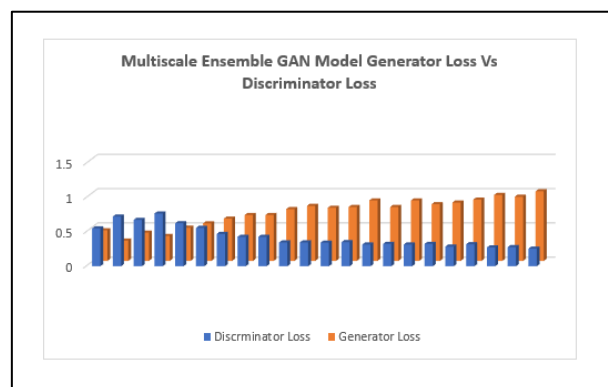


Figure 7: Line Plots of Multiscale Ensemble GAN Model Loss Function of generator and discriminator.

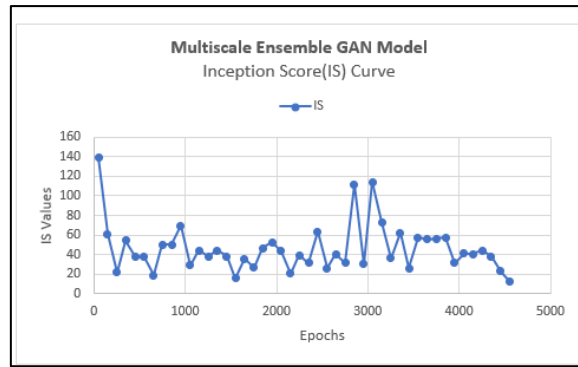


Figure 8: Line Plots of Multiscale Ensemble GAN Model Inception Score (IS).

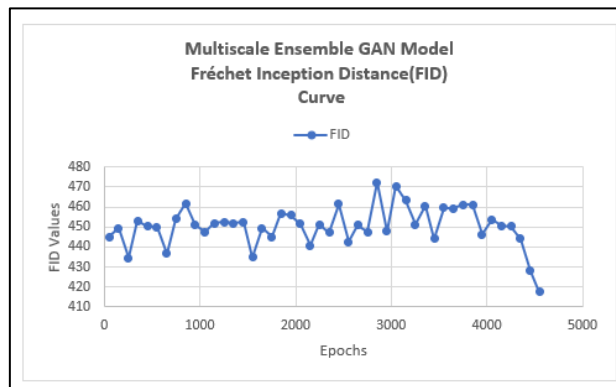


Figure 8: Line Plots of Multiscale Ensemble GAN Model Fréchet Inception Distance (FID).

The model's generalization and robustness are provided by the IS (1.4) and FID (412) scores' optimized values. The model performs excellently and achieves a respectable accuracy rating of 99.91%. On Basisof achieved scores it is observed that the architecture works like a multi-scale denoiser to improve image output quality and ensure noise-free reproduction of images. Also, mode training problems can also be eliminated using this ensemble approach. Performance studies suggest that model analyses the image with many typical facial features and detects facial details rather well. Additionally, the model can conduct deconvolution on human faces in a respectably acceptable manner. It is observed to be better than the other approaches as shown in Table 1 below.

Research	Dataset	Accuracy Score
Ours	Google Facial Expression Comparison (FEC)	99.91%
Wang et al. [19]	CIFAR10	99.5%
Benny et al. [20]	CIFAR10 / MIST	69.51% 98.90%

When compared to current models, our suggested model outperforms them. Our model achieves an accuracy of 99.91% using the Google Facial Expression Comparison (FEC). When applied to CIFAR10, Wang et al. attain an accuracy of 99.5%, which is surpassed by our model. Furthermore, when compared to Benny et al.'s models on the CIFAR10 and MIST datasets, which reach 69.51% and 98.90% accuracy, respectively, our model exceeds both in terms of accuracy, indicating its usefulness in accurately capturing and recognizing facial emotions. The

findings demonstrate our suggested model's improved performance, establishing it as a strong challenger in the domain of facial expression comparison.

6. CONCLUSION

The presented research provides a strong technique to denoising and super-resolution via the combination of multiscale picture representation with an ensemble Generative Adversarial Network (GAN) model. Using a hybrid discriminator architecture designed for multiscale analysis, the model exhibits a surprising capacity to manage various types of noise inherent in the input data. The training approach, which includes simultaneous updates to the generator and discriminator models, optimises the learning technique. Furthermore, dynamically choosing the discriminator with the lowest loss value improves overall output quality. Inception Score (IS) and Fréchet Inception Distance (FID) evaluation measures highlight the performance of the proposed ensemble GAN model, with an excellent accuracy of 99.91%. This study not only enhances the state-of-the-art in denoising techniques, but also demonstrates the versatility of ensemble GANs for dealing with diverse forms of noise in multiscale picture representation, resulting in high-quality, noise-free images.

In terms of future work, there are still several restrictions on the suggested GAN model, despite its ability to provide realistically synthesised discrete data, continuous data, and even time series data to address the problems of fewer labels and imbalanced classifications. It has been difficult to analyse these created data and figure out how to apply them to current applications, such as medical research, which requires genuine data to corroborate.

REFERENCES

- [1] Y. Intrator, G. Katz, and A. Shabtai, "Multi-Discriminator Generative Adversarial Networks".
- [2] R. A. Khan, Y. Luo, and F. X. Wu, "Multi-scale GAN with residual image learning for removing heterogeneous blur," *IET Image Process.*, vol. 16, no. 9, pp. 2412–2431, 2022, doi: 10.1049/ipr2.12497.
- [3] P. Gong, J. Liu, and S. Lv, "Image Denoising with GAN Based Model," *J. Inf. Hiding Priv. Prot.*, vol. 2, no. 4, pp. 155–163, 2020, doi: 10.32604/jihpp.2020.010453.
- [4] T. Pang, H. Zheng, Y. Quan, and H. Ji, "Recorrupted-to-Recorrupted: Unsupervised Deep Learning for Image Denoising," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2043–2052, 2021, doi: 10.1109/CVPR46437.2021.00208.
- [5] L. D. Tran, S. M. Nguyen, and M. Arai, "GAN-Based Noise Model for Denoising Real Images," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12625 LNCS, pp. 560–572, 2021, doi: 10.1007/978-3-030-69538-5_34.
- [6] I. Goodfellow *et al.*, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020, doi: 10.1145/3422622.
- [7] C. Yuan, K. Deng, C. Li, X. Zhang, and Y. Li, "Improving Image Super-Resolution Based on Multiscale Generative Adversarial Networks," 2022.
- [8] S. Li, P. Qian, X. Zhang, and A. Chen, "Reconstruction Technology of Multiscale-Fusion Images," vol. 2021, 2021.
- [9] C. C. By, "Adversarial Gaussian Denoiser for Multiple-Level," 2021.
- [10] [A. Tavakoli and A. Pourmohammad, "Image Denoising Based on Compressed Sensing," *Int. J. Comput. Theory Eng.*, vol. 4, no. 2, pp. 266–269, 2012, doi: 10.7763/ijcte.2012.v4.463.

- [11] R. Rajni, "Image Denoising using Wavelet Packet Transform and Optimal Wavelet Basis Image Denoising using Wavelet Packet Transform and Optimal Wavelet Basis," *Proceeding Natl. Conf. Comput. Commun. Electr. Syst. 2016*, no. November 2016, 2018.
- [12] J. Choudhary and A. Choudhary, "Enhancement in Morphological Mean Filter for Image Denoising Using GLCM Algorithm," *Int. J. Comput. Theory Eng.*, vol. 13, no. 4, pp. 134–137, 2021, doi: 10.7763/ijcte.2021.v13.1302.
- [13] M. Tico, "MULTI-FRAME IMAGE DENOISING AND STABILIZATION," no. Eusipco, pp. 1–4, 2008.
- [14] H. Yan, X. Chen, V. Tan, W. Yang, and J. Wu, "Unsupervised Image Noise Modeling with Self-Consistent GAN".
- [15] S. Lee, M. Negishi, H. Urakubo, H. Kasai, and S. Ishii, "Mu-net : Multi-scale U-net for two-photon microscopy image denoising and restoration," *Neural Networks*, vol. 125, pp. 92–103, 2020, doi: 10.1016/j.neunet.2020.01.026.
- [16] J. Li, B. Li, Y. Jiang, and W. Cai, "MSAt-GAN : a generative adversarial network based on multi-scale and deep attention mechanism for infrared and visible light image fusion," *Complex Intell. Syst.*, vol. 8, no. 6, pp. 4753–4781, 2022, doi: 10.1007/s40747-022-00722-9.
- [17] Z. Wang, L. Wang, S. Duan, and Y. Li, "An Image Denoising Method Based on Deep Residual GAN," *J. Phys. Conf. Ser.*, vol. 1550, no. 3, 2020, doi: 10.1088/1742-6596/1550/3/032127.
- [18] R Vemulapalli, A Agarwala, " A Compact Embedding for Facial Expression Similarity ", CoRR, abs/1811.11283, 2018.
- [19] Z. Wang, Q. She and T. Ward, "Generative Adversarial Networks in Computer Vision", (2022), *ACM Computing Surveys*, vol. 54, no. 2, pp. 1-38. Available: 10.1145/3439723.
- [20] Y. Benny, T. Galanti, S. Benaim and L. Wolf, "Evaluation Metrics for Conditional Image Generation" (2021), *International Journal of Computer Vision*, Vol. 129, no. 5, pp. 1712-1731. Available: 10.1007/s11263-020-01424-w.

USING AUGMENTED REALITY INTERFACES FOR ARTIFICIAL INTELLIGENCE SYSTEMS

Büşra Öztürk^{1,2} and Yakup Genç¹

¹Computer Engineering Department, Gebze Technical University,
Kocaeli, Turkey

²Augmented and Virtual Reality Team, HAVELSAN Inc., Kocaeli, Turkey

ABSTRACT

Augmented reality interfaces offer users an effective environment. In this study, a visualization approach with 3D augmented reality interfaces was introduced to enable users to understand and analyze complex deep learning models in a short time. It has been investigated whether the immersive experience that augmented reality creates on the user in other systems has the same effect when analyzing these models. Two-dimensional studies on deep learning models were examined and what could be done in three dimensions was emphasized. By adding another dimension with augmented reality interfaces, a three-dimensional experience is offered to the user and the results are observed. A CNN model is visualized in the application. When test data was given to the model, the feature maps, filters and connections in the layers were displayed. The application was first run in 2 dimensions, then as a desktop application, and then in 3 dimensions, on Microsoft Hololens-2, a mixed reality headset. Tasks are given to users. Usability was measured with a test called the SUM model, which included completion or non-completion situations, errors, completion times and satisfaction. Here, satisfaction was measured using ASQ(After-Scenario Questionnaire), a user satisfaction measurement questionnaire. The usability of augmented reality 3D interfaces was found to be 80%. The conclusion reached with the answers; It has been stated that users are willing to use this system, their awareness in 3 dimensions is undeniable, and these systems can be used as a feature that increases human ability in artificial intelligence systems.

KEYWORDS

Visualization, Deep Learning Models, Augmented Reality Interfaces

1. INTRODUCTION

People can perceive information differently. People can perceive any situation through hearing, touching, seeing, or listening. By using these perception methods together, complexity is reduced when there is an abundance of information. For many people, visualization is the most important of these perception methods. According to Guzman, visualization can be divided into 4 classes. For him, visualization is not only what we perceive with our eyes, but it also includes the psychological aspect of vision. These are isomorphic, homomorphic, analogical, and diagrammatic visualizations [1].

User interfaces enable human-computer interaction and communication. Users can use systems through user interfaces. To design these interfaces for ease of use and to understand users, it is necessary to get to know users, be familiar with their experiences, and design the overall functionality of the system.

From the beginnings of machine learning to the present day, there has been a growing interest in artificial intelligence, which has led to the emergence of deep learning architectures, the most widely used artificial intelligence algorithms today. Many deep learning approaches have been developed to solve artificial intelligence problems along with deep learning architectures. Various industries such as industry, medicine, robotics, image processing, computer vision, object detection, speech processing-recognition, translation, future prediction, and finance are producing intelligent solutions in numerous fields [2].

Deep learning models are created using artificial neural networks. A neural network consists of input values to be trained, and weight values adjusted during training in the hidden layers. After training is complete, the model makes predictions based on new input values. The weights in the hidden layers are updated through the backpropagation method to make better predictions and improve the model's performance. It is important to understand where the model is working as expected and where it is failing by performing these operations. There are many methods that can be used to understand this, and in these methods, inferences are drawn based on evaluation metrics. Visualization is one of the most important methods. Visualizing the metrics in the model makes it easier for the user, i.e., the data scientist, to make inferences while using the model.

User interfaces can be expressed in 2D, or by adding another dimension, they can be represented in 3D. This allows for cognitive, perceptual, and motor performance for people interacting in the physical world. In augmented reality, a 3D representation of the world is created to allow digital objects to coexist with physical objects. Augmented Reality (AR) interfaces enhance the user's perception with additional information by merging parts of the real world with synthetic, computer-generated images [3].

In the second section of this study, we will describe the visualization of deep learning models. In the third section, we will provide general information about augmented reality interfaces. The fourth section will detail our work on visualizing the deep learning model using the HoloLens 2 device with the MRTK library. In the fifth section, we will present the results.

1.1. Related Work

When we look at studies that combine augmented reality and artificial intelligence applications, it is evident that they have been used in various fields. In radiology, Trestioreanu et al.[4] combined machine learning algorithms for image segmentation using Microsoft HoloLens for use in live medical operating rooms.

Another area where augmented reality (AR) technology is applied is education. Lin and Chen [5] developed a deep learning recommendation system that combines AR technology and learning theories. This system is designed for students with different learning backgrounds and different majors.

Additionally, Bermejo et al.[6] have carried out studies on the combination of big data with augmented reality and their visualization, especially in areas such as tourism, health and public services.

Looking at the visualization of artificial intelligence models; Inkarbekov et al. [7] discusses the exploration of virtual reality as a tool for visualizing AI systems in the context of human-computer interaction. It also highlights the diverse methods within the field of AI visualization, emphasizing the power of interactive 3D visualization and other specialized approaches in making information about AI models more accessible and engaging.

Bock and Schreiber [8] made model visualization in virtual reality using the unreal engine. Yosinski et al. [9] created the Deep Visualization Toolbox based on the development of better tools for visualizing and interpreting neural networks in order to understand how the models work and what calculations they perform in the intermediate layers. It is used to create neuron-by-neuron visualizations using regular optimization.

Kath and Lüers et al. [10] have prepared a virtual reality tool to automate the process of assigning data entries to different categories.

Linse, Alshazly and Martinetz [11] applied visualization in virtual reality by addressing the problem of how to create complex CNNs.

Visualization is of great importance in the implementation of deep learning technologies, which are a subset of artificial intelligence, to facilitate ease of use for the user in data collection, training, setup, and usage. Whether it's monitoring data and models during the application development phase or providing an effective environment for users through augmented reality interfaces in the final application, visualization plays a crucial role in enhancing the user experience.

Many methods are available to visualize deep learning models in 2D for users. Some of these methods use graphical representations when visualizing the model, while others focus on visualizing the network structure. This study has been conducted to answer the question of how augmented reality interfaces will impact users when used to visualize deep learning models. Two-dimensional studies on deep learning models have been examined, and the possibilities of three-dimensional visualization have been explored. Visualization has been achieved through an application using Microsoft HoloLens 2, and the benefits of three-dimensional visualization to users has been tested with users working with deep learning models.

The topics we contributed to the literature with this study can be listed as follows:

- We enabled the visualization of the deep learning model on the HoloLens device, which is a mixed reality glasses that combines the real environment and the virtual environment.
- We measured the usability of this system by measuring users' perception of 3D interfaces with augmented reality on the HoloLens device and 2D interfaces on the monitor.

2. VISUALIZATION OF DEEP LEARNING MODELS ON 2D INTERFACES

Deep learning is a learning approach designed to perform tasks such as creating complex data representations and pattern recognition using multi-layered neural networks. Deep learning is the automatic learning of features by processing large amounts of data and is particularly successful at tasks such as image recognition, text analysis and speech recognition.

A deep learning model is a machine learning model designed to perform complex learning tasks using multi-layered neural network architectures. Such models are particularly successful in tasks that require large datasets and high computational power. The visualization of deep learning models aims to represent the functioning and learning process within complex structures in a more understandable and transparent manner. Deep learning models involve multi-layered neural networks, and these networks perform tasks such as feature extraction and complex data processing.

Deep learning models can be visualized as;

Weight and Activation Maps: Visualizing the weights and activations in each layer of a deep learning model can help you better understand the model's learning process. Heatmaps of weight matrices or activation maps can be used to visualize how the layers are functioning.

Feature Visualization: Visualizing the features learned by the model is actually a method used to understand how the model works on the data and what types of patterns it has learned. Feature visualization reveals how data is represented and processed.

Graphic and Network Structure Visualization Tools: As Graphic Visualization Tools ; TensorBoard, developed for TensorFlow users, is used to visualize deep learning models. It provides information on model performance, learning curves, and graphical representations. Neptune, aims to store, organize, display and compare all metadata created during the model development process. Comet.ml is for researchers and practitioners who want to understand NLP model behavior visually, interactively, and through an extensible tool. Weights and Biases (WandB), focuses on deep learning and allows tracking of training runs with information such as loss, accuracy (learning curves). It also enables the visualization of weight and bias histograms. During training, rich objects like charts, videos, audio, or interactive charts can be logged. Visdom, is a flexible tool for creating, organizing, and sharing visualizations of live, rich data. Hiplot is a straightforward interactive visualization tool that aids AI researchers in discovering correlations and losses in high-dimensional data.

As Network Structure Visualization Tools ; CNNVis, is useful for analyzing a snapshot of a CNN model during training and is focusing on offline analysis. Neural Networks Playground aims to make neural networks more accessible and easier to learn. Neutron is a viewer for neural networks, deep learning, and machine learning models. ANN Visualizer provides visualization by creating a presentable graph of the neural network being constructed.

Qiang Hu, Lei Ma, and Jianjun Zhao, introduced a Pycharm tool called *DeepGraph* that combines visualization and code mapping features to understand and visualize deep learning models[12].

3. VISUALIZATION WITH AUGMENTED REALITY INTERFACES

Augmented Reality (AR) refers to the technology of enriching the real world with digital information. AR adds information generated by computers, such as sound, video, graphics, or GPS data, without obstructing our view and perception of the real environment. This is typically achieved using devices like smartphones, tablets, glasses, or other AR equipment. With Augmented Reality, users can interact with information and other elements that constitute the real environment. Artificial information and objects related to the surroundings can be seamlessly integrated with the real world. AR applications enable users to recognize physical objects in the real world and enhance them with digital content.

Augmented Reality (AR) interfaces refer to graphical or user interfaces that assist users in interactively experiencing the real world and AR content. AR interfaces enable users to interact between the physical world and digital content. Some common examples of AR interfaces include:

Motion Detection and Tracking: AR applications can enable users to interact with objects in the real world by detecting and tracking their movements. This goes beyond simply interacting with touch screens or the physical world, offering a higher level of interaction.

Image Recognition: Users can use the cameras on AR applications to view real-world objects. The application can recognize these objects and overlay digital content on them. For example, during a museum visit, when you view paintings or sculptures, an AR application can recognize these works and display additional information on the screen.

Voice Commands: AR interfaces provide users with the ability to interact with voice commands. Users can perform specific tasks by giving commands to AR devices. For example, you can use voice commands with AR glasses to review a restaurant's menu.

Touch Screens: Touch screens, commonly used in portable devices and tablets, allow users to interact with digital AR objects by tapping and dragging.

Holograms and Virtual World Layers: AR devices can create holographic images or virtual world layers combined with the real world. This allows users to overlay digital information onto or around real-world objects.

Head-Mounted Displays (HMDs): Head-mounted displays for AR provide users with interactive access to AR content in the real world. Examples include Microsoft HoloLens and Google Glass.

Real-Time Data Display: AR allows users to view real-time data by integrating it with the real world. For example, an AR weather application can overlay real-time weather conditions onto the scene the user is viewing.

AR interfaces represent a continuously evolving field aimed at enhancing the user experience and providing more interactions. They create a seamless bridge between the real world and the digital world for users. These technologies are used in many areas, ranging from education to gaming, healthcare services to marketing.

When using augmented reality interfaces, it's important to ensure that these interfaces are user-friendly for the comfort of users. For this purpose, user experience is crucial. User experience (UX) refers to the experience a user has while interacting with a product or service throughout the process. A good user experience helps users achieve their goals quickly and easily, interact with a user-friendly interface, and overall have a positive experience. User-centered designs are required.

4. VISUALIZATION DEEP LEARNING MODELS WITH AUGMENTED REALITY INTERFACES

As a result of the conducted analyses, it has been decided that visualizing deep learning models greatly contributes to understanding the model. In this context, the aim is to provide an immersive environment to the user through augmented reality by presenting a three-dimensional interface. It has been deliberated whether this visualization should be graphical or based on the network structure, and the decision has been made to visualize the network structure. The study has been conducted in this direction.

For the network structure, it is essential to first extract the architecture of the network. This architectural structure was investigated, and it was observed that the study can be divided into

two separate sections: model design and model training. Model design includes the number of hidden layers, activation functions, weight determination, and dropout parameters. Model training encompasses the learning rate, optimization algorithm, epoch, and batch size.

The question to be explored is, "What is the difference in two or three dimensions regarding the benefits of creating awareness for users of the system using these parameters?"

This study is a mixed reality application and will be showcased on the Microsoft HoloLens 2 device.

Microsoft HoloLens-2

The HoloLens-2 device developed by Microsoft combines the virtual and real worlds. It is a wearable headset that allows users to enrich the real world with digital information, providing them with a mixed reality experience. It offers wireless functionality for user convenience. With the help of its sensors, it can track hand and eye movements, and it is capable of accepting voice commands. HoloLens is widely used in visualization applications and research. By merging the real world with digital information, it offers opportunities to visualize data and improve business processes. It is particularly beneficial in the fields of education, medicine, construction, military applications, and industrial use.

4.1. Stages of the Work

In the study, a path was followed as shown in Figure 1. First, a txt file was created to obtain the model's information, and this file was provided to the Unity application to visualize the model.

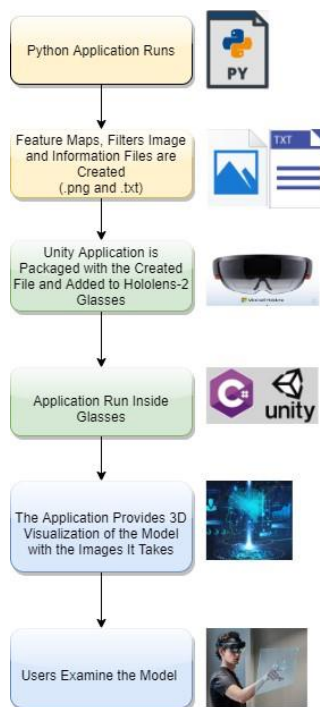


Figure 1 Stage of the Work

For visualization, mnist dataset was used on the lenet model like Figure 2.

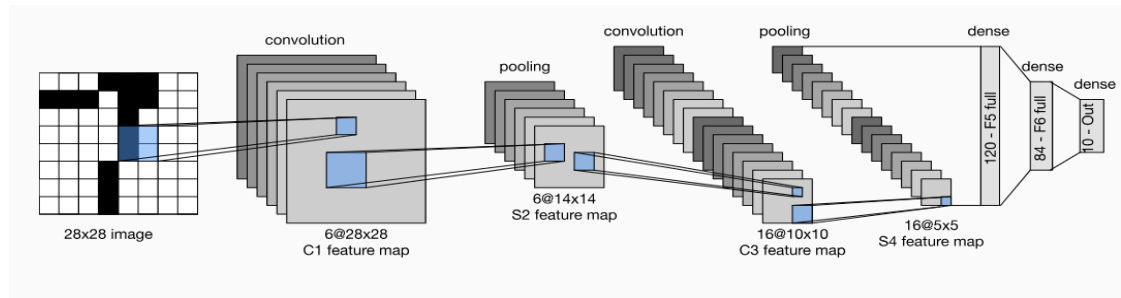


Figure 2 Lenet Model Architecture

4.2. Python Application

First of all, the curse model was created. It was trained with the Mnist dataset. It was then tested with sample data. The filters used at this stage were saved as png files (Figure 3). In addition, the weight values of these filters are stored in the txt file (Figure 4). During the testing phase with sample data, the feature maps in the layers were also recorded. Each numerical value of the Mnist data set was tested. The feature maps of all these values have been recorded .

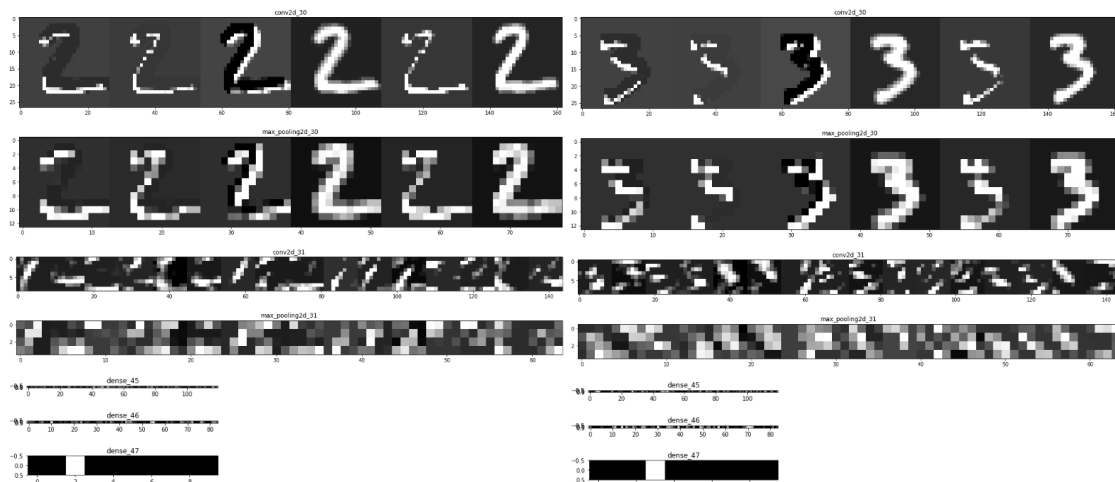


Figure 3 Saved Feature Maps In Each Layer

```

conv_1 [[[ 0.1436145  0.36392194 -0.35958698  0.2831421  0.42979705
-0.17651269]]]

[[[ 0.41274172  0.5671656  0.31445003  0.40183255  0.23169805
-0.06156456]]]

[[[ 0.37006384  0.33062437  0.3566598 -0.17446826 -0.11316572
0.24071108]]]

[[[-0.3338682  0.2797875  0.4647417 -0.1130608  0.47212687
0.58496  ]]]

conv_2 [[[-0.0035638  0.01857955  0.00345579 -0.03389946  0.08544827
-0.02641048  0.04425801 -0.07087898 -0.13192365  0.09264576
-0.02786475 -0.10374412  0.03902938 -0.08340353  0.03945175
-0.05052324]
[-0.11797303 -0.02329039  0.06379768 -0.05451531 -0.12430871
-0.0365213  0.10436592 -0.05729429 -0.15864901  0.09168959
0.07641833 -0.01234336 -0.10829927 -0.02704247 -0.03722082
0.01436257]
[-0.04601551  0.00257322 -0.04086462  0.08437762 -0.02252564
0.04007699 -0.01719958  0.05571664 -0.00258061  0.09509603
-0.02107869 -0.09936313 -0.05063814  0.11793681 -0.03177359
0.01895607]
[-0.10492108  0.04813968 -0.01049864 -0.10914589  0.04594256
-0.09280792  0.01913452 -0.08232185 -0.18332808 -0.11007209
-0.07392413 -0.06140407 -0.10403623 -0.01369665  0.01005093
-0.02693376]

```

Figure 4 Weights Of Saved Filters

4.3. Unity Application

For Fully Connected Layers, the information received is given to the variables held by the PerceptronManager script created in Unity, as shown in Table 1. When the application is run, the perceptron prefab, which represents the layers, the number of neurons in the layers and their weights.

Table 1 Entering Parameters for Network Structure in Unity Editor

PerceptronManager	
Perceptron Prefab	GameObject
Layer Padding	Float
Perceptron Padding	Float
Input Layer Perceptron Count	Int
Output Layer Perceptron Count	Int
Hidden Layer Perceptron Count	Int
Hidden Layer Count	Int
Start Position	Vector3

- MRTK library was used to run on Microsoft Hololens-2.
- The user was offered a choice of test data. (In the next study, the user will be allowed to draw with his own hand.)
- Volume was given to visualize the size of each layer (Figure 5 – Figure 6).
- Layer information was written over the layers.
- Filters visualized(Figure 7).
- Weight information was written as each filter scanned.
- When each layer was clicked, the feature maps of the layer were visualized (Figure 8 – Figure 10).
- To make it more impressive, animations were used while filters scanned inputs and layers were clicked.
- When focusing on the fully connected layers, connections can be visualized (Figure 9).

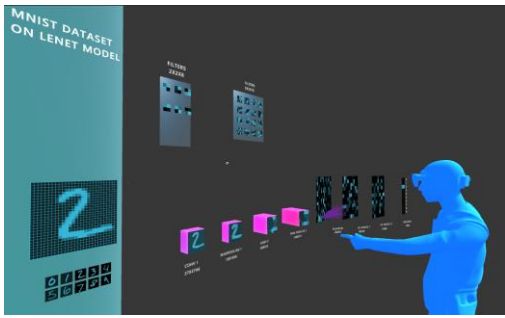


Figure 5 Visualization All Layers on LenetModel

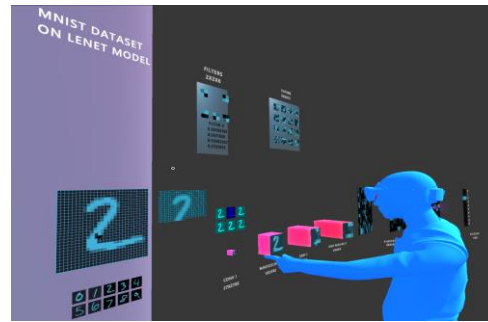


Figure 6 Visualization All Layers on LenetModel

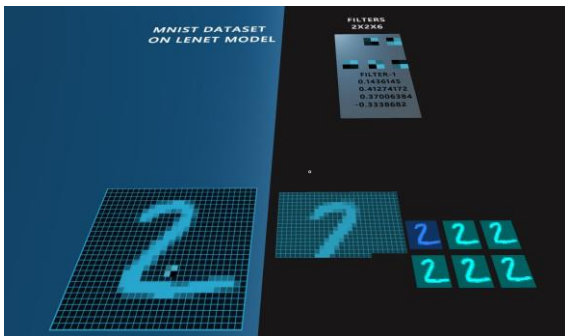


Figure 7 Filters and Opened Feature Maps After Click



Figure 8 Feature Maps

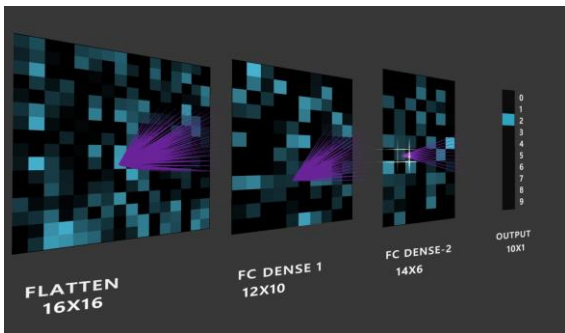


Figure 9 Fully Connected Layers and Output Layer Visualization

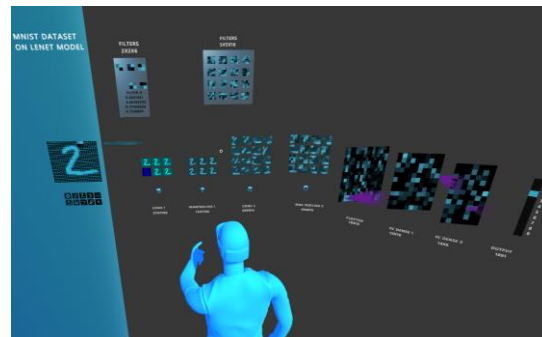


Figure 10 All Layers Opened



Figure 11 User Viewing the Network Structure on Hololens-2

4.4. User Tests

By including real user data in the design process and measuring the benefits or harms of each change, it is necessary to fully understand and explain what effects the changes have on users' behavior.

User research should identify the most important statistical information and tools needed to measure user experience, providing guidelines on the practical aspects of quantitative analysis. Statistical theory should be linked to practice and user research should be measured. It can be done by the following methods:

- summarize data and calculate margins of error,
- determine appropriate sample sizes,
- standardize usability surveys,
- resolve controversies in measurement and statistics [12].

In this study, the user group has been created that include 20 people working in artificial intelligence. The network structure was tested them on Hololens 2. Some questions are asked the users. The system was compared with a 2D version of this application running on a computer. Sum(The Single Usability Metric) was used for user testing. It is the representation of four basic parameters that facilitate the usability of a task or system with a single average statistic.

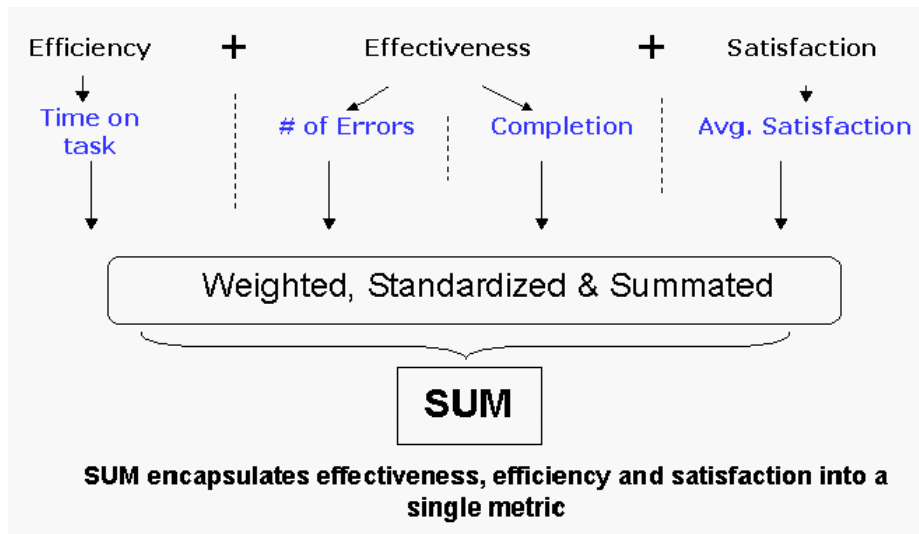


Figure 12 SUM Model

Table 2 Title Distribution of Employees

	Groups	Frequency	Percent (%)
Title	Software Engineer Interested In Data Science	10	50
	Software Engineer Interested In Augmented Reality	4	20
	Software Engineer Interested In Augmented Reality and Data Science	2	10
	Software Engineer not Interested In Augmented Reality and Data Science	4	20
	Total	20	100

Table 3 Professional Experience Distribution of Employees

	Groups	Frequency	Percent (%)
Professional Experience	1-5 years	10	50
	6-10 years	10	50
	Total	20	100

First of all, 2D application was tested by users and results were saved. There are 2 tasks that users do. After that 3D application was tested. SUM values were obtained.

Task 1: Open the application on the PC and select the test data. Which part of the test data is focused on in the 2nd convolution layer?

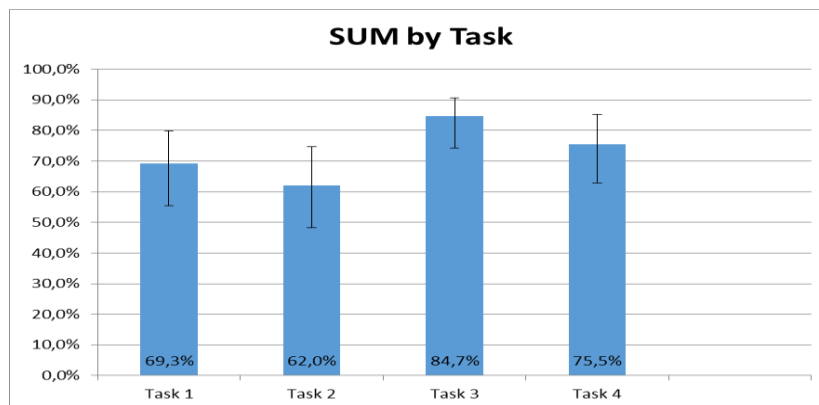
Task 2 : Which convolution filter detects certain features in the data more effectively on PC?

Task 3 : Open the application on the Hololens 2 and select the test data. Which part of the test data is focused on in the 2nd convolution layer?

Task 4: Which convolution filter detects certain features in the data more effectively on Hololens 2.

Every single user did this task. Completion, error size and times were saved. For Satisfaction ASQ (After-Scenario Questionnaire) model was applied as shown in the Table 4. An After-Scenario Questionnaire is a tool designed to gather feedback and insights from participants or observers after they have experienced a particular scenario, event, training session, or any other structured activity. The purpose of the questionnaire is to assess the effectiveness of the scenario, identify strengths, pinpoint areas for improvement, and collect subjective opinions from those involved. By collecting responses to these questions, organizers or facilitators can gain valuable insights into the strengths and weaknesses of the scenario. The feedback can be used to refine future scenarios, enhance facilitation techniques, and ensure that participants' needs and expectations are met. Overall, an After-Scenario Questionnaire is a valuable tool for continuous improvement and refining the design and delivery of scenarios or events [13].

Table 7 Result of Tasks



As a result of the answers received, it was determined that users generally had difficulty in using Hololens in the beginning. However, it was seen that everyone was eager to use the new system. Looking at the answers received from the questions, it was seen that the users gave answers that were open to improvement in 3D interfaces.

They commented that being 3D and interactive provides ease of understanding for more complex networks. In this sense, they stated that they wanted to use 3D.

In the continuation of the study, it is aimed to develop the system to visualize the model structure based on this foundation. At this stage, a structure has been presented to users to form ideas. UX studies will continue.

5. CONCLUSIONS

As with every information system, deep learning systems require interaction with the user. This interaction can occur at any stage of an artificial intelligence system. For example, there is a close relationship with the expert during data collection for the development of supervised learning systems. While some systems use classical 2D (two-dimensional) interfaces in this regard [14, 15, 16], augmented reality can play an important role, especially in spatial data collection applications [17, 18]. Virtual and augmented reality methods can also provide useful interaction in cases where the data is partially created by artificial methods.

There is also machine-human interaction during model training. Manual adjustment of training parameters by the expert requires effective presentation of the training result to the user [19, 20]. Visual interaction is frequently used in this field [21] and it is thought that the use of augmented reality will be especially useful in evaluating the performance of models that produce spatial data [23].

On the other hand, augmented reality interfaces can also be used during the application of deep learning algorithms. While the outputs of deep learning-based systems cannot often be used directly in automatic decision-making (in case of errors), user interfaces have the potential to increase the usability of such systems. These systems [22, 23], which use artificial intelligence systems as a feature that enhances human capabilities, emerge as an area with high potential for augmented reality, especially in cases where spatial data is processed.

As a result of the research, 2-dimensional interfaces were examined, focusing on their intended use and what could be done in 3 dimensions. In this sense, considering the graphical and network structure visualization, it was decided that visualizing the network structure would be more useful

for users. By visualizing the layers, weights and hyperparameters of the model with Hololens, it draws the user into the event, increasing his awareness and making it easier to make decisions.

Additionally, users will not only be able to visualize the network in this system's limitless virtual space but also find its graphical representation in the same environment afterward.

For these purposes, an example network structure was created with the application made on Unity. For this purpose, firstly, the Lenet model was created with Python code, trained with the mnist data set, and tested with sample data. The feature maps that emerged during this testing phase were recorded. In addition, the filters used in the layers, weights in the Full connected layers, and hyperparameters were also recorded. Visualization was made on Unity with all these saved files. The application was first run as a desktop application and then on Hololens-2, a mixed reality headset. Applications were tested separately with 20 people. The usability of the application was measured using the SUM model. The conclusion reached through the answers was that users were willing to use this system and their awareness of 3 dimensions was undeniable. As a result, these systems can be used as artificial intelligence systems that increase human capabilities.

REFERENCES

- [1] YILMAZ R., ARGÜN Z. (2013) “Matematiksel Genelleme Sürecinde Görselleştirme ve Önemi” Hacettepe Üniversitesi Eğitim Fakültesi Dergisi (H. U. Journal of Education) 28(2), 564-576
- [2] DOĞAN F., Adıyaman Üniversitesi, TÜRKÖĞLU İ. Fırat Üniversitesi, ” Derin Öğrenme Modelleri ve Uygulama Alanlarına İlişkin Bir Derleme” , 2018
- [3] A. Fuhrmann, H. Loffelmann, D. Schmalstieg and M. Gervautz, "Collaborative visualization in augmented reality," in IEEE Computer Graphics and Applications, vol. 18, no. 4, pp. 54-59, July-Aug. 1998, doi: 10.1109/38.689665.
- [4] Trestioreanu, Lucian & Glauner, Patrick & Meira, Jorge & Gindt, Max & State, Radu. (2020). Using Augmented Reality and Machine Learning in Radiology. 10.1007/978-3-030-41309-5_8.
- [5] P. Lin and S. Chen, "Design and Evaluation of a Deep Learning Recommendation Based Augmented Reality System for Teaching Programming and Computational Thinking," in IEEE Access, vol. 8, pp. 45689-45699, 2020, doi: 10.1109/ACCESS.2020.2977679.
- [6] Bermejo, Carlos & HUANG, Zhanpeng & Braud, Tristan & Hui, Pan. (2017). When Augmented Reality meets Big Data. 10.1109/ICDCSW.2017.62.
- [7] Medet Inkarbekov, Rosemary Monahan and Barak A. Pearlmutter, “Visualization of AI Systems in Virtual Reality: A Comprehensive Review” International Journal of Advanced Computer Science and Applications(IJACSA), 14(8), 2023. <http://dx.doi.org/10.14569/IJACSA.2023.0140805>
- [8] Bock, Marcel and Andreas Schreiber. “Visualization of neural networks in virtual reality using Unreal Engine.” Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (2018)
- [9] Yosinski, J., Clune, J., Nguyen, A., Fuchs, T. and Lipson, “Understanding neural networks through deep visualization” International Conference on Machine Learning (ICML), 2015.
- [10] Kath, H., Luers, B., Gouvêa, T.S., & Sonntag, D. (2023). A Virtual Reality Tool for Representing, Visualizing and Updating Deep Learning Models. ArXiv, abs/2305.15353.
- [11] C. Linse, H. Alshazly, and T. Martinetz, “A walk in the black-box: 3D visualization of large neural networks in virtual reality,” Neural Computing and Applications, vol. 34, no. 23, pp. 21 237–21 252, 2022.
- [12] Q. Hu, L. Ma and J. Zhao, "DeepGraph: A PyCharm Tool for Visualizing and Understanding Deep Learning Models," 2018 25th Asia-Pacific Software Engineering Conference (APSEC), Nara, Japan, 2018, pp. 628-632, doi: 10.1109/APSEC.2018.00079.
- [13] Sauro, Jeff and Erika Kindlund. “Using a Single Usability Metric (SUM) to Compare the Usability of Competing Products.” (2005).
- [14] James R. Lewis. 1991. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ. SIGCHI Bull. 23, 1 (Jan. 1991), 78–81. <https://doi.org/10.1145/122672.122692>

- [15] Wspanialy, P., Brooks, J., & Moussa, M. (2020). An image labeling tool and agricultural dataset for deep learning. arXiv preprint arXiv:2004.03351.
- [16] Roh, Y., Heo, G., & Whang, S. E. (2019). A survey on data collection for machine learning: a big data-ai integration perspective. *IEEE Transactions on Knowledge and Data Engineering*.
- [17] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare. *Nature medicine*, 25(1), 24-29.
- [18] Alhaja, H. A., Mustikovela, S. K., Mescheder, L., Geiger, A., & Rother, C. (2017, September). Augmented reality meets deep learning for car instance segmentation in urban scenes. In *British machine vision conference* (Vol. 1, p. 2).
- [19] ElSayed, N. A., Thomas, B. H., Marriott, K., Piantadosi, J., & Smith, R. T. (2016). Situated analytics: Demonstrating immersive analytical tools with augmented reality. *Journal of Visual Languages & Computing*, 36, 13-23.
- [20] Carmona, Kim, Erin Finley, and Meng Li. "The Relationship Between User Experience and Machine Learning." Available at SSRN 3173932 (2018).
- [21] Hartson, R., & Pyla, P. S. (2012). *The UX Book: Process and guidelines for ensuring a quality user experience*. Elsevier.
- [22] Gehrmann, S., Strobelt, H., Krüger, R., Pfister, H., & Rush, A. M. (2019). Visual interaction with deep learning models through collaborative semantic inference. *IEEE transactions on visualization and computer graphics*, 26(1), 884-894
- [23] Wang, S., Zargar, S. A., & Yuan, F. G. (2020). Augmented reality for enhanced visual inspection through knowledge-based deep learning. *Structural Health Monitoring*, 1475921720976986.
- [24] Hamid, Oussama H., Norris Lee Smith, and Amin Barzanji. "Automation, per se, is not job elimination: How artificial intelligence forwards cooperative human-machine coexistence." 2017 IEEE 15th International Conference on Industrial Informatics (INDIN). IEEE, 2017.

AUTHORS

Büşra Öztürk received her BSc degree in Computer Engineering from the University of Gazi, Turkey, in 2020. She started her master's degree in 2020 at Gebze Technical University, Institute of Science, Department of Computer Engineering. Currently, she is employed as software engineer in HAVELSAN Inc, Naval Combat Management Technologies Center. Her research interests include Mixed Reality, data science and user experience(UX).



Yakup Genc received his PhD in Computer Science from the University of Illinois at Urbana-Champaign. Right after graduation, Dr. Genc joined Siemens Corporate Research (SCR) in September 1999. As research scientist, project manager, program manager and group manager, he developed technology and research strategy in the areas of computer vision, augmented reality and machine learning. His tenure at SCR produced numerous publications and patents. Since September 2012, as a member of the faculty of the Computer Engineering department at the Gebze Technical University, he continues to conduct research in fields of computer vision, augmented reality, autonomous vehicles, machine learning and deep learning while maintaining close ties with the industry for practical applications of his research.



INTRUSION DETECTION IN A STAND-ALONE 5G NETWORK USING MACHINE LEARNING EVALUATION

Hafiz Bilal Ahmad¹ Haichang Gao², and Fawwad Hassan Jaskani³

^{1,2}Department of Computer Science and Technology, Xidian University, Xi'an,
China

³Department of Computer Engineering, Islamia University, Bahawalpur,
Pakistan

ABSTRACT

In order to meet the specific requirements of various industries and the stringent demands of 5G, the control and management of 5G networks will heavily depend on the integration of Software Defined Networking, Network Function Virtualization, and Machine Learning. Machine learning can play a crucial role in addressing challenges such as slice type prediction, route optimization, and resource management. To effectively evaluate the use of machine learning in 5G networks, a suitable testing environment is necessary. This study proposes a lightweight testbed that leverages container virtualization technologies to support the development of machine learning network functions within 5G networks. The Deep Slice 5G dataset from Kaggle was utilized to predict the type of communication between users based on packet loss and delay budget ratio, with the goal of making 5G systems more efficient. To accomplish this, we applied several Boosted Machine Learning models such as XGBoost, Gradient Boost, AdaBoost, LightGradientBoosting, CatBoost, and HistGradientBoosting. After evaluation, the Catboost model demonstrated the highest accuracy of 99% in identifying the correct slice of 5G based on the selected features of the dataset.

KEYWORDS

5G Network, Machine Learning, Intrusion Detection System, Slice Type Prediction.

1. INTRODUCTION

The Internet of Things (IoT) and mission-critical communications applications are only two examples of the kinds of use cases that will place new demands on the capabilities of 5G networks [1]. Efficient, intelligent, and agile network administration is essential in the face of the increased problems posed by these needs and use cases. In addition, 5G will foster an environment that encourages the development of cutting-edge software in a wide variety of sectors, including the industrial, medical, media, financial, public safety, transportation, agricultural, dietary, and municipal sectors [2]. Latency, throughput, availability, dependability, coverage, mobility, and so on are only some of the metrics that must be met [3]. 5G will offer a versatile network that can meet these wide-ranging needs. More software, virtualization, and automation in the network is required for greater portability [4]. Software Defined Networking (SDN) and Network Function Virtualization (NFV) are often viewed as the realization of the softwarization notion and the virtualization paradigm, respectively, from a networking perspective [2].

Network slicing is a crucial part of allowing network flexibility because it allows us to build specialized logical networks on top of a shared physical infrastructure to more effectively meet the unique requirements of each individual business. The network functions and supporting infrastructure that make up a network slice. Network slices can be implemented with the help of SDN and NFV because of their programmability, flexibility, and modularity [5].

Software Defined Networks (SDN) and Network Function Virtualization (NFV) give the network more adaptability and configurability by enabling network services to run in software rather than being hardwired into the system [6]. With this flexibility, network operations can be moved, upgraded, and installed at any node. However, manual provisioning, maintenance, and control of network slices is impractical due to the dynamic behaviour of network operations [7]. Due to the ever-changing nature of the environment, network analytics and constant monitoring are now necessities for gaining insight into how networks function. In a similar vein, the provision of automation capabilities to the network is crucial for network operation and management. By eliminating the potential for human mistake and accelerating the time it takes to bring a service to market, automation in the network helps to keep operational costs down [8]. In addition, when Machine Learning (ML) is applied to network analytics, the network gains the ability to learn and make decisions for itself. In order to govern and maintain networks autonomously and provide services, ML approaches can extract useful information from the data collected by the networks [9]–[11]. Allocating the necessary amount of network resources without overprovisioning, ML methods may predict network behavior based on historical and real-time data and adapt to the changing network conditions [12], [13]. ML can also be used to optimize for energy efficiency. To save money on energy expenditures, it may be able to turn off unused components or move services to areas with lower demand. ML has the potential to be successfully employed in automatically orchestrating and managing networks, which would pave the way for self-organizing networks. To rephrase, ML serves as a critical enabler of automation and helps solve the issue of delivering network intelligence. In this light, SDN, NFV, and ML all play important roles in facilitating the deployment of 5G networks [14].

As a result, organizations like the 3rd Generation Partnership Project (3GPP), European Telecommunications Standards Institute (ETSI), and International Telecommunication Union Telecommunication Standardization Sector are all working to ensure that AI and ML are properly represented in 5G and B5G mobile networks (ITU-T). For instance, the ITU-Focus T's Group on Machine Learning for Future Networks, Including 5G (FG-ML5G) designed an ML architecture to be incorporated into future mobile networks, standardizing the terminology used to describe ML-related mechanisms while maintaining compatibility with existing network infrastructure [15]. As has been widely explored in the context of other communications paradigms, such as mesh networks [2]–[4], testbeds provide a viable alternative to simulators for evaluating and integrating the newly created AI algorithms. Since testbeds offer more realistic simulation settings, and solutions built on testbeds have a quicker path from the research stage to products, they are crucial for creating and evaluating network technologies [16]. Due to the complex nature of 5G and B5G networks, simulations often fail to capture key details. This makes testbeds increasingly crucial.

This article details the machine learning-based testbed architecture necessary to rapidly establish realistic 5G scenarios with varying degrees of network slicing based on predetermined budgets for packet loss and latency. The main objectives of this study are:

We have collected the Deep Slice 5G dataset from Kaggle in order to make 5G systems lighter by predicting the sort of communication that will occur between users based on the ratio of packet loss to delay budget. This will allow us to reduce the amount of time that data will be lost.

Boosted models of machine learning, such as XGBoost, Gradient Boosting, AdaBoost, LightGradientBoosting, CatBoost, and HistGradientBoosting, have been utilized on our end for the purpose of making predictions regarding the slices.

2. METHODOLOGY

In this study we have collected the Deep Slice 5G dataset from Kaggle in order to make 5G systems more lightweight. This was done in order to forecast the sort of communication that takes place between users based on the amount of packet loss and the delay budget ratio. We have employed boosted models of machine learning, such as XGBoost, Gradient Boost, AdaBoost, LightGradientBoosting, CatBoost, and HistGradientBoosting, for the prediction of slices.

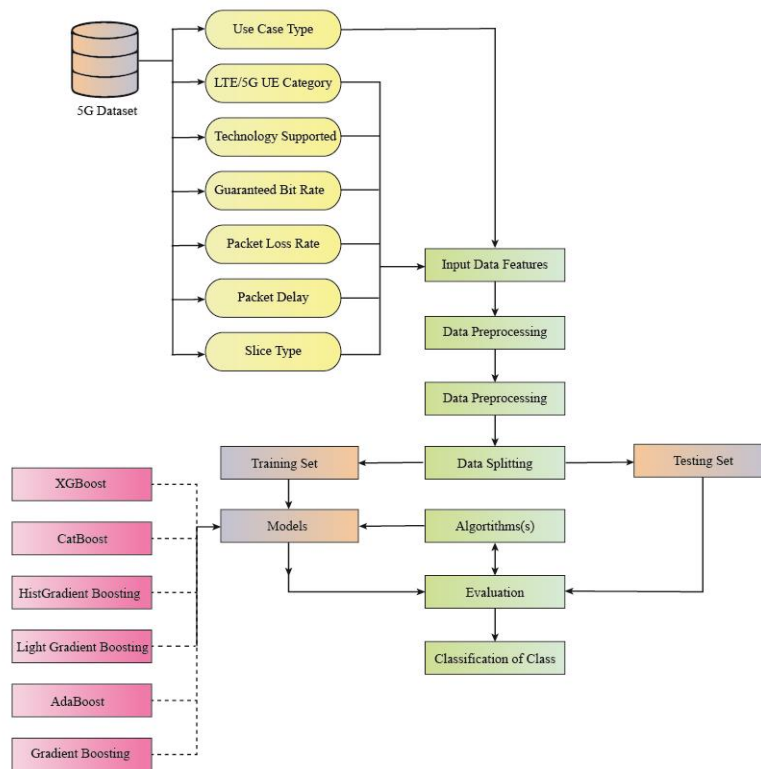


Figure 1. Proposed Model Structure

2.1. Dataset Description

The dataset contains number of features with 6 inputs and 1 output variable. Each feature contains multiple scenarios. List of input and output features has been given in Table 1:

Table 1. Units for Magnetic Properties

Feature	Description	Division	Type
Use Case Type	Real-World 5G Applications: With high capacity and ultra-low latency, 5G will drive AI and IoT applications	Gaming, Healthcare, Industry, IoT Devices, Public Safety, Smart	Input

	across industries and use cases.	City, Home Smart Transportation Smartphone	
LTE/5G UE Category	Existing equipment can be confidently transferred from old 2G/GSM and 3G/UMTS systems to 4G/LTE networks using the LTE Cat M or LTE Cat NB-IoT standards. 5G coverage is currently confined to high-user locations.	1: LTE 2: 5G	Input
Technology Supported	5G uses OFDM to modulate a digital signal across many channels to reduce interference. 5G NR air interface leverages OFDM concepts.	1: LTE 2: 5G 3: IoT	Input
GBR	GBR services with minimum GBR requirements and non-GBR services are supplied in enhanced mobile Broadband usage scenarios.	0: Non GBR 1: GBR	Input
Packet Loss Rate	The packet loss ratio is the percentage of sent packets lost.	0.001 to 0.00001	Input
Packet Delay Budget	The Packet Delay Budget (PDB) limits packet delay between the UE and the N6 termination point at the UPF.	50ms to 1000ms	Input

2.2. Classification Models

We have collected the Deep Slice 5G dataset from Kaggle in order to make 5G systems lighter by predicting the sort of communication that will occur between users based on the ratio of packet loss to delay budget. This will allow us to reduce the amount of time that data will be lost. Boosted models of machine learning, such as XG Boost, Gradient Boost, AdaBoost, Light Gradient Boosting, Cat Boost, and Hist Gradient Boosting, have been utilised on our end for the purpose of making predictions regarding the slices. As a consequence of this, the Cat boosting Model has demonstrated the maximum accuracy, which is 99.8%, in determining the appropriate slice of 5G on the basis of specified aspects of the dataset.

2.2.1. XG Boost

Extreme Gradient Boosting (XG Boost) is a machine learning toolkit that provides a scalable implementation of a gradient-boosted decision tree (GBDT). It is the most popular machine learning library, and it offers parallel tree boosting, which can be applied to issues of regression, classification, and ranking. Let's approximate a function $f(x)$ with the easiest linear approximation we can calculate as:

$$f(x) = f(a) + f'(a)(x - a) \dots (1)$$

X can be considered as change in a prediction function when we have multiple outputs:

$$f(\Delta x) = f_t x_i \dots (2)$$

In this scenario, the loss function l is denoted as $f(x)$, the anticipated value from the previous iteration $(t - 1)$ and x is the new learner to be introduced in iteration t . To optimize in Euclidean space, we can use the above at each iteration t to define the objective (loss) function as a simple function of the newly added learner. To recap, in step (t) , the prediction from step $(t - 1)$ is a ,

and in step (t) , the new learner we need to add in order to greedily reduce the goal is $(x - a)$. In this case, if we adopt the Taylor approximation of the second order, we obtain:

$$f(\Delta x) = \sum f(a) + f'(a)(x - a) + \left[\frac{1}{2} g \right] (x - a)^2 + f''(a)(x - a)^2 \dots (3)$$

Where g is the second order gradient used for multi classification.

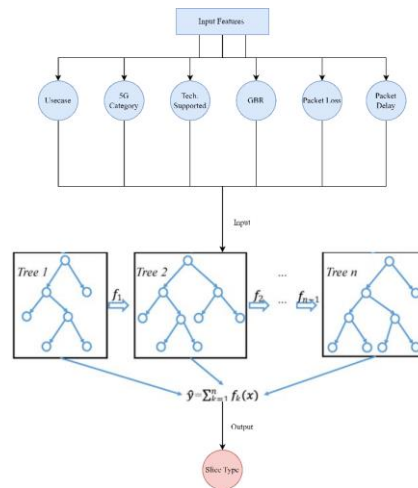


Figure 2. XGBoost Architecture

2.2.2. Cat Boost

Cat Boost can automatically manage features that can be categorized. Despite its popularity, one-hot encoding can no longer be used when there are too many features in the data set. To address this issue, features are classified into groups based on the statistics used to evaluate success (estimate target value for each category). The desired estimate for the i th categorical variable in the k th element of D can be written in mathematical notation as,

In order to reduce prediction time, CatBoost uses oblivious decision trees, which are binary trees in which the same features are utilized to create left and right splits for each level of the tree. It efficiently deals with categorical features by using sorted target statistics.

In the first step we will initialize the model,

$$F_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, \gamma) \dots (4)$$

For $m = 1$ to M , we will compute the residuals.

$$\gamma_{im} = - \left[\frac{\partial L[y_i, F(x_i)]}{\partial F x_i} \right]_{F(x) = F_{m-1}(x)} \dots (5)$$

Then we will fit the base learner to compute it with pseudo residuals:

$$y_{im} = \underset{y}{\operatorname{argmin}} \sum_{x^i}^n L(y, F_{M-1}(x)) \dots (6)$$

Updated Model will be:

$$y = F_m(x) = F_{M-1}(x) + \alpha \sum_{i=1}^n y_{im} \dots (7)$$

The architecture of CBC Model has been shown in Figure 3:

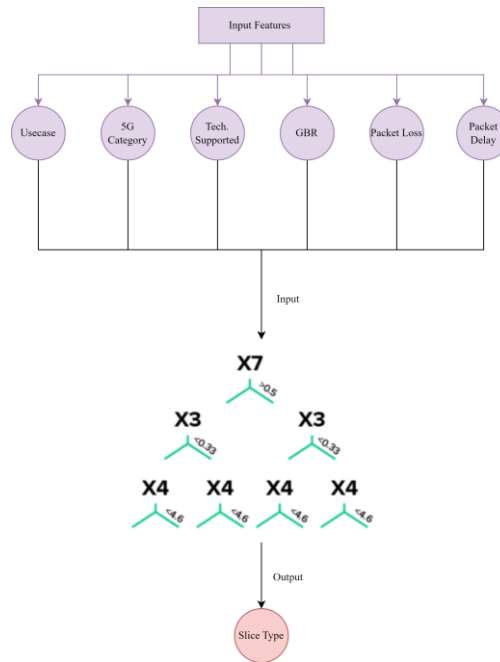


Figure 3. CatBoost Architecture

2.2.3. Gradient Boosting

Specifically, Gradient Boosting is an iterative functional gradient algorithm, which minimizes a loss function by repeatedly selecting a function that tends toward the negative gradient, or a hypothesis with low statistical support. Mathematical model of GBC Classification model is as follows:

$$y = y^i = y^i + \alpha * \frac{\partial \sum (y_i - y_i^p)^2}{\partial y_p^i} \dots (8)$$

The architecture of GBC Model has been shown in Figure 4.

2.2.4. AdaBoost

AdaBoost, which stands for "Adaptive Boosting," is an algorithm that employs the Boosting technique as part of an Ensemble Method for machine learning. Adaptive boosting gets its name from the fact that it re-assigns weights to each instance, giving more weight to examples that were mistakenly classified. This model has been developed by ensembling trees model into

AdaBoost Classifier to improve accuracy. Mathematical model of ABC Classification model is as follows

$$y = \text{significance} \sum_{t=1}^T \alpha_t h_t(x) \dots (9)$$

The architecture of ABC Model has been shown in Figure 5.

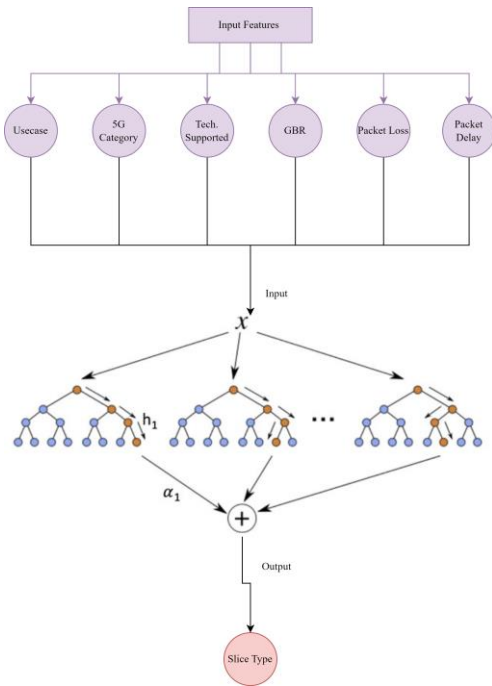


Figure 4. Gradient Boosting Architecture

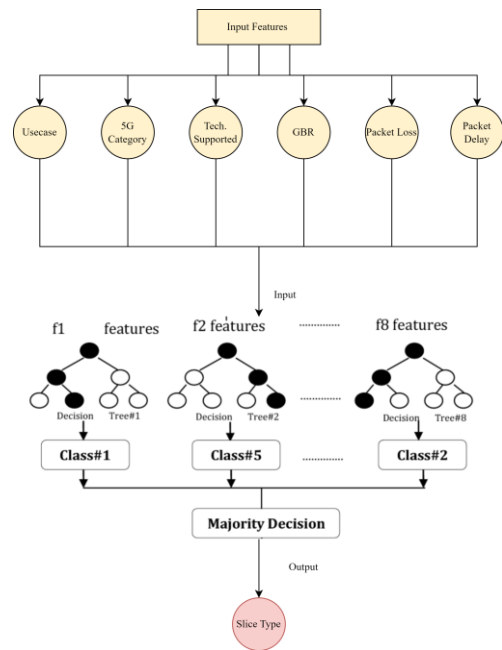


Figure 5. Adaboosting Architecture

2.2.5. Hist Gradient Boosting

A histogram can be used to visualize or tally the frequency of data (number of occurrences) over intervals (bins). The histogram approach is essentially straightforward, with each bin representing the frequency of the corresponding value. Mathematical model of HGBC Classification model is as follows

$$y = \frac{\text{sum of residuals}}{\text{sum of each } (1 - p) \text{ for each sample in the leaf}} \dots (10)$$

The architecture of HGBC Model has been shown in Figure 6.

2.2.6. Light Gradient Boosting

In contrast to traditional boosting algorithms, LightGBM divides the tree at each leaf level as it expands. In order to maximize its delta loss, it selects the leaf that is the most advantageous for growth. The loss of the leaf-wise approach is less than that of the level-wise algorithm because the leaf is always known. Mathematical model of LGBM Classification model is as follows

$$y = \alpha \sum_{t_i \in Tree} \eta^i * leaf(t_i) \dots (11)$$

The architecture of LGBC Model has been shown in Figure 7.

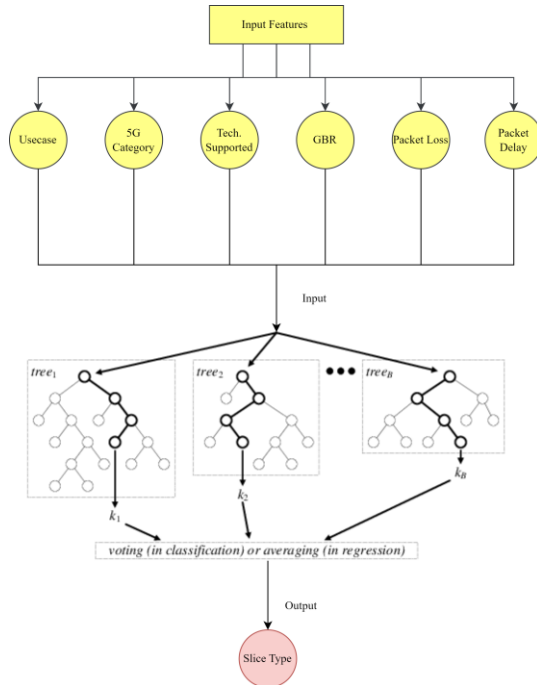


Figure 6. HGBC Architecture

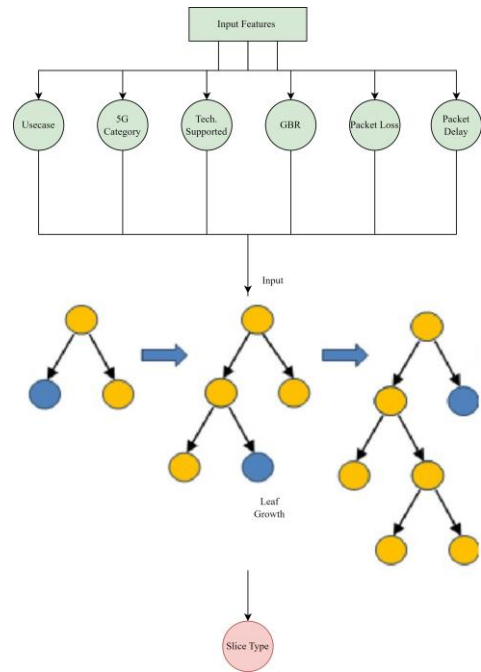


Figure 7. Light Gradient Boosting Architecture

2.3. Performance Metrics

2.3.1. TPR (True Positive Rate):

True positives divided by total positives and false negatives, or $TP/TP+FN$, is a method for determining a test's sensitivity. Probability of a positive test is known as the TPR (Test Prevalence Ratio). For example, a test's true negative rate (also known as specificity) is the number of times a patient who is genuinely negative will be incorrectly identified as a negative by the test's results.

$$TPR = \frac{TP}{TP + FN}$$

2.3.2. False Positive Rate

FPR, commonly known as the false positive rate, is a measure of how reliable a test is and how many false positives it has. Medical diagnostic tests, machine learning models, or any other kind of test can be used for this. The term "false positive rate" in statistics refers to the probability of rejecting the null hypothesis wrongly.

$$FPR = 1 - spcificity = \frac{FP}{TN + FP}$$

2.3.3. Accuracy

All projected data points are counted to determine Accuracy. True positives and true negatives divided by the total number of true positives, true negatives, false positives, and false negatives is a more formal definition of it. It is calculated as:

$$Accuracy = \frac{TP}{TP + FP}$$

2.3.4. Macro Average/ Weighted Average

Basically, it's just the average of the scores from each class. A macro-average recall is the sum of individual class recollections, A, B, and C combined. Weighted average is the sum of combined classes as A, B and C.

2.3.5. Evaluation of ROC and AUC

A good model has an AUC that is relatively close to 1, which indicates that it has a good measure of separability. A bad model will have an AUC that is close to 0, which indicates that it has a weak measure of separability.

3. RESULT AND DISCUSSION

3.1. XG Boost (XGB)

Extreme Gradient Boosting, often known as XG Boost, is a machine learning toolkit that offers a gradient-boosted decision tree that may be implemented in a scalable manner (GBDT). It is the most popular machine learning library, and it includes both sequence and parallel tree boosting, which may be applied to challenges of regression, classification, and ranking. It can be seen from Figure 8 that XGB has shown the highest accuracy for class 4 as 93% with a macro average of 77% and micro average of 88%.

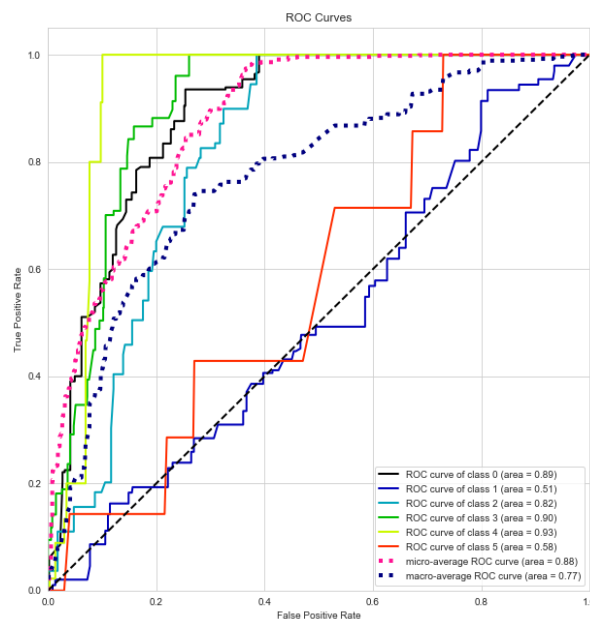


Figure 8. ROC of XGB

3.2. Cat Boost (CBC)

Cat Boost has the ability to manage in an automated fashion features that can be categorized. In spite of its widespread use, one-hot encoding cannot be applied when there are an excessive number of characteristics contained inside the data set. In order to solve this problem, features are divided into categories according to the data that are used to assess their level of success (estimate target value for each category). It can be seen from Figure 9 that CBC has shown 99% accuracy for class 0 and 1 while 100% accuracy for class 2, 3 and 4. CBC has shown the highest accuracy with micro average of 99% and macro average of 90%

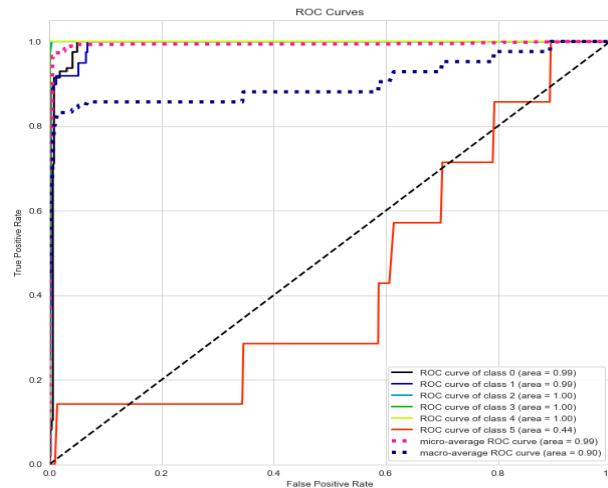


Figure 9. ROC of CBC

3.3. Gradient Boosting (GBC)

It is an iterative functional gradient algorithm that minimizes a loss function by continually selecting a function that tends toward the negative gradient or a hypothesis with poor statistical support. Gradient boosting is an example of this type of technique. GBC has shown highest 92% accuracy for class 3 and 4 with micro average 93% and macro average of 79% as shown in Figure 10.

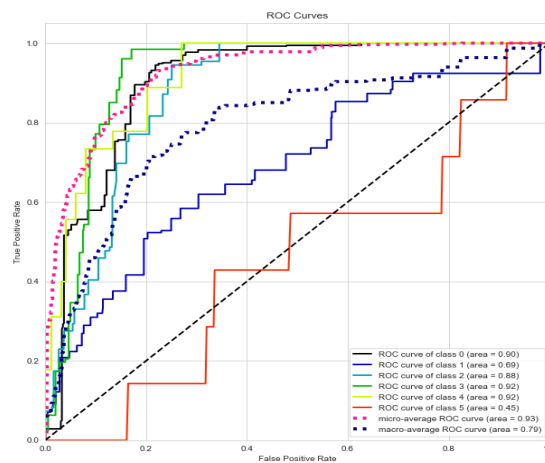


Figure 10. ROC of GBC

3.4. Ada boosting (ABC)

It uses the Boosting method as a component of an Ensemble Method for the purpose of machine learning. Adaptive boosting gets its name from the fact that it reassigns weights to each instance, providing examples that were incorrectly categorized a greater amount of weight than others in the process. It can be seen from Figure 11 that ABC model has shown highest accuracy of 91% for class 4 with micro average of 88% and macro average of 78%.

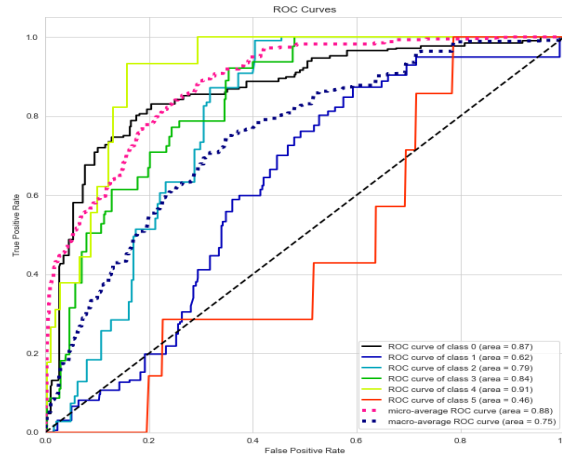


Figure 11. ROC of ABC

3.5. Light Gradient Boosting (LGBC)

LGBC is an alternative to more conventional boosting algorithms since it divides the tree at each leaf level as it grows larger. It chooses the leaf that offers the most potential for development in order to achieve the greatest possible increase in its delta loss. Because the leaf is always known, the loss that occurs when using the leaf-wise approach is far lower than when using the level-wise algorithm. It can be seen from figure 12 that LGBC has shown highest accuracy of 92% for class 2 and 3 respectively, with a micro average of 93% and macro average of 79%.

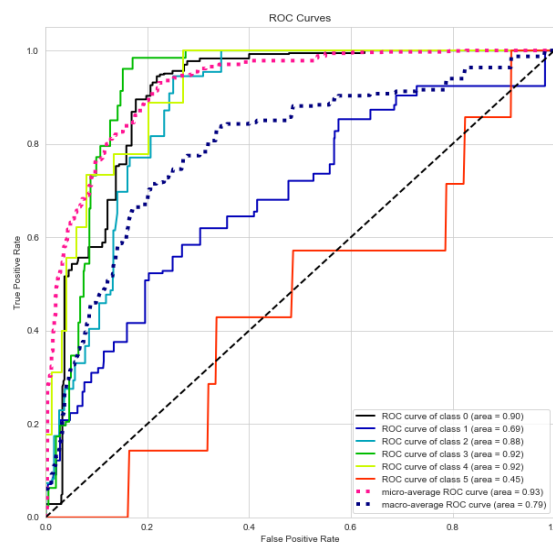


Figure 12. ROC of LGBC

3.6. HistGradient Boosting (HGBC)

The frequency of data (the number of occurrences) over intervals can be seen with the help of a histogram, which can also be used to tabulate the data (bins). The histogram method is rather easy to understand, as each box in the graph displays the frequency of the value to which it corresponds. It can be seen from Figure 13 that HGBC has shown highest accuracy of 91 for Class 4 with a micro average of 88% and macro average of 75%.

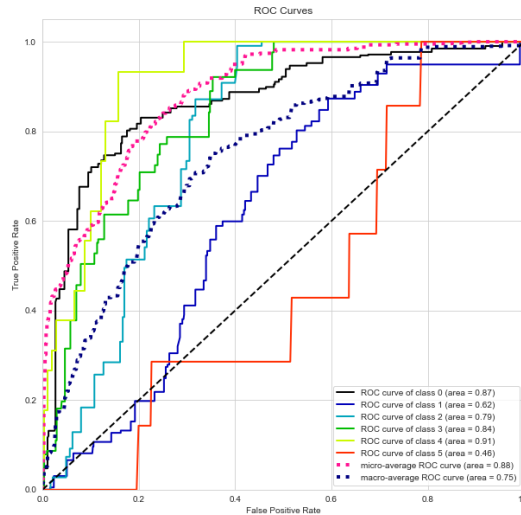


Figure13. ROC of HGBC

3.7. Comparative Analysis

CBC model has shown the highest accuracy for the classification of Deep Slices in 5G Network for Standalone Testbeds. CBC has shown 99% accuracy for class 0 and 1 while 100% accuracy for class 2, 3 and 4. CBC has shown the highest accuracy with micro average of 99% and macro average of 90%. While, LGBC has shown highest accuracy of 92% for class 2 and 3 respectively, with a micro average of 93% and macro average of 79%. On the other hand, HGBC has shown highest accuracy of 91 for Class 4 with a micro average of 88% and macro average of 75%. ABC model has shown highest accuracy of 91% for class 4 with micro average of 88% and macro average of 78%, and XGB has shown the highest accuracy for class 4 as 93% with a macro average of 77% and micro average of 88% while GBC has shown highest 92% accuracy for class 3 and 4 with micro average 93% and macro average of 79%.

4. CONCLUSIONS

Fifth-generation (5G) networks will rely largely on the implementation of Software Defined Networking, Network Function Virtualization, and Machine Learning in order to fulfil the particular needs of vertical industries and the stringent standards of 5G. Slice type forecasting, route optimization, and resource management are all areas where 5G networks could benefit from the application of machine learning. To evaluate machine learning's role in 5G networks, ideal test conditions are required. Using container lightweight virtualization technologies, this study proposes a lightweight testbed to aid in the development of machine learning network functionalities across the 5G network. To reduce the burden on 5G networks, we have collected the Deep Slice 5G dataset from Kaggle and are using it to create predictions about the nature of user-user communication based on packet loss and delay budget ratio. Machine Learning Boosted

models (XG Boost, Gradient Boost, Ada Boost, Light Gradient Boosting, Cat Boost, and Hist Gradient Boosting) were employed for slice prediction. Therefore, while selecting the proper 5G slice using a subset of the dataset's features, CBC model has shown the highest accuracy for the classification of Deep Slices in 5G Network for Standalone Testbeds. CBC has shown 99% accuracy for class 0 and 1 while 100% accuracy for class 2, 3 and 4. CBC has shown the highest accuracy with micro average of 99% and macro average of 90%. While, LGBC has shown highest accuracy of 92% for class 2 and 3 respectively, with a micro average of 93% and macro average of 79%. On the other hand, HGBC has shown highest accuracy of 91 for Class 4 with a micro average of 88% and macro average of 75%. ABC model has shown highest accuracy of 91% for class 4 with micro average of 88% and macro average of 78%, and XGB has shown the highest accuracy for class 4 as 93% with a macro average of 77% and micro average of 88% while GBC has shown highest 92% accuracy for class 3 and 4 with micro average 93% and macro average of 79%. In future studies this study can be implemented on 6G and can be transformed to deep learning in real time slicing prediction. The limitations associated with the use of Boosted Machine Learning models, including computational complexity and the interpretability of results. We recognize that these factors play a crucial role in the practical implementation and deployment of machine learning models, and their consideration is vital for a comprehensive understanding of the proposed methodology. Future work should focus on mitigating computational complexities associated with Boosted Machine Learning models, exploring methods for performance optimization without compromising accuracy. Enhancing the interpretability of machine learning models is pivotal. Research efforts should be directed towards developing methodologies that provide insights into the decision-making processes of complex models.

ACKNOWLEDGEMENTS

The authors wish to thank the editors and anonymous reviewers for their valuable comments and helpful suggestions which greatly improved the paper's quality. This work was supported in part by the National Key R&D Program of China (SQ2023YFB3100028), in part by the Natural Science Foundation of China (61972306, 62302371), and in part by SongShan Laboratory (YYJC012022005).

REFERENCES

- [1] C. Ssengonzi, O. P. Kogeda, and T. O. Olwal, "A survey of deep reinforcement learning application in 5G and beyond network slicing and virtualization," *Array*, vol. 14, no. January, p. 100142, 2022, doi: 10.1016/j.array.2022.100142.
- [2] C.V. Nahum et al., "Testbed for 5G Connected Artificial Intelligence on Virtualized Networks," *IEEE Access*, vol. 8, no. M1, pp. 223202–223213, 2020, doi: 10.1109/ACCESS.2020.3043876.
- [3] G. P. Koudouridis, Q. He, and G. Dán, "An architecture and performance evaluation framework for artificial intelligence solutions in beyond 5G radio access networks," vol. 2022, no. 1. 2022.
- [4] W. Package and D. Level, "5G Mobile Network Architecture Testbed setup and 5G-MoNArch technologies demonstrated," no. 761445, 2016.
- [5] K. Saeedi, "Machine Learning for Ddos Detection in Packet Core Network for IoT," *Comput. Sci. Eng.*, 2019.
- [6] Y. Dai, D. Xu, S. Maharjan, G. Qiao, and Y. Zhang, "Artificial Intelligence Empowered Edge Computing and Caching for Internet of Vehicles," *IEEE Wirel. Commun.*, vol. 26, no. 3, pp. 12–18, 2019, doi: 10.1109/MWC.2019.1800411.
- [7] A. A. Al-habob and O. A. Dobre, "Mobile Edge Computing and Artificial Intelligence: A Mutually-Beneficial Relationship," 2020, [Online]. Available: <http://arxiv.org/abs/2005.03100>.
- [8] S. M. Kumar and D. Majumder, "Healthcare Solution based on Machine Learning Applications in IOT and Edge Computing," *Int. J. Pure Appl. Math.*, vol. 119, no. 16, pp. 1473–1484, 2018.

- [9] M. Ayaz, M. Ammad-Uddin, Z. Sharif, A. Mansour, and E. H. M. Aggoune, "Internet-of-Things (IoT)-based smart agriculture: Toward making the fields talk," *IEEE Access*, vol. 7, pp. 129551–129583, 2019, doi: 10.1109/ACCESS.2019.2932609.
- [10] Y. Tsai and D. Chang, "applied sciences Edge Computing Based on Federated Learning for Machine Monitoring," 2022.
- [11] F. Ali et al., "An intelligent healthcare monitoring framework using wearable sensors and social networking data," *Futur. Gener. Comput. Syst.*, vol. 114, pp. 23–43, 2020, doi: 10.1016/j.future.2020.07.047.
- [12] Y. Wang, H. Zen, M. F. M. Sabri, X. Wang, and L. C. Kho, "Towards Strengthening the Resilience of IoV Networks—A Trust Management Perspective," *Futur. Internet*, vol. 14, no. 7, pp. 1–21, 2022, doi: 10.3390/fi14070202.
- [13] C. K. Leung, Y. Chen, S. Shang, and D. Deng, "Big Data Science on COVID-19 Data," *Proc. - 2020 IEEE 14th Int. Conf. Big Data Sci. Eng. BigDataSE 2020*, pp. 14–21, 2020, doi: 10.1109/BigDataSE50710.2020.00010.
- [14] Y. Dong and Y. D. Yao, "IoT platform for covid-19 prevention and control: A survey," *IEEE Access*, vol. 9, pp. 49929–49941, 2021, doi: 10.1109/ACCESS.2021.3068276

AUTHORS

Hafiz Bilal Ahmad received the B.S. degree from Mirpur University of science and technology, Azad Kashmir, Pakistan in 2017 and M.S. degrees from North University of China, Taiyuan, China, in 2022. Currently He is pursuing Ph.D. degree in computer science and technology from Xidian University, Xi'an, China. His research interests include Network security, ML security, and privacy protection.



Haichang Gao (Member, IEEE) received the Ph.D. degree in computer science and technology from Xi'an Jiaotong University, Xi'an, Shaanxi, China, in 2006. He is currently a Professor with the School of Computer Science and Technology, Xidian University, Xi'an. He has published more than 75 papers. He is currently in charge of a project of the National Natural Science Foundation of China. His current research interests include Captcha, computer security, and machine learning.



Fawwad Hassan Jaskani has done his engineering degree from Islamia University of Bahawalpur then switched to MS Engineering from Islamia university of Bahawalpur. He has published more than 40 papers. Currently he is defending his thesis of PhD from UTHM Malaysia. His current research interests include Machine Learning, Neural Networks, image processing and Security.



EXPLORING DAG-BASED ARCHITECTURE AS AN ALTERNATIVE TO BLOCKCHAIN FOR CRITICAL IOT USE CASES

Ledesma, O.; Sánchez, M.A.; Lamo, P.

Escuela Superior de Ingeniería y Tecnología, Universidad Internacional de
la Rioja (UNIR), 26004 Logroño, Spain

ABSTRACT

This paper analyzes the Directed Acyclic Graph (DAG)-based architecture as an alternative to Blockchain technology for critical Internet of Things (IoT) use cases. The speed of transactions and the scalability of Blockchain technology are limitations for critical IoT applications such as vaccine cold chain monitoring. A pilot project has been developed to analyze the speed of the DAG architecture. It simulates monitoring the vaccine cold chain, recording temperatures and alarms. Using the same architecture, two cases with different IoT connectivity technologies in the pilot project are defined: LoRaWAN and Sigfox. The results of these two cases show the comparison between both technologies that show that the DAG architecture can provide the necessary time delays for critical IoT use cases. The main limitation found after the execution of the two cases of the pilot project is related to the need for worldwide coverage of the communications technologies used. For this reason, the study of communications through IoT satellites with global coverage is proposed as future work.

KEYWORDS

Distributed Ledger Technology, Blockchain, Directed Acyclic Graph, Internet of Things, IOTA

1. INTRODUCTION

Distributed Ledger Technology (DLT) is a decentralized distributed ledger technology that enables transactions and assets to be recorded, authenticated, and processed on a distributed ledger [1]. Unlike traditional distributed database architectures [2], DLTs are decentralized, trusted in untrusted environments, and cryptographically encrypted. The most well-known property of DLTs is their immutability, which means that all the information stored in them cannot be modified, deleted, or altered. DLTs also offer several different architectures, such as Blockchain and Directed Acyclic Graph (DAG) based architectures [3]. In this sense, although Blockchain is presented as a handy tool for implementing various examples in all areas of Industry 4.0 and the Industrial Internet of Things (IoT) in general, it does not always guarantee the best solution to cover the needs and requirements of the use cases.

One of the main requirements for critical IoT use cases is the speed of transactions, both for recording data and alarms in real-time, as well as for micropayments in cryptocurrencies between machines [4]. Likewise, registering sensor data in a DLT must be done immediately so users can

have this information as soon as possible. However, two of the Blockchain's limitations are the speed of transactions and the low scalability [5]. For this reason, this article analyzes the architecture based on DAG [6] as an alternative in IoT that guarantees the speed of transactions in cases where this is critical for correct operation.

A pilot has been developed to analyze the speed of DAG-based DLT architectures. It simulates monitoring the vaccine cold chain and recording temperatures and alarms in a DAG. Given this critical nature, the information and alarms must be recorded practically in real-time. The users (in this case, hospitals, pharmaceuticals, or vaccination centers) can access temperature and alarm data when they occur, thus acting as soon as possible to avoid breaking the cold chain and discarding the monitored vaccine batch.

2. CRITICAL CASE: COLD CHAIN OF VACCINES

Historically, the control and monitoring of the cold chain have been carried out manually without significant changes. For this, labels with chemical reagents or thermometers have been used that recorded the break in the cold chain, but their reading or verification was always carried out manually [7]. Therefore, in most cases, it was impossible to verify whether the cold chain had been broken until the vaccines arrived at the health center, without having the possibility of checking the temperature recorded throughout the entire chain. Likewise, the current vaccine cold chain verification system, which is not connected or sends temperature or cold chain break data in real-time, and does not record these data in DLTs, entails a high percentage of batches of vaccines in refrigerators that are lost due to break in the cold chain during transport [8].

Control of the cold chain of vaccines requires knowing and having visibility of the temperature in real time [9], as well as alarms that indicate an excess of temperature or a break in the cold chain of vaccines at the instant they occur. In order to corroborate the correct operation of the DAG and examine the registration times in the proposed DLT, the experimental test of two environments that present different IoT platforms and Low-Power Wide-Area Network (LPWAN) communications has been carried out, such as LoRaWAN and Sigfox.

An experimental test was used as the research method. The test aims to analyze the speed of DAG-based DLT architectures in monitoring the vaccine cold chain. The experimental test simulates the recording of temperatures and alarms in a DAG, with two different IoT connectivity technologies: LoRaWAN and Sigfox. The test results compare both technologies to demonstrate if the DAG architecture can provide the necessary time delays for critical IoT use cases.

2.1. Proposed architecture

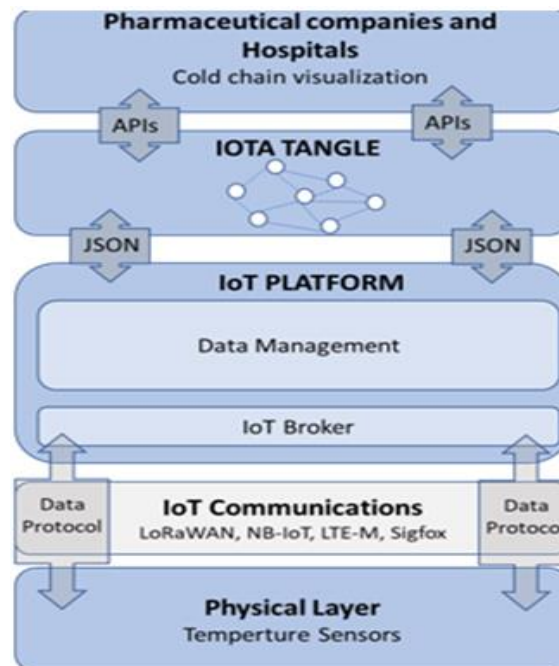


Figure. 1. Architecture of the pilot project used For Lorawan And Sigfox Cases

The architecture proposed to develop the pilot project is presented in Figure 1. It is used for the two cases with different connectivity technologies: LoRaWAN and Sigfox. Each of the defined layers fulfils a specific function in the process of collecting, processing, and displaying temperature data and alarms from the installed sensors:

- **Physical Layer:** Houses the IoT sensors that measure the temperature inside refrigerators or cooling boxes. These devices, characterized by low consumption and ARM architecture, use LPWAN connectivity to transmit the temperature to the IoT platform periodically. The pilot project will use temperature sensors measuring between -80 °C and +70 °C and LoRaWAN and Sigfox communications from the manufacturer SLB System. Also, QR codes are included that are placed on the fridge and on the sensor itself, and that will be redirected by reading them to the web or address of the DAG viewer that shows the recorded data.

- **Iot Communications:** This layer is responsible for transporting the data from the sensors to the IoT platform. Depending on the IoT device and the communication technology, different data transport protocols exchange information between the IoT platform and the sensors.

- **Iot Platform:** On it, the ingestion and processing of the data from the sensors installed in the vaccine refrigerators is carried out. When the platform receives the temperature data from the sensors, the information is automatically recorded in the DAG to know the temperature recording times or the alarms in said DLT.

- **Dag-Iota Tangle:** It is the DLT where the temperature sensors and alarm information are stored. It allows open access to pharmaceutical companies, hospitals, or health centers to view the real-time status of vaccines or the cold chain.

- **Visualization Applications:** Layer for the software tools that allow the visualization of the temperature data and alarms registered in the IOTA Tangle.

2.2. Case 1: Lorawan

This first case uses SLB Systems sensors and a private LoRaWAN network at 868 MHz with a LoRaWAN gateway from the manufacturer Multitech model Conduit AP. From this gateway, an IPsec VPN tunnel is established to an instance of Microsoft Azure where the Chirpsstack server is located, in charge of managing and administering the private LoRaWAN network. IoT devices and LoRaWAN gateways are enlisted from this server, and the initial ingestion of temperature data sent by the LoRaWAN gateway and its processing and storage is done on the Thingsboard IoT platform. Instead of storing the data internally, they are uploaded through a Rest API to the DAG to the corresponding address where the temperature data of each sensor is saved as a transaction in the DLT with value 0 and the "Message" field of the transaction to save the value of the temperature measured by the sensors. Once the information is stored, it can be accessed immediately by hospitals, pharmaceutical companies, or patients.

2.3. Case 2: Sigfox

In this second case of the pilot project, Sigfox connectivity technology and the IoT platform of SLB IoT Site, hosted in a public cloud in Germany, have been used. The same SLB Systems temperature sensor model has been used, but with Sigfox connectivity, which makes it possible to use an operator network and not have to implement a private network, contracting only connectivity with the Sigfox operator. The sensor periodically transmits the temperature of the vaccines or refrigerators through the Sigfox connectivity to the nearest base station of the Sigfox operator. This data is immediately available on the Sigfox backend, which integrates with the IoT platform of the SLB IoT Site to collect the temperature data. Instead of storing the data in an internal database, through an API, a transaction is performed with value 0 and with the temperature value in the message field to the DAG, to save the temperature data in the corresponding address for the sensor. Similarly to the previous case, once stored, different users (for example, hospitals, pharmaceutical companies, or patients) will have immediate access to the information stored in the DAG. This approach allows effective traceability of vaccines and guarantees their integrity during transport and storage, which is critical for the protection of public health.

2.4. Administrative Management

Regardless of the architecture and technology used, a public or private organization or administration is responsible for coordinating the implementation of a cold chain monitoring solution for a health system at a regional, national, or international level. This coordination must include the definition of the technical requirements of the IoT temperature sensors, the connectivity protocols allowed for communication with the IoT platform, and the information exchange protocols between the IoT platform and the sensors. In addition, it is necessary to generate addresses in the DAG to store the information, register the IoT sensors in the IoT platform, and associate them with an address in the DAG. Also, the methods must be defined to access the cold chain information stored in the DAG, manage and centralize all the information received from the IoT sensors, and guarantee the operation of the solution from the moment the information is received from the sensors until it is stored in the DAG. Lastly, it is essential to coordinate with the logistics teams of the pharmaceutical companies and the healthcare teams in charge of receiving and maintaining the vaccines to ensure the efficacy and safety of the system. The initial sensitization of refrigerators and other vaccine refrigeration systems should fall to pharmaceutical companies, replacing the temperature control systems used by IoT temperature

sensors. They will be responsible for generating the QR code placed on the fridge and the sensor itself. The rest of the process would remain as currently used, with the sensor installed inside the refrigerator or refrigeration system. Vaccine transport and storage need not undergo any operational changes.

3. RESULTS

The mentioned cases of the pilot project (Figure 2) have been running for 15 days. During this time, messages from the sensors have been transmitted to the platform with 100% success in both cases. The response times, which represent the speed of the transaction and depend on the number of hops between the backend where the IoT platform of each of the pilots is hosted, have been analyzed in depth.

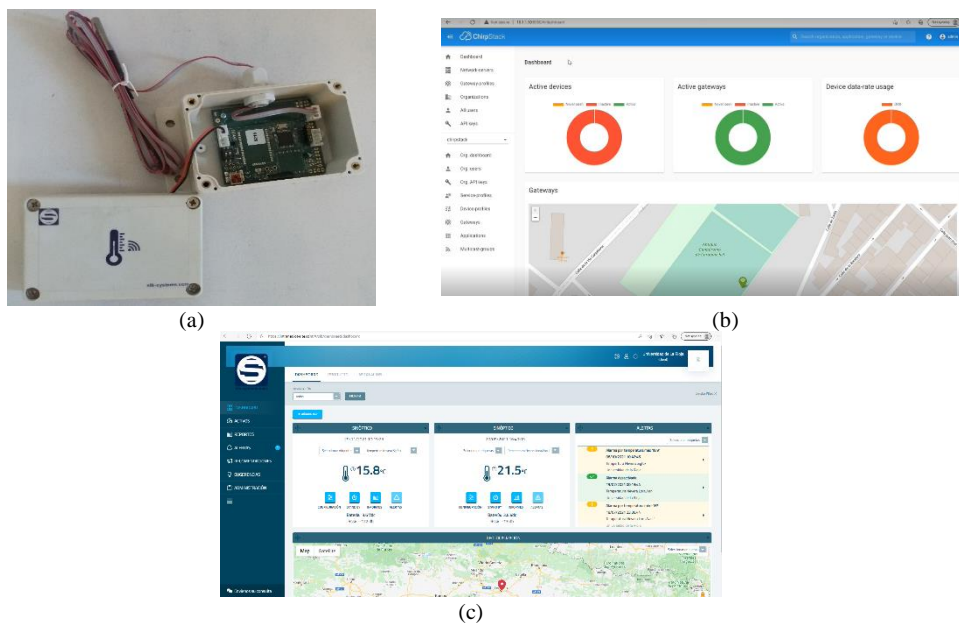


Figure 2. Experimental setup: (a) Sensorization, (b) Dashboard of Chipstack Network Server from Thingsboard and (c) IoT Site platform from SLB Systems developed

Table 1 shows an exhaustive statistical analysis of timed delay in two communication protocols: Sigfox and LoRaWan. These data result from time measurements in seconds obtained in various observations. In the case of the Sigfox protocol, a minimum value of 1 second is observed, while the maximum reaches 5 seconds. This generates a 4 s time range where the events unfold. In terms of mode, the most frequent value is identified as 1 second, which reveals a clear trend in the data. With the measures of central tendency, it is determined that the mean for the Sigfox protocol is 1.76 s. This figure allows to understand the average value of the delay times. On the contrary, the median, representing the midpoint in an ordered data set, is at 2 s. This value gives a more precise view of the central location of the data. To evaluate the dispersion of the data, the variance, and the standard deviation are calculated.

The variance obtained for the Sigfox protocol is 1.0233 s squared. This value shows how the data moves away from the mean and gives us an idea of the present variability. Likewise, the standard deviation, which is the square root of the variance, is calculated to be 1.0116 s. This value indicates how much the data is spread out about the mean. In addition, the asymmetry of the data distribution is analyzed. In the case of the Sigfox protocol, an asymmetry value of 1.7265 is obtained. This figure reveals a skewed distribution to the right, meaning that high values can

affect the mean and move it away from the central tendency. This information, collected visually, is also presented in Figure 3.

Table 1. Statistical analysis of timed delayed in the two cases of the pilot project.

Evaluated Statistics	Sigfox	Lorawan
Minimum	1	1
Maximum	5	8
Range	4	7
Mode	1	1
Mean	1.76	2.72
Median	2	2
Variance	1.0233	3.1267
Standrad deviation	1.0116	1.7682
Asymmetry	1.7265	1.2659

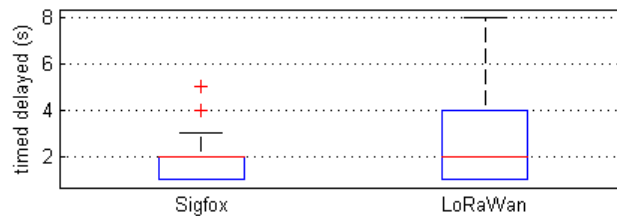


Figure 3. Probability distribution of the data recording time in the DLT using Sigfox and Lorawan

Although Blockchain is still the most widely used DLT today, it is important to highlight the existence of other architectures, such as IOTA, which has been specifically designed to address certain problems associated with the use of the IoT. Compared to Blockchain, the DAG architecture stands out for its high scalability since Blockchain blocks act as bottlenecks that can generate network congestion and delay transactions. DAG greatly mitigates this problem, allowing greater transaction processing efficiency [10]. Another advantage of DAG is its lower energy consumption in terms of computing resources, significantly reducing network maintenance costs [11]. In addition, in DAG, the payment of commissions per transaction is not required [12] since there are no miners involved in the validation of transactions, and they are processed with greater speed [13], which improves the ability of the network to record and manage large volumes of data. The results obtained in these pilots show that the DAG architecture, regardless of the communications protocol and the IoT platform used, can provide the necessary speed and scalability for critical IoT use cases.

This solution focuses on taking advantage of the capacity of satellites dedicated to IoT communications to establish broader and more reliable connectivity. Using IoT satellites seeks to guarantee global and more extensive coverage, thus allowing data exchange in areas where terrestrial connectivity is insufficient. To adapt the existing architecture to this emerging trend, a ground station would be required, which facilitates communication between IoT devices and satellites [14]. The IoT satellite architecture with IOTA Tangle is presented in Figure 4.

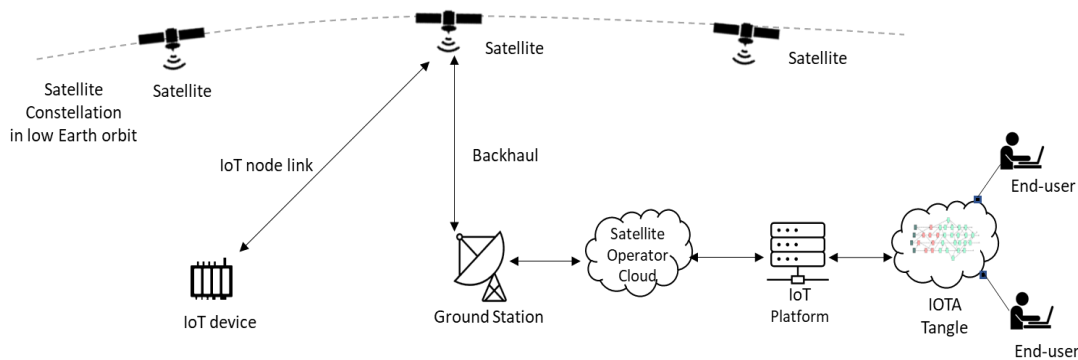


Figure 4. Iot Satellite Architecture With Iota Tangle

In vaccine supply chain monitoring with satellites, temperature data from IoT monitoring devices is transmitted through LPWAN communication channels to a satellite. The satellite serves as a relay transmitting the data to a ground station. Subsequently, the ground station forwards this information to an IoT platform for systematic processing. The processed data is then securely recorded on the IOTA Tangle, ensuring a secure and immutable storage mechanism.

4. CONCLUSIONS

The study focuses on the feasibility of monitoring the cold chain of vaccines through DLT technologies based on DAG architecture as an alternative to Blockchain to solve the lack of information in transport. The main improvements obtained in this critical case are remote and real-time monitoring of the temperature throughout the cold chain, secure and immutable storage of data, checking the status of the cold chain using QR codes, and detection of breaks in the cold chain during storage or transport. This solution will reduce the number of vaccine consignments spoiled by cold chain breaks, increase the general availability of vaccines for patients, and have centralized management and monitoring of the vaccine cold chain with reliable information stored in the DLT of the IOTA DAG. Furthermore, this solution can be used for similar use cases in healthcare and food environments. The work presents two cases of a pilot project based on the use of Sigfox and LoRaWAN and statistically evaluates the time delay that occurs due to the backend of the IoT platform used. Finally, it is worth mentioning that one limitation of the pilot project is the lack of global coverage in all cases, which hinders the continuous transmission of data over extended periods. As a suggestion for future work, the utilization of IoT satellite communication technologies is proposed to address this limitation.

ACKNOWLEDGMENTS

This research was funded by Universidad Internacional de La Rioja (UNIR), grant number JT-2022-04 (Project: “Arquitectura y sistema de comunicaciones para la conectividad con nanosatélites de baja órbita”). Furthermore, the authors would like to thank the Doctoral Office, EDUNIR and the UNIR Vice-Rector for Research for funding attendance at this conference.

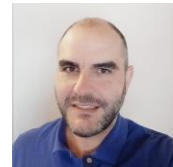
REFERENCES

- [1] Lima, C. (2018). Developing Open and Interoperable DLT/Blockchain Standards [standards]. *Computer*, 51(11), 106-111.
- [2] Shareef, A. A. A., Yannawar, P. L., Ahmed, Z. A., & Al-Madani, A. M. (2021, February). Applying Blockchain Technology to Secure Object Detection Data. In *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)* (pp. 874-879). IEEE.

- [3] Alshaikhli, M., Elfouly, T., Elharrouss, O., Mohamed, A., & Ottakath, N. (2021). Evolution of Internet of Things from blockchain to IOTA: A survey. *IEEE Access*, 10, 844-866.
- [4] Mercan, S., Kurt, A., Akkaya, K., & Erdin, E. (2021). Cryptocurrency solutions to enable micropayments in consumer IoT. *IEEE Consumer Electronics Magazine*, 11(2), 97-103.
- [5] Ensor, A., Schefer-Wenzl, S., & Miladinovic, I. (2018, December). Blockchains for IoT payments: A survey. In *2018 IEEE Globecom Workshops (GC Wkshps)* (pp. 1-6). IEEE.
- [6] Hellani, H., Sliman, L., Samhat, A. E., & Exposito, E. (2021, October). Tangle the blockchain: towards connecting blockchain and DAG. In *2021 IEEE 30th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)* (pp. 63-68). IEEE.
- [7] World Health Organization. (2015). How to monitor temperatures in the vaccine supply chain (No. WHO/IVB/15.04). World Health Organization.
- [8] Kitamura, T., Bouakhasith, V., Phounphenghack, K., Pathammavong, C., Xeuatvongsa, A., Norizuki, M., ... & Hachiya, M. (2018). Assessment of temperatures in the vaccine cold chain in two provinces in Lao People's Democratic Republic: a cross-sectional pilot study. *BMC Research Notes*, 11, 1-6.
- [9] Kartoglu, U., & Ames, H. (2022). Ensuring quality and integrity of vaccines throughout the cold chain: the role of temperature monitoring. *Expert Review of Vaccines*, 21(6), 799-810
- [10] Alshaikhli, M., Elfouly, T., Elharrouss, O., Mohamed, A., & Ottakath, N. (2021). Evolution of Internet of Things from blockchain to IOTA: A survey. *IEEE Access*, 10, 844-866.
- [11] Sherman, A. T., Javani, F., Zhang, H., & Golaszewski, E. (2019). On the origins and variations of blockchain technologies. *IEEE Security & Privacy*, 17(1), 72-77.
- [12] Conti, M., Kumar, G., Nerurkar, P., Saha, R., & Vigneri, L. (2022). A survey on security challenges and solutions in the IOTA. *Journal of Network and Computer Applications*, 203, 103383.
- [13] Aljahdali, A. O., Habibullah, A., & Aljohani, H. (2023). Efficient and Secure Access Control for IoT-based Environmental Monitoring. *Engineering, Technology & Applied Science Research*, 13(5), 11807-11815.
- [14] Capez, G. M., Henn, S., Fraire, J. A., & Garello, R. (2022). Sparse satellite constellation design for global and regional direct-to-satellite IoT services. *IEEE Transactions on Aerospace and Electronic Systems*, 58(5), 3786-3801

AUTHORS

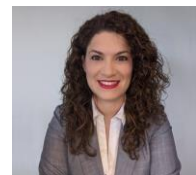
Óscar Ledesma García is a Computer Engineer with a Master's degree in Industry 4.0 and a Master's degree in Astrophysics. He works as a professor in the Master's program of Industry 4.0 at UNIR and is responsible for IoT infrastructures at Eviden. Oscar's research focuses on developing IoT solutions with Blockchain and infrastructure solutions for smart buildings and cities. He is also involved in research projects on implementing Industry 4.0 solutions in space projects and digital transformation in astrophysics.



Miguel Ángel Sánchez Vidales is a Doctor in Computer Science with an MBA in Project Management. He is the CEO of Mobente and the Research Manager at CoperniCO2. Miguel has extensive experience as a Project Manager at Indra and as the IT Director at the University of Salamanca. He has been involved in 35 R&D projects and is a university professor with expertise in mobile technologies. His current research interests lie in the field of Industry 4.0, and he has published numerous articles, books, and participated in conferences.



Paula Lamo Anuarbe holds a PhD in Industrial Engineering and is a professor at UNIR. She has multiple academic degrees, including four Master's degrees. Paula collaborates with research groups at UNIR and the University of Cantabria, focusing on topics such as Industrial Design and Production Technologies, Renewable Energies, and Quality Management. She has experience as a secondary school teacher and has conducted research on Industry 4.0 and blended learning.



A MULTIFACETED SWIM TRAINING APP ON ENHANCING SKILLS AND PERFORMANCE

Xiuhan (Daniel) Fu

Santa Margarita Catholic High School, 22062 Antonio Pkwy, Rancho Santa Margarita, CA

ABSTRACT

This research develops a swimmer's fitness journal along with a nutritional guidance application. The method employs journaling, expert suggestions, and detailed nutrition tips aimed at tackling precision issues using strictly-controlled research trials. The application exhibits remarkable accuracy in suggesting proteins and carbohydrates. It promises of more refinement according to personal goals and level of activity. The paper begins by presenting a brief overview of the issue in the swimming community. This highlights the significance of individualized dietary guidelines and coaching services tailored for swimmers.

Integration of BMI and BMR calculations for accurate nutritional recommendations is explained in this section of the methodology [13]. The author emphasizes on the need for accuracy in this work. It is also validated to previous standards and methodologies. The section of experiments and results reveals that several experiments were carried out in different situations, which prove the accuracy of the application's recommendations on nutrition. More nutrients can be incorporated as well. Another part of this article explores possible scenarios where this application applies, including swimmer focused design, more generalized fitness advice and nutrition recommendations targeting a more extensive audience.

Existing solutions are compared within a framework of a methodology comparison where one looks at features, effectiveness, and drawbacks. Limitations and avenue of future improvements are outlined in the summary section in addition to extending nutrient recommendations. This has bearing on application's importance in health of the population. Finally, this detailed fitness diary helps improve swimmer's training and performance. It might even benefit others who are looking for customized fitness and dietary guidelines.

KEYWORDS

Swimming, Fitness, Journaling, Advice

1. INTRODUCTION

There are a range of challenges that swimmers encounter in their training and competition journey. One significant issue is often the overlooked aspect of nutrition. Many swimmers struggle to grasp and implement effective nutritional strategies that can optimize their performance, increase recovery, and mitigate risks of injury. According to the National Library of Medicine, neglecting proper nutrition poses significant risks to athletes, potentially impairing their performance, delaying recovery, and increasing susceptibility to injuries and illnesses. Inadequate intake of essential nutrients can compromise energy levels, muscle function, and

David C. Wyld et al. (Eds): ICCSEA, NLAI, SCAI, CSIA, IBCOM, SEMIT, NECO, SPPR, MLDS -2023
pp. 245 - 262, 2023. CS & IT - CSCP 2023 DOI: 10.5121/csit.2023.132419

overall health, undermining an athlete's ability to excel in their sport [1]. Additionally, countless people have trouble with planning, and swimmers encounter the difficulty of keeping track of training progress, techniques, and goals. Manual tracking on paper can be prone to oversight and forgotten, potentially hindering improvement. For decades, the sport of swimming has grappled with these issues. While professional athletes have access to specialized coaching and nutritionists, half of the total number of swimmers rely on schools to host swimming lessons, not to mention private coaching [2].

The importance of proper nutrition, consistent tracking of performance, expert advice, and knowledge of safe training environments has long been recognized in the swim community. In the long run, lack of proper nutrition may result in impaired performance, reduced muscle mass and strength, delayed recovery, and long-term growth impairments, especially in young athletes. Furthermore, with many affected by the cost of lessons, a lot of swimmers are swimming with poor techniques without a coach to correct them. Common causes of injuries are wrong technique and poorly planned training. The most prevalent overuse injury is the shoulder, which could result in a long-term permanent injury. Additionally, knowing safe training locations around you can reduce the risk of accidents or encounters with hazardous conditions. A survey conducted by USA Swimming found that 75% of swimmers who consistently tracked their progress in a log journal and constantly improved their techniques reported seeing significant improvements in their performance over time [3].

1.1. Method Proposal

The proposed solution is a comprehensive swimming app that integrates a nutrition calculator, a training journal, and a map feature to locate nearby pools, addressing the challenges swimmers face in nutrition management, progress tracking, and safe training environments. The app will have three main features:

Nutrition calculator: It will provide personalized nutrition recommendations based on the user's gender, weight, BMI, activity level, and more. Users can input their statistics and receive tailored advice on nutrient intake for carbohydrates, protein, and calories to ensure they meet their energy and nutrient needs for optimal performance.

Log Journal: Swimmers can digitally record their training progress, techniques, and goals. The app will offer reminders for consistent tracking after every practice and/or meet, based on the user's preference. Users only need to input a date and location to fully organize each journal entry they create, which will eliminate the potential for oversight and allow for more effective and consistent training planning.

AI Advice: This section revolutionizes the way users access expert guidance. By integrating cutting-edge artificial intelligence, swimmers can now receive personalized and instant responses to any questions they have about training or technique. With 24/7 availability and the ability to understand natural languages, the AI Advice Section provides a level of support and empowerment that sets our app apart from others. It's like having a team of experts on call, always ready to offer expert insights whenever swimmers need them, boosting their motivation, confidence, and ultimately their performance in the pool.

By combining nutrition guidance, progress tracking, and pool location services in a single app, swimmers have a centralized tool that addresses multiple facets of their training journey. The nutrition calculator provides tailored recommendations, accounting for individual preferences and training objectives. The app also offers a user-friendly platform that can be accessed

anytime, anywhere, eliminating the need for manual tracking methods and providing easy access to important information.

Compared to traditional paper tracking, the app provides a digital platform that is less prone to oversight and offers reminders for consistent recording. Additionally, it integrates nutrition guidance and pretty much every type of advice one can think of, creating a one-stop solution for swimmers' needs. This app addresses the challenges comprehensively, making it a more efficient and effective solution for swimmers looking to optimize their training and performance

2. CHALLENGES

2.1. Challenge A

There were several problems with the nutrition page, which is supposed to have the user input their "weight," "height," "gender," and other information, and return their body mass index as well as suggested intake of various nutrients. Firstly, it lacks comprehensive error handling, particularly in the `calculateNutrition` function, where it doesn't handle null or zero values for carbohydrates and protein. Input validation is also missing, making it vulnerable to invalid user inputs. Furthermore, hardcoded string literals should be replaced with constants or localized strings for maintainability. Unused code blocks like `_showNutrientAdviceDialog` should be removed for clarity. Most of those problems were rooted in the long and complex code, so it is planned to include more comments to clarify complex logic and enhance accessibility through labels, which would also improve the code's quality.

2.2. Challenge B

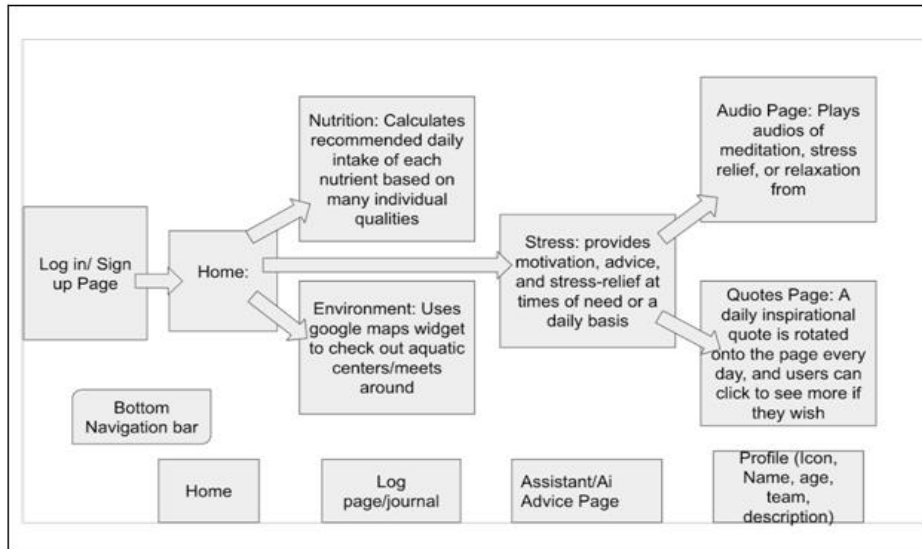
One key problem in designing the AI Advice Page for the swimmer's app is to provide accurate and dependable solutions to user queries. To begin, the model's training data must be large and relevant to swimming to minimize misrepresentation. Second, potential biases or skewed information generated by AI must be addressed. It would be necessary to use a strong pre-processing method to screen out any unsuitable or untrustworthy content. Furthermore, regular monitoring and updating of the training data, as well as user feedback loops, could assist refine the AI's responses over time. Finally, integrating a reporting system for consumers to indicate any wrong advice would add a high degree of safety and accuracy assurance.

2.3. Challenge C

A potential issue with the current implementation of the audio page could arise if new audio files are added to the Firebase Storage after the initial app launch. The `audioFiles` list is populated during the `initState` function and does not dynamically update if new files are added. This means that users might not see the latest audio files without restarting the app. To address this, the project could implement a mechanism to refresh the list of audio files at runtime. One approach could be to include a "Refresh" button on the UI that triggers a call to `fetchAudioFiles` when pressed. This way, users can manually update the list of available audio files. Furthermore, if immediate updates are critical, for this project, it is planned to explore Firebase Database or Firestore to track changes in the storage and automatically push updates to the app. This would ensure that the list of audio files is always up-to-date without requiring manual intervention from the user.

3. METHOD ANALYSIS

3.1. A – Diagram



3.1. System Overview

The Log Page serves as a platform for users to record their swimming activities. It encompasses various fields like journal title, date, location, strokes, distance, results, and goals. The entered data is then saved to a Firebase Firestore database, providing a structured and accessible record of the user's swim sessions. This component is crucial for users to keep track of their progress and performance over time.

The Nutrition Page is another central feature, where users input key details like weight, height, age, gender, carbohydrates, and protein intake. This information is managed through specific controllers for easy retrieval. The BMI is calculated using the `calculateBMI()` function, updating the display accordingly. When the user clicks "Calculate Nutrition," `calculateNutrition()` verifies that all necessary information is provided. If so, it computes BMI and offers tailored advice on carbohydrate and protein intake based on user specifics. The `calculateBMR()` function calculates Basal Metabolic Rate using weight, height, age, and gender. This is vital for generating precise nutrition advice.

`Get Carbohydrates Advice()` and `get Protein Advice()` then provide personalized recommendations based on the calculated metrics and user inputs. `_show Validation Error Dialog()` ensures users are prompted if all necessary information isn't provided. This page streamlines the user experience, making it user-friendly and efficient.

The assistance page comprises a Chat gpt model 3.5 widget in a Flutter application. The widget facilitates user interaction with the AI. Messages are managed through controllers, and chat history is stored. Communication with the Open AI API is moderated to generate responses. The user can ask any question or advice related to swimming or sports in general and they will receive a professional answer right away. The program's structure allows for seamless interaction between the user and AI.

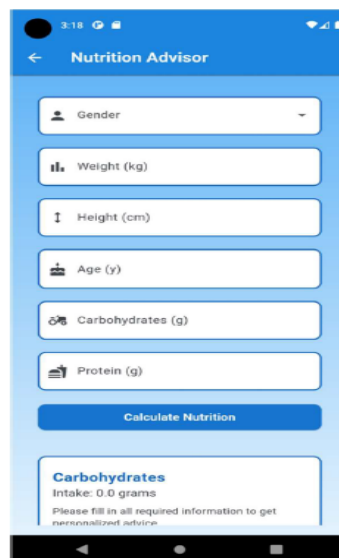
After logging in, the user is taken to the Nutrition Page, one of your application's three key components. They provide their personal information such as weight, height, age, gender, and nutrition intake here. The program validates this information and calculates the BMI, offering personalized advice on carbohydrates and protein consumption based on the user's particulars. In the home screen, the user can also navigate to the AI Advice Page, another major component, where they can seek advice on various health and fitness topics. This feature leverages AI-powered recommendations, providing users with valuable insights and tips. The Log Page, Nutrition Page, and AI Advice Page are interlinked through navigation buttons or tabs, allowing seamless movement between them. The program ensures data integrity by utilizing controllers and functions for calculations. The user can log out from any page, returning them to the Log Page, effectively concluding their session. Overall, this flow ensures a smooth user experience, guiding them through the application's features in a logical sequence. It balances user input, data processing, and output of personalized advice effectively, providing a comprehensive health and fitness tool.

Flutter was utilized to create this program.

3.2. Component Analysis A

The "Nutrition Page" is a vital component of the application. It allows users to input personal details like weight, height, age, gender, and nutrition intake. Its primary purpose is to calculate the user's BMI and provide tailored advice on carbohydrate and protein consumption. This component relies on Flutter for the user interface and Dart for logic and calculations. It doesn't use specialized concepts like NLP or neural networks. Instead, it employs basic nutrition principles. In summary, it processes user input to offer personalized nutrition advice, enhancing the user's health journey.

3.3. Ui Screenshot



3.4. Code Sample

```

String getCarbohydratesAdvice() {
    if (bmi != null &&
        weight != null &&
        height != null &&
        age != null &&
        gender != null) {
        double bmr;
        if (gender == Gender.Male) {
            bmr = 88.362 + (13.397 * weight!) + (4.799 * height!) - (5.677 * age!);
        } else {
            bmr = 447.593 + (9.247 * weight!) + (3.098 * height!) - (4.338 * age!);
        }

        double tdee = bmr;
        double carbsNeeded = (tdee * 0.6) / 4.0;

        double carbsIntake = carbohydrates ?? 0.0;

        if (carbsIntake < carbsNeeded * 0.8) {
            return "Consider increasing your carbohydrate intake for energy.";
        } else if (carbsIntake > carbsNeeded * 1.2) {
            return "Your carbohydrate intake seems high. Adjust based on your activity level.";
        } else {
            if (bmi < 18.5) {
                return "Your carbohydrate intake is appropriate for energy, but ensure you are getting enough nutrients to support your weight.";
            } else if (bmi < 24.9) {
                return "Your carbohydrate intake is balanced for your weight and energy needs.";
            } else if (bmi < 29.9) {
                return "Your carbohydrate intake is balanced. Focus on portion control and a balanced diet to maintain a healthy weight.";
            }
        }
    }
}

String getProteinAdvice() {
    if (bmi != null && bmi != 0 &&
        weight != null &&
        height != null &&
        age != null) {
        double bmr;
        if (gender == Gender.Male) {
            bmr = 88.362 + (13.397 * weight!) + (4.799 * height!) - (5.677 * age!);
        } else {
            bmr = 447.593 + (9.247 * weight!) + (3.098 * height!) - (4.338 * age!);
        }

        double tdee = bmr;
        double proteinNeeded = (tdee * 0.15) / 4.0;

        double proteinIntake = protein ?? 0.0;

        if (proteinIntake < proteinNeeded * 0.8) {
            return "Consider increasing your protein intake to support your body's needs.";
        } else if (proteinIntake > proteinNeeded * 1.2) {
            return "Your protein intake seems high. Adjust based on your activity level.";
        } else {
            if (bmi < 18.5) {
                return "Your protein intake is balanced, but ensure you are meeting your nutrient needs for your weight.";
            } else if (bmi < 24.9) {
                return "Your protein intake is appropriate for your weight and activity level.";
            } else if (bmi < 29.9) {
                return "Your protein intake is balanced. Focus on maintaining your weight with a balanced diet.";
            }
        }
    }
}

void calculateBMI() {
    if (weight != null && height != null) {
        double heightInMeters = height! / 100;
        double bmiValue = weight! / (heightInMeters * heightInMeters);
        setState(() {
            bmi = bmiValue;
        });
    }
}

void calculateNutrition() {
    if (weight != null &&
        height != null &&
        age != null &&
        gender != null &&
        (carbohydrates != null && carbohydrates != 0) &&
        (protein != null && protein != 0)) {
        calculateBMI();
        setState(() {
            carbohydratesAdvice = getCarbohydratesAdvice();
            proteinAdvice = getProteinAdvice();
        });
    } else {
        _showValidationErrorDialog();
    }
}

double calculateBMR() {
    if (weight != null && height != null && age != null && gender != null) {
        if (gender == Gender.Male) {
            return 88.362 + (13.397 * weight!) + (4.799 * height!) - (5.677 * age!);
        } else {
            return 447.593 + (9.247 * weight!) + (3.098 * height!) - (4.338 * age!);
        }
    } else {
        return 0.0;
    }
}

```


3.5. Code Explanation

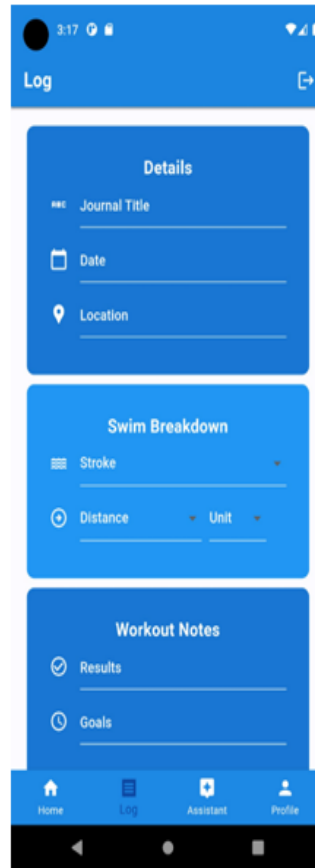
The code includes a number of essential procedures for computing and giving the user-specific nutritional advice. First, the `calculate BMI()` function uses the user's provided height and weight to calculate the Body Mass Index (BMI). The state of the program is then updated with this value. The complete process of providing nutritional advice is coordinated by the `calculate Nutrition()` method. It confirms the accessibility of crucial user data, such as body mass index, height, age, gender, and protein and carbohydrate intake. If all necessary information is provided, `calculate BMI()` starts the BMI computation, ensuring current BMI values. It then uses the functions `get Carbohydrates Advice()` and `get Protein Advice()` to update the advice for both carbs and protein. It prompts the user with an error message using `_show Validation Error Dialog()` if any important information is missing. The Basal Metabolic Rate (BMR), which is based on variables including weight, height, age, and gender, is calculated using the `calculate BMR()` function. Taking into account the user's BMI, weight, height, age, and gender, the functions `get Carbohydrates Advice()` and `get Protein Advice()` provide tailored advice on carbohydrate and protein intake, respectively.

Based on determined variables, these approaches compare the user's actual intake to the suggested quantities. In the end, this collection of techniques successfully directs the user toward selecting a diet that is in line with their unique health objectives. It starts working when the user interacts with the nutrition page, entering their information and starting the computation. User privacy and data security are ensured because the program runs locally on the device and does not communicate with a backend server.

3.6. Component B

The Log Page component's goal is to provide a user interface for logging swim-related data. It allows users to enter information such as the journal title, date, location, stroke type, distance, unit, results, and swim goals. This data is then saved to a Firestore database linked to the current user.

The code relies on several packages to build this system, including `cloud_firestore` for communication with the Firestore database and `firebase_auth` for user authentication. Furthermore, it makes use of standard Flutter libraries such as `material.dart` for UI components and `intl.dart` for date formatting. In general, this component serves by providing a structured interface via which users can input and save swim-related data. Before attempting to save the data to the Firestore database, it ensures that all required fields are completed. Overall, this component is critical in supporting the logging and tracking of swim-related data throughout the program.



```

void _saveEntry() async {
  try {
    _getCurrentUser();

    if (!_areRequiredFieldsFilled()) {
      showDialog(
        context: context,
        builder: (context) => AlertDialog(
          title: Text('Error'),
          content: Text('Please fill out all the fields before saving.'),
          actions: [
            TextButton(
              onPressed: () {
                Navigator.of(context).pop();
              },
              child: Text('OK'),
            ), // TextButton
          ],
        ), // AlertDialog
      );
      return;
    }

    final collectionReference = FirebaseFirestore.instance
      .collection("Users")
      .doc(_currentUser.uid)
      .collection("Journal Entries");
  }
}

```

```

await collectionReference.add({
  'Journal Title': _journalTitleController.text.trim(),
  'Date': _dateController.text.trim(),
  'Location': _locationController.text.trim(),
  'Strokes': _selectedStroke,
  'Distance': '$_selectedDistance$_selectedUnit',
  'Results': _resultsController.text.trim(),
  'Goals': _goalsController.text.trim(),
});

// Clear input fields after saving
_journalTitleController.clear();
_dateController.clear();
_locationController.clear();

_resultsController.clear();
_goalsController.clear();

setState(() {
  _selectedStroke = null;
  _selectedDistance = null;
  _selectedUnit = null;
});

showDialog(
  context: context,
  builder: (context) => AlertDialog(
    title: Text('Success'),
    content: Text('Journal entry saved!'),
  ),
);

```

```

showDialog(
  context: context,
  builder: (context) => AlertDialog(
    title: Text('Success'),
    content: Text('Journal entry saved!'),
    actions: [
      TextButton(
        onPressed: () {
          Navigator.of(context).pop();
        },
        child: Text('OK'),
      ), // TextButton
    ],
  ), // AlertDialog
);
} catch (error) {
  showDialog(
    context: context,
    builder: (context) => AlertDialog(
      title: Text('Error'),
      content: Text('Journal entry failed to save. Please try again.'),
      actions: [
        TextButton(
          onPressed: () {
            Navigator.of(context).pop();
          },
          child: Text('OK'),
        ), // TextButton
      ],
    ), // AlertDialog
  );
}

```

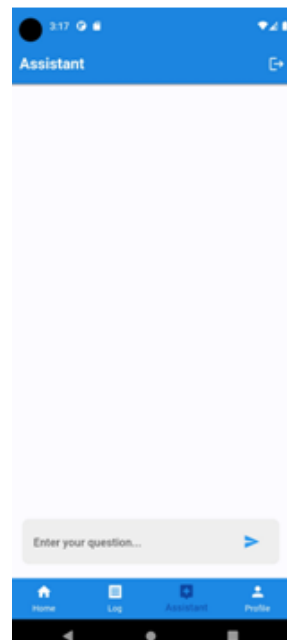
When a user interacts with the "Save" button in the user interface, the `_saveEntry()` method in the given code is called. This function is in charge of handling the logic when a user wants to save a swim log entry. To begin, it checks to see if all of the required fields, such as journal title, date, location, stroke, distance, unit, outcomes, and goals, are filled out. Following successful input validation, the procedure retrieves the currently authenticated user using Firebase Authentication. Following that, it connects to the Firestore database, creating a link to the user's journal entry collection. It conducts housekeeping duties after saving by resetting input fields, clearing controllers, and unselecting options.

Several controllers for managing text input are initialized by the `_Log Page State` class, including `_journal Title Controller`, `_date Controller`, `_location Controller`, `_results Controller`, and `_goals Controller`. There is also a variable called `_current User` that contains information about the currently authorized user. The code also creates lists of options for stroke, distance, and unit selection, which are used in the user interface.

This code interfaces with a backend server, particularly Firestore, which is a Firebase NoSQL cloud database. When `_save Entry()` is called, the server receives a request to add a new document to the user's Firestore collection. The server responds to this request by producing and saving a new document with the user's swim log details, essentially persisting the user's data in the cloud database.

3.7. Component C

The Assistant Page component allows users to communicate with an Open AI-driven chat bot that is powered by the GPT-3.5 Turbo model. It uses the HTTP package for API communication and is written in Flutter. Users enter questions, which causes an OpenAI API request for a produced answer, imitating a discussion. For authentication and selection, the API key and model information are used. The `_sendMessage()` function handles updating the chat history by adding the user's message, formatted as "You: [message]", followed by processing the message through `_getChatResponse()`. The generated response from the chatbot is appended to the chat history as "ChatGPT: [response]". If any errors occur during the process, an error message is displayed in the chat history.



```

Future<String> _getChatResponse(String input) async {
  final apiKey = 'API-KEY-HERE';
  final model = 'gpt-3.5-turbo';
  final apiEndpoint = 'https://api.openai.com/v1/chat/completions';

  final response = await http.post(
    Uri.parse(apiEndpoint),
    headers: {
      'Content-Type': 'application/json',
      'Authorization': 'Bearer $apiKey',
    },
    body: json.encode({
      'input': input,
      'model': model,
    })),
  );

  if (response.statusCode == 200) {
    final Map<String, dynamic> data = json.decode(response.body);
    return data['output'];
  } else {
    print('Reponse Status Code: ${response.statusCode}');
    print('Response Body: ${response.body}');
    throw Exception('Failed to load response');
  }
}

```

The code segment is a Flutter application that works with OpenAI's GPT-3.5 Turbo model to build a conversation interface. The AssistantPage class is the primary component in charge of handling the chat UI. It includes methods such as `_getChatResponse` and `_sendMessage`.

The contact with OpenAI's server is handled by the `_getChatResponse` method. This function is called when a user submits a message. It creates an API request using the user's input, the model selected, and the API key. The request is forwarded to OpenAI's server, which extracts and returns the chat bot's output if it receives a successful response (status code 200). If an error occurs, it logs the status code and response body before throwing an exception.

When the user sends a message, the `_sendMessage` method is called. It adds the user's message to the conversation history before calling `_get Chat Response` to receive a response from the chat bot. When an answer is received, it is added to the chat history.

The variables `_input Controller` and `_chat History` govern the user's input in the chat, while `_chat History` records the conversation history. The user interface is created with Flutter widgets such as Scaffold, Column, Expanded, List View. builder, and Text Field.

The application sends an API request to Open AI's server when a user delivers a message. This request is processed by the server, and a response is generated. This response is then returned to the program and shown in the chat UI.

4. Experiments

4.1. Experiment A

For the nutrition page, one blind spot is the user not providing valid input for weight, height, age, or nutrition intake values, which could lead to incorrect advice or errors in calculation.

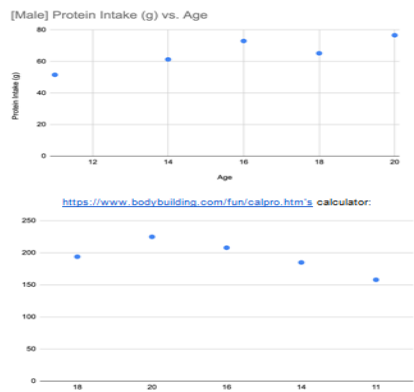
4.2. Design

First, the program would simulate invalid input for one or more of the fields with blindspot. In the application, deliberately input incorrect or incomplete data. The control data are sourced from

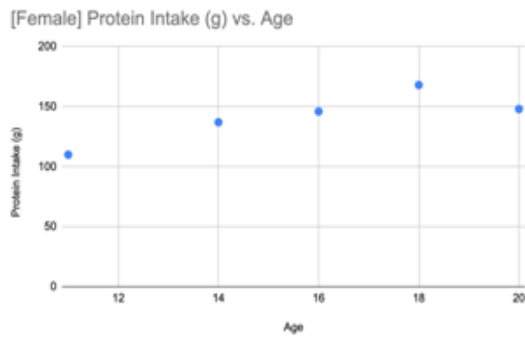
primarily two formulas used to calculate Body Mass Index and Basal Metabolic Rate (BMR). The formula for BMR is sourced by Garnet Health. It will also consist of valid inputs that are within the expected range [4]. It will serve as a baseline to compare against the results obtained from the experiment. We can record the behavior of the program when faced with invalid input. The program should detect and notify the user of the invalid input, providing clear error messages. Additionally, it should not crash or produce unexpected behavior. Finally, the app should guide the user to correct the input or prevent them from proceeding until valid data is provided.

4.3. Data and Visualization

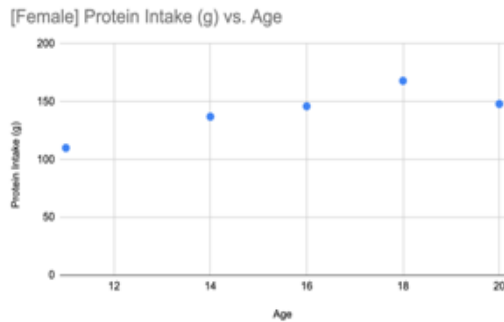
My nutrition calculator



My calculator:



Theirs:



4.4. Analysis

This code calculates daily protein intake according to age, height, and body weight using BMR as a basis. Determining the correct BMR is crucial and entails attributes such as the age, sex, height and body weight that are drawn using established calculations. This gap between the estimated protein requirement and actual values suggests that there may be a mismatch of BMR and protein consumption. This precision is essential, usually expressed in terms of grams per kilogram of bodyweight (g/kg).

It is recommended to debug systematically, involving thorough examination of the conversion procedure, unit checking and controlled tests with sample data. Diagnostics, such as cross-referencing with authoritative sources and sensitivity analysis, are also useful. Specialized opinions could be sought from expert consultation if it is necessary for resolution. A critical assessment of the conversion process should be carried out together with specific tests towards locating where the discrepancy in protein calculation was generated.

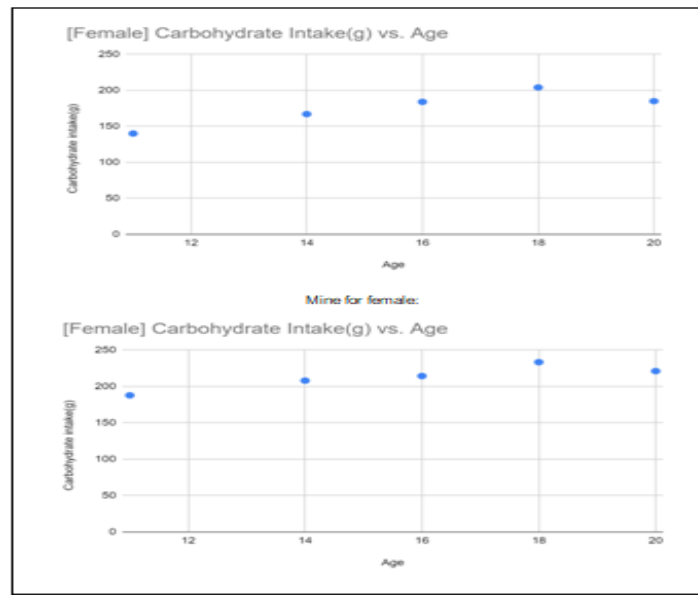
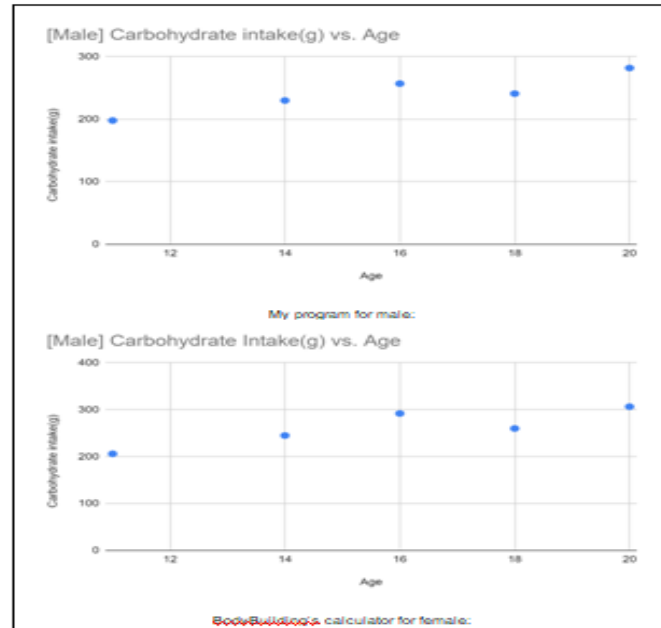
It is also important to consider sensitivity analysis to find out how much error in BMR calculation could cause significant differences in protein need. If these measures do not resolve the issue, consulting a nutritionist or any expert is the next step.

4.5. Experiment 2

Another possible blind spot is the calculation for carbohydrates. Even though the information collected had different sources for the formula of protein and carbohydrate, uncertainty remains on whether the calculations are correct for the carbohydrates data possessed.

In order to evaluate the application's performance, deliberately erroneous or incomplete data was introduced, creating controlled scenarios akin to blind spots. This involves intentionally inputting inaccurate or partial information. The foundational data is sourced from two critical formulas used for computing Body Mass Index (BMI) and Basal Metabolic Rate (BMR), with BMR sourced from Garnet Health. Alongside, valid inputs falling within the expected range will be included to serve as a reference for comparison. Following this, I will systematically document the program's response when confronted with instances of invalid input. The application's functionality must encompass the capacity to swiftly detect and communicate the presence of invalid data, providing lucid error notifications. Equally important is the program's stability, ensuring it refrains from crashing or displaying unforeseen behavior under these circumstances.

Body Building's calculator for male



The code experiment was designed to establish daily amounts of carbohydrates based on submitted information such as age, height, and weight. Interesting insights in the comparison with a professional carbohydrate calculator. The predicted carbohydrate requirements for the five tests for men were off by less than 20 grams on average from those of professional calculator calculations. This shows an acceptable rate of precision, implying that the underlying programming is sound and fit for the male population. Nonetheless, the situation is dramatically different for female users. Women's carbohydrate needs displayed much more variability than the pro calculator, with an average difference amounting to nearly 40 grams. Although this difference appears more noticeable in the case of females, it is still acceptable due to the natural variability of metabolism and biology in people.

However, several reasons might account for these variations. However, one should not forget that nutrition requirements may be affected by many individual-specific factors other than age, height, and bodyweight, such as changes in hormone level and physical activity. On top of that, minor differences in the base calculations or premises used in the specialized calculator and your code may explain some of these differences. The accuracy of the carbohydrates can be improved by checking the algorithms and formulas against current nutrition practices. Also, more variables or making adjustments on the existing ones would yield a higher level of accuracy, especially with the females.

5. METHODOLOGY COMPARISON

5.1. Methodology A

In an article regarding the nutritions for swimmers by Swimming World Magazine offers practical advice for swimmers on optimizing their nutrition. It highlights the significance of carbohydrates and proteins in their diet, providing specific intake recommendations based on body weight. The emphasis on hydration and electrolyte balance, particularly before competitions, is valuable. However, it could benefit from addressing individual dietary needs, micronutrient considerations, and the psychological impact of nutrition on an athlete's performance. Additionally, long-term nutritional planning and the potential use of supplements are aspects that could be explored further [5].

This project allows users to input their specific details, such as gender, weight, height, and age, enabling it to provide highly personalized nutrition advice. This level of customization is not possible with a generic article. Also, the app offers personalized advice on carbohydrate and protein intake, as well as BMI information. In contrast, the first article primarily focuses on general recommendations without considering individual factors.

5.2. Methodology B

An article published by Runners World offers a set of swimming techniques for triathletes, drawing from the expertise of Michael Phelps, Terry Laughlin, and David Marsh. It emphasizes streamlining the body, effective hand placement, body rotation, and head positioning to reduce resistance and optimize propulsion. This approach is effective in improving triathletes' swimming proficiency, particularly given the demands of their multi-discipline training. However, it's important to note that individual variability and the need for ongoing practice may influence progress. Additionally, the guide does not address open water specifics, race strategies, or equipment considerations, which are important for a well-rounded training regimen [6]. The project code creates a journal with search functionality, which is much more efficient and easier than journaling on paper. Helper methods such as `_buildErrorMessage` and `_BuildLoadingIndicator` for displaying different types of message and formatted text to ensure readability and have the entries easy to understand. Lastly, 'StreamBuilder' listens to changes in the user's journal entries and updates the UI accordingly.

5.3. Methodology C

Finally, Your Swim Log talks about how crucial journaling daily entries and logging individual times are for athletics. The journaling solution works by having swimmers write down their ideal workouts and races, helping them visualize success and focus their mental approach. It's effective in reducing pre-race nerves, as studies show it improves performance under pressure. However,

it's not a replacement for physical training and may not cover all aspects of performance improvement, such as external factors or specific technical skills [7].

This digital platform enhances the journaling process by providing a convenient and organized way for swimmers to document their practices, races, and reflections. Additionally, it integrates Firebase services for authentication and Firestore for easy data management. Compared to the earlier passage which discussed the benefits of journaling, this code provides a practical tool that enables swimmers to directly apply the described techniques in a digital format.

6. CONCLUSIONS

6.1. Limitations And Improvements

Under the “Nutrition Advisor” page, users are able to fill in the most important data concerning their condition (gender, weight, height and age) as well as diet analysis. Afterward, the app provides personalized advice on carbs, protein, and weight based on such information input. It provides the opportunity for addition of more specific useful products such as fats, vitamins, minerals, and fiber. It is expected that this growth will enable consumers to understand their own diet in detail. Additionally, for enhancing the reliability and relevance of recommended guidelines that are specific to particular nutrients needs should be embraced.

This can be done through quoting from respected sources such as the DRIs and RDAs recommended by recognized health bodies. Since our studies used a calculator that incorporates weight goals and activity level, we may discuss their possible inclusion into this process. This would increase the accuracy of the advice, for body weight goals and physical activity greatly influence nutrient requirements. With the inclusion of these details, the tool is set to provide more individualized and persuasive recommendations, thus raising the app’s relevance to people who wish to follow customized eating plans.

6.2. Concluding Remarks

This project is a key step toward developing a comprehensive fitness journal and nutritional guidance app. The Journal Page allows for easy access to entries, whereas the Nutrition Page gives personalized help via BMI calculation along with customized counsel. On the advice/assistant page, the ChatGPT widget improves user interaction, while the quotations, stress, and audio features promote well-being. In essence, the software not only meets but exceeds its purpose by providing a user-centric platform that tackles all aspects of health and wellness. It blends fitness tracking with mindful health habits to provide a well-rounded approach to fitness.

7. SUMMARIES

7.1. Experiment Recap

Experiment 1 aimed to identify potential blind spots related to invalid user input in the nutrition calculator. Invalid data for weight, height, age, or nutrition values could lead to errors. The experiment involved deliberately inputting incorrect and comparing it with valid inputs within expected ranges. Control data were sourced from established formulas for Body Mass Index (BMI) and Basal Metabolic Rate (BMR), with the BMR formula from Garnet Health. The program's behavior in handling invalid input was observed, emphasizing the need for clear error messages and program stability.

Experiment 2 focused on assessing the accuracy of carbohydrate calculations. Data was sourced from critical BMI and BMR formulas, with BMR from Garnet Health. The results revealed acceptable precision for men, with minor discrepancies. However, variability was more pronounced for female users, indicating the influence of individual-specific factors beyond basic demographics. Suggestions included refining algorithms, incorporating additional variables, and aligning calculations with current nutrition practices to enhance accuracy, particularly for females.

7.2. Methodology Comparison

Methodology A focuses on optimizing nutrition for swimmers, as outlined in an article from Swimming World Magazines. It places a strong emphasis on the significance of carbohydrates and proteins in a swimmer's diet, providing specific intake recommendations based on body weight. It also highlights the importance of maintaining proper hydration and electrolyte balance, particularly before competitions. While this approach offers valuable foundational advice for swimmers, it doesn't really address individual dietary needs and micronutrient considerations [8]. Methodology B, outlined in an article from Runners World, talks about swimming techniques tailored for triathletes. The approach draws advice from accomplished swimmers like Michael Phelps, Terry Laughlin, and David Marsh. It places a strong emphasis on refining techniques such as streamlining the body, ensuring effective hand placement, mastering body rotation, and optimizing head positioning. While this is highly effective in enhancing triathletes' swimming proficiency, it lacks coverage on crucial aspects like open water specifics, race strategies, and equipment considerations.

Methodology C, advocated by Your Swim Log, explains the significance of journaling daily entries and logging individual times for athletes, particularly swimmers. The approach encourages swimmers to write down their ideal workouts and races, which aids in visualizing success and improving their mental approach. This practice proves highly effective in reducing pre-race nerves and enhancing performance under pressure, according to studies. However, it's important to note that while journaling is a valuable tool, it cannot serve as a complete replacement for physical training.

REFERENCES

- [1] Beck, K. L., Thomson, J. S., Swift, R. J., & von Hurst, P. R. (2015, August 11). Role of nutrition in performance enhancement and postexercise recovery. U.S. National Library of Medicine. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4540168/>
- [2] Moore, C. (2021, August 31). Millions of children can't swim because lessons are costly, survey suggests. Fox Business. <https://www.foxbusiness.com/lifestyle/millions-children-cant-swim-lessons-costly-survey-says>
- [3] Department of Health & Human Services. (2007, August 7). Swimming - preventing injury. Better Health Channel. <https://www.betterhealth.vic.gov.au/health/healthyliving/swimming-preventing-injury>
- [4] Garnet Health. (2016, July 1). Basal Metabolic Rate Calculator. Garnet Health. <https://www.garnethealth.org/news/basal-metabolic-rate-calculator#:~:text=Calculate%20Basal%20Metabolic%20Rate&text=Your%20basal%20metabolism%20rate%20is>
- [5] Duran, B. (2018, July 10). The Big Deal About a Swimmer's Nutrition. Swimming World News. <https://www.swimmingworldmagazine.com/news/the-big-deal-about-a-swimmers-nutrition/>
- [6] Bean, M. (2009, November 23). Seven Steps To Better Swimming Technique. Runner's World. <https://www.runnersworld.com/uk/training/triathlon/a764520/seven-steps-to-better-swimming-technique/>

- [7] The Power of Journaling for Swimmers. (2017, July 7). YourSwimLog.com. <https://www.yourswimlog.com/journaling-for-swimmers/>
- [8] Duran, B., & Duran, B. (2023, October 18). Wellness Wednesday: The Big Deal About a Swimmer's Nutrition: What to Know. Swimming World News. <https://www.swimmingworldmagazine.com/news/the-big-deal-about-a-swimmers-nutrition/>
- [9] Sabouchi, N. S., Rahmandad, H., & Ammerman, A. S. (2013). Best-fitting prediction equations for basal metabolic rate: informing obesity interventions in diverse populations. *International Journal of Obesity*, 37(10), 1364–1370. <https://doi.org/10.1038/ijo.2012.218>
- [10] Poirier-Leroy, O. (2020, June 4). The power of journaling for swimmers. SwimSwam. <https://swimswam.com/power-journaling-swimmers/>

COMMUNICATION SIGNALS MODULATIONS CLASSIFICATION BASED ON NEURAL NETWORK ALGORITHMS

Yahya Benremdane ¹, Said Jamal ¹, Oumaima Taheri ¹, Jawad Lakziz ¹
and Said Ouaskit ¹

¹ Faculty of science Ben M'Sik, University Hassan II, Casablanca, Morocco

ABSTRACT

This paper aims to find an automatic solution for the modulation's classification of different types of radio signals by relying on Artificial Intelligence. This project is part of a long process of Communications Intelligence looking for an automatic solution to demodulate, decode and decipher communication signals. Our work therefore consisted in the choice of the database needed for supervised deep learning, the evaluation of existing techniques on raw communication signals, and the proposal of a solution based on deep learning networks allowing to classify the types of modulation with an optimal ratio (computation time / accuracy). We first carried out a research work on the existing models of automatic classification in order to use them as a reference. We consequently proposed an ensemble learning approach based on tuned ResNet and Transformer Neural Network that is efficient at extracting multi-scale features from the raw I/Q sequence data and also considers the challenge of predicting in low Signal Noise Ratio (SNR) conditions. In the end, we delivered an architecture that is easy to handle and apply to communication signals. This solution has an optimal and robust architecture that automatically determines the type of modulation with an accuracy up to 95%.

KEYWORDS

Automatic modulation classification, Modulation recognition, Artificial Intelligence & Deep Learning

1. INTRODUCTION

Communication signals, marked by diverse modulations to achieve high data rates while mitigating interference, present a formidable challenge for Intelligence Systems tasked with monitoring the communications spectrum. As the complexity of modulations increases, the identification and demodulation processes become progressively intricate, particularly for extracting valuable information in the realm of Communications Intelligence (COMINT).

In the domain of COMINT, where the primary objective is to extract meaningful information, the study focuses on the intricate task of recognizing and classifying modulations in intercepted signals. This is pivotal for understanding the type of transmission and subsequently facilitating the demodulation process. Unlike Electronic Intelligence (ELINT), which predominantly deals with radars, COMINT involves decoding communication signals, whether voice or data.

In the transmission of information through communication signals, modulation is a fundamental process. The information is modulated into a specific frequency, enabling high-speed transmission and overcoming atmospheric attenuation challenges. Intercepting an unknown

signal within the vast spectrum of communications initiates the Intelligence process, involving measurements of frequency and signal levels. However, the initial and critical challenge lies in determining the modulation used for transmitting the radio signal. Traditionally, intelligence approaches involved employing various demodulators iteratively, a method proven to be slow and ineffective, particularly with modern modulations. The advent of Artificial Intelligence (AI) has significantly transformed this landscape. Automatic classification of modulation types at the receiver has garnered substantial attention in the wireless research community, notably improving spectrum utilization efficiency. Early efforts utilized spectrogram images generated by different modulations and applied Convolutional Neural Network (CNN) architectures. Recent studies have taken a novel approach, leveraging the Inphase and Quadrature signals (I/Q) of the signal—referred to as the "DNA" of any signal. I/Q data has demonstrated superior performance in automatic modulation recognition compared to traditional approaches. Essentially, any signal comprises two components: the In-phase component (Cosine) and the Quadrature component (Sinus). These I/Q samples describe a complex baseband signal, where the real and imaginary parts are represented by the waveforms $I(t)$ and $Q(t)$.

The complete signal description, $X(t) = I(t) + jQ(t)$, encapsulates the essence of the signal, encoded into a matrix of two rows representing I and Q. This I/Q-based approach proves to be a powerful methodology for modulation classification, providing a comprehensive and effective means of deciphering the intricate modulations present in modern communication signals.

2. BACKGROUND

Automatic Modulation Classification (AMC) techniques encompass a spectrum of methodologies, broadly categorized into traditional approaches, where most of them are basically categorized into the likelihood-based (LB) and feature-based (FB) approaches and advanced techniques leveraging deep learning.

2.1. Traditional Approaches

2.1.1. Likelihood-Based Methods

In the early stages of Automatic Modulation Classification (AMC), likelihood-based methods were prevalent. These methods involve the precise derivation of likelihood functions for different modulation types. The fundamental idea is to match the received signal against a set of predefined likelihood functions to determine the most probable modulation type. Likelihood-based methods employ probability theories and hypothetical models to address modulation identification challenges in scenarios with both known and unknown channel information [1]. While these approaches can achieve optimal classification accuracy under the assumption of perfect knowledge of signal and channel models, they demand considerable computational complexity for estimating model parameters [2], [3].

2.1.2. Feature-Based Approaches

In the realm of AMC, feature-based techniques serve as a foundational approach for distinguishing modulation patterns. This method hinges on feature extraction and classifier building, offering a pragmatic balance between computational efficiency and classification accuracy. The fundamental premise is to capture the distinctive characteristics of various signals without the need to intricately derive the likelihood function of the signal. The feature-based approach unfolds in two critical steps: pre-processing the signal and extracting relevant features. Subsequently, a classifier is applied to categorize the signal based on these features. The success

of this approach crucially depends on the careful selection of signal attributes and the construction of robust classifiers. Feature-based techniques are particularly advantageous in scenarios where algorithm complexity needs to be minimized, making them suitable for real-time applications and resource-constrained environments [4].

Although feature-based methods exhibit adaptability to various channel models, they encounter significant limitations, including the weak discriminatory capability of manually crafted features and the constrained learning capacity of conventional classification algorithms [5], [6].

2.2. Advanced Approaches

Deep learning (DL), with its exceptional data processing capabilities, has drawn extensive interest and been applied in a variety of sectors because of the rapid growth of Artificial Intelligence (AI) technology including radio signal processing for communications. The use of deep learning for AMC is an active area of research, with new techniques and architectures being proposed to improve classification accuracy and reduce computational complexity. Indeed, applications of DL as a solution to conventional feature-based signal classification issues provide an efficient and cost-effective alternative for AMC. Several recent AMC methods utilizing deep networks such as deep neural networks (DNNs), convolutional neural networks (CNNs) [7], recurrent neural networks (RNNs), and long short-term memory networks (LSTMs) [8], have been proposed to address the existing limitations of traditional approaches. However, the performance of these deep learning-based AMC methods may still be affected by the over-fitting issue brought on by a considerable number of network parameters [9].

2.3. Ensemble Learning for AMC

Ensemble learning has emerged as a powerful paradigm in machine learning, demonstrating significant success in various domains. The concept involves combining predictions from multiple models to enhance overall performance, providing improved robustness and accuracy [10]. The application of ensemble models in Automatic Modulation Classification (AMC) has garnered attention due to its ability to address the complex and dynamic nature of communication signals. Ensemble models integrate diverse sources of information, enabling them to capture intricate patterns inherent in modulation types and varying SNR conditions [11],[12]. Ensemble models offer several advantages in the context of AMC. They excel in handling diverse modulation types, adapting to variations in SNR, and providing enhanced accuracy in classification outcomes. Recent advancements in ensemble models for AMC include innovative architectures and methodologies. Noteworthy examples include ensemble models based on deep learning, leveraging architectures such as deep neural networks (DNNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs). These models demonstrate the potential to improve classification accuracy and reduce computational complexity [13].

Despite their success, challenges exist in designing effective ensemble models for AMC. Striking the right balance between model diversity and coherence is crucial. Additionally, addressing issues related to overfitting and ensuring the generalization of ensemble models across different signal scenarios are ongoing research areas. The proposed ensemble model in this study draws inspiration from the advancements and challenges outlined in the literature on ensemble models in AMC. The choice of combining ResNet and Transformer neural networks is motivated by the need to leverage complementary strengths. A critical analysis of existing ensemble models in AMC reveals gaps and opportunities for improvement. The proposed model aims to address these gaps by integrating state-of-the-art architectures and refining the ensemble learning process for more effective modulation classification.

3. THE PROPOSED APPROACH

In this section, we present our innovative approach (Figure 1) to modulation classification, leveraging an ensemble of two powerful neural network models: Residual Network (ResNet) and Transformer Neural Network (TNN): one optimized for accurately predicting signals with high SNRs, and the other for predicting signals with low SNRs. The key objective is to address challenges posed by varying Signal-to-Noise Ratios (SNRs) by tailoring each model to excel in specific SNR conditions. The ensemble design aims to capitalize on the complementary strengths of ResNet, proficient in spatial feature extraction, and TNN, adept at handling sequential data and capturing temporal dependencies.

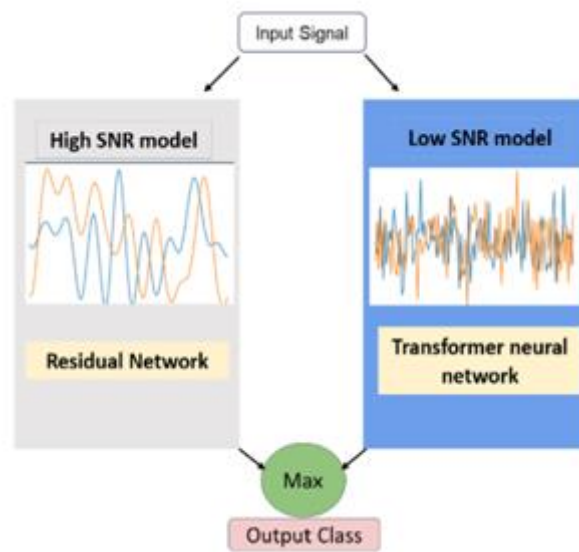


Figure 1. Proposed ensemble model (Resnet with TNN).

3.1. High SNR Model: Residual Net

ResNet (Residual Network) is a type of convolutional neural network (CNN) architecture that was introduced in 2015 by Microsoft researchers [14]. The key innovation of ResNet is the use of "residual connections," which allow the network to learn a residual mapping rather than an explicit mapping from the input to the output. This makes it possible to train much deeper networks than was previously possible, while still maintaining satisfactory performance and without encountering the vanishing gradient problem. ResNet has been used to achieve state-of-the-art results on a variety of image classification tasks and can also be utilized for AMC [15].

The ResNet architecture is composed of two main parts: the residual stack and the residual unit. The residual stack is a sequence of residual unit, where each unit contains multiple layers. The residual stack is responsible for learning a residual function with reference to the layer inputs. This can be accomplished by adding the input of a layer to the output of the same layer, before passing it through the next layer. It is responsible for deeping network and allowing it to learn complex characteristics from the data. The residual unit is the core building block of the ResNet architecture. It consists of two or more convolutional layers, with the output of the first layer being added to the input of the second layer. This helps to save information from the source and allows the network to learn a residual function. The residual unit also includes a batch that

normalize the layer, which is used to normalize the output of the convolutional and improve the stability of the network.

3.2. Low SNR Model: Transformer Neural Network

The TNN is an architecture that is solving easily sequence to sequence tasks in the long-range dependencies [16]. Transformer models apply a set of mathematical procedures known as attention or self-attention, to detect influence and dependency of distant data elements. Attention mechanisms are used to weight the various parts of the input signal differently, which can help the network focus on the most important parts of the signal, like the signal of interest, and disregard the noise. The core function of this mechanism is to determine which features in the input are significant for the target and which features are not by generating a weighting coefficient to weight the input to sum up for a given target.

The Transformer neural network architecture comprises several layers, including encoding and decoding. The encoder is composed of multiple layers of self-attention and feed-forward neural networks. The self-attention mechanism enables the model to weigh the importance of different input components when making predictions. The feed-forward neural network is used to process the output of the self-attention layer.

The decoder is also composed of multiple layers of self-attention and feed-forward neural networks. The decoder also uses a mechanism called “masked self-attention” which prevents the model from “peeking” at future tokens in the input sequence when making predictions. The transformer architecture also contains a Multi-Head Attention mechanism, which allow the model to attend to various parts of the input at the same time, improving its ability to understand the input. It is highly parallelizable and computationally efficient. The architecture used is as follows:

- Transformer Block: contains a Feed-Forward neural network (FFN) with 256 nodes, used to increase the capacity of the model by introducing non-linearity.
- Global Average Pooling: average of all the values in the input tensor.
- Alpha Dropout (0.3): which randomly drops out certain proportion of the activations to prevent overfitting. It maintains the mean and variance of the input by keeping them at their original values.
- Two fully connected network along with Alpha Dropout (0.2): the activation function applied is SeLU that stands for Scaled Exponential Linear Unit.

The Transformer neural network has been chosen for low SNR signals as it is able to handle sequential data such as time series, and also it has been shown to be effective in tasks that require understanding the context and dependencies among different inputs. Indeed, our method involves using a transformer encoder to extract features from a low SNR signal, which are then used by a transformer decoder to reconstruct the signal. The encoder and decoder are jointly trained to reduce the error of reconstitution between input and out-put signals.

3.3. Ensemble Model Integration

The ensemble model proposed in this study leverages the synergies between two distinct deep learning architectures: Residual Network (ResNet) and Transformer Neural Network (TNN). This integration is designed to harness ResNet’s proficiency in capturing spatial features and TNN’s effectiveness in handling sequential data and temporal dependencies.

ResNet, optimized for high Signal-to-Noise Ratio (SNR) environments, excels in distinguishing modulation signals in clear, noise-free conditions. To seamlessly integrate ResNet into the

ensemble, its spatial feature extraction output becomes a crucial input. The model is trained to make predictions based on spatial characteristics. Conversely, the Transformer Network is tailored for low SNR scenarios and adeptly processes sequential data, making it suitable for capturing temporal dependencies.

In the ensemble, the TNN's output, enriched with its self-attention mechanisms, contributes predictions based on sequential patterns in the signal. Unique to our ensemble model is the simultaneous prediction capability of both ResNet and TNN. Each model independently processes the input signal and generates a prediction. The ensemble decision-making mechanism is then employed, where the maximum prediction among the two is selected as the final output. This strategy ensures that the ensemble benefits from the strengths of both models, providing a robust and adaptive classification approach. During the joint training of the ensemble, the models are fine-tuned collaboratively. This involves optimizing the parameters of ResNet and TNN while incorporating the decision-making mechanism that selects the maximum prediction. The ensemble learns to dynamically adapt to varying challenges posed by different SNR conditions, making it a versatile solution. Architecturally, the ensemble model is enhanced to accommodate the dual predictions and the decision making process. Additional layers and connections are introduced to facilitate the flow of information between ResNet and TNN, preserving their unique contributions to the overall classification process.

The proposed ensemble model uniquely involves the simultaneous predictions of ResNet and TNN, with the final output determined by selecting the maximum prediction. This dynamic approach ensures that the ensemble is robust and capable of capitalizing on the strengths of both models, ultimately enhancing the accuracy of modulation classification across diverse SNR conditions.

4. EXPERIMENTAL RESULTS

4.1. Experimental Setting: Dataset Selection and Characteristics

To validate the efficacy of our proposed model, we curated a comprehensive dataset that combines synthetic and real-world gathered data. This dataset is carefully designed to encompass a diverse range of modulation scenarios, including both synthetic and simulated channel effects.

4.1.1. Dataset Composition

The dataset consists of the following key components:

- **Synthetic Data:** Our synthetic dataset incorporates twenty-four different modulation types, reflecting the complexities of real-world communication. Notably, high-order modulations prevalent in high-SNR low-fading channel environments are included.
- **Real Gathered Data:** To further enhance the realism of our dataset, we incorporated real-world gathered data with 44,876 frames, each representing different modulations at varying noise levels. The presence of real-world noise introduces challenges that closely mimic practical communication scenarios.

4.1.2. Dataset Structure

The dataset is structured as follows:

- **Size:** In total, our dataset comprises 2,600,780 samples, ensuring a robust representation of diverse modulation scenarios.
- **Split:** We partitioned the dataset into training (80%) and testing (20%) sets, maintaining a balanced distribution to ensure unbiased model evaluation.

4.1.3. Modulation types

Our dataset covers a spectrum of modulation types (See Figures 3 and 4), including:

- **PSK Modulations:** QPSK, 8PSK, 16PSK, 32PSK, 16APSK, 32APSK, 64APSK, 128APSK.
- **QAM Modulations:** 16QAM, 32QAM, 64QAM, 128QAM, 256QAM.
- **Others:** AM-SSB-WC (Amplitude Modulation - Single Sideband – Wideband Carrier).

4.1.4. Synthetic Dataset Details

The synthetic dataset is characterized by:

- **SNR Levels:** Featuring twenty-six levels of Signal-to-Noise Ratio (SNR) for each modulation type, providing a comprehensive range of noise conditions.
- **Frame Composition:** Comprising 4,096 frames for each modulation-SNR combination, with each frame containing 1,024 complex time-series samples.
- **Data Format:** Samples are represented as floating-point in-phase and quadrature (I/Q) components.

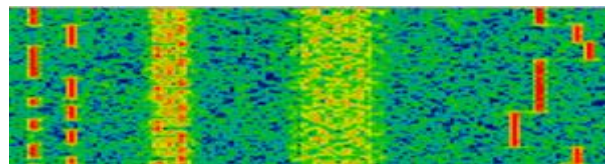


Figure 2. FSK and PSK modulations.

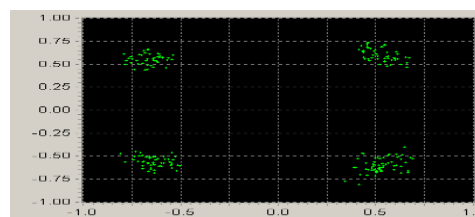


Figure 3. 4 DPSK modulation in constellation representation.

4.1.5. Real Dataset Characteristics

The real dataset introduces authentic challenges:

- **Frame Count:** Containing 44,876 frames, each representing different modulations in the presence of real-world noise.

- **Classification Challenges:** The noise component in the real dataset enhances the difficulty of modulation classification, reflecting the complexities encountered in practical applications. Both synthetic and real datasets were thoughtfully merged into a unified dataset of 2,600,780 samples, ensuring a holistic representation of diverse modulation scenarios.

4.1.6. Technical Implementation

All neural network implementations are constructed using Keras, with Tensorflow serving as the backend, ensuring a robust and standardized framework for model development and evaluation. In summary, our dataset composition, structure, and inclusion of both synthetic and real-world data positions it as a robust foundation for evaluating the performance of our proposed ensemble model under varied and realistic conditions.

4.2. Results for high SNR (ResNet)

After conducting various experiments, it was observed that the ResNet model achieved almost perfect accuracy in classifying the high signal-to-noise ratio (SNR) dataset. The highest accuracy attained by the model was 95.9%, which was achieved at 30dB (Figure 5). Nevertheless, the classification over signals at a low SNR was too modest (35% for -4 dB). This is due to the effect of noise, and it is also related to certain modulation signals which are clearly more difficult to classify due to signal characteristics.

The consistency of our results across all test cases indicates that this deep learning model is robust and generalizable for predicting high SNR signals rather than those in low SNR environments (Figures 7, 6 and 8). Furthermore, when comparing our results to those of other state-of-the-art techniques for high SNR conditions, our proposed ResNet-based method surpasses existing approaches in terms of accuracy.

This showcases the potential of our method as a dependable solution for automatic modulation classification tasks in high SNR conditions.

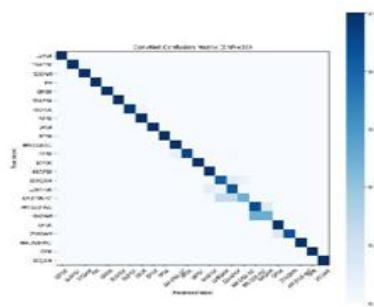


Figure 5. Confusion matrix of the ResNet model at +30 dB SNR.

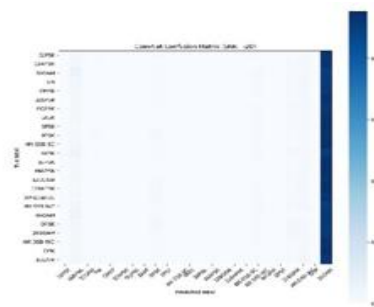


Figure 6. Confusion matrix of the ResNet model at -20dB SNR.

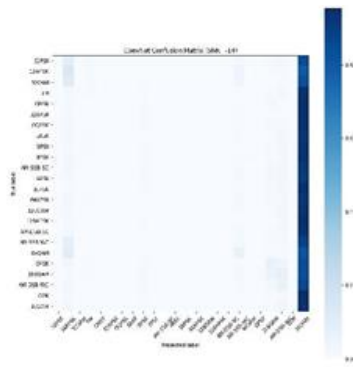


Figure 7. Confusion matrix of the ResNet model at -14dB SNR.

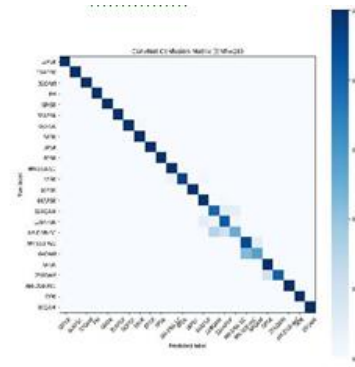


Figure 8. Confusion matrix of the ResNet model at +26dB SNR.

4.3. Results for low SNR (Transformer network)

The Transformer Network (TNN) underwent rigorous testing under challenging low SNR conditions, and the outcomes revealed its impressive capabilities. The model exhibited a remarkable ability to not only reconstruct input signals but also to accurately recognize them, achieving an outstanding accuracy rate of 72.6%. This performance signifies a substantial advancement over previous methodologies that struggled to attain high accuracy rates in low SNR conditions. The Transformer Network's efficacy in such challenging environments establishes it as a breakthrough in the realm of modulation classification under adverse signal-to-noise scenarios. In direct comparison with the ResNet model, which encountered difficulties in detecting signals in low SNR conditions (ranging from -20 dB to 0 dB) and achieved a maximum accuracy of only 20%, the Transformer Network's superiority becomes apparent. This notable contrast highlights the inherent limitations of traditional deep learning models, such as ResNet, when tasked with signal processing in environments with low SNR.

The robustness of the Transformer Network in low SNR conditions can be attributed to its architectural features, particularly the incorporation of self-attention mechanisms. These mechanisms empower the model to selectively focus on relevant components of the input signal while effectively filtering out noise. By intelligently attending to significant parts of the signal, the Transformer Network demonstrates a unique resilience to the challenges posed by low SNR conditions, resulting in a substantial boost in classification accuracy. The success of the Transformer Network in low SNR conditions holds promising implications for real-world applications, particularly in communication systems where noise interference is a prevalent concern. The model's ability to navigate through challenging signal environments positions it as a valuable tool for modulation classification tasks in scenarios where maintaining signal integrity amidst low SNR is crucial.

In conclusion, the Transformer Network's outstanding performance in low SNR conditions, coupled with its architectural strengths, marks a significant stride forward in the development of robust and accurate modulation classification models, particularly in the face of challenging noise-laden communication channels.

4.4. Results of ensemble model

The experimental results of the deep ensemble learning model, depicted in Figures 9, 10, 11, and 12, offer a comprehensive insight into the model's performance across a spectrum of Signal-to-

Noise Ratios (SNRs). The chosen architecture consistently outperforms baseline models, showcasing superior results for both low and high SNR conditions.

The proposed ensemble architecture excels in achieving higher overall accuracy, a notable advantage that becomes apparent when considering diverse SNR scenarios. Figure 11 and Figure 10 depict the model's robust performance in low SNR conditions, while Figure 12 and Figure 9 highlight its proficiency in high SNR environments.

These findings underscore the efficacy of the ensemble learning approach in enhancing the stability and accuracy of the model across varied SNR conditions. The ensemble model's ability to consistently outperform individual baseline models reflects its capacity to adapt and perform optimally under different signal challenges. Notably, our observations reveal a remarkable trend: when the signal-to-noise ratio (SNR) is lower, the classification performance of the ensemble model is approximately 50% greater than that of the single baseline model, ResNet. This substantial performance gain in low SNR conditions highlights the inherent strength of ensemble learning in mitigating the impact of noise and improving classification accuracy when signal clarity is compromised. The observed performance of the ensemble model has significant implications for modulation classification tasks in practical communication scenarios.

The model's ability to maintain high accuracy across a range of SNR conditions positions it as a robust solution for real-world applications, where signal quality can vary widely.

In conclusion, the ensemble learning model's superior performance across different SNR levels signifies its adaptability and resilience in the face of varying signal challenges. These results strengthen the case for employing ensemble learning as an effective strategy for improving the stability and accuracy of modulation classification models, particularly in dynamic communication environments where SNR fluctuations are prevalent.

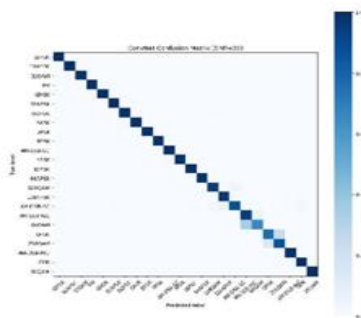


Figure 9. Confusion matrix of the ensemble model at +30dB SNR.

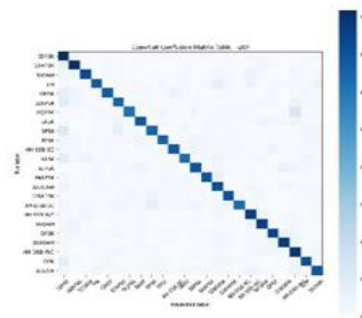


Figure 10. Confusion matrix of the ensemble model at -20dB SNR.

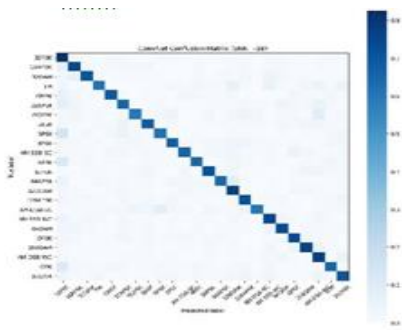


Figure 11. Confusion matrix of the ensemble model at -18dB SNR.

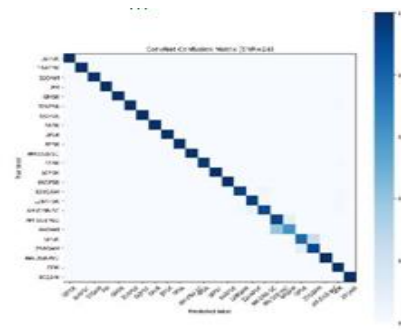


Figure 12. Confusion matrix of the ensemble model at +24dB SNR.

4.5. Advantages in Practical Applications

To elucidate the advantages of our chosen models in practical applications, we consider the following factors:

4.5.1. ResNet in High SNR Environments:

Capturing Spatial Features in Rich Detail:

ResNet's effectiveness in high Signal-to-Noise Ratio (SNR) scenarios is underpinned by its proficiency in capturing spatial features from high-dimensional data. The architecture's unique use of residual connections enables the network to learn intricate patterns and structures in the data. In modulation classification tasks characterized by high SNR and minimal fading, ResNet excels at extracting and interpreting spatial features. This capability is crucial for accurately distinguishing modulation signals within clear, noise-free conditions.

Robust Signal Discernment in Noise-Free Conditions:

In pristine environments with high SNR, ResNet showcases a remarkable ability to discern subtle nuances in modulation signals. The model's capacity to navigate through intricate spatial patterns ensures a high level of accuracy in identifying modulation schemes, contributing to its reliability in scenarios where signal clarity is paramount. The robustness of ResNet in noise-free conditions positions it as a dependable solution for applications where the integrity of the transmitted signal is of utmost importance, such as in high-quality communication channels.

Applicability in Real-World High-SNR, Low-Fading Channels:

Furthermore, ResNet's aptness extends to real-world high-SNR, low-fading channel environments. Its adaptability to varying signal complexities makes it well-suited for scenarios where signal strength is consistently high. This adaptability enhances its applicability in communication systems where maintaining a high SNR is a priority, ensuring reliable performance in conditions akin to those encountered in stable communication channels.

4.5.2. TNN in Low SNR Environments:

Handling Sequential Data with Precision:

The Transformer Neural Network (TNN) emerges as a robust solution for modulation classification tasks in low Signal-to-Noise Ratio (SNR) environments. Its strength lies in its adept handling of sequential data, a characteristic particularly valuable in scenarios marked by low SNR and heightened noise levels. TNN's architecture, based on attention mechanisms, enables it to analyze sequential input signals with precision, allowing for effective extraction of temporal dependencies.

Selective Focus on Relevant Signal Components:

The distinctive feature of attention mechanisms within TNN empowers the model to selectively focus on relevant parts of the input signal. In low SNR conditions, where noise can obfuscate crucial signal components, TNN's ability to discern and prioritize informative sections of the signal proves advantageous. This selective focus contributes to the model's resilience against noise interference, enhancing its accuracy in classifying modulation schemes in challenging, low SNR environments.

Adaptability to Real-World Noisy Communication Channels:

TNN's suitability for modulation classification in low SNR conditions extends to real-world communication channels characterized by noise and interference. Its ability to effectively handle sequential data, coupled with the attention mechanisms, positions TNN as a viable solution for applications where signal degradation due to noise is a prevalent challenge. The model's adaptability in such noisy communication channels highlights its potential for deployment in practical scenarios with varying degrees of signal clarity.

In summary, our choice of ResNet and Transformer Neural Network is informed by a nuanced understanding of their strengths and limitations. While ResNet excels in high SNR conditions, TNN demonstrates superiority in low SNR environments. The ensemble of these models leverages their respective strengths, resulting in a robust solution that exhibits improved stability and accuracy across a spectrum of SNR scenarios.

CONCLUSION

As key part of communication signal processing, automatic modulation classification (AMC) has become increasingly crucial in areas such as cognitive electronic warfare and cognitive radio (CR) with the development of Artificial Intelligence, including Deep Learning, neural networks and others. Its primary goal is to accurately classify the modulated modes of the received signals. This paper proposes an end-to-end deep learning model for modulation signal classification, which uses an ensemble learning network to boost the model's stability and integrate the prediction capacity of several features. Ensemble learning techniques are commonly employed for managing multi-class classification problems and enhancing the overall accuracy of classification. These methods work by improving the functionality of features and promoting each model. Our approach involves leveraging the strengths of two deep learning architectures: ResNet and Transformer network and learning from each other to form a robust algorithmic framework with strong adaptability. Through our experiments, we demonstrated that the proposed deep ensemble method achieves high classification recognition accuracy and stability for both high and low SNRs.

REFERENCES

- [1] W. Su, J. L. Xu, and M. Zhou, "Real-time modulation classification based on maximum likelihood," *IEEE Commun. Lett.*, vol. 12, no. 11, pp. 801–803, Dec. 2008.

- [2] J. L. Xu, W. Su, and M. Zhou, "Likelihood-ratio approaches to automatic modulation classification," *IEEE Trans. Syst., Man, Cybern. C, Appl.Rev.*, vol. 41, no. 4, pp. 455–469, Jul. 2011.
- [3] O. Ozdemir, R. Li, and P. K. Varshney, "Hybrid maximum likelihood modulation classification using multiple radios," *IEEE Communication. Lett.*, vol. 17, no. 10, pp. 1889–1892, Oct. 2013
- [4] S. Majhi, R. Gupta, W. Xiang, and S. Glisic, "Hierarchical hypothesis and feature-based blind modulation classification for linearly modulated signals," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 11057–11069, Dec. 2017.
- [5] V. D. Orlic and M. L. Dukic, "Automatic modulation classification algorithm using higher-order cumulants under real-world channel conditions," *IEEE Commun. Lett.*, vol. 13, no. 12, pp. 917–919, Dec. 2009.
- [6] W. Su, "Feature space analysis of modulation classification using very high-order statistics," *IEEE Commun. Lett.*, vol. 17, no. 9, pp. 1688–1691, Sep. 2013.
- [7] T. O'Shea, J. Corgan, and T. Clancy, "Convolutional radio modulation recognition networks," in *Proc. International conference on engineering applications of neural networks*, Aberdeen, UK, pp. 213-226, Sept. 2016.
- [8] Z. Zhang, HH. Luo, C. Wang, C. Gan, and Y. Xiang, "Automatic modulation classification using CNN-LSMT based dual-stream structure," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 69-70, Nov. 2020.
- [9] S. Huang, Y. Jiang, Y. Gao, Z. Feng, and P. Zhang, "Automatic modulation classification using contrastive fully convolutional network," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1044-1047, Aug. 2019.
- [10] Dietterich, Thomas G. "Ensemble methods in machine learning." *International workshop on multiple classifier systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000.
- [11] Roy, Chirag, et al. "An Ensemble Deep Learning Model for Automatic Modulation Classification in 5G and Beyond IoT Networks." *Computational Intelligence and Neuroscience 2021* (2021).
- [12] Venkata Subbarao, M., and P. Samundiswary. "Automatic modulation classification using cumulants and ensemble classifiers." *Advances in VLSI, Signal Processing, Power Electronics, IoT, Communication and Embedded Systems: Select Proceedings of VSPICE 2020*. Springer Singapore, 2021.
- [13] Le, Ha-Khanh, Van-Sang Doan, and Van-Phuc Hoang. "Ensemble of Con-volution Neural Networks for Improving Automatic Modulation Classification Performance." (2022)
- [14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE conference on computer vision and pattern recognition*, Las Vegas NV, USA, Dec. 2016, pp. 770-778.
- [15] H. Zhang, L. Yuan, G. Guangyu Wu, F. Zhou, and Q. Wu, " Automatic modulation classification using involution enabled residual networks," *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp. 2417-2420, Nov. 2021.
- [16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Sukhin, "Attention is all you need," in *Proc. Conference on Advances in neural information processing systems*, Long Beach California, USA, pp. 30-31, Dec. 2017.

AUTHORS

Mr Yahya Benremdane Engineer and Phd Student, Faculty of Science Ben M'sik, University Hassan II, Casablanca, Morocco As a Telecommunications engineer, Yahya BENREMDANE has been working on signal analysis and proprieties of electromagnetic waves. He's interested in new technologies of radio transmissions, radars and receiver's technologies. Since 2019 he's preparing a thesis on classification of the modulations of telecommunications based on artificial Intelligence, at the Faculty of Science Ben M'sik in Casablanca.



Mr Said JAMAL Engineer and Phd Student, Faculty of Science Ben M'sik, University Hassan II, Casablanca, Morocco Said JAMAL is an engineer and Phd student at the university Hassan II. He's preparing his thesis on the application of AI in the context of the classification of submarine acoustic waves.



Mrs Oumaima TAHERI Engineer and Phd Student, Faculty of Science Ben M'sik ,University Hassan II, Casablanca, Morocco Oumaima Taheri holds a Master's in Modeling and Data Science. With a passion for AI and data-driven solutions, she excels in applying predictive modeling techniques. Oumaima's expertise lies in navigating complex challenges through innovative data analysis, making her a valuable asset in the evolving landscape



Dr Jawad LAKZIZ Phd, Faculty of Science Ben M'sik, University Hassan II, Casablanca, Morocco Dr Jawad LAKZIZ holds a Phd in Modeling acoustic waves in submarine environment. He is currently an Associate researcher at the Faculty of Science Ben M'sik, University Hassan II, Casablanca, Morocco, and guest Editor of the Journal of Marine Technology and Environment. His research interests are modeling acoustic waves, and signal processing, including underwater acoustic classification based on IA .



Pr Said OUASKIT Professor, Faculty of Science Ben M'sik, University Hassan II, Casablanca, Morocco Professor Said OUASKIT is a professor at the Faculty of Science Ben M'sik, University of Hassan II Casablanca. He's an expert in experimental physics, condensed Matter, atomic, molecular and Optical Physics.



A SECURED IMAGE COMMUNICATION WITH DUAL ENCRYPTION AND REVERSIBLE WATERMARKING

Surya Boppanaa, William Kane, and Long Ma

Department of Computer Science, Troy University, Troy, Alabama, USA

ABSTRACT

Secured communication is the optimal means of exchanging information without the risk of data leakage. Data encryption serves as a crucial method for safeguarding and fortifying sensitive information. This paper introduces a groundbreaking approach known as reversible information concealment, specifically designed for digital images. It employs an integer-to-integer wavelet transformation alongside a companding process to embed and retrieve confidential data, restoring the image to its pristine state. Furthermore, the paper explores the utilization of genetic operators in cryptography. Given the prevalence of general information tampering within networks, it becomes imperative to protect messages during transmission. To address this concern, the paper proposes a novel encryption technique leveraging genetic operators such as crossover and mutation. This method ensures message confidentiality during the transmission process, contributing to an enhanced and more secure environment. The overarching goal is to establish a robust security framework through the integration of encryption and digital image watermarking for discreet data concealment within images.

KEYWORDS

Genetic Algorithm, Cryptography, Adaptive Thresholding, Companding Technique, Integer Wavelet Transform, Reversible Watermarking work Protocols

1. INTRODUCTION

Secured communication is achieved when two corresponding entities maintain no exposure, a pivotal factor in our contemporary lifestyle where data transmission serves as the linchpin for numerous societal functions. Acknowledging the profound impact of this interconnection, the imperative of securing data takes center stage as a primary technological responsibility. Tackling this challenge involves the strategic application of encryption to restore the original, un-watermarked image—an indispensable process highly valued in document imaging services catering to diverse enterprises. A related facet is cryptography, ensuring exclusive access to encrypted information for understood and predefined authorized users. Cryptography establishes qualified security across federal, economic, and social network domains [1]. Integrating these functions guarantees comprehensive, state-of-the-art image security, facilitating optimal utilization.

Watermarking in a digital image serves to obscure and safeguard encoded data linked to the original documenter, ensuring comprehensive accreditation. However, issues often arise, particularly with complex documents, where the original shape becomes distorted during the data installation phase. Mitigating this challenge proves challenging due to various factors such as bit substitution, quantization, and truncation. While some models acknowledge these struggles,

certain fields, including medical and military enterprises, cannot afford such inaccuracies. Inconsistencies may lead to legal troubles, altering the perception of a doctor's findings or causing diplomatic and physical consequences for military maps. Reversible watermarking offers a solution in such scenarios, securely and accurately encoding and displaying a digital document with its original intent intact. A genetic algorithm [3] is a heuristic search based on natural hypothesized selection. This algorithm has five stages, which consist of three critical operators. These key functions are understood as population generation, crossover, and mutation. To kickstart this process is population generation, which brings about communicated binary or hex chromosomes for the introduction. The crossover phase is concerned with applying the operator of the same name to people in their current state to result in a parent solution. A mutation process is incorporated to achieve genetic diversity across species using a mutation director to achieve a fixated genetic standard. Some research has been done in this field as in 2012, Ankita Agarwal presented an encryption method that pairs the highlighted genetic algorithm with a genetic operator to achieve a fulfilled encrypted secret key image [4]. With research continuing in 2014, Sindhuja K and Pramila Devi S proposed a method to encrypt data using the operator's right shift, matrix addition, modulo operation, and genetic modifiers [5]. While advancements have been made in completely reversing a watermark on an image, in 2011, N.A. Memon brought about a stable method for showing the opposite hand of clear-cutting a watermark for an image [16].

This study introduces a strategic approach to reverse watermarking digital images using cryptography and paired algorithms for enhanced security and noise reduction. The outcome is a digitally watermarked image of superior quality and reliability. The work employs genetic operators and encryption methods in cryptography to fortify data movement. The PSNR block calculates the peak signal-to-noise ratio between two images, serving as a quality measure. MATLAB, a versatile numerical-performance language, is utilized for its computational, visualization, and programming capabilities. Our proposed method utilizes two-dimensional images for increased reliability. Digital image watermarking conceals information about the content creator or owner for authentication or copyright protection. However, the standard limitation of distortion in the original image is addressed through decryption and reversible watermarking, ensuring lossless data transfer and recovery, particularly in medical images.

2. PROPOSED APPROACH

Our proposed methodology elucidates the principles of fixed threshold-based companding, encompassing the establishment of a threshold framework, watermark creation, and the seamless embedding of the watermark into an image. These fundamental components form the groundwork for achieving reversible watermarking and adeptly concealing data within the target image. Each block within this process encounters an adaptive resolution of limits, strategically crafted to incorporate thresholds specific to individual blocks. The subsequent section provides a comprehensive breakdown of the constituent elements in our approach:

2.1. Fixed Threshold-Based Companding

Companding is the method of compressing a signal followed by extending said signal. Let Y be a compression function and X be an expansion function. For a signal f , Y and X have the accompanying relationship: $Y(X(f)) = f$. For a digital signal, Y_q and X_q represent the quantized versions for Y and X individually, and q indicates the quantization work. The compression function Y_q is given by [16]:

$$1. \quad f_q = Y_q = \begin{cases} f & |f| < TH \\ \text{sgn}(f) \times (|f| - TH/2) + TH & |f| \geq TH \end{cases}$$

where $\text{sign}(\cdot)$ is the sign function, and TH is a pre-defined fixed threshold. The expansion function Y_q is given by:

$$2. \quad Y_q(f) = \begin{cases} f & |f| < TH \\ \text{sgn}(f) \times (2|f| - TH) & |f| \geq TH \end{cases}$$

Companding values via equations (1) and (2) produce expected results if $|f| < TH$. However, when $|f| \geq TH$, the companding error is produced:

$$3. \quad z = |f| - |x_q(Y_q(f))| \quad \text{where } Z \in \{0,1\}$$

A structured understanding of the Companding watermarking technique is as follows:

- 1) Compression function Y is applied to the original signal f to obtain a new signal $l = Y(f)$. Assume the binary expression of l is $p_1 p_2 \dots p_n$, where $p_i \in \{0,1\}$.
- 2) A bit $b \in \{0,1\}$ is appended after the least significant bit (LSB) of l . In this way, l becomes $l' = p_1 p_2 \dots p_n b$ which can mathematically be expressed as: $l' = 2 \times l + b$.
- 3) In the data extraction stage, we only need to extract the LSB bit from the received signal h' , which means $b = \text{LSB}(l')$. The signal l can thus be recovered by the expression $l = l' - b/2$.
- 4) After obtaining the signal h , we can recover the original signal by expression $f = Y_q(l) + Z$, where Z is a companding error.

2.2. Watermark Creation

The prospective method creates the watermark from the following segments.

2.2.1. The Error Vector (Z)

If the coefficient estimation is more unique than or equivalent to the client-characterized limit, it is compressed (equation 1) and extended upon (equation 2). To have the capacity to recoup the first picture precisely, it is vital to gather the companding mistakes (equation 3) to recoup the entirety of the first picture. These mistakes are aggregated in vector Z [16].

2.2.2. Payload (P)

The payload can be defined as the communication to client characterized data, which can be any amount of mystery data identified with a picture [16].

2.3. Embedding Watermark into an Image

Partitioning the info picture understood as "I" into squares of size $S \times S$. Figure the 2D IWT (Integer Wavelet Transform) of each block (i,j) up to level 2 [16]. Acquire the threshold $T(i,j)$ from THMAP for the corresponding block (i,j) and therefore, apply the compression function on the coefficients of each sub-band (HL1, LH1, HH1) given the limit $TH(i,j)$. Thus, implant the watermark W into a block (i,j) utilizing condition $l' = 2 \times l + b$. and figure backward IWT to get the watermarked block (i,j) . The operation proceeds until the point where W is installed is put into the blocks, and THMAP should be embedded in the picture to facilitate active recovery. This action is performed because the embedding has been drawn out in each block with an alternate edge, which leads to TMAP being compressed and utilizing math encoding to decrease its size

fundamentally. Finally, compacted THMAP is embedded in level 2 coefficients (HL2, LH2, HH2) regardless of blocks using THMIN as well as other data implanted alongside THMAP as the extent of block, i.e., S . The stamped picture “I” in this manner acquired is the last watermarked picture.

Author names are to be written in 13 pt. Times New Roman format, centred and followed by a 12pt. paragraph spacing. If necessary, use superscripts to link individual authors with institutions as shown above. Author affiliations are to be written in 12 pt. Times New Roman, centred, with email addresses, in 10 pt. Courier New, on the line following. The last email address will have an 18 pt. (paragraph) spacing following.

2.4. Watermark Extraction

Process 2D IWT of image ‘I’ engages with the CDF channels and breaks down the image to level 2 to get HL2, LH2, and HH2 wavelet sub-groups. Demanding the realized threshold estimation of TMIN extract as the figured TMAP and data utilization of in the relation $b = \text{LSB}(l')$. Uncompressed TMAP-related data to discover TMAP and square size will lead to introducing B with block-size data. Separation of the watermarked image “I” into $S \times S$ measure blocks. Presently the process of the IWT of each block (i,j) what's more, perform disintegration up to level 1, and for each block (i,j) , locate the corresponding threshold $T(i,j)$ from THMAP. Concentrate the watermark from the wavelet coefficients of each block utilizing $b = \text{LSB}(l')$ and collect all slightest significant bits utilized before recuperating the watermark bitstream W' . Decompress the bitstream W' utilizing mathematical unraveling calculation to restore the first-bit stream, and when “W” has been decompressed, the blunder vector Z' and payload P' can be recovered to form. Next would be restoring coefficients by utilizing condition, $l = l' - b/2$ and getting the first coefficients by acting on the condition $f = Y_q(h) + Z$. The first picture is then acquired by taking the backward IWT of reestablished coefficients.

2.5. Creating Threshold Outline (THMAP)

Each block threshold can have its underlying details checked through the proposed method ahead of the actual watermark. To find the THMAP which is embellished as pursues is to use a block graph as follows: Introduce THMAP to an $(M/S) \times (N/S)$ zero framework, as S is the client characterized in the size of the block and M and N are the tallness and broadest of the info picture individually. Instate THINIT, THMIN, and PSNRMAX (Peak Signal to Noise Ratio) with the client-characterized values and set the estimation of TH as THINIT. Apply the compression work utilizing equation (1) on all parameters, even on vertical and slanting sub-band coefficients that are not as much as TH. Next, install the watermark in all sub-band coefficients independent of TH utilizing the equation $l' = 2 \times l + b$, where l speaks to the first coefficient while b is the watermark bit to be implanted. Process the converse IWT of the block (i,j) to get block (i,j) where I, j are line and section files of a block individually. [16] The PSNR of block (i,j) is figured on the possibility of being malleable. If PSNR is observed to be more prominent or equivalent to the most extreme permitted PSNRMAX, the limit TH is recorded in THMAP. Presently decline the estimation of TH by one and this decrement in TH will increment the PSNR of the block in the wake of embedding the watermark. This is because of the way that when TH is little, to an ever-increasing extent coefficients are companded, and in this manner, installing twisting will be negligible, and the excellent visual nature of the stamped picture is accomplished. The emphasis will proceed till TH is equivalent to THMIN; thus, we obtain the lattice THMAP containing edge estimations of each block of information picture depending upon the properties of that block [16].

3. A SECURED PLATFORM

A comprehensive MATLAB tool consolidates all necessary components onto a unified platform, offering a seamless representation of complete information transmission with the added security of a Graphics Processing Unit (GPU) or a graphical user interface.

3.1. Watermarked Image Encryption

Textual data is given a designated space where it will undergo encryption. An image is selected to employ encrypted textual data. Here, the encrypted data is embedded into the image through the watermarking method. At last, this watermarked image will again go through another image encryption. Therefore, the encrypted image with hidden encrypted textual data will be transmitted to the desired destination.

3.2. Decryption and Data Recovery

The watermarked encrypted image is selected to wrest the image and data. Initially, the image is decrypted, and the original image is extracted through a reversible watermarking method. Now, the encrypted textual data is extracted from the image. The decryption process takes place using a key, which is generated during the time of encryption. At the time of decryption, selection of this key to decrypt the data. Therefore, we can see the original textual data that is transmitted originally.

This proposed approach is calculated only for the textual data. If any numerical data is brought into the equation, it will generate an error. Therefore, no numerical data is to be used in this approach. Figures 1 and Figure 2 are examples of the above-written data.

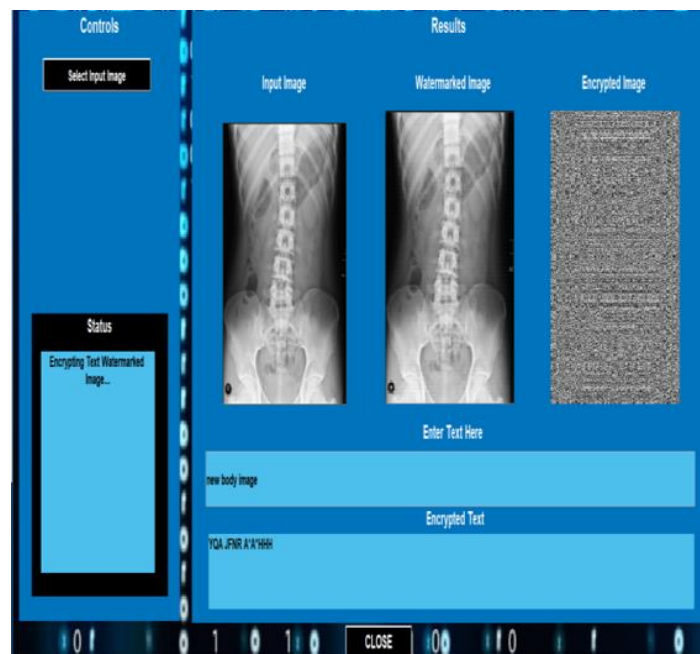


Figure 1. Embedding text and Encrypting the Image

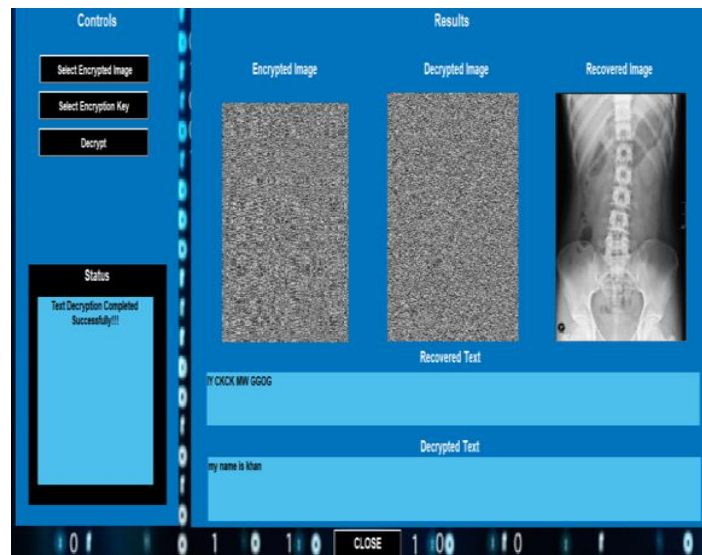


Figure 2. Decryption and extracting the original textual data

3.3. Recovery of Original Image

The following steps recover the original image:

1. Read the watermarked image I'
2. Iteration begins
3. Divide the watermarked image I' into blocks of size $S \times S$
4. Compute the 2D IWT of a block (i,j) using CDF filters (Cohen, Daubechies, Feauveau) and decompose it up to 2nd level
5. Recover the coefficients by using $l = l' - b/2$
6. Expand the coefficients by using
7.
$$Y_q(f) = \begin{cases} f & |f| < TH \\ \text{sgn}(f) \times (2|f| - TH) & |f| \geq TH \end{cases}$$
8. The first-level coefficients are expanded by using threshold $TH(i,j)$ and second-level coefficients with $THMIN$
9. Obtain the original coefficients by adding the companding errors in recovered coefficients using $f = Y_q(h) + Z$.
- 10.. Compute the inverse IWT to get the original block (i,j)
11. Iteration ceases
12. The process will be continued until all the blocks are processed. Therefore, the resultant image will be the same as the original image

4. EXPERIMENTAL RESULTS

Our experiment uses standard 2D pictures of X-ray, CT scan, Ultrasound, and MRI with size 512×512 . The first encrypted and watermarked adaptations of these pictures appear in Fig. 1, and the settled qualities of $THMIN$ and $PSNRMAX$ are heuristically set to 2 and 42.0 dB separately. Be that as it may, the estimation of $THINIT$ is chosen in the scope of $\{2-15\}$. The value of $PSNRMAX$ can be divided according to a requested dimension of imperceptibility utilized in a specific application. The evaluation of $PSNRMAX$ straightforwardly controls the nature of the watermarked picture. For watermarked pictures appearing in Fig. 1, the initial threshold ($THINIT$) is 15, and the block size is set to 8×8 .

The suggested technique grants improved ambiguity regarding the PSNR for a similar payload. This area of enhancement is prominent at low- and high-level payloads. The distinction or enhancement is more if there should be an existence of finished pictures. This claims the installation is performed on high recurrence, which gives an abundance of inserting space in finished pictures and is utilized by the nearby limit adjustment. Altogether, we connected every one of the strategies into one stage with a GUI in which we can take the necessary steps from encryption of literary information, watermarking of pictures, information inserting, and encryption of printed information as every yield will be unmistakable on a similar GUI screen. Therefore, in Fig. 3, we can see the original image and watermarked images with embedded textual data, including encrypted images and encrypted textual data. Following this, we can send a message through the encryption and watermarking process to send the data. At the same time, we can decrypt the same message by extracting the original image from the watermarked image, resulting in the finalized uncovered image. Fig. 4 displays the comparative results of X-ray images for different block sizes.

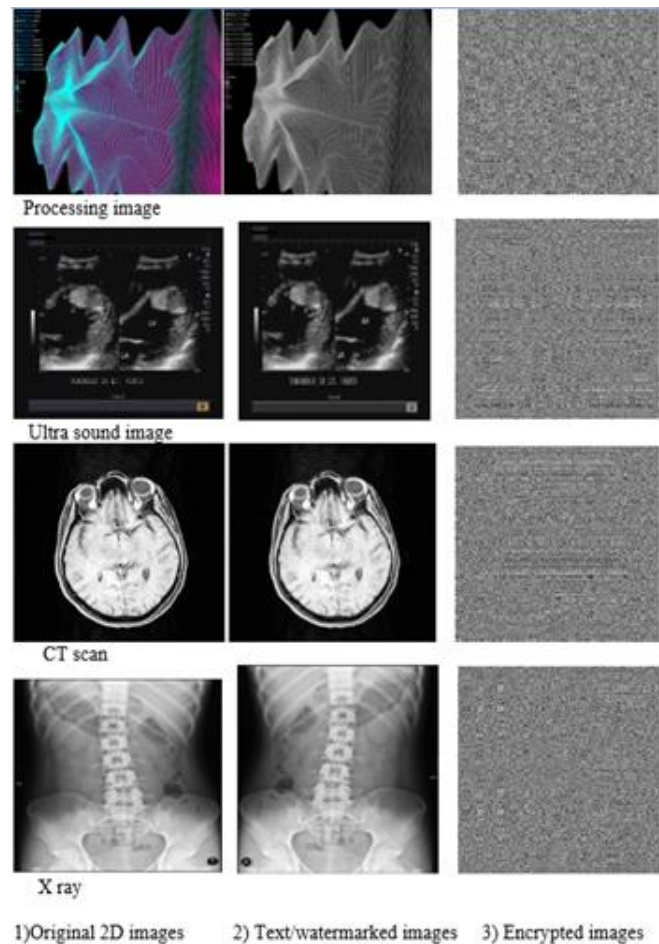


Figure 3. Types of 2-Dimensional Images with Watermarking and Encryption

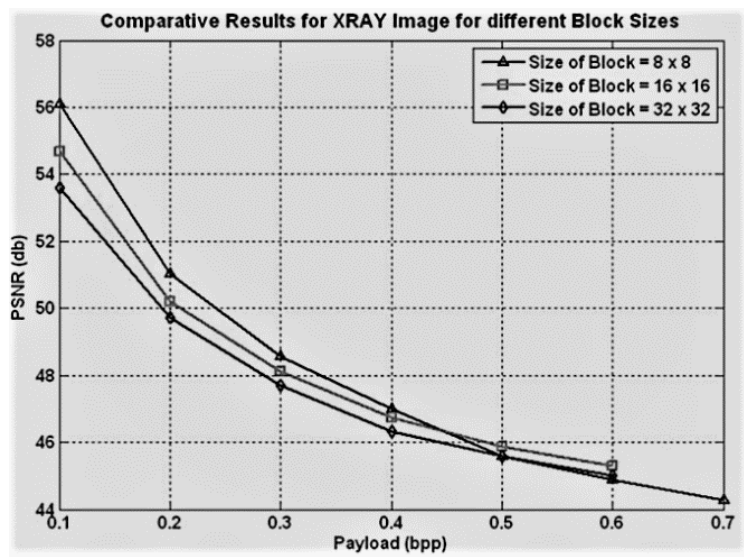


Figure 4. Results of X-ray images on different block sizes

5. CONCLUSIONS

This paper introduces a straightforward cryptographic framework employing genetic algorithm operators for encryption, ensuring robust security. The incorporation of frequent binary and decimal transformations enhances its overall capability. The suggested reversible watermarking system, reliant on companding and adaptable thresholding, exhibits potential applications in image comparison within military, medical, and law enforcement domains. The approach involves block-based watermarking and iterative threshold progression to uphold histogram quality efficiently. The method ensures secure data coverage, rendering transactions impervious to external threats, making it versatile across diverse industries. Additionally, it can be customized to integrate both textual and numerical data. This conceptual development has the potential for extension into the realm of artificial intelligence, offering a reduction in data transfer time complexity.

REFERENCES

- [1] Behrouz A. Forouzan, *Cryptography & Network Security*—, Tata McGraw – Hill, 2007.
- [2] William Stallings, *Cryptography and Network Security*l, 3rd Edition.
- [3] S., N. Sivanandan, S. N. Deepa, *Introduction to Genetic Algorithms*, Springer Verlag Berlin Heidelberg, 2008.
- [4] Ankita Agarwal, —Secret Key Encryption Algorithm Using Genetic Algorithml, *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 2, Issue 4, April 2012.
- [5] Sindhuja K, Pramila Devi S, A Symmetric Key Encryption Technique Using Genetic Algorithml, *International Journal of Computer Science and Information Technologies*, Vol. 5 (1 2014).
- [6] Amritha Thekkumbadan Veetil, An Encryption Technique Using Genetic Operators, *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 4, ISSUE 07, JULY 2015*
- [7] S. Hai-mei, M. Tian-can, Q. Qian-ging, “Spread Spectrum Watermark based on Wavelet Transform for Still Digital Image,” *Wuhan University Journal of Natural Sciences*, vol 9, No. 2, pp. 203-208, 2004.
- [8] A. Khan, A.M. Mirza, “Genetic perceptual shaping: utilizing cover image and conceivable attack information during watermarking embedding, *Information*,” *Fusion*, vol. 8, pp. 354-365, 2007.
- [9] A. Khan, “Intelligent perceptual shaping of a digital watermark.” PhD Thesis, Faculty of Computer Science and Engineering, GIK Institute, Pakistan, 2006.

- [10] J. H. K. Wu, R.F. Chang, C. J. Chen, C. L. Wang, T. H. Kuo, W. K. Moon, W. K. Chen," Tamper detection and recovery for medical images using near-lossless information hiding technique," *Journal of Digital Imaging*, vol. 21, no. 1, pp-59-76, 2008.
- [11] N. A. Memon, S.A.M . Gilani, " NROI watermarking of medical images for content authentication," In the Proceedings of 12th IEEE International Mutitopic Conference (INMIC'08), Karachi, Pakistan, 2008. p. 106- 110.
- [12] S.I. Fraser, A.R. Allen," A high capacity reversible watermarking technique based on difference expansion," In the Proceedings of Signal and Image Processing, Kailua-Kona, HI, USA, 2008.
- [13] J. Tian, "Reversible watermarking by difference expansion," In Proceedings of Workshop on Multimedia and Security, 2002. p. 19-22.
- [14] G. Xuan, Y.Q. Shi, C. Yang, Y. Zheng, D. Zou, P. Chai," Lossless data hiding using wavelet transform and threshold embedding technique," *Proceedings of IEEE International Conference on Multimedia and Expo*, 2005. p. 1520-1523.
- [15] G. Xuan, C. Yang, Y. Zhen, Y.Q. Shi, Z. Ni," Reversible data hiding using IWT and companding technique," I. J. Cox et al. (Eds.) *IWDW 2004, LNCS 3304*. p. 115-124. World Academy of Science, Engineering and Technology 55 2011 620.
- [16] Nisar Ahmed Memon, A Novel Reversible watermarking method based on adaptive thresholding and companding technique, *World Academy of Science, Engineering and Technology* 55 2011

G-KMM: A flexible kernel mean matching optimization method for density ratio estimation involving multiple train & test datasets

Cristian Daniel Alecsa

Technical University of Cluj-Napoca, Cluj-Napoca, Romania
Romanian Institute of Science and Technology, Cluj-Napoca, Romania

Abstract. In the present paper we introduce new optimization algorithms for the task of density ratio estimation. More precisely, we consider extending the well-known KMM (kernel mean matching) method using the construction of a suitable loss function, in order to encompass more general situations involving the estimation of density ratio with respect to subsets of the training data and test data, respectively. The codes associated to our Python implementation can be found at <https://github.com/CDAlecsa/Generalized-KMM>.

Keywords: Kernel mean matching, quadratic optimization, density ratio estimation, loss function.

1 Introduction

In statistical data processing, the comparison of two distributions is of paramount importance. In general, the problem of assessing whether two probability distributions are equivalent or not is addressed through the so-called two-sample tests. There exists classical methods that tackle this issue, such as the *t-test* which compares the means of two Gaussian distributions with common variance, and the well-known non-parametric *Kolmogorov-Smirnov test*. Recently, in [1], Gretton et. al. introduced the *maximum mean discrepancy* (MMD) statistic which compares the similarities through a positive-defined kernel across two samples in an universal reproducing kernel Hilbert space (universal RKHS), and is commonly used for multivariate two-sample testing. In [2], Kirchler et. al. extended the MMD statistic by using a neural network trained on an auxiliary dataset which defines the so-called *deep maximum mean discrepancy* (DMMD) statistic, where the mapping from the input domain to the network's last hidden layer is utilized as the kernel used in the MMD statistic.

A different approach to the problem of two-sample testing is based upon the evaluation of a divergence between two distributions, such as the *f-divergence* which includes the case of the *Kullback-Leibler divergence* and the *Pearson divergence*, respectively. Due to the fact that the density estimation is a hard task, a practical approach to the divergence estimation is to directly approximate the density ratio function. One of the most well known method is the *kernel mean matching* (KMM) algorithm from [3] based on infinite-order moment matching, and which represents the MMD statistic in the particular case when a distribution is weighted accordingly to the density ratio model. There are several extensions of the KMM method such as the *Ensemble KMM* introduced in [4] where, instead of a single train-test split, one uses multiple non-overlapping test datasets and a single train dataset. A more generalized version was introduced in [5] where the main idea is to divide the training data due to the fact that the *Ensemble KMM* is not suitable with large training datasets. Consequently, this method which we will call it *Efficient Sampling KMM*, takes a bootstrap approach for the training data and then merges the results with

an aggregation process. A combination between the aforementioned two methodologies was done in [6] where Haque and his coauthors considered a KMM-type density ratio estimation called SKMM based on using bootstrap generation method for the training data and a partitioning of the test data. An extension of the classical *KMM* method is the neural network technique introduced [7] where the loss function is the actual MMD objective and the bandwidth of the underlying kernel is considered as a hyper-parameter. Due to the inherent flexibility of neural architectures and that the training is done on randomized batches this Deep Learning approach outperforms the classical KMM algorithms.

Different alternatives to MMD-based methods mostly rely on solving different *optimization problems*. Such examples are the *unconstrained least squares importance fitting* (uLSIF) from [8] and the *relative uLSIF* (RuLSIF) method introduced in [9], where in the latter one the density ratio model involves a mixture of densities. The uLSIF method was successfully employed in [10] for the comparison of distributions using a permutation test approach. In more detail, by utilizing a weighted Gaussian kernel model at test samples, the weights of the density ratio model are learned through a *quadratic optimization problem*, after which the *Pearson divergence* is employed in order to compare the train and test distributions. A general method encompassing the constrained variant of the uLSIF method, namely LSIF, is the Bregman formulation given in [11]. The unified method which utilizes the Bregman divergence contains as a particular case the well-known *KL importance estimation procedure* (KLIEP) from [12] for which the *objective function* is given with respect to the test samples, while the constraints depend on the training samples. When the true density ratio is approximated through a linear or kernel model, then one obtains a *convex optimization problem with constraints*. On the other hand, for the situation when the choice of the density ratio model is the *log-linear model* as in [13] then the underlying density fitting framework reduces to a *unconstrained convex optimization problem* and therefore the global solution can be obtained by iterative methods such as *gradient descent*.

As previously mentioned, the KMM method introduced in [7] uses the MMD statistic as the *loss function* for the density ratio approximation which is modeled through a neural network. A similar approach was developed recently in [14] in which a neural-type approach was developed for the RuLSIF method in the setting of change point detection. In both these works, the true density ratio is represented as a neural network and the learning of such a network depends on a loss function suitable for density ratio estimation. A different approach for the RuLSIF method is the one from [15] where the density ratio is not directly represented as a neural network but is considered as a weighted feed-forward neural model (instead of a weighted kernel model). By employing the RuLSIF approach one finds at each iteration the weights as the global optimum solution for the quadratic optimization problem. After finding the weights, the classical backpropagation algorithm is applied to the density ratio model with respect to the parameters of the feed-forward neural model. In addition to this, the aim of the method introduced in [15] is to estimate the density ratio from a few training samples, by using instances in different but related datasets (also called source datasets).

As we formerly emphasized, the framework regarding the density ratio models *Ensemble KMM* from [4] and *Efficient Sampling KMM* from [5] depends on multiple train or test datasets. On the other hand, the density fitting methodology proposed in [16] (which we shall briefly call it *MultiDistribution DRE*) is related to the idea that one has access to i.i.d. samples from multiple distributions. As a particular case, this can be perceived as having multiple test datasets and a single reference train dataset. The purpose is to estimate the density ratios between all pairs of distributions. This can be efficiently done by employing the Bregman divergence with respect to the reference density function and thus optimizing

a vector density ratio. Consequently, the optimization function can be written as a sum of multiple objective mappings, where each of them depends on a particular density ratio component. Accordingly, this approach generalizes the LSIF and KLIEP optimization algorithms to the so-called Multi-LSIF and Multi-KLIEP methods, respectively.

There are various alternatives to the aforementioned density ratio methods. The most well known approach to the previously mentioned techniques is the probabilistic density fitting method where, as described in [17], one learns a probabilistic classifier that separates the train and test samples. The methodology behind these classification algorithms is that the density ratio is approximated by the ratio of the sample sizes multiplied by the class posterior probabilities, the latter ones being obtained from the classifier's output. Furthermore, the main advantage of the probabilistic classification technique is that it is easy to implement it in a real world situation.

We end this section with the table (1), in which we describe the investigation that was done in different papers which are in connection to our theoretical and empirical results.

Articles	Methodology
Sugiyama et al. (2011) [10]	- comparison of distributions using permutation tests based on the uLSIF method
Miao et al. (2015) [4]	- <i>Ensemble KMM</i> method which utilizes multiple non-overlapping test datasets and a single train dataset
Chandra et al. (2016) [5]	- <i>Efficient Sampling KMM</i> method based upon a bootstrap approach for the training data which merges the results with an aggregation proces
Haque et al. (2016) [6]	- SKMM method which uses a using bootstrap generation method for the training data and a partitioning of the test data
Hushchyn & Ustyuzhanin (2021) [14]	- a RuLSIF type neural network model for change point detection tasks
Yu et al. (2021) [16]	- the aim is to estimate the density ratios between all pairs of distributions.
de Mathelin et al. (2022) [7]	- an extension of the classical KMM method to neural networks

Table 1: Recent research contributions

2 Motivation

Let's consider two sample sets $\mathcal{X} = \{x_i \mid x_i \in \mathbb{R}^d\}_{i=1}^n$ and $\mathcal{X}' = \{x'_j \mid x'_j \in \mathbb{R}^d\}_{j=1}^{n'}$ such that $\mathcal{X} \stackrel{i.i.d.}{\sim} P$ and $\mathcal{X}' \stackrel{i.i.d.}{\sim} P'$, where P and P' are probability distributions with densities p, p' , respectively.

The MMD (maximum mean discrepancy) statistic between \mathcal{X} and \mathcal{X}' is defined as

$$MMD_{\phi, \psi}(\mathcal{X}, \mathcal{X}') = \|\mathbb{E}_{p'(x)}[\psi(x)] - \mathbb{E}_{p(x)}[\phi(x)]\|^2,$$

where $\phi, \psi : \mathbb{R}^d \rightarrow \mathbb{R}^p$ are two given feature maps. By defining the density ratio function $r(x) = \frac{p(x)}{p'(x)}$, where we assume that $p'(x) > 0$ for all x , and choosing $\psi(x) = r(x)\phi(x)$, then we obtain the loss function used in the KMM (kernel mean matching) approach from [3] with respect to an approximation \hat{r} of r :

$$MMD_{\phi, \hat{r}\phi}(\mathcal{X}, \mathcal{X}') = \|\mathbb{E}_{p'(x)}[\hat{r}(x)\phi(x)] - \mathbb{E}_{p(x)}[\phi(x)]\|^2$$

By using that

$$1 = \int p(x)dx = \int r(x)p'(x)dx = \mathbb{E}_{p'(x)}[r(x)]$$

and taking into account the above objective function, one obtains the following *optimization problem with constraints* (that needs to be solved by considering an approximation of the density ratio model \hat{r}):

$$\left\{ \begin{array}{l} \min_{\hat{r}} \|\mathbb{E}_{p'(x)}[\hat{r}(x)\phi(x)] - \mathbb{E}_{p(x)}[\phi(x)]\|^2 \\ \text{subject to } \begin{cases} \hat{r}(x) \geq 0 \text{ for all } x \\ \mathbb{E}_{p'(x)}[\hat{r}(x)] = 1 \end{cases} \end{array} \right. \quad (\text{OptPb-KMM})$$

For simplifying the formulation of (OptPb-KMM), we introduce the kernel map $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ such that $K(x, y) = \langle \phi(x), \phi(y) \rangle$. Furthermore, let's consider $h \in \mathbb{R}^{n'}$ such that $h_j = \frac{n'}{n} \sum_{i=1}^n K(x'_j, x_i)$ for $j = \{1, \dots, n'\}$. At the same time, define $H \in \mathbb{R}^{n' \times n'}$ such that $H_{jk} = K(x'_j, x'_k)$ for $j, k \in \{1, \dots, n'\}$, along with $r_{\mathcal{X}'} \in \mathbb{R}^{n' \times 1}$, where $r_{\mathcal{X}'} = (r(x'_1), \dots, r(x'_{n'}))^T$. If we define $\hat{r}(x)$ as a model approximating the true density ratio $r(x)$, and ignoring irrelevant constants with respect to $\hat{r}(x)$ then (OptPb-KMM) becomes the following *quadratic optimization problem with constraints*:

$$\left\{ \begin{array}{l} \min_{\hat{r}_{\mathcal{X}'}} \left(\frac{1}{2} \hat{r}_{\mathcal{X}'}^T H \hat{r}_{\mathcal{X}'} - h^T \hat{r}_{\mathcal{X}'} \right) \\ \text{subject to } \begin{cases} (\hat{r}_{\mathcal{X}'})_j \geq 0 \text{ for } j \in \{1, \dots, n'\} \\ \frac{1}{n'} \sum_{j=1}^{n'} (\hat{r}_{\mathcal{X}'})_j = 1 \end{cases} \end{array} \right.$$

A different kind of density ratio technique is the RuLSIF method from [9] which uses a generalization of the density ratio i.e., the α -relative density ratio $r_\alpha(x) = \frac{p(x)}{q_\alpha(x)}$ for $\alpha \in [0, 1)$, where $q_\alpha(x)$ represents the α -mixture density of $p(x)$ and $p'(x)$ i.e., $q_\alpha(x) = \alpha p(x) + (1 - \alpha)p'(x)$. In the RuLSIF method, one models the true density ratio as a linear kernel method with weights $w \in \mathbb{R}^{n \times 1}$, namely $\hat{r}(x) = w^T k(x)$, where $k(x) = (K(x, x_1), \dots, K(x, x_n))^T \in \mathbb{R}^{n \times 1}$ and $K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\tilde{\sigma}^2}\right)$ is the corresponding Gaussian kernel with the variance $\tilde{\sigma}^2$, respectively. Let's define $H \in \mathbb{R}^{n \times n}$ and $h \in \mathbb{R}^{n \times 1}$, namely $H_{l,k} = \frac{\alpha}{n} \sum_{i=1}^n K(x_i, x_l)K(x_i, x_k) + \frac{1 - \alpha}{n'} \sum_{j=1}^{n'} K(x'_j, x_l)K(x'_j, x_k)$ and $h_l = \frac{1}{n} \sum_{i=1}^n K(x_i, x_l)$, where $l, k \in \{1, \dots, n\}$. By ignoring constants irrelevant to the weights w and using a regularization parameter λ , we obtain the following *unconstrained regularized quadratic optimization problem* corresponding to the RuLSIF loss:

$$\min_w \left(\frac{1}{2} w^T H w - h^T w + \frac{\lambda}{2} w^T w \right) \quad (\text{OptPb-RuLSIF})$$

In order to generalize the KMM approach to multiple sample sets, we observe that the objective function belonging to (OptPb-KMM) leads to

$$\mathbb{E}_{p(x)}[\phi(x)] = \int \phi(x)p(x)dx = \int r(x)\phi(x)p'(x)dx = \mathbb{E}_{p'(x)}[r(x)\phi(x)], \quad (1)$$

therefore, under the true density ratio model $r(x)$, the loss function is minimized. We will use this simple technique for extending in a precise manner the KMM algorithm to multiple sample sets.

- The first case we investigate is when we have, for $i \in \{1, \dots, N\}$, the sets $\mathcal{X}_i = \{x_{k,(i)} \mid x_{k,(i)} \in \mathbb{R}^d\}_{k=1}^{n_i}$ and $\mathcal{X}' = \{x'_j \mid x'_j \in \mathbb{R}^d\}_{j=1}^{n'}$ such that $\mathcal{X}_i \stackrel{i.i.d.}{\sim} P_i$ and $\mathcal{X}' \stackrel{i.i.d.}{\sim} P'$, where P_i and P' are probability distributions with densities p_i, p' , respectively. If we consider \mathcal{X}' to be the training dataset and \mathcal{X}_i to represent the non-overlapping partitions of a given test dataset \mathcal{X} , then the technique proposed in *Ensemble KMM* uses the fact that $p(x \in \mathcal{X}) = \sum_{i=1}^N \frac{n_i}{n} p(x \mid x \in \mathcal{X}_i)$, where $n = \sum_{j=1}^N n_j$. Therefore, the probability associated to the test sample set \mathcal{X} can be written as a mixture between non-overlapping partitions with the weights given by the ratio between the size of the partition and the total size of the test dataset. This is in accordance with the definition from the formulation of *Ensemble KMM* of the density ratio corresponding to \mathcal{X} and \mathcal{X}' which is given by a mixture of density ratios between \mathcal{X}_i and \mathcal{X}' , respectively. By dropping the notation of conditional probability density concerning the partitions of \mathcal{X} , let's consider the general case when $r(x) = \sum_{i=1}^N \omega_i r_i(x)$ where $r_i(x) = \frac{p_i(x)}{p'(x)}$ for $i \in \{1, \dots, N\}$ and where the weights satisfy $\sum_{i=1}^N \omega_i = 1$ with $\omega_i \in [0, 1]$ for $i \in \{1, \dots, N\}$. Therefore

$$r(x) = \sum_{i=1}^N \omega_i r_i(x) = \sum_{i=1}^N \omega_i \frac{p_i(x)}{p'(x)} = \frac{\sum_{i=1}^N \omega_i p_i(x)}{p'(x)} = \frac{p(x)}{p'(x)},$$

where $p(x)$ represents the mixture density defined as $p(x) = \sum_{i=1}^N \omega_i p_i(x)$. Inspired by the identity (1) from the case of KMM, we infer the loss function which is minimized under the true density ratio $r(x)$:

$$\begin{aligned} \mathbb{E}_{p'(x)}[r(x)\phi(x)] &= \int r(x)\phi(x)p'(x)dx = \int \left(\sum_{i=1}^N \omega_i p_i(x) \right) \phi(x)dx \\ &= \sum_{i=1}^N \omega_i \int p_i(x)\phi(x)dx = \sum_{i=1}^N \omega_i \mathbb{E}_{p_i(x)}[\phi(x)], \end{aligned}$$

thus we consider the following loss function which needs to be solved with respect to the approximate model \hat{r} of the density ratio r :

$$\mathcal{L} = \left\| \mathbb{E}_{p'(x)}[\hat{r}(x)\phi(x)] - \sum_{i=1}^N \omega_i \mathbb{E}_{p_i(x)}[\phi(x)] \right\|^2 \quad (2)$$

An alternative of the above computations is to consider the approach of *MultiDistribution DRE* where we define as before, for $i \in \{1, \dots, N\}$, $r_i(x) = \frac{p_i(x)}{p'(x)}$. In this case we compute for every $i \in \{1, \dots, N\}$ the following:

$$\mathbb{E}_{p'(x)}[r_i(x)\phi(x)] = \int r_i(x)\phi(x)p'(x)dx = \int \phi(x)p_i(x)dx = \mathbb{E}_{p_i(x)}[\phi(x)],$$

hence we can define the i^{th} loss mapping with respect to the approximation \hat{r}_i of r_i :

$$\mathcal{L}_i = \|\mathbb{E}_{p'(x)}[\hat{r}_i(x)\phi(x)] - \mathbb{E}_{p_i(x)}[\phi(x)]\|^2,$$

therefore one can propose the mixture loss function \mathcal{L} where the weights w_i represent the contribution of each particular loss function \mathcal{L}_i , namely

$$\mathcal{L} = \sum_{i=1}^N \omega_i \mathcal{L}_i = \sum_{i=1}^N \omega_i \|\mathbb{E}_{p'(x)}[\hat{r}_i(x)\phi(x)] - \mathbb{E}_{p_i(x)}[\phi(x)]\|^2. \quad (3)$$

It is worth pointing out that the loss function (3) resembles the approach of *Ensemble KMM* structure, where one solves simultaneously N optimization problems, with the condition that the objective function of the i^{th} problem is related to the approximation $\hat{r}_i(x)$ of its associated density ratio model r_i .

- Now we turn our attention to our second case which we shall analyze it, where from a practical point of view one has multiple non-overlapping training datasets and a single test dataset. In order to do this we consider, for $i \in \{1, \dots, N'\}$, the sets $\mathcal{X} = \{x_k \mid x_k \in \mathbb{R}^d\}_{k=1}^n$ and $\mathcal{X}'_i = \{x'_{j,(i)} \mid x'_{j,(i)} \in \mathbb{R}^d\}_{j=1}^{n'_i}$ such that $\mathcal{X} \stackrel{i.i.d.}{\sim} P$ and $\mathcal{X}'_i \stackrel{i.i.d.}{\sim} P'_i$, where P and P'_i are probability distributions with densities p, p'_i , respectively. In a similar fashion with the previous case, let's consider $r_i(x) = \frac{p(x)}{p'_i(x)}$ and the weights $\tilde{\omega}_i \in [0, 1]$, for each

$i \in \{1, \dots, N'\}$ such that $\sum_{i=1}^{N'} \tilde{\omega}_i = 1$. Then, it follows that

$$\begin{aligned} \mathbb{E}_{p(x)}[\phi(x)] &= \int \phi(x)p(x)dx = \int \phi(x)p(x) \left(\sum_{i=1}^{N'} \tilde{\omega}_i \right) dx \\ &= \int \phi(x) (\tilde{\omega}_1 p(x) + \dots + \tilde{\omega}_{N'} p(x)) dx \\ &= \int \phi(x) (\tilde{\omega}_1 r_1(x) p'_1(x) + \dots + \tilde{\omega}_{N'} r_{N'}(x) p'_{N'}(x)) dx \\ &= \int \phi(x) \left(\sum_{i=1}^{N'} \tilde{\omega}_i r_i(x) p'_i(x) \right) dx = \sum_{i=1}^{N'} \tilde{\omega}_i \int \phi(x) r_i(x) p'_i(x) dx \\ &= \sum_{i=1}^{N'} \tilde{\omega}_i \mathbb{E}_{p'_i(x)}[r_i(x)\phi(x)], \end{aligned}$$

hence it is natural to propose the following loss function:

$$\mathcal{L} = \left\| \sum_{i=1}^{N'} \tilde{\omega}_i \mathbb{E}_{p'_i(x)}[\hat{r}_i(x)\phi(x)] - \mathbb{E}_{p(x)}[\phi(x)] \right\|^2 \quad (4)$$

For the previous case we shall present an alternative method in order to infer a suitable loss function. We proceed by considering, for each $i \in \{1, \dots, N'\}$ the weights $\omega_i \in [0, 1]$ that satisfy $\sum_{i=1}^{N'} \omega_i = 1$. Along with these we define the mixture probability density $p'(x) =$

$\sum_{i=1}^{N'} \omega_i p'_i(x)$ and the corresponding true density ratio $r(x) = \frac{p(x)}{p'(x)}$. Then, it follows that

$$\begin{aligned} \mathbb{E}_{p(x)}[\phi(x)] &= \int \phi(x)p(x)dx = \int \phi(x)r(x)p'(x)dx = \int \phi(x)r(x) \left(\sum_{i=1}^{N'} \omega_i p'_i(x) \right) dx \\ &= \sum_{i=1}^{N'} \omega_i \int \phi(x)r(x)p'_i(x)dx = \sum_{i=1}^{N'} \omega_i \mathbb{E}_{p'_i(x)}[r(x)\phi(x)], \end{aligned}$$

which defines the following loss function:

$$\mathcal{L} = \left\| \sum_{i=1}^{N'} \omega_i \mathbb{E}_{p'_i(x)}[\hat{r}(x)\phi(x)] - \mathbb{E}_{p(x)}[\phi(x)] \right\|^2 \quad (5)$$

In what follows we will show an equivalence between (4) and (5) under certain assumptions on the weights with respect to the true density ratio. By using that $p(x) = r_i(x)p'_i(x)$, it follows for each $i \in \{1, \dots, N'\}$ that

$$r(x) = \frac{r_i(x)p'_i(x)}{\sum_{j=1}^{N'} \omega_j p'_j(x)}.$$

So, for every $k \in \{1, \dots, N'\}$ we have that

$$\begin{aligned} \sum_{i=1}^{N'} \omega_i \mathbb{E}_{p'_i(x)}[r(x)\phi(x)] &= \sum_{i=1}^{N'} \omega_i \int r(x)\phi(x)p'_i(x)dx = \int \phi(x)r(x) \left(\sum_{i=1}^{N'} \omega_i p'_i(x) \right) dx \\ &= \int \phi(x)r_k(x)p'_k(x)dx = \mathbb{E}_{p'_k(x)}[r_k(x)\phi(x)]. \end{aligned}$$

Therefore, by summing over k , it follows that

$$\sum_{i=1}^{N'} \omega_i \mathbb{E}_{p'_i(x)}[r(x)\phi(x)] = \sum_{k=1}^{N'} \left(\frac{1}{N'} \right) \mathbb{E}_{p'_k(x)}[r_k(x)\phi(x)],$$

for which we can select the uniformly distributed weights $\tilde{\omega}_i = \frac{1}{N'}$ for every $i \in \{1, \dots, N'\}$.

• We end the present section with a brief remark about a particular case regarding the inference of (5). By using the form of the true density ratio, we obtain the following computations:

$$r(x) = \frac{p(x)}{p'(x)} = \frac{p(x)}{\sum_{i=1}^{N'} \omega_i p'_i(x)} = \frac{p(x)}{\sum_{i=1}^{N'-1} \omega_i p'_i(x) + \omega_{N'} p'_{N'}(x)}.$$

Let's consider the situation when $p'_{N'}(x) = p(x)$ and $\omega_{N'} = \alpha \in [0, 1)$. Then, it follows that

$$r(x) = \frac{p(x)}{\sum_{i=1}^{N'-1} \omega_i p'_i(x) + \alpha p(x)},$$

where $\sum_{i=1}^{N'-1} \omega_i + \alpha = \sum_{i=1}^{N'} \omega_i = 1$, so $\sum_{i=1}^{N'-1} \omega_i = 1 - \alpha$. This can be considered as a generalization of the relative density ratio due to the fact that, when $N' = 2$, one has access to the test dataset \mathcal{X} and the train dataset \mathcal{X}'_1 , respectively. Therefore, one obtains the α -relative density ratio $r_\alpha(x) = \frac{p(x)}{\alpha p(x) + (1 - \alpha)p'(x)}$. Finally, we highlight that, for $i \in \{1, \dots, N'\}$, the sample sets $\mathcal{X}'_i = \{x'_{j,(i)} \mid x'_{j,(i)} \in \mathbb{R}^d\}_{j=1}^{n'_i}$ corresponding to $p'_i(x)$ form a partition of non-overlapping sets. Hence, the situation when $p'_{N'}(x) = p(x)$ is equivalent to the fact that \mathcal{X} is non-overlapping with any other training subsets \mathcal{X}'_i for $i \in \{1, \dots, N' - 1\}$. In order to avoid this limitation, when dealing with the particular case of the generalized relative density ratio, we propose to formally use the same formulas as above despite the fact that the non-overlapping condition does not hold in general between train and test datasets, respectively.

3 Structure of the paper

The aim of the previous section (2) was to present in a step-by-step manner the main stimulus behind our **Generalized** KMM method. For this, we investigated an approach for devising a *quadratic optimization problem with constraints* based on the situation of multiple non-overlapping training datasets, along with the case of multiple non-overlapping test datasets. Our algorithmic framework is different than the methodologies from *Ensemble KMM* introduced in [4] and *Efficient Sampling KMM* from [5] since our technique is constructed using a loss-function approach. In section (4) we will actually construct our optimizer. By introducing an extended density ratio function using mixtures of probability densities, our technique is based upon the construction of a non-negative loss mapping which attains its minimum value of 0 under the true density ratio. The empirical version is obtained when one uses an approximate linear kernel model using the points of the whole test dataset, by utilizing some constraints suitable to numerical implementations. In section (5) we present some numerical simulations developed with a custom implementation made in `Python` regarding the comparison of probability densities under the learned density ratio weights, along with importance reweighting examples. Finally, in the last section (6), we discuss about the advantages of our generalized method along with the underlying limitations.

4 Proposed optimization method

In what follows we consider our general KMM-type optimization technique based upon the computations made in the previous section. In the usual case of (OptPb-KMM) and its extensions, one considers optimizing the MMD-based loss function involving the density ratio model only at the training points. Inspired by the techniques utilized in (OptPb-RuLSIF) we propose to use a linear kernel model for the density ratio in order to obtain an optimization problem with respect to the underlying parameters of the model. It is worth pointing out that this methodology is similar to the one proposed in [7] where a neural network was used for the density ratio model. Furthermore, the training of the neural network model is made at each epoch with respect to non-overlapping shuffled data batches thus the setting from [7] is similar to ours (see also the alternative of the bootstrap aggregation technique from [5]). But, the main difference is that, in our case, the train and test partitions are given at the beginning of the algorithm and are not randomly created at each iteration.

Let's consider N' training sets $\mathcal{X}'_l = \{x'_{j,(l)} \mid x'_{j,(l)} \in \mathbb{R}^d\}_{j=1}^{n'_l}$ such that $\mathcal{X}'_l \stackrel{i.i.d.}{\sim} P'_l$, where P'_l is the probability distribution with the underlying density p'_l for every $l \in \{1, \dots, N'\}$. At the same time, let's suppose that we have N test sets $\mathcal{X}_i = \{x_{k,(i)} \mid x_{k,(i)} \in \mathbb{R}^d\}_{k=1}^{n_i}$ such that $\mathcal{X}_i \stackrel{i.i.d.}{\sim} P_i$ where P_i is the probability distribution corresponding to the density p_i for each $i \in \{1, \dots, N\}$. In what follows we will consider the test and train mixtures of probability densities $p(x) = \sum_{i=1}^N \omega_i p_i(x)$ and $p'(x) = \sum_{j=1}^{N'} \gamma_j p'_j(x)$, where the weights satisfy $\sum_{i=1}^N \omega_i = \sum_{j=1}^{N'} \gamma_j = 1$, with $\omega_i \in [0, 1]$ for every $i \in \{1, \dots, N\}$ and $\gamma_j \in [0, 1]$ for each $j \in \{1, \dots, N'\}$. The general density ratio between the train and test samples is defined as

$$r(x) = \frac{p(x)}{p'(x)} = \frac{\sum_{i=1}^N \omega_i p_i(x)}{\sum_{j=1}^{N'} \gamma_j p'_j(x)}.$$

One observes that

$$\begin{aligned} \mathbb{E}_{p(x)}[\phi(x)] &= \int \phi(x) p(x) dx = \int \phi(x) \left(\sum_{i=1}^N \omega_i p_i(x) \right) dx \\ &= \sum_{i=1}^N \int \phi(x) \omega_i p_i(x) dx = \sum_{i=1}^N \omega_i \int \phi(x) p_i(x) dx = \sum_{i=1}^N \omega_i \mathbb{E}_{p_i(x)}[\phi(x)]. \end{aligned} \quad (6)$$

At the same time we have that

$$\begin{aligned} \mathbb{E}_{p'(x)}[r(x)\phi(x)] &= \int r(x)\phi(x)p'(x)dx = \int r(x)\phi(x) \left(\sum_{j=1}^{N'} \gamma_j p'_j(x) \right) dx \\ &= \sum_{j=1}^{N'} \gamma_j \int r(x)\phi(x)p'_j(x)dx = \sum_{j=1}^{N'} \gamma_j \mathbb{E}_{p'_j(x)}[r(x)\phi(x)]. \end{aligned} \quad (7)$$

Also

$$\mathbb{E}_{p'(x)}[r(x)\phi(x)] = \int r(x)\phi(x)p'(x)dx = \int \phi(x)p(x)dx = \mathbb{E}_{p(x)}[\phi(x)] \quad (8)$$

By combining (6), (7) and (8) we arrive at

$$\sum_{j=1}^{N'} \gamma_j \mathbb{E}_{p'_j(x)}[r(x)\phi(x)] = \sum_{i=1}^N \omega_i \mathbb{E}_{p_i(x)}[\phi(x)],$$

and taking into account that

$$\mathbb{E}_{p'(x)}[r(x)] = \int r(x)p'(x)dx = \int p(x)dx = 1,$$

along with

$$\mathbb{E}_{p'(x)}[r(x)] = \int r(x)p'(x)dx = \int r(x) \left(\sum_{j=1}^{N'} \gamma_j p'_j(x) \right) dx = \sum_{j=1}^{N'} \gamma_j \mathbb{E}_{p'_j(x)}[r(x)],$$

we therefore consider the following optimization problem:

$$\left\{ \begin{array}{l} \min_{\hat{r}} \left\| \sum_{j=1}^{N'} \gamma_j \mathbb{E}_{p'_j(x)} [\hat{r}(x) \phi(x)] - \sum_{i=1}^N \omega_i \mathbb{E}_{p_i(x)} [\phi(x)] \right\|^2 \\ \text{subject to } \begin{cases} \hat{r}(x) \geq 0 \text{ for all } x \\ \sum_{j=1}^{N'} \gamma_j \mathbb{E}_{p'_j(x)} [\hat{r}(x)] = 1 \end{cases} \end{array} \right. \quad (\text{OptPb-G-KMM})$$

where $\hat{r}(x)$ represents a density ratio model which approximates the true density ratio $r(x)$. In the following we shall show that the optimization problem presented in (OptPb-G-KMM) can be written as a *quadratic optimization problem with constraints*. Then, the underlying loss function can be written as

$$\begin{aligned} \mathcal{L} := \left\| \sum_{j=1}^{N'} \gamma_j \mathbb{E}_{p'_j(x)} [\hat{r}(x) \phi(x)] - \sum_{i=1}^N \omega_i \mathbb{E}_{p_i(x)} [\phi(x)] \right\|^2 &= \left\| \sum_{i=1}^N \omega_i \mathbb{E}_{p_i(x)} [\phi(x)] \right\|^2 \\ &+ \left\langle \sum_{j=1}^{N'} \gamma_j \mathbb{E}_{p'_j(x)} [\hat{r}(x) \phi(x)], \sum_{k=1}^{N'} \gamma_k \mathbb{E}_{p'_k(x)} [\hat{r}(x) \phi(x)] \right\rangle \\ &- 2 \left\langle \sum_{j=1}^{N'} \gamma_j \mathbb{E}_{p'_j(x)} [\hat{r}(x) \phi(x)], \sum_{i=1}^N \omega_i \mathbb{E}_{p_i(x)} [\phi(x)] \right\rangle. \end{aligned}$$

Ignoring constants irrelevant with respect to $\hat{r}(x)$, the objective function defined above can be taken as

$$\mathcal{L} = \left\langle \sum_{j=1}^{N'} \gamma_j \mathbb{E}_{p'_j(x)} [\hat{r}(x) \phi(x)], \sum_{k=1}^{N'} \gamma_k \mathbb{E}_{p'_k(x)} [\hat{r}(x) \phi(x)] \right\rangle - 2 \left\langle \sum_{j=1}^{N'} \gamma_j \mathbb{E}_{p'_j(x)} [\hat{r}(x) \phi(x)], \sum_{i=1}^N \omega_i \mathbb{E}_{p_i(x)} [\phi(x)] \right\rangle,$$

hence

$$\mathcal{L} = \sum_{j=1}^{N'} \sum_{k=1}^{N'} \left\langle \gamma_j \mathbb{E}_{p'_j(x)} [\hat{r}(x) \phi(x)], \gamma_k \mathbb{E}_{p'_k(x)} [\hat{r}(x) \phi(x)] \right\rangle - 2 \sum_{j=1}^{N'} \sum_{i=1}^N \left\langle \gamma_j \mathbb{E}_{p'_j(x)} [\hat{r}(x) \phi(x)], \omega_i \mathbb{E}_{p_i(x)} [\phi(x)] \right\rangle.$$

By consider employing empirical averages, we therefore obtain that $\mathbb{E}_{p_i(x)} [\phi(x)] \approx \frac{1}{n_i} \sum_{l=1}^{n_i} \phi(x_{l,(i)})$

and $\mathbb{E}_{p'_j(x)} [\hat{r}(x) \phi(x)] \approx \frac{1}{n'_j} \sum_{t=1}^{n'_j} \hat{r}(x'_{t,(j)}) \phi(x'_{t,(j)})$, which implies that the empirical loss

function $\hat{\mathcal{L}}$ which approximates \mathcal{L} takes the form

$$\begin{aligned} \hat{\mathcal{L}} &= \sum_{j=1}^{N'} \sum_{k=1}^{N'} \left\langle \gamma_j \left(\frac{1}{n'_j} \sum_{t=1}^{n'_j} \hat{r}(x'_{t,(j)}) \phi(x'_{t,(j)}) \right), \gamma_k \left(\frac{1}{n'_k} \sum_{s=1}^{n'_k} \hat{r}(x'_{s,(k)}) \phi(x'_{s,(k)}) \right) \right\rangle \\ &- 2 \sum_{j=1}^{N'} \sum_{i=1}^N \left\langle \gamma_j \left(\frac{1}{n'_j} \sum_{t=1}^{n'_j} \hat{r}(x'_{t,(j)}) \phi(x'_{t,(j)}) \right), \omega_i \left(\frac{1}{n_i} \sum_{l=1}^{n_i} \phi(x_{l,(i)}) \right) \right\rangle. \end{aligned}$$

Taking $n'_{max} := \max_{j \in \{1, \dots, N'\}} \{n'_j\}$ and multiplying $\widehat{\mathcal{L}}$ with $\frac{1}{2} (n'_{max})^2$, we simplify the previous identity as follows:

$$\begin{aligned} \widehat{\mathcal{L}} &= \sum_{j=1}^{N'} \sum_{k=1}^{N'} \left(\frac{(n'_{max})^2}{n'_j n'_k} \frac{\gamma_j \gamma_k}{2} \left\langle \sum_{t=1}^{n'_j} \hat{r}(x'_{t,(j)}) \phi(x'_{t,(j)}), \sum_{s=1}^{n'_k} \hat{r}(x'_{s,(k)}) \phi(x'_{s,(k)}) \right\rangle \right) \\ &\quad - \sum_{j=1}^{N'} \sum_{i=1}^N \left(\frac{(n'_{max})^2}{n_i n'_j} \gamma_j \omega_i \left\langle \sum_{t=1}^{n'_j} \hat{r}(x'_{t,(j)}) \phi(x'_{t,(j)}), \sum_{l=1}^{n_i} \phi(x_{l,(i)}) \right\rangle \right). \end{aligned}$$

By utilizing the linearity of the inner product, we obtain

$$\begin{aligned} \widehat{\mathcal{L}} &= \sum_{j=1}^{N'} \sum_{k=1}^{N'} \left(\frac{(n'_{max})^2}{n'_j n'_k} \frac{\gamma_j \gamma_k}{2} \sum_{t=1}^{n'_j} \sum_{s=1}^{n'_k} \hat{r}(x'_{t,(j)}) \langle \phi(x'_{t,(j)}), \phi(x'_{s,(k)}) \rangle \hat{r}(x'_{s,(k)}) \right) \\ &\quad - \sum_{j=1}^{N'} \sum_{i=1}^N \left(\frac{(n'_{max})^2}{n_i n'_j} \gamma_j \omega_i \sum_{t=1}^{n'_j} \sum_{l=1}^{n_i} \hat{r}(x'_{t,(j)}) \langle \phi(x'_{t,(j)}), \phi(x_{l,(i)}) \rangle \right). \end{aligned}$$

From the definition of the kernel mapping as an inner product of the feature maps i.e., $K(x, y) = \langle \phi(x), \phi(y) \rangle$, it follows that

$$\begin{aligned} \widehat{\mathcal{L}} &= \sum_{j=1}^{N'} \sum_{k=1}^{N'} \left(\frac{(n'_{max})^2}{n'_j n'_k} \frac{\gamma_j \gamma_k}{2} \sum_{t=1}^{n'_j} \sum_{s=1}^{n'_k} \hat{r}(x'_{t,(j)}) K(x'_{t,(j)}, x'_{s,(k)}) \hat{r}(x'_{s,(k)}) \right) \\ &\quad - \sum_{j=1}^{N'} \sum_{i=1}^N \left(\frac{(n'_{max})^2}{n_i n'_j} \gamma_j \omega_i \sum_{t=1}^{n'_j} \sum_{l=1}^{n_i} \hat{r}(x'_{t,(j)}) K(x'_{t,(j)}, x_{l,(i)}) \right). \end{aligned} \quad (9)$$

In order to simplify the previous computations we consider the following notations:

$$\begin{aligned} \hat{r}_{\mathcal{X}'_j} &:= \left(\hat{r}(x'_{1,(j)}), \dots, \hat{r}(x'_{n'_j,(j)}) \right)^T \in \mathbb{R}^{n'_j \times 1}, \quad j \in \{1, \dots, N'\} \\ \hat{r}_{\mathcal{X}_i} &:= \left(\hat{r}(x_{1,(i)}), \dots, \hat{r}(x_{n_i,(i)}) \right)^T \in \mathbb{R}^{n_i \times 1}, \quad i \in \{1, \dots, N\}. \end{aligned}$$

At the same time we define for $i \in \{1, \dots, N\}$ and $j \in \{1, \dots, N'\}$ the vector $h^{[i,j]} \in \mathbb{R}^{n'_j \times 1}$, such that

$$h_t^{[i,j]} = \frac{(n'_{max})^2}{n_i n'_j} \gamma_j \omega_i \sum_{l=1}^{n_i} K(x'_{t,(j)}, x_{l,(i)}) \quad \text{for each } t \in \{1, \dots, n'_j\}.$$

Also, for every $j, k \in \{1, \dots, N'\}$ we define the matrix $H^{[j,k]} \in \mathbb{R}^{n'_j \times n'_k}$, such that

$$H_{t,s}^{[j,k]} = \frac{(n'_{max})^2}{n'_j n'_k} \frac{\gamma_j \gamma_k}{2} K(x'_{t,(j)}, x'_{s,(k)}) \quad \text{for each } t \in \{1, \dots, n'_j\} \text{ and } s \in \{1, \dots, n'_k\}.$$

By using the above notations we obtain the following computations:

$$\begin{aligned} \frac{(n'_{max})^2}{n_i n'_j} \gamma_j \omega_i \sum_{t=1}^{n'_j} \sum_{l=1}^{n_i} \hat{r}(x'_{t,(j)}) K(x'_{t,(j)}, x_{l,(i)}) &= \sum_{t=1}^{n'_j} \hat{r}(x'_{t,(j)}) \left(\frac{(n'_{max})^2}{n_i n'_j} \gamma_j \omega_i \sum_{l=1}^{n_i} K(x'_{t,(j)}, x_{l,(i)}) \right) \\ &= \sum_{t=1}^{n'_j} \hat{r}(x'_{t,(j)}) h_t^{[i,j]} \\ &= \left(h^{[i,j]} \right)^T \hat{r}_{\mathcal{X}'_j}. \end{aligned} \quad (10)$$

On the other hand, by denoting

$$C^{[j,k]} := \frac{(n'_{max})^2}{n'_j n'_k} \frac{\gamma_j \gamma_k}{2} \sum_{t=1}^{n'_j} \sum_{s=1}^{n'_k} \hat{r}(x'_{t,(j)}) K(x'_{t,(j)}, x'_{s,(k)}) \hat{r}(x'_{s,(k)}), \quad (11)$$

we find that

$$\begin{aligned} C^{[j,k]} &= \sum_{t=1}^{n'_j} \sum_{s=1}^{n'_k} \hat{r}(x'_{t,(j)}) \left(\frac{(n'_{max})^2}{n'_j n'_k} \frac{\gamma_j \gamma_k}{2} K(x'_{t,(j)}, x'_{s,(k)}) \right) \hat{r}(x'_{s,(k)}) \\ &= \sum_{t=1}^{n'_j} \hat{r}(x'_{t,(j)}) \sum_{s=1}^{n'_k} \left(\frac{(n'_{max})^2}{n'_j n'_k} \frac{\gamma_j \gamma_k}{2} K(x'_{t,(j)}, x'_{s,(k)}) \hat{r}(x'_{s,(k)}) \right) \\ &= \sum_{t=1}^{n'_j} \hat{r}(x'_{t,(j)}) \sum_{s=1}^{n'_k} H_{t,s}^{[j,k]} \hat{r}(x'_{s,(k)}) \\ &= \sum_{t=1}^{n'_j} \hat{r}(x'_{t,(j)}) \left(H^{[j,k]} \hat{r}_{\mathcal{X}'_k} \right)_t \\ &= \left(\hat{r}_{\mathcal{X}'_j} \right)^T H^{[j,k]} \left(\hat{r}_{\mathcal{X}'_k} \right). \end{aligned} \quad (12)$$

By merging (9), (10), (11) and (12), we find a simpler formulation of the empirical loss, namely

$$\hat{\mathcal{L}} = \sum_{j=1}^{N'} \sum_{k=1}^{N'} \left(\left(\hat{r}_{\mathcal{X}'_j} \right)^T H^{[j,k]} \left(\hat{r}_{\mathcal{X}'_k} \right) \right) - \sum_{j=1}^{N'} \sum_{i=1}^N \left(\left(h^{[i,j]} \right)^T \left(\hat{r}_{\mathcal{X}'_j} \right) \right). \quad (13)$$

In order to make our method more similar to RuLSIF we consider modeling the true density ratio as a linear model i.e., $\hat{r}(x) = \langle \theta, \xi(x) \rangle$ where $\theta = (\theta_1, \dots, \theta_b) \in \mathbb{R}^b$ and $\xi : \mathbb{R}^d \rightarrow \mathbb{R}^b$ such that $\xi(x) = (\xi_1(x), \dots, \xi_b(x))$ for each $x \in \mathbb{R}^d$. For $j \in \{1, \dots, N'\}$ we define the matrix $A^{[j]} \in \mathbb{R}^{n'_j \times b}$ such that

$$A^{[j]} = \begin{pmatrix} \xi_1(x'_{1,(j)}) & \dots & \xi_b(x'_{1,(j)}) \\ \vdots & \vdots & \vdots \\ \xi_1(x'_{n'_j,(j)}) & \dots & \xi_b(x'_{n'_j,(j)}) \end{pmatrix} \quad (14)$$

Some easy computations reveal that

$$A^{[j]} \theta = \begin{pmatrix} \xi^T(x'_{1,(j)}) \theta \\ \vdots \\ \xi^T(x'_{n'_j,(j)}) \theta \end{pmatrix} = \hat{r}_{\mathcal{X}'_j} \in \mathbb{R}^{n'_j \times 1} \quad (15)$$

Therefore, (15) implies that

$$\left(\hat{r}_{\mathcal{X}'_j} \right)^T H^{[j,k]} \left(\hat{r}_{\mathcal{X}'_k} \right) = (A^{[j]} \theta)^T H^{[j,k]} (A^{[k]} \theta) = \theta^T \left((A^{[j]})^T H^{[j,k]} A^{[k]} \right) \theta. \quad (16)$$

Using again (15), it follows that

$$\left(h^{[i,j]} \right)^T \left(\hat{r}_{\mathcal{X}'_j} \right) = \left(h^{[i,j]} \right)^T A^{[j]} \theta = \left(\left(h^{[i,j]} \right)^T A^{[j]} \right) \theta. \quad (17)$$

Consequently, (13), (16) and (17) imply that

$$\hat{\mathcal{L}} = \theta^T \left(\sum_{j=1}^{N'} \sum_{k=1}^{N'} (A^{[j]})^T H^{[j,k]} A^{[k]} \right) \theta - \left(\sum_{j=1}^{N'} \sum_{i=1}^N (h^{[i,j]})^T A^{[j]} \right) \theta. \quad (18)$$

For the particular case of kernel methods we can choose $\xi(x)$ with the same technique as in the case of RuLSIF, namely we will use the test dataset defined as the reunion of the non-overlapping test sample datasets: $\mathcal{X} = \bigcup_{i=1}^N \mathcal{X}_i$. Therefore, we will select $\xi_k(x) = K(x, x_k)$

where $x_k \in \mathcal{X}$ for each $k \in \{1, \dots, b\}$, thus $b = \sum_{i=1}^N n_i$.

In what follows we will select K as a kernel endowed with non-negative values, such as the *RBF kernel* or the *Laplacian kernel*. In [3] the value of $\hat{r}(x)$ was bounded (only with respect to the training samples) in the interval $[0, B]$, where $B > 0$. In our case, in order to simplify this condition, we consider $\hat{r}(x) = \langle \theta, \xi(x) \rangle \geq 0$ and using that K takes non-negative values, we shall impose that $\theta_k \in [0, B]$ for every $k \in \{1, \dots, b\}$, where $B > 0$ is a constant chosen up to our choice. On the other hand, we define $\Xi = (\Xi_1, \dots, \Xi_b) \in \mathbb{R}^b$ as

$$\Xi := \sum_{j=1}^{N'} \left(\frac{\gamma_j}{n'_j} \right) \sum_{k=1}^{n'_j} \xi(x_{k,(j)}).$$

The constraint $\sum_{j=1}^{N'} \gamma_j \mathbb{E}_{p'_j(x)} [\hat{r}(x)] = 1$ from (OptPb-G-KMM) leads to its empirical counterpart, namely

$$1 \approx \sum_{j=1}^{N'} \gamma_j \left(\frac{1}{n'_j} \sum_{k=1}^{n'_j} \hat{r}(x_{k,(j)}) \right) = \sum_{j=1}^{N'} \left(\frac{\gamma_j}{n'_j} \right) \sum_{k=1}^{n'_j} \hat{r}(x_{k,(j)}) = \sum_{j=1}^{N'} \left(\frac{\gamma_j}{n'_j} \right) \sum_{k=1}^{n'_j} \langle \theta, \xi(x_{k,(j)}) \rangle = \langle \theta, \Xi \rangle.$$

Using the above identity $\langle \theta, \Xi \rangle = 1$, similar to the numerical description of KMM from [3], we consider $\varepsilon > 0$ such that $|\langle \theta, \Xi \rangle - 1| \leq \varepsilon$, hence $\sum_{k=1}^b \theta_k \Xi_k \leq \varepsilon + 1$ and $-\sum_{k=1}^b \theta_k \Xi_k \leq \varepsilon - 1$, respectively.

By combining (18) with the constraints presented above, we finally obtain our **Generalized KMM** method represented by the following *empirical generalized KMM optimization problem*:

$$\left\{ \begin{array}{l} \min_{\theta} \left[\theta^T \left(\sum_{j=1}^{N'} \sum_{k=1}^{N'} (A^{[j]})^T H^{[j,k]} A^{[k]} \right) \theta - \left(\sum_{j=1}^{N'} \sum_{i=1}^N (h^{[i,j]})^T A^{[j]} \right) \theta \right] \\ \text{subject to} \left\{ \begin{array}{l} \theta_k \in [0, B] \text{ for } k \in \{1, \dots, b\} \\ + \sum_{k=1}^b \theta_k \Xi_k \leq \varepsilon + 1 \\ - \sum_{k=1}^b \theta_k \Xi_k \leq \varepsilon - 1. \end{array} \right. \end{array} \right. \quad (\text{OptPb-Empirical-G-KMM})$$

5 Results & experiments

In this section we present some numerical simulations based on our implementation of the **Generalized KMM** optimizer concerning certain experiments made on some synthetic

datasets. We highlight that our codes hinge on `SKLearn` [18] and the `CVXPY` package [19] and [20], respectively. At the same time, all the details about our implementation and the corresponding experiments can be found in our `GitHub` link presented in the *Abstract* of the present paper. It is of utmost importance to mention that our parameter σ which will appear in the experiments from the following sequel and in the underlying implementation, is denoted as γ in the `SKLearn` implementation (and when $\gamma \approx \tilde{\sigma}^{-2}$ then the aforementioned parameter is associated with the variance $\tilde{\sigma}^2$ of the kernel K).

Our first experiment is related to an application of the **Generalized KMM** described through the optimization problem (OptPb-Empirical-G-KMM), along with the classical KMM method, respectively. The implementation for the underlying KMM algorithm is inspired by the codes belonging to [21] and [22], respectively. For this experiment we have considered 4 clusters (two of them belonging to the training dataset and the other two representing the test samples) consisting of different number of samples, i.e. 200, 1000, 1000 and 300, respectively. The clusters were generated using the function `make_blobs` from `SKLearn` with the following standard deviation values: 0.6, 0.6, 0.9 and 0.6, respectively. For both numerical methods, the parameter B was set to 1000 while the values of the parameter σ were chosen as 1.0, 3.5, 2.0 and *None*. We mention that *None* is equivalent to the default value used in the definition of the kernels from `SKLearn`. In the case of the **Generalized KMM** algorithm each vector and matrix that is defined through a kernel has the same default value defined as `1/n_features`, but we have chosen to write it as *None* since this value is already shown in the plots related to the KMM method (due to the fact that we use the same datasets for both methods hence we have the same number of features). From the results depicted in figure (1) we observe that the classical KMM method has weight values smaller than those of the **Generalized KMM** algorithm. Our optimization method gives weights a higher value near the boundary of the two training clusters, while KMM emphasize also the samples belonging near the center of the training data. One also observes that by increasing the σ parameter the values of the weights also increase. On the other hand, the particular case when σ is attributed the value of *None* (depicted in the bottom right plot) shows that KMM leads to fewer weights with high values in contrast with the **Generalized KMM** method.

Our next experiments are related to the comparison of various distributions by employing the cases of multiple train and test datasets. In figures (2), (3) and (4), for each simulation that we have made, a custom selection of the train and test distributions is represented in the corresponding left plot while in the associated right plot we have considered visualizing the predictions of a basic `SGDRegressor` with and without the sample weights generated by the **Generalized KMM** algorithm. In order to inspect more closely the comparison of the effect of the density ratio weights, the title of each right plot shows the *MAE* with and without the sample weights. For all the plots containing the regression results, the target is generated using a *sinc* function at which we added a noise term following a normal distribution. Also, the B term involved in (OptPb-Empirical-G-KMM) was set to 1000.

The first simulation we have done is related to the case of multiple train datasets and it is shown in figure (2). For this, we have generated 3 random normal train datasets with sizes 200, 150 and 100, with the means -0.5 , 0.5 and 1.5 , and with the standard deviation equal to 0.1 . At the same time, the test dataset is composed of only 30 samples generated using a normal distribution with mean 1.0 and standard deviation 0.4 . In the left plots of the first row we have chosen a lower value of σ , namely 0.1 which implies that the weighted train distribution is uniformly distributed with respect to the train partitions. This can be easily visualized in the corresponding right plot where the weighted and un-

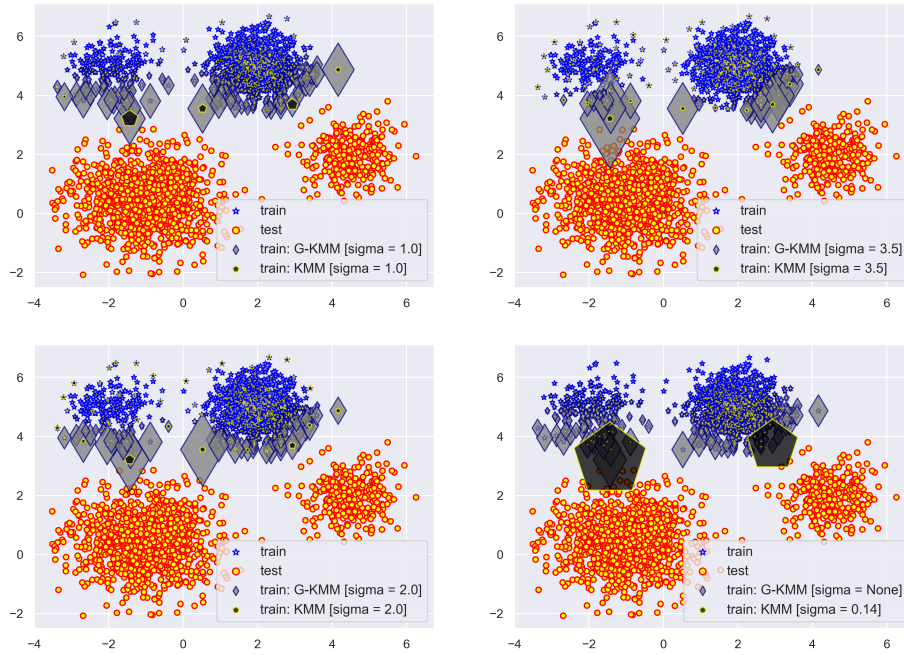


Fig. 1: KMM vs. Generalized KMM

weighted predictions behave similarly. In the right pair of plots from the first row of figure (2) we took σ equal to 1.0 but we have chosen the case of the α -relative density ratio with $\alpha = 0.25$, and where the γ_j weights of the training subsets were considered as 0.5, 0.2 and 0.05, respectively. The effect of the α -mixture density can be seen through the visualization of the distortion of the weighted train distribution towards the skewed test dataset. For the last case, which is represented in the second row, we have set σ to 100, α to 0.5 and the γ_j values as 0.05, 0.2 and 0.25, respectively. These choices shows a similar effect as in the previous case regarding the γ_j mixture values.

Our next simulations presented in figure (3) correspond to the case of multiple test datasets. We have generated a single train dataset of size 300 from a normal distribution with mean 1.0 and standard deviation 0.25 for the results depicted in the first row, while for the second row the train was generated using a normal distribution with mean 0.5 and standard deviation 0.25, respectively. On the other hand, the test datasets, both of size 100, were generated from normal distributions with means -0.5 and 1.5 , with the corresponding standard deviations equal to 0.15. Furthermore, for all our simulations depicted in figure (3) the parameter σ was chosen as 100. The left plots belonging to the first row shows that the weighted train distribution becomes closer to the test subset which overlaps the train dataset. On the other hand, the right plots from the first row shows the effect of the α mixture coefficient which was set to 0.75 along with the γ_1 coefficient of the single train dataset which was eventually chosen as 0.25. Here, we see that the mixture coefficient emphasize much more the test dataset which is closer to the training dataset, and it eventually leads to a worse approximation of the *SGDRegressor*. Finally, the simulation made in the plots from the second rows are based upon the same choice of

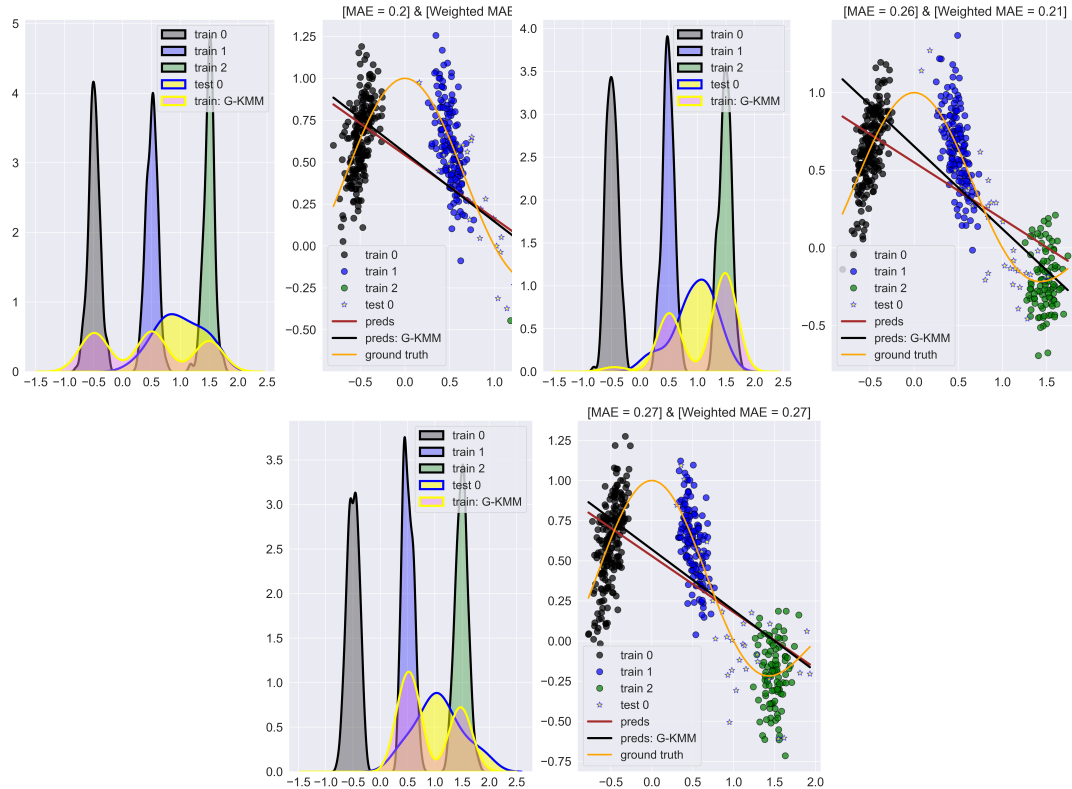


Fig. 2: Multiple train datasets

the coefficients as in the previously described simulation, namely α is 0.75, γ_1 was set to 0.25 and σ to 100, respectively. The main difference is that the training dataset is shifted to the left hence it is located between the two test datasets. One can observe that the α -mixture density ratio approach is suitable for this regression problem setting, by leading to a uniform-like distribution of the weighted train dataset.

The last experiment that we will present involves multiple train and also multiple test datasets and it is shown in figure (4). As before, we generate the train and test datasets using a random normal distribution. In the corresponding simulations we created 3 train datasets of sizes 200, 150 and 100, along with 2 test datasets of sizes equal to 100. The training subsets were generated from random normal distributions with means -0.5 , 0.5 and 1.5 , and standard deviation 0.1. On the other hand, the two test datasets were generated using random normal distribution with means -0.5 and 1.5 , with a standard deviation equal to 0.15. In the plots further to the left from the first row of (4), we have chosen the value 0.1 for σ . Similar to the left-most plots from the first row of figure (2), the low value for σ implies a weighted train distribution with 3 peaks uniformly distributed, along with a weighted MAE equal to the MAE obtained from the unweighted predictions. On the other hand, in the right-most plots from the first row of (4) the value for σ was increased to 10. This leads to only 2 peaks, uniformly distributed and centered at the test distributions. Consequently, the MAE metric decreases if one uses the weighted predictions of the **SGDRegressor**.

Now, let's turn our attention to the simulations made in the second row of figure (4). For the results shown in the further to the left plots we have chosen σ equal to 100, and the ω_i test weights 0.85 and 0.15, respectively. Along with these we have considered an α -mixture density approach, where α was not defined directly, i.e. at first the γ_j weights of the train

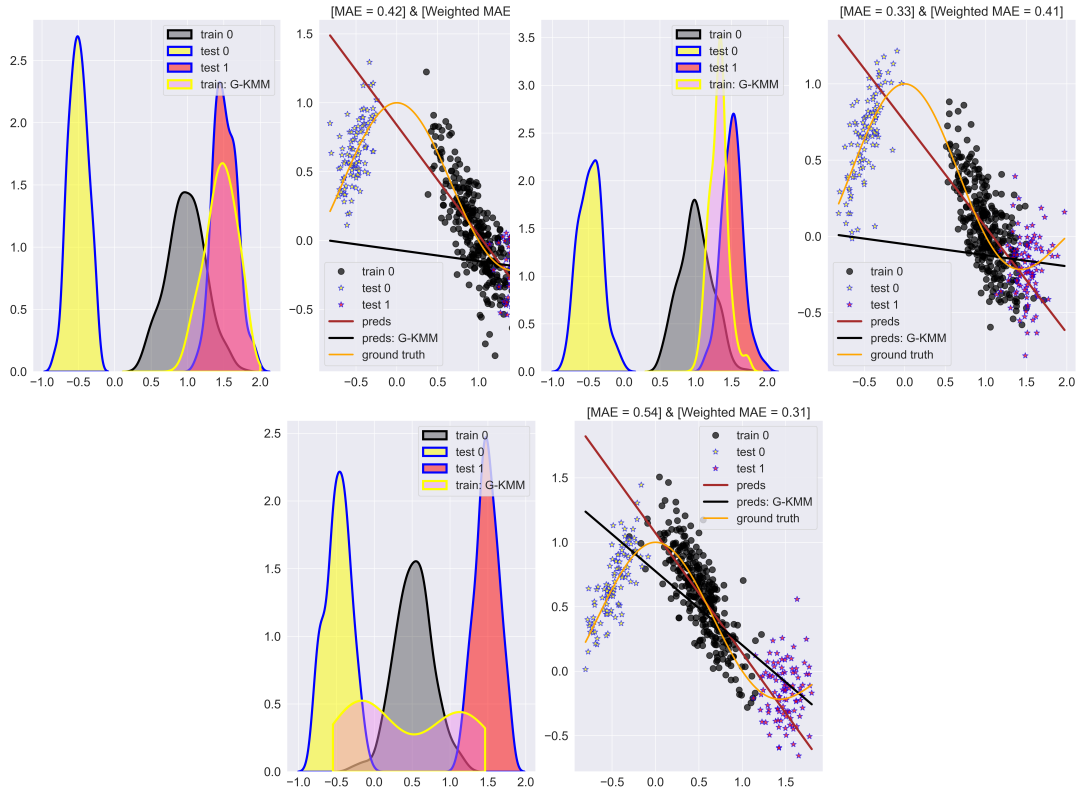


Fig. 3: Multiple test datasets

subsets were constructed using the ratio of each train subset size and the size of the total training dataset, and then α was determined such that α and the sum of γ_j add to the value of 1 (for this see the basic example belonging to the ending part of section (2)). In this case one observes that the mixture density technique modifies the distribution of the peaks of the weighted training data. Furthermore, the value of the weights ω_i related to the test datasets shows that the higher the ω_i weight is then the higher is the peak pointing to the corresponding test dataset. Similar to the case of multiple test datasets which were depicted in the last row of figure (3), the α -mixture density approach is crucial in the learning process of the optimal density ratio weights.

For the right-most plots shown in the second row of figure (4) we considered also σ equal to 100. But, the weights corresponding to the training subsets, namely γ_j were chosen this time as 0.25, 0.2 and 0.05 while the weights ω_i for the test subsets were selected with the values 0.15 and 0.85, respectively. As explained before, since α and the sum of all the γ_j coefficients must sum up to 1, we set α to the value of 0.5. Due to the fact that the weight of the second test subset is higher than the coefficient corresponding to the first test subset, the peak of the weighted train dataset is higher in the location of the second test subset.

Finally, we conclude the present section by highlighting that the experiment made in figure (1) reveals a qualitative comparison between the classical KMM algorithm and our **Generalized KMM** density ratio optimization method. At the same time, the simulations presented in figures (2), (3) and (4) shows the versatility of our method through the choices of the coefficients, especially for the case of the α -mixture density ratio.

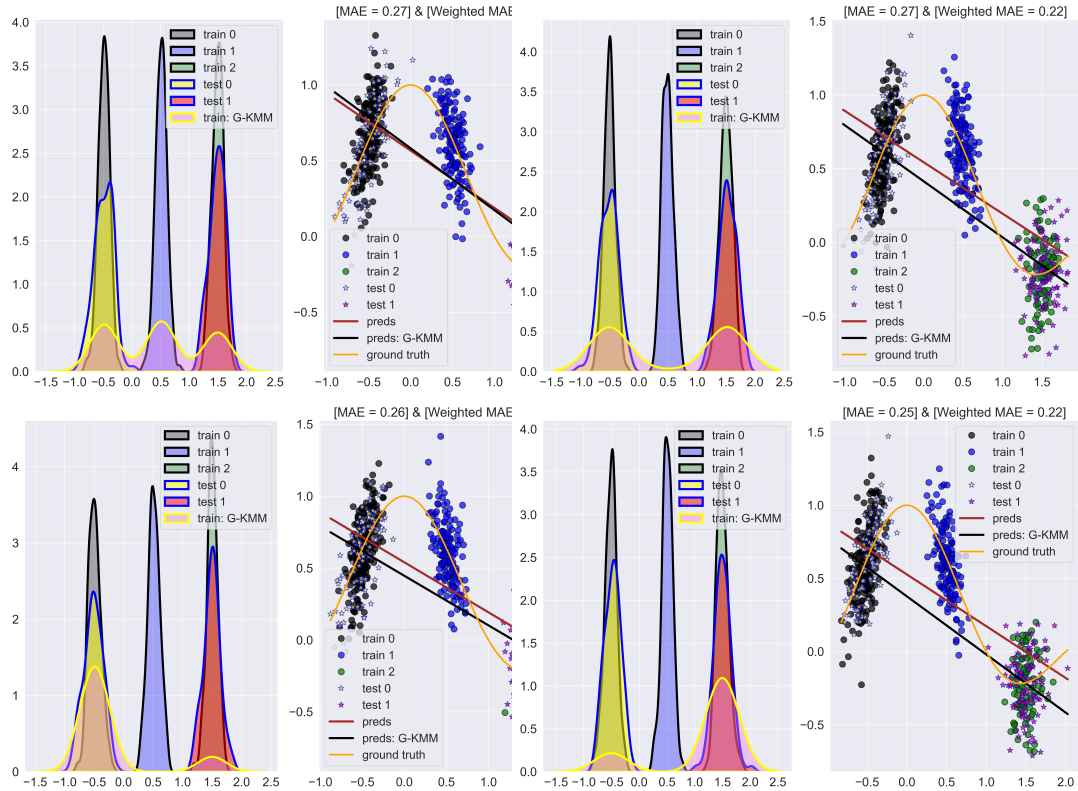


Fig. 4: Multiple train & test datasets

6 Conclusions & perspectives

In this final section we present a brief overview of our **Generalized KMM** algorithm given through the *quadratic optimization problem with constraints* (OptPb-Empirical-G-KMM) along with the underlying limitations and the possible extensions for future research.

6.1 Novelty

In the present study, our main contribution is the introduction of a new type of density ratio estimation technique entitled **Generalized KMM**, which is an extension of the classical KMM algorithm. From both a theoretical and a practical point of view our proposed method is completely novel from the following perspectives:

- The *Ensemble KMM* method from [4] is based on the idea of dividing the test dataset into multiple non-overlapping test sets, while *Efficient Sampling KMM* introduced in [5] is associated with the idea of a bootstrap aggregation approach for the training data. In contrast, our method is not developed using heuristic arguments, but it relies on the construction of a suitable loss function which attains its minimum value in the theoretical situation when one uses the true density ratio.
- In [3], the classical KMM algorithm uses directly the density ratio model with respect to the training samples. But, in our work we employed the approach used in RuLSIF where the density ratio is approximated with a linear kernel model, where the underlying kernel depends on the test points. Hence we minimize a loss function with respect to some weights belonging to a lower-dimensional space, where the dimension is given by the total number of test samples.

- Although we have constructed our minimization problem in connection to the cases consisting of non-overlapping train/test datasets, our approach contains as a particular case also a generalized version of the α -relative density ratio, which is unique from the point of view of KMM-type methods. On the other hand, it is worth emphasizing that the theoretical construction of our method was done using the idea of non-overlapping sets, while the case of the α -relative density ratio is devised only through a formal and mimetic approach.

6.2 Research limitations

Our method has not only advantages but it is also constrained by our inherent methodology as shown below:

- Despite the fact that the **Generalized KMM** is rigorously developed, one loses the parallelization property of the *Ensemble KMM* and *Efficient Sampling KMM*, respectively.
- Similar to the classical KMM algorithm, our method has the same dependence on the hyper-parameters ε , B and σ , respectively.

6.3 Recommendations for future research

For future research, we propose the following methods to enlarge our KMM-type framework:

- In a similar manner with [7] we can extend our algorithm to the case of neural networks. More precisely, one can utilize the objective function given in (OptPb-Empirical-G-KMM) along with the constraints which can be applied directly into the forward propagation process. Consequently, we can make our method faster using randomized batch learning, hence we may utilize the KMM-type algorithm for adjusting the probability densities associated to data augmentation sample sets.
- Similar to the classical KMM algorithm, our method is suitable for the estimation of the density ratio weights. In general, only a few training samples contribute to the reweighting process, due to the fact that the density ratio estimation and the regression/classification learning are separated. In order to alleviate this, we can proceed as in [23] by simultaneously training our **Generalized KMM** density ratio model and the underlying weighted loss function, in the framework of supervised learning.

References

1. A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schölkopf, A. Smola, A kernel two-sample test, *The Journal of Machine Learning Research*, vol. 13, no. 1, 2012, pp. 723-773.
2. M. Kirchler, S. Khorasani, M. Kloft, C. Lippert, Two-sample testing using deep learning, in: *International Conference on Artificial Intelligence and Statistics*, PMLR, 2020, pp. 1387-1398.
3. A. Gretton, A. Smola, J. Huang, M. Schmittfull, K. Borgwardt, B. Schölkopf, Covariate shift by kernel mean matching, *Dataset shift in machine learning*, vol. 3, no. 4, 2009, pp. 5.
4. Y-Q. Miao, A.K. Farahat, M.S. Kamel, Ensemble kernel mean matching, in: *2015 IEEE International Conference on Data Mining*, IEEE, 2015, pp. 330-338.
5. S. Chandra, A. Haque, L. Khan, C. Aggarwal, Efficient sampling-based kernel mean matching, in: *2016 IEEE 16th International Conference on Data Mining (ICDM)*, IEEE, 2016, pp. 811-816.
6. A. Haque, Z. Wang, S. Chandra, Y. Gao, L. Khan, C. Aggarwal, Sampling-based distributed kernel mean matching using Spark, in: *2016 IEEE International Conference on Big Data (Big Data)*, IEEE, 2016, pp. 462-471.
7. A. de Mathelin, F. Deheeger, M. Mougéot, N. Vayatis, Fast and Accurate Importance Weighting for Correcting Sample Bias, *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Cham: Springer International Publishing, 2022, pp. 659-674.

8. T. Kanamori, S. Hido, M. Sugiyama, A least-squares approach to direct importance estimation, *The Journal of Machine Learning Research*, vol. 10, 2009, pp. 1391-1445.
9. M. Yamada, T. Suzuki, T. Kanamori, H. Hachiya, M. Sugiyama, Relative density-ratio estimation for robust distribution comparison, *Neural computation*, vol. 25, no. 5, 2013, pp. 1324-1370.
10. M. Sugiyama, T. Suzuki, Y. Itoh, T. Kanamori, M. Kimura, Least-squares two-sample test, *Neural networks*, vol. 24, no. 7, 2011 pp. 735-751.
11. M. Sugiyama, T. Suzuki, T. Kanamori, Density-ratio matching under the Bregman divergence: a unified framework of density-ratio estimation, *Annals of the Institute of Statistical Mathematics*, vol. 64, 2012, pp. 1009-1044.
12. M. Sugiyama, T. Suzuki, S. Nakajima, H. Kashima, P. Von Büna, M. Kawanabe, Direct importance estimation for covariate shift adaptation, *Annals of the Institute of Statistical Mathematics*, vol. 60, 2008, pp. 699-746.
13. Y. Tsuboi, H. Kashima, S. Hido, S. Bickel, M. Sugiyama, Direct density ratio estimation for large-scale covariate shift adaptation, *Journal of Information Processing*, vol. 17, 2009, pp. 138-155.
14. M. Hushchyn, A. Ustyuzhanin, Generalization of change-point detection in time series data based on direct density ratio estimation, *Journal of Computational Science*, vol. 53, 2021, pp. 101385.
15. A. Kumagai, T. Iwata, Y. Fujiwara, Meta-learning for relative density-ratio estimation, *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 30426-30438.
16. L. Yu, Y. Jin, S. Ermon, A unified framework for multi-distribution density ratio estimation, *arXiv preprint arXiv:2112.03440*, 2021.
17. M. Sugiyama, T. Suzuki, T. Kanamori, *Density ratio estimation in machine learning*, Cambridge University Press, 2012.
18. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, *Scikit-learn: Machine learning in Python*, the *Journal of machine Learning research*, vol. 12, 2011, pp. 2825-2830.
19. A. Agrawal, R. Verschueren, S. Diamond, S. Boyd, A rewriting system for convex optimization problems, *Journal of Control and Decision*, vol. 5, no. 1, 2018, pp. 42-60.
20. S. Diamond, S. Boyd, *CVXPY: A Python-embedded modeling language for convex optimization*, *The Journal of Machine Learning Research*, vol. 17, no. 1, 2016, pp. 2909-2913.
21. A. de Mathelin, F. Deheeger, G. Richard, M. Mougeot, N. Vayatis, *ADAPT: Awesome domain adaptation python toolbox*, *arXiv preprint arXiv:2107.03049*, 2021.
22. T. Fang, N. Lu, G. Niu, M. Sugiyama, Rethinking importance weighting for deep learning under distribution shift, *Advances in neural information processing systems*, vol. 33, 2020, pp. 11996-12007.
23. S. Chen, X. Yang, Tailoring density ratio weight for covariate shift adaptation, *Neurocomputing*, vol. 333, 2019, pp. 135-144.

Authors

C.D. Alecsa received his PhD in Mathematics from Babeş-Bolyai University and has a broad experience in both academia and industry. Currently, he is a researcher on two national grants: at the Technical University of Cluj-Napoca (on Optimization) and at the Romanian Institute of Science and Technology (on Machine Learning), respectively. His research interests include Pattern Recognition, Machine Learning, Statistics, and Applied Mathematics.

Inhance deep customizations in a multi-tenant SaaS application using the BPMN

Amira Ksiksi

REsearch Groups in Intelligent Machines (REGIM Lab) University of Sfax,
National Engineering School of Sfax (ENIS), Sfax, Tunisia

Abstract. A multi-tenant application aims to provide a single instance of an application with the capability for each organization to have its own specific functionalities. Recent researches have proved the efficiency of the intrusive and non-intrusive approaches in providing deep customizations. However, deep customizations are still limited to the features provided for each organization. In order to enhance the deep customization, we propose a BPMN-based customizations to provide to each organization the capability to create its own features. such a method requires an administration module to provide to the organization to create forms, scripts and notifications to be integrated in a BPMN workflow's tasks. Such a method has proved its capability to introduce new functionalities using understandable graphical representations which reduce the need for the vendors' intervention.

Keywords: Multi-tenancy, BPMN, deep customizations, workflow

1 Introduction

A multi-tenant SaaS application aims to provide a single instance of an application with the capability for each organization to have its own specific functionalities. Providing the customizations to each organization as microservices is a costly task as it requires the from the organization that provides the service to provide the implementation of the new functionalities and add the needed configuration to make it accessible by the specific tenant that has asked for those customizations. In order to reduce the cost and time of multi-tenant SaaS customizations, we propose to use the Business Process Management and Notation (BPMN) to provide for each organization to create its own customizations by introducing and modifying executable workflows by simply doing a drag and drop of components. However, the BPMN tasks require to be configured to ensure the workflow deployment. For this reason, we proposed to define an administration module that provides UIs to the tenant to facilitate the preparation of the configuration relevant to the tasks and facilitate its integration in the workflow by customizing the BPMN.io¹ extension. Our method has proved its capability in defining new functionalities in a short time and is cheaper than intrusive and non-intrusive customizations due to the SaaS pay-as-you-go feature.

The major contributions of this paper are twofold. First, we propose a BPMN-based customization method. Then, we provide a demonstration of deep customizations through a BPMN-based multi-tenant SaaS application as a case study.

The rest of the paper is organized as follows, section 2 provides the background of our research work. Section 3 presents the proposed method. The demonstration of BPMN-based multi-tenant SaaS application case study is provided in section 4. Finally, we conclude with the conclusion and future work.

¹ <https://bpmn.io/>

2 Related work

2.1 Multi-tenancy

Providing tenant-specific customization in a multi-tenant SaaS is a challenging task [1]. By using intrusive custom microservices, Song et al. [2] proposed an architecture for multi-tenant SaaS customization. The intrusive microservices aim to provide accessibility to the customizations that are introduced in separate microservices by registration and mapping of the customizations using tenant managers. So, once the tenant request reaches this part, it will be redirected to the registered microservice. However, Nguyen and Muller [3] propose a non-intrusive customization framework called MISC-CLOUD. MISC-CLOUD allows the Multi-tenant SaaS customization through an API gateway to manage the authority of customization's API calls. Both intrusive and non-intrusive approaches allow deep customization of multi-tenant SaaS applications [4]. However, to provide the flexibility to adapt to the specific-tenant's changing requirements, the BPMN represents a good solution due to its capability in creating executable workflows using easy graphical customizations [5, 6].

2.2 BPMN

The BPMN is a visual modeling language for business processes that helps to create executable workflows [7]. In BPMN, pools are used to present a participant in a collaboration. Each pool can contain multiple lanes which helps to further organize the workflow's activities [8]. Recently, several researches have been conducted to improve the BPMN capabilities. On one hand, Ribeiro et al. [10] introduced a new BPMN extension that aims to Model Inter-Organizational Processes. On the other hand, Delgado et al. [9] proposed a model-driven approach to deal with the BPMN variants. The use of the BPMN to model executable business logics enhances the tenant-specific deep customization in a multi-tenant SaaS application by reducing cost and time of production.

Form.io Form.io is a framework that facilitates the creation of forms by using the drag and drop feature [11]. Due to the use of APIs, Form.io created forms can be easily shared and exploited in customizable applications. Moreover, as the BPMN's 'start event' and 'user task' must contain a form to be filled, the Form.io created forms can be easily integrated. To this end, we exploited the Form.io form builder to create forms for the BPMN workflow.

3 Methodology

In order to propose deep customizations in a multi-tenant SaaS application, that reduce the cost and time of production, we propose to provide an administration module that provides each tenant with the capability to create its own customizations. The main component of this module is the BPMN workflows constructor. It helps to create executable workflows. However, as the different tasks in the BPMN require to be configured so it can be deployed and executed, we provide the necessary components to prepare the needed configurations. First of all, we have the form builder that uses the Form.io extension to provide to the tenant the capability of creating forms using the drag and drop feature. The created forms can be easily integrated in a 'user task' or a 'start event'. Then, we provide the notification builder which helps to configure the emails so it can be integrated in the 'send task'. Also, we provide a script builder to introduce a groovy script that can be integrated in a 'script

task'. Otherwise, the service provider can provide the other customizations in microservices using JavaDelegate that can be called from a 'service task'.

To proceed to the workflow deployment and execution, we provide a microservice that aims to add the executable process to the menu and configure the notification of each user task as well as the fields to be shown in the table relevant to this process. At that time, by accessing the menu relevant to the process, a tenant can create a new instance of the process and access the historical variables relevant to completed as well as uncompleted processes' executions.

4 Experimentation

As experimentation, we provide an illustration of a BPMN-based multi-tenant SaaS application developed from scratch. In order to illustrate the BPMN-based customization, we provide an example of a workflow that contains a 'start event', a 'user task', a 'send task' and a 'script task' to explain their configuration.

Fig. 1: The form builder UI

First, we proceed to the Form creation for the 'start event' and the 'user task' as shown in figure 1. Then, we proceed to create the configuration of the emails in the notification builder for the 'send task' as represented in figure 2. Thereafter, we present the creation of a simple script using the script creator for the 'script task' as shown in figure 3. Finally we proceed to the workflow creation and the tasks configuration as represented in figure 4.

Once the workflow is ready, we start to prepare the deployment and the configuration of the notification and table of instances which is represented in the figure 5a , figure 5b and figure 5c respectively. The execution of the workflow now becomes possible by accessing the process from the menu and filling the start event form and all historical data still accessible from the table of instances which is represented in figure 6.

Fig. 2: The notification builder UI

Fig. 3: The script builder UI

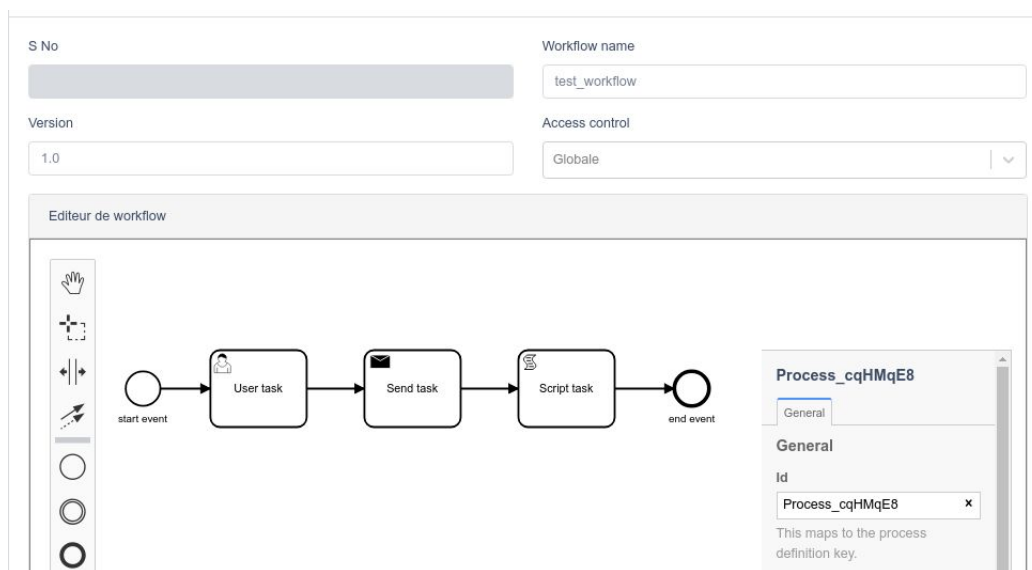


Fig. 4: The workflow builder UI

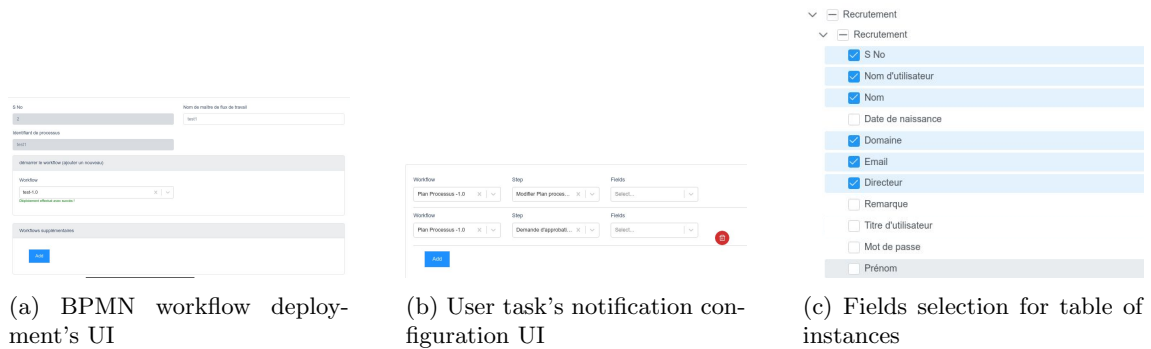


Fig. 5: Process' preparation from the BPMN workflow

S No	Type de système de gestion	Départements	Nom du processus	statut approuvé	Autorité d'approbation
1				approuve	amira23
7					amira23

Fig. 6: List of process instances' historical variables

5 Conclusion and future work

In this paper, we proposed a BPMN-based customizations in a multi-tenant SaaS application. A demonstration of the proposed method is illustrated in a BPMN-based multi-tenant customizable SaaS application. Due to the UIs used to introduce the new executable business logics and its configuration, our method has proved its capability to reduce the time of production as well as the cost of customizations.

As a future work, we aim to provide the semantic interoperability over the BPMN processes instances and integrate the artificial intelligence to further enhance the capability of our method.

References

1. Olabanji, D., Fitch, T., & Matthew, O. (2023). Multi-tenancy in Cloud-native Architecture: A Systematic Mapping Study. *WSEAS Transactions on Computers*, 22, 25-43.
2. Song, H., Chauvel, F., & Solberg, A. (2018, May). Deep customization of multi-tenant SaaS using intrusive microservices. In *Proceedings of the 40th International Conference on Software Engineering: New Ideas and Emerging Results* (pp. 97-100).
3. Nguyen, P. H., Song, H., Chauvel, F., Muller, R., Boyar, S., & Levin, E. (2019, August). Using microservices for non-intrusive customization of multi-tenant SaaS. In *Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (pp. 905-915).
4. D. Ying Hong, S. Kwee Teck, L. Tong Ming and C. Hoong Jack, "Customizing of ERP In Microservice SaaS Architecture: An Overview of Intrusive & Non-Intrusive Approach," 2023 IEEE 8th International Conference On Software Engineering and Computer Systems (ICSECS), Penang, Malaysia, 2023, pp. 134-138, doi: 10.1109/ICSECS58457.2023.10256303.
5. Fraj, I. B., Hlaoui, Y. B., & BenAyed, L. (2020, October). A control system for managing the flexibility in BPMN models of cloud service workflows. In *2020 IEEE 13th International Conference on Cloud Computing (CLOUD)* (pp. 537-543). IEEE.

6. Ding, W., Liu, J., Yang, Z., Lv, B., Li, H., & Xu, H. (2022). A multi-tenancy and robust workflow management system. *Expert Systems*, 39(6), e12878.
7. Lyttbacka, T. (2022). Modeling & Designing Toolkit For BPMN Processes: Toolkit for creating a visual representation of BPMN workflows.
8. da Silva, C. E., Gomes, E. L., & Basu, S. S. (2022). BPM2DDD: A Systematic Process for Identifying Domains from Business Processes Models. *Software*, 1(4), 417-449.
9. Delgado, A., Calegari, D., García, F., & Weber, B. (2022). Model-driven management of BPMN-based business process families. *Software and Systems Modeling*, 1-37.
10. Ribeiro, V., Barata, J., & Rupino da Cunha, P. (2021). A BPMN Extension to Model Inter-Organizational Processes in Industry 4.0.
11. Namee, K., Phoarun, R., Albadrani, G. M., Polpinij, J., Tanessakulwattana, S., & Sphanphong, P. (2019, November). A Form and API Data Management Platform for Progressive Web Application and Serverless Application Architecture. In *Proceedings of the 2019 2nd International Conference on Computational Intelligence and Intelligent Systems* (pp. 144-149).
12. Hanoka G, Shuldt J.C.N, "On signatures with tight security in the multi-user setting" (2017) in : *Proceedings of 2016 International Symposium on Information Theory and Its Applications, ISITA 2016*, art. no. 7840392, pp. 91-95.

AUTHOR INDEX

<i>Abderezak Touzene</i>	11
<i>Ahmed Al Farsi</i>	11
<i>Akansha Akansha</i>	93
<i>Amira Ksiksi</i>	307
<i>Ang Li</i>	117
<i>Aoi Nagatani</i>	45
<i>Ashima Anand</i>	129
<i>Bakul Gupta</i>	129
<i>Büşra Öztürk</i>	207
<i>Cristian Daniel Alecsa</i>	287
<i>David Broneske</i>	59
<i>Dharmender Salian</i>	35
<i>Emily X. Ding</i>	155
<i>Fawwad Hassan Jaskani</i>	223
<i>Gita Alaghband</i>	171
<i>Gurleen Kaur</i>	129
<i>Hafiz Bilal Ahmad</i>	223
<i>Haichang Ga</i>	223
<i>Hayastan Avetisyan</i>	59
<i>Henok Ghebrechristos</i>	171
<i>Hiro Horie</i>	01
<i>Hitesh Kumar Sharma</i>	193
<i>Jawad Lakziz</i>	263
<i>Joseph (Yu) Liu</i>	155
<i>Josephine (Hsin) Liu</i>	155
<i>Kiyoshi Yasuda</i>	01
<i>Lamo P</i>	237
<i>Ledesma O</i>	237
<i>Licheng Xiao</i>	117
<i>Long Ma</i>	277
<i>Manoj Kumar</i>	193
<i>Marin Shoda</i>	45
<i>Masahide Nakamura</i>	01, 21, 45
<i>Nasser Al Zeid</i>	11
<i>Oumaima Taheri</i>	263
<i>Parisa Safikhani</i>	59
<i>Phoebe (Yun) Liu</i>	155
<i>Preeti Sharma</i>	193
<i>Robert J. Hou</i>	155
<i>Rohit Khankhoje</i>	109, 143
<i>Ruijin Deng</i>	77
<i>Sachio Saiki</i>	21
<i>Said Jamal</i>	263
<i>Said Ouaskit</i>	263
<i>Sánchez M.A</i>	237

<i>Shun Sato</i>	45
<i>Sinan Chen</i>	01, 21
<i>Stuart So</i>	93
<i>Surya Boppanaa</i>	277
<i>Takuya Nakata</i>	21
<i>Tasuku Watanabe</i>	45
<i>Victor Phan</i>	77
<i>William Kane</i>	277
<i>Xiuhan(Daniel) Fu</i>	245
<i>Yahya Benremdane</i>	263
<i>Yakup Genç</i>	207
<i>Yoshifumi Kamae</i>	45
<i>Yuya Tarutani</i>	45