

Computer Science & Information Technology 151

Advances in Machine Learning

David C. Wyld,
Dhinaharan Nagamalai (Eds)

Computer Science & Information Technology

3rd International Conference on Machine Learning & Applications (CMLA 2021),
September 25 ~ 26, 2021, Toronto, Canada

3rd International Conference on Internet of Things (CIoT 2021)

8th International Conference on Computer Science, Engineering and Information
Technology (CSEIT 2021)

13th International Conference on Network and Communications Security (NCS 2021)

2nd International Conference on NLP & Big Data (NLPD 2021)

8th International Conference on Signal, Image Processing and Multimedia (SPM 2021)

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: 978-1-925953-49-7

DOI: 10.5121/csit.2021.111501 - 10.5121/csit.2021.111520

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

The International Conference on 3rd International Conference on Machine Learning & Applications (CMLA 2021), September 25 ~ 26, 2021, Toronto, Canada, 3rd International Conference on Internet of Things (CIoT 2021), 8th International Conference on Computer Science, Engineering and Information Technology (CSEIT 2021), 13th International Conference on Network and Communications Security (NCS 2021), 2nd International Conference on NLP & Big Data (NLPD 2021) and 8th International Conference on Signal, Image Processing and Multimedia (SPM 2021) was collocated with International Conference on 3rd International Conference on Machine Learning & Applications (CMLA 2021). 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 CMLA 2021, CIoT 2021, CSEIT 2021, NCS 2021, NLPD 2021 and SPM 2021 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, CMLA 2021, CIoT 2021, CSEIT 2021, NCS 2021, NLPD 2021 and SPM 2021 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 CMLA 2021, CIoT 2021, CSEIT 2021, NCS 2021, NLPD 2021 and SPM 2021.

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,
Abdelhadi Assir,
Abdellatif I. Moustafa,
Abderrahim Siam,
Abdulhamit Subasi,
Abhishek Shukla,
Adedeji Jelili kunle,
Adnan Saher Mohammed,
Afaq Ahmad,
Ahmad A. Saifan,
Ahmed Farouk AbdelGawad,
Ahmed Yaseen Mjhool,
Ahmet CIFCI,
Ajay Anil Gurjar,
Alexander Gelbukh,
Ali A. Al-Zuky,
Ali A. Amer,
Ali Abdrhman Mohammed Ukasha,
Ali Asghar Anvary Rostamy,
Ali Karkeh Abadi,
Allel Hadjali,
Almir Pereira Guimarães,
Alper Ugur,
Amari Houda,
Amel Ourici,
Amina El murabet,
Amizah Malip,
Amosa Babalola,
Ana Leal,
Ana Luísa Varani Leal,
Anamika Ahirwar,
Anita Yadav,
António Moreira,
Anuj Singal,
Archit Yajnik,
Arti Noor,
Ashraf Elnagar,
Assem abdel hamied moussa,
Assia Djenouhat,
Atanu Nag,
Attila Kertesz,
Atul Garg,
Aymen Ben Said,

Ain Shams University, Egypt
Hassan 1st University, Morocco
Umm AL-Qura University, Saudi Arabia
University of Khenchela, Algeria
Effat University, Saudi Arabia
R D Engineering College, India
Adekunle Ajasin University, Nigeria
Karabuk University, Turkey
Sultan Qaboos University, Oman
Yarmouk University, Jordan
Zagazig University, Egypt
University of Kufa, Iraq
Burdur Mehmet Akif Ersoy University, Turkey
Sipna College of Engineering & Technology, India
Instituto Politécnico Nacional, Mexico
Mustansiriyah University, Iraq
Taiz University, Yemen
Sebha University, Libya
Tarbiat Modares University (TMU), Iran
University of Tehran, Iran
LIAS/ENSMA, France
Federal University of Alagoas, Brazil
Pamukkale University, Turkey
Networking & Telecom Engineering, Tunisia
Badji Mokhtar University of Annaba, Algeria
Abdelmalek Essaadi University, Morocco
University of Malaya, Malaysia
Federal Polytechnic Ede, Nigeria
University of Macau, China
University of Macau, China
Jayoti Vidhyapeeth Women's University, India
Harcourt Butler Technical University, India
University of Aveiro, Portugal
GJU S&T, India
Sikkim Manipal University, India
CDAC Noida, India
University of Sharjah, UAE
GGA,GP&HR, Asdf-SCRA AFRICA, Egypt
University Badji Mokhtar Annaba, Algeria
Modern Institute of Engineering & Technology, India
University of Szeged, Hungary
Chitkara University, India
University of Regina, Canada

Ayodele Periola,	University of Johannesburg, South Africa
Balagadde,	Kampala International University, Uganda
Benyamin Ahmadnia,	University of California at Davis, USA
Benyettou Noria,	Ecole national polytechnique of Oran, Algeria
Beshair Alsiddiq,	Prince Sultan University, Saudi Arabia
Bibudhendu Pati,	Rama Devi Women's University, India
Bin Zhao,	JD.com Silicon Valley R&D Center, USA
Bouchra Marzak,	Hassan II University, Morocco
Boukari Nassim,	Skikda Universiy, Algeria
Brahmi Menaouer,	National Polytechnic School of Oran, Algeria
Brahim lejdel,	University of El-oued, Algeria
Caitong Yue,	Zhengzhou University, China
Carlos Guardado da Silva,	Universidade de Lisboa, Portugal
Casalino Gabriella,	University of Bari, Italy
Chahinez Mérièm Bentaouza,	Mostaganem University, Algeria
Chemesse ennehar Bencheriet,	University of Guelma, Algeria
cheng siong chin,	Newcastle University, Singapore
Cherkaoui Leghris,	Hassan II University of Casablanca, Morocco
Ching-Nung Ynag,	National Dong Hwa University, Taiwan
Christian Mancas,	Ovidius University, Romania
Dakshina Ranjan Kisku,	National Institute of Technology Durgapur, India
Daniel Rosa Canedo,	Federal Institute of Goias, Brazil
Daniela López De Luise,	CI2S lab director, Argentina
Dário Ferreira,	University of Beira Interior, Portugal
Debjani Chakraborty,	Indian Institute of Technology, India
Dhanamma Jagli,	University of Mumbai, India
Diab Abuaiadah,	Waikato Institute of Technology, New Zealand
Diogo S. Carvalho,	INESC-ID & University of Lisbon, Portugal
Djalila belkebir,	Kasdi Merbah Ouargla Univeristy, Algeria
Ekbal Rashid,	RTC Institute of Technology, India
Elżbieta Macioszek,	Silesian University of Technology, Poland
Endre Pap,	University Singidunum, Serbia
Essam Sourour,	Alexandria University, Egypt
Ez-Zahout Abderrahmane,	Mohamed V University, Morocco
F. Abbasi,	Islamic Azad University, Iran
Fatma Outay,	Zayed University, UAE
Felix J. Garcia Clemente,	University of Murcia, Spain
Fernando Zacarias Flores,	Universidad Autonoma de Puebla, Mexico
Francesco Zirilli,	Sapienza Universita Roma, Italy
Fulvia Pennoni,	University of Milano-Bicocca, Italy
Gabriel Badescu,	University of Craiova, Romania
Gajendra Sharma,	Kathmandu University, Nepal
Ghasem Mirjalily,	Yazd University, Iran
Giuliani Donatella,	University of Bologna, Italy
Grigorios N. Beligiannis,	University of Patras, Greece
Gururaj H L,	Vidyavardhaka College of Engineering, India
Habil Gabor Kiss,	Obuda University, Hungary
Hakan Altinpulluk,	Anadolu University, Turkey
Hala Abukhalaf ,	Palestine Polytechnic University, Palestine
Hamed Taherdoost,	West University, Canada
Hamid Ali Abed AL-Asadi,	Basra University, Iraq
Hamid Khemissa,	USTHB University Algiers, Algeria

Hamid Rastegari,	Islamic Azad University, Iran
Hamza Ouarnoughi,	Université des Hauts-de-France, France
Hang Su,	Politecnico di Milano, Italy
Hao-En Chueh,	Chung Yuan Christian University, Taiwan
Haqi Khalid,	University Putra Malaysia, Malaysia
Harmandeep singh Gill,	GAD khalsa College, India
Héctor Migallón Gomis,	Miguel Hernández University, Spain
Hemavathi P,	Bangalore Institute of Technolgy, India
Henrique Vicente,	University of Évora, Portugal
Hichem Houassi,	University of Abbes Laghrour Khenchela, Algeria
Himani mittal,	GGDSD College, India
Hiromi Ban,	Nagaoka University of Technology, Japan
Hlaing Htake Khaung Tin,	University of Computer Studies, Myanmar
Hossein Bavarsad,	Mechanical Design Engineer & Project Manager, Iran
Hussein Yousif Aziz,	Al Muthanna University, Iraq
Ibtesam Al-Saedi,	University of Technology, Iraq
Ilham Huseyinov,	Istanbul Aydin University, Turkey
Isa Maleki,	Islamic Azad University, Iran
Islam Atef,	Alexandria University, Egypt
Israa Shaker Tawfic,	Ministry of Science and Technology, Iraq
Iyad Alazzam,	Yarmouk University, Jordan
J.Naren,	Sastra University, India
Jamal Zraqou,	Isra University, Jordan
Janet Walters,	University of Technology, Jamaica
Jasmin Cosic,	Deutsche Bahn (R&D), Germany
Jawad K. Ali,	Microwave Research Group, Iraq
Jayavignesh T,	Vellore Institute of Technology, India
Jehoiada Jackson,	University of electronic science and technology, China
Jesuk Ko,	Universidad Mayor de San Andres, Bolivia
Jeyanthi,	VIT University, India
Jia Ying Ou,	York University, Canada
Jiajun Sun,	Huaiyin Normal University, China
Joao Antonio Aparecido Cardoso,	The Federal Institute of São Paulo, Brazil
João Calado,	Instituto Superior de Engenharia de Lisboa, Portugal
Jonah Lissner,	technion - israel institute of technology, Israel
Jong-Ha Lee,	Keimyung University, South Korea
Juntao Fei,	Hohai University, P. R. China
Kabid Hassan Shibly,	Dhaka International University, Bangladesh
Kamel Benachenhou,	Blida University, Algeria
Kamel Hussein Rahouma,	Minia University, Egypt
Kanga Koffi,	Ecole supérieure Africaine des TIC, Côte d'Ivoire
Kanniga Devi,	Kalasalingam University, India
Kanstantsin MIATLIUK,	Bialystok University of Technology, Poland
Kanwalvir Singh Dhindsa,	BBSBEngg. College, India
Karim Mansour,	University Salah Boubenider, Algeria
Kazi Sultana Farhana Azam,	Friedrich-Schiller University Jena, Germany
Ke-Lin Du,	Concordia University, Canada
Kelvin Nnamani,	Nnamdi Azikiwe University, Nigeria
Keneilwe Zuva,	University of Botswana, Botswana
Khader Mohammad,	Birzeit University, Palestine
Khalid M.O Nahar,	Yarmouk University, Jordan
Kire Jakimoski,	FON University, Republic of Macedonia

Kiril Alexiev,	Bulgarian Academy of Sciences, Bulgaria
Kirtikumar Patel,	Chemic Engineers, USA
Klenilmar L. Dias,	Federal Institute of Amapa - IFAP, Brazil
Koh You Beng,	University of Malaya, Malaysia
li>Mu-Song Chen,	Da-Yeh University, Taiwan
Luca Virgili,	University of Marche, Italy
Luis Miguel Nunes Corujo,	Universidade de Lisboa, Portugal
Luisa Maria Arvide Cambra,	University of Almeria, Spain
M V Ramana Murthy,	Osmania University, India
MA.Jabbar,	Vardhaman College of Engineering, India
Maad M. Mijwil,	Baghdad College of Economic Sciences University, Iraq
Mabroukah Amarif,	Sebha University, Libya
Mahbuba Afrin,	Swinburne University of Technology, Australia
Mahdi Abbasi,	Aix-Marseille University, Iran
Malka N.Halgamuge,	The University of Melbourne, Australia
Marco Javier Sarez Barón,	UPTC, Colombia
Maria Hallo,	Escuela Politécnica Nacional, Ecuador
Mariem Haoues,	University of Sfax, Tunisia
Mario Versaci,	DICEAM - University Mediterranea, Italy
Masoomah Mirrashid,	Semnan University, Iran
Maumita Bhattacharya,	Charles Sturt University, Australia
Mehdi Gheisari,	IAU, Iran
Mehdi Nezhadnaderi,	Tonekabon Branch Islamic Azad University, Iran
Meisam Abdollahi,	University of Tehran, Iran
Mihai Carabas,	University POLITEHNICA of Bucharest, Romania
Mirsaeid Hosseini Shirvani,	Islamic Azad University, Iran
Mohamed A.M.Ibrahim,	Taiz University, Republic of Yemen
Mohamed Abdelaziz Hassan Eleiwa,	Electrical Engineering, Egypt
Mohamed Ismail Roushdy,	Ain Shams University, Egypt
Mohamed Yacoab,	The New college, India
Mohamed-Khireddine,	Echahid Hamma Lakhdar d'El-Oued, Algeria
Mohammad A. Alodat,	Sur University College, Oman
Mohammad Jafarabad,	Iran University of Science & Technology, Iran
Mohammad Mahmoud Abu Omar,	Al-Quds Open University, Palestine
Mohammad Mehraeen,	Ferdowsi University of Mashhad, Iran
Mohammad Reza Ghavidel Aghdam,	University of Tabriz, Iran
Mohammed Bouhorma,	Abdelmalek Essadi University, Morocco
Mohammed Mahmoud,	Beijing Institute of Technology, China
Morris Riedel,	University of Iceland, Iceland
Morteza Alinia Ahandani,	University of Tabriz, Iran
Mourad Chabane Oussalah,	University of Nantes, France
M-Tahar Kechadi,	University College Dublin, Ireland
Mudhafar Hussein Ali,	College of Engineering, Al-Iraqia University, Iraq
Muge Erel-Ozcevik,	Manisa Celal Bayar University, Turkey
Muhammad Asif Khan,	Qatar University, Qatar
Muneer Masadeh Bani Yassein,	Jordan University of Science and Technology, Jordan
Mu-Song Chen,	Da-Yeh University, Taiwan
Mussa Turdalyuly,	Satbayev University, Kazakhstan
Mustafa S. Abd,	Baghdad University, Iraq
N Md Jubair Basha,	KHIT, India
Nadia Abd-Alsabbour,	Cairo University, Egypt
Nahlah Shatnawi,	Yarmouk University, Jordan

Nancy Alonistioti,
 Naveen Chandra,
 Neamtu Iosif Mircea,
 Neeraj kumar,
 Ngoc Hong Tran,
 Nihar Athreyas,
 Nikolai Prokopyev,
 Omar Khadir,
 Omid Mahdi Ebadati E,
 Osman Toker,
 Otilia Manta,
 P.V.Siva Kumar,
 Panagiotis Fotaris,
 Paulo Jorge dos Mártires Batista,
 Pavel Loskot,
 Peiyan Yuan,
 Przemyslaw Falkowski-Gilski,
 Qi Zhang,
 R. Vadivel,
 Raghad Ghalib Alsultan,
 Rahul M Mulajkar,
 Rahul Saha,
 Rajesh Bose,
 Rajini Kanth,
 Ramakrishnan,
 Ramana Murthy,
 Ramgopal Kashyap,
 Rashmi Kushwah,
 Ravi Kumar,
 Ren-Song Ko,
 Rinku Datta Rakshit,
 Saeed Iranmanesh,
 Said Agoujl,
 Said Nouh,
 Saif aldeen Saad Obayes,
 Samrat Kumar Dey,
 Sarra Nighaoui,
 Sathyendra Bhat J,
 Sebastian Fritsch,
 Sébastien Combéfis,
 Seif-Eddine Benkabou,
 Seppo Sirkemaa,
 Shadan Sadigh Behzadi,
 Shah Khalid Khan,
 Shahid Ali,
 Shahram Babaie,
 Shankar Gangisetty,
 Sharon Andrews,
 Shashikant Patil,
 Shereena V B,
 Shubham Sharma,
 Siddhartha Bhattacharyya,

National and Kapodistrian University of Athens, Greece
 Uttarakhand Technical University, India
 Lucian Blaga University, Romania
 Chitkara University, India
 Vietnamese-German University, Vietnam
 CTO, Spero Devices, Inc. USA
 Kazan Federal University, Russia
 Hassan II University of Casablanca, Morocco
 Kharazmi University, Tehran
 Yildiz Technical University, Turkey
 Romanian American University (RAU), Romania
 VNR VJIET, India
 University of Brighton, UK
 University of Évora, Portugal
 ZJU-UIUC Institute, China
 Henan Normal University, China
 Gdansk University of Technology, Poland
 Shandong University, China
 Phuket Rajabhat Univeristy, Thailand
 Northern Technical University, Iraq
 JCOE, India
 University of Padova, Italy
 Brainware University, India
 Snist, India
 Drmgr Educational and Research Institute, India
 Osmania University, India
 Amity University Chhattisgarh, India
 Jaypee Institute of Information Technology, India
 VIT University, India
 National Chung Cheng University, Taiwan
 Asansol Engineering College, India
 Shahid Bahonar University of Kerman, Iran
 Moulay Ismail University, Morocco
 Hassan II University of Casablanca, Morocco
 Shiite Endowment Office, Iraq
 Dhaka International University, Bangladesh
 National Engineering School of Tunis, Tunisia
 St Joseph Engineering College, India
 IT and CS enthusiast, Germany
 ECAM Brussels Engineering School, Belgium
 LIAS/Poitiers University, Poitiers, France
 University of Turku, Finland
 Islamic Azad University, Iran
 RMIT University, Australia
 AGI Education Ltd, New Zealand
 Islamic Azad University, Iran
 KLE Technological University, India
 University of Houston, United States
 SVKMs NMIMS Mumbai, India
 MES College Marampally, India
 University of Regina, Canada
 CHRIST (Deemed to be University), India

Sidi Mohammed Meriah,	University of Tlemcen, Algeria
Sikandar Ali,	China University of Petroleum, China
Simanta Shekhar Sarmah,	Alpha Clinical Systems, USA
Siva Kumar,	VNR VJIET, India
Smmain Femmam,	UHA University, France
Sobia Wassan,	Nanjing University, China
Sobin C C,	SRM University, India
Somya Goyal,	Manipal University Jaipur, India
Sridhar Iyer,	S.G. Balekundri Institute of Technology, India
Stefano Michieletto,	University of Padova, Italy
Sumit Kumar Debnath,	National Institute of Technology Jamshedpur, India
Sun-yuan Hsieh,	National Cheng Kung University, Taiwan
Taha Mohammed Hasan,	University of Diyala, Iraq
Taleb Zouggar Souad,	Oran 2 University, Algeria
Tanzila Saba,	Prince Sultan University, Saudi Arabia
Tara Singh Kamal,	Council member Institution of Engineers, India
Thamer Al-Rousan,	Isra University, Jordan
Tobey H. Ko,	The University of Hong Kong, Hong Kong
Tri Minh Tran,	Gunma University, Japan
Umesh Kumar Singh,	Vikram University, India
Usman Naseem,	University of Sydney, Australia
Vafa Andalibi,	Indiana University Bloomington, United States
Valerianus Hashiyana,	School of Computing University of Namibia, Namibia
Venkata Duvvuri,	Northeastern University, USA
Venkata Ramana Rao Puram,	ANGR Agricultural University, India
Venkatalakshmi,	Velammal Engineering College, India
Virupakshi Patil,	Sharnbasva University Kalaburagi, India
Vivek D,	PSG College of Arts & Science, India
Wahbi Azeddine,	Hassan II University, Morocco
Wajid Hassan,	Indiana State University, USA
Waseem Alshanti,	Jubail University College, Saudi Arabia
William R. Simpson,	Institute for Defense Analyses, USA
Wilson Enriquez López,	Escuela politécnica Nacional, Ecuador
WU Yung Gi,	Chang Jung Christian University, Taiwan
Xiao-Zhi Gao,	University of Eastern Finland, Finland
Yanrong Lu,	Tianjin University, China
Yexiong Lin,	Hunan University, China
Yogendra Kumar Jain,	Samrat Ashok Technological Institute, India
Yousef Farhaoui,	Moulay Ismail University, Morocco
Yousef J. Al-Houmaily,	Institute of Public Administration, Saudi Arabia
Yu-Chen Hu,	Providence University, Taiwan
Yuchen Zheng,	Shihezi University, China
Zahra Pezeshki,	Shahrood University of Technology, Iran
Zewdie Mossie,	Debre Markos University, Ethiopia
Zhang Ziwen,	Guangzhou Maritime University, China
Zhu Wang,	SANY Heavy Industry Co. LTD, China
Zoran Bojkovic,	LSM IEEE University of Belgrade, Serbia
Zoran Bojkovic,	University of Belgrade, Serbia
Zulqarnain Siddiqui,	Iqra University, Iraq
Zurab Kiguradze,	Missouri S&T, EMC Laboratory, USA

Technically Sponsored by

Computer Science & Information Technology Community (CSITC)



Artificial Intelligence Community (AIC)



Soft Computing Community (SCC)



Digital Signal & Image Processing Community (DSIPC)



3rd International Conference on Machine Learning & Applications (CMLA 2021)

How Different Text-Preprocessing Techniques using the Bert Model Affect the Gender Profiling of Authors.....	01-08
<i>Esam Alzahrani and Leon Jololian</i>	
Pedestrian Attribute Recognition using Gabor Wavelet Layers.....	09-19
<i>Imran N. Junejo</i>	
Online Obstructive Sleep Apnea Detection Based on Hybrid Machine Learning and Classifier Combination for Home-Based Applications.....	21-35
<i>Hosna Ghandeharioun</i>	
Artificial Intelligence & Machine Learning Role in Financial Services.....	37-42
<i>Prudhvi Parne</i>	
The Difference of Machine Learning and Deep Learning Algorithms.....	249-257
<i>Yew Kee Wong</i>	

3rd International Conference on Internet of Things (CIoT 2021)

Open Lorawan Sensor Node Architecture for Agriculture Applications.....	43-62
<i>Philipp Bolte, Ulf Witkowski and Rolf Morgenstern</i>	
Deep Learning for Identifying Malicious Firmware.....	63-70
<i>David Noever and Samantha E. Miller Noever</i>	
Limiting Factors in Widespread Adoption of Active Queue Management in the Philippines' Consumer Electronics Space.....	71-80
<i>Min Guk I. Chi</i>	
Understanding the Features of Internet of Things (IoT) and Big Data Analysis.....	259-266
<i>Yew Kee Wong</i>	

8th International Conference on Computer Science, Engineering and Information Technology (CSEIT 2021)

Development of an Autism Screening Classification Model for Toddlers.....	81-92
<i>Afef Saihi and Hussam Alshraideh</i>	
Revisiting Mobile Crowdsensing: An Open Challenge.....	93-101
<i>Vitalis Ayu</i>	

Predicting Consumer Purchasing Decisions in the Online Food Delivery Industry.....103-117
Batool Madani and Hussam AlShraideh

13th International Conference on Network and Communications Security (NCS 2021)

SHAPEIoT: Secure Handshake Protocol for Autonomous IoT Device Discovery and Blacklisting using Physical Unclonable Functions and Machine Learning.....119-137
Cem Ata Baykara, Ilgın Şafak and Kübra Kalkan

2nd International Conference on NLP & Big Data (NLPD 2021)

Arabic Poems Generation using LSTM, Markov-LSTM and Pre-Trained GPT-2 Models.....139-147
Asmaa Hakami, Raneem Alqarni, Mahila Almutairi and Areej Alhothali

Extractive Text Summarization using Recurrent Neural Networks with Attention Mechanism.....233-248
Shimirwa Aline Valerie and Jian Xu

8th International Conference on Signal, Image Processing and Multimedia (SPM 2021)

Aliasing Free For Mixed Spectra for Stable Processes.....149-173
Rachid Sabre

An Efficient Method for a Specific Case of Detecting Impulse Noise on Scanned Documents.....175-187
Petar Prvulović, Jelena Vasiljević and Dhinakaran Nagamalai

DCT based Fusion of Variable Exposure Images for HDRI.....189-201
Vivek Ramakrishnan and D. J. Pete

The Combination of Narrative News and VR Games: Comparison of Various Forms of News Games.....203-217
Xiaohan Feng and Makoto Murakami

Feature Fusion-Based Siamese Region Proposal Network for Ultrasound Tracking.....219-232
Xinglong Zhu, Ruirui Kang, Yifan Wang, Danni Ai, Tianyu Fu and Jingfan Fan

HOW DIFFERENT TEXT-PREPROCESSING TECHNIQUES USING THE BERT MODEL AFFECT THE GENDER PROFILING OF AUTHORS

Esam Alzahrani^{1, 2} and Leon Jololian¹

¹Department of Electrical and Computer Engineering,
University of Alabama at Birmingham, Birmingham, AL, USA

²Department of Computer Engineering, Al-Baha University,
Alaqiq, Saudi Arabia

ABSTRACT

Forensic author profiling plays an important role in indicating possible profiles for suspects. Among the many automated solutions recently proposed for author profiling, transfer learning outperforms many other state-of-the-art techniques in natural language processing. Nevertheless, the sophisticated technique has yet to be fully exploited for author profiling. At the same time, whereas current methods of author profiling, all largely based on features engineering, have spawned significant variation in each model used, transfer learning usually requires a preprocessed text to be fed into the model. We reviewed multiple references in the literature and determined the most common preprocessing techniques associated with authors' genders profiling. Considering the variations in potential preprocessing techniques, we conducted an experimental study that involved applying five such techniques to measure each technique's effect while using the BERT model, chosen for being one of the most-used stock pretrained models. We used the Hugging face transformer library to implement the code for each preprocessing case. In our five experiments, we found that BERT achieves the best accuracy in predicting the gender of the author when no preprocessing technique is applied. Our best case achieved 86.67% accuracy in predicting the gender of authors.

KEYWORDS

Authorship profiling, NLP, digital forensics, transfer learning

1. INTRODUCTION

Forensic author profiling has proven to be an important yet complicated task that requires further investigation. According to Keretna et al. [1], writing styles are affected by factors like a person's culture, educational background, and the environment he/she has been raised in. Historically, Criminal Forensic profiling is established by a former Federal Bureau of Investigation (FBI), John Edward Douglas [2]. He started studying and analyzing serial killers' crimes. Moreover, he has interviewed some of the serial killers to find a pattern that is associated with their criminal activities and their characteristics profiling. Further, this approach is believed to help to direct the course of investigation towards the most possible suspect who committed the crime. As the use of the Internet grows, the need to adopt this change is inevitable. The content of the internet contains a big amount of unstructured text especially in Online Social Networks (OSNs). One of the tasks of analyzing textual content is authorship profiling by which demographic information

about anonymous authors can be revealed. Beyond that, authorship profiling can contribute to the deanonymization of anonymous malicious texts. Even though most OSNs are regarded as an auxiliary in which users use their real names, the option of anonymity is still available. Criminals are known to seek anonymity to avoid getting caught by law enforcement [3], [4]. Moreover, the considered attributes that are used to create similarity among different profiles, can be used to build profile clusters by which users' profiles can be assigned.

However, other gaps need to be addressed as well, including that no method of classifying age or gender—both of which are aspects of author profiling—that also considers the genre or nature of the analyzed text has been widely endorsed. Beyond that, to the best of our knowledge, no research has involved surveying or comparing the different approaches used in author profiling. In fact, current knowledge about the task is largely based on small-scale experiments conducted to find a reliable classification method with a near-zero error rate. In order to use forensic authorship profiling as admissible evidence in courts, the proposed methods must have about 100% accuracy. The average accuracy of good proposed methods are ranging from 70-85% [5]–[9]. However, the proposed methods still suffer from low accuracy. A more specific gap is that transfer learning, an emerging technique in natural language processing (NLP) proven to be state-of-the-art for many NLP tasks [10], has not been fully tested in the context of author profiling. As a consequence, literature on author profiling using transfer learning has remained slight, and how the technique contributes to and affects such profiling remains poorly understood, at least in a systematic sense. For those reasons, the applicability of the advanced, reliable technique of transfer learning to author profiling is worth investigating.

According to Devlin and Chang, a serious challenge in NLP is the limited amount of training data [11]. To address that challenge, researchers have developed transformer-based models that are trained on enormous unlabelled datasets—for instance, Wikipedia's dataset—so that researchers can use the pretrained models on smaller datasets instead of having to develop training models from scratch. Although the technique has been shown to afford more accuracy in executing different NLP tasks [11], [12], the pretrained models require so-called “fine-tuning” before being used with smaller datasets. An example of such a pretrained model is the bidirectional encoder representations from transformers (BERT) model, which is distinguished from other pretrained models by virtue of its bidirectionality—that is, it considers context when words have dual valence. For example, when processing the word bank, which can mean a financial institution or the shore of a river, the BERT model examines all words in the sentence at both valences and generates a score that indicates the best representation of the meaning of the words in their given context. In its implementation, the BERT model is based on transformer model architecture developed by researchers at Google in 2017 [11].

The purpose of this study is to examine the impact of the most used preprocessing techniques in profiling the age and gender of the author if a pretrained model is used, BERT. In the remainder of this paper, the next section introduces past work related to the preprocessing techniques in author profiling, followed by an experimental section that details the implementation of the five cases of preprocessing techniques that we considered, and the steps performed in each experiment. After that, a section presents the results of each experiment and discusses the effect of each preprocessing technique on the model's accuracy in predicting the gender of authors. In the paper's conclusion, we restate the important findings of our study and indicate directions for future work.

2. RELATED WORK

To conduct a thorough literature review, we considered the valuable contribution of PAN's shared tasks in author profiling, which provides a broad range of approaches and methods in the

field. We reviewed PAN’s shared tasks from 2013 to 2017 to identify the types of preprocessing techniques that researchers consider in their efforts to profile authors [13]–[17]. Besides, we also included papers that investigate author profiling using English datasets. Some papers used uncommon preprocessing techniques e.g., extending shortened texts such as slang words, contractions, and abbreviations [18]. Lundeqvist & Svensson removed HTML tags, and used Twitter custom tokenizer (nltk.tokenize package — NLTK 3.6.2 documentation)[19]. However, some papers considered common preprocessing techniques, similar to those used in PAN shared tasks [20]. At least five research groups represented in PAN’s shared tasks from 2013 to 2017 removed retweet tags from the texts during preprocessing [13]–[17]; 17 groups removed hashtags [13]–[17], [20]; and 19 teams considered removing URLs. For the removal of the mentioned tags, 17 research groups considered removing them from the processed text. Stop words were removed only four times [21]–[23] [24], and 29 teams did not apply any preprocessing technique whatsoever [13]–[17]. In 11 instances, retweet tags, URLs, and mentions were all removed [14]–[17], [25].

The effectiveness of preprocessing techniques in machine learning approaches depends on the selection of features and classifiers. In transfer learning, the extensive training of pretrained models on large data equips them with the needed power to model the language and capture most of its contextualized aspects. Because transfer learning uses the previously learned knowledge to tokenize the downstream text [26], it uses fine-tuning to train and classify the downstream task [27].

Some techniques in NLP, including transfer learning, have been proven to be state-of-the-art for many NLP tasks [10], however, not fully tested for author profiling. A systematic understanding of how transfer learning contributes to author profiling is also lacking, despite the clear value of investigating the applicability of such an advanced, reliable technique in author profiling. Although transfer learning is a technique with growing interest, publications on author profiling using it have been few. With this publication, we aim to narrow all of those gaps, at least in part.

3. EXPERIMENTS

3.1. Dataset

Table 1. The distribution of the dataset per class. The raw data was extracted from the URLs that were sent by PAN. The table illustrates the number of tweets per men and the number of tweets per women.

Number of authors	436
Number of tweets	363,031
Number of tweets per men	149,059
Number of tweets per women	113,972

The dataset we used is from PAN’s 2016 shared tasks involving author profiling[16]. We chose to conduct our experiments on the English corpus only due to the focus and scope of our study. The most studied datasets in the literature are collected from Twitter. Twitter texts represent the characteristics of today’s text e.g., unstructured, short, and colloquial. In PAN shared task in 2016, the participated research groups were sent the URLs of the tweets with a truth table containing the authors’ gender and age labels. The URLs of the tweets have relatively smaller size and easier to share compared to the complete textual dataset. Age was categorized as follows: 1) 18–24, 2) 25–34, 3) 35–49, 4) 50–64, and 5) 65 and older. Table 1 shows the distribution of the dataset per class. Given the aim of our study, we profiled the author’s gender only.

3.2. Experimental Setups

The adopted experimental setups were based on the most common techniques observed in the literature [13]–[17], all of which have been extensively tested in the context of author profiling using machine and deep learning techniques. The effect of preprocessing techniques on author profiling using transfer learning techniques has not yet been studied, however. As illustrated in Table 2, we considered five cases for the preprocessing techniques. In Case 1, we included three basic techniques: mentions removal, retweet tags removal, and hashtags removal. In Case 2, we added URLs removal to the techniques from Case 1, and in Case 3, we added the removal of punctuation. In Case 4, we applied a well-known technique in NLP, stop words removal, which involves eliminating extremely common words that are liable to be repeated in many texts and that some researchers characterize as noise, not as markers. Last, in Case 5, we chose not to apply any preprocessing technique in order to gauge its effect. We built all five cases using regex in Python and the Hugging Face transformer library on Google Colab.

Table 2. Preprocessing cases. The preprocessing techniques are explained for each case.

Case	Preprocessing techniques
Case 1	Mentions removal
	Retweet tags removal
	Hashtags removal
Case 2	Mentions removal
	Retweet tags removal
	Hashtags removal
	URLs removal
Case 3	Mentions removal
	Retweet tags removal
	Hashtags removal
	URLs removal
	Punctuation removal
Case 4	Mentions removal
	Retweet tags removal
	Hashtags removal
	URLs removal
	Punctuation removal
	Stop words removal
Case 5 *	None (i.e., each text as-is)

*No preprocessing technique was applied

All of the experiments were carried out using Google Colab’s graphical processing unit (GPU) to optimize the time efficiency. We chose to run each case’s code for three epochs, as suggested by the BERT model’s developers [11], and we separated each code for each case and manually double-checked the effect of the preprocessing technique performed on the dataset. The rest of the code concerns the implementation of the BERT model for binary classification: 0 for authors who are men, 1 for authors who are women. The only difference in each experiment was in preprocessing; the rest of the experiment parameters were controlled and the same.

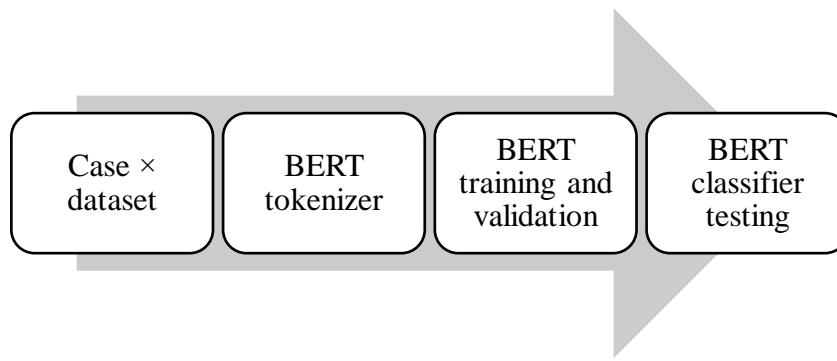


Figure 1. Experiments pipeline. The sequence of the conducted experiments. The case number changes based on the case we are considering. The rest of the experiment is the same for all cases.

Figure 1 illustrates the steps of our implementation which can be summarized as follows:

1. Dataset handling and preprocess:
 - 1.1. Reading the dataset and storing in dataframe
 - 1.2. Preprocessing based on the target case x ($x=1-5$), see table 2
 - 1.3. Loading BERT tokenizer
 - 1.3. Building PyTorch dataset
 - 1.4. Splitting Dataset into 90% training dataset and 10% testing dataset.
 - 1.5. Building dataloader using BERT tokenizer, Batch_size = 32, and MAX_LENGTH = 100
2. Model fine-tuning:
 - 2.1. Loading bert-base-cased from Huggingface library
 - 2.2. Building a binary classifier using BERT pretraining parameters with learning rate = 2^{-5}
 - 2.3. Implementing functions for training and evaluation
 - 2.5 Iterating the training and evaluation for the number of epochs = 3
3. Testing the model:
 - 3.1. Implementing testing function using the trained classifier
 - 3.2. Classifying testing dataset and compare prediction with real values
 - 3.3. Creating a classification report for f1-score

The model that we used is the uncased BERT base model with hyperparameters listed in Table 3. For our study, the values of the hyperparameters and the type of the BERT model considered were not in focus; therefore, we did not dedicate much time to experimenting with different values of the hyperparameters. To better measure the effect of the various preprocessing techniques, we controlled the values of the hyperparameters in all five experiments.

Because transfer learning does not require much consideration of features engineering, the only obvious parameter that can vary from one study to another is the preprocessing techniques. The obvious advantage of examining those cases is to determine the best preprocessing technique for author profiling when transfer learning is used.

Table 3. Values of the hyperparameters.

Parameter	Value
Model	BERT-base-cased
Epochs	3
Batch size	32
Text max length	100
Learning rate	2^{-5}

4. RESULTS AND DISCUSSION

In our study, we sought to examine the effect of the most commonly used preprocessing techniques on the gender profiling of authors when using a pretrained model: the BERT model. To distinguish our five cases of preprocessing techniques from each other, we conducted an experiment for each case. Because transfer learning models were trained on a relatively large dataset, we thought that the pretrained models would perform better in downstream tasks when the downstream dataset was larger. After performing the five experiments, we found that the best case for BERT was not applying any preprocessing technique and that the worst was Case 4, when we applied five preprocessing techniques (i.e., mentions removal, retweet tags removal, hashtags removal, URLs removal, and punctuation removal). Removing stop words also negatively affected the use of the BERT model. The rest of the cases differed slightly in accuracy, as shown in Table 4. The difference between Case 4 and Case 5 contributed to a significant difference overall (i.e., approx. 8%). The fewer preprocessing techniques we applied, the higher accuracy we observed. However, the time needed to train the model for Case 5 was the longest.

Table 4. Results of experiments. We consider using cross-validation to test the accuracy of the built models. We split the dataset into 90% for training and 10% for testing.

Case	Accuracy
Case 1	0.8229
Case 2	0.8074
Case 3	0.7946
Case 4	0.7886
Case 5	0.8667

A possible explanation for those results is that pretrained models perform better on larger texts and need every token that they might learn from. Even though stop words might not be used as markers in machine learning methods [28], the BERT model performed better when stop words were not removed. As mentioned, transfer learning techniques are features-independent, and their capability in contextually model language makes them powerful enough to understand natural language. Our goal was to test aspects that can affect the performance of such pretrained models, namely the most commonly used preprocessing techniques for profiling the age and gender of authors. As a result, our study offers valuable findings that shed light on the best preprocessing techniques that can be applied when using a pretrained model to profile authors by gender.

5. CONCLUSIONS

Our experiments on how preprocessing techniques impact the gender profiling of authors when using a transfer learning model confirmed that the BERT performs best when no preprocessing techniques are applied. They also revealed that removing stop words lowers the accuracy by 1%. Those results indicate that pretrained models perform better when longer texts are present. Other common preprocessing techniques in the literature were included in the experiments and showed that they affect the pretrained model performance negatively. On top of that, our findings suggest that the use of pretrained models could be standardized, for it does not rely on many dependent parameters such as preprocessing and variable features.

For the future, the study has provided the groundwork for using transfer learning techniques to advance the field of author profiling, with findings that can serve as a starting point for using transfer learning in author profiling. Although the scope of the study was limited in terms of the

pretrained model and the number of preprocessing techniques used, future work could involve using more than one pretrained model in a bid to better generalize the findings. Researchers could also apply more preprocessing techniques to cover more preprocessing possibilities.

REFERENCES

- [1] S. Keretna, A. Hossny, and D. Creighton, "Recognising user identity in twitter social networks via text mining," *Proc. - 2013 IEEE Int. Conf. Syst. Man, Cybern. SMC 2013*, pp. 3079–3082, 2013, doi: 10.1109/SMC.2013.525.
- [2] T. Yepes, Ray (ATX Forensics LLC, Austin, "The Art of Profiling in a Digital World.pdf." International Association of Chiefs of Police, p. 8, 2016.
- [3] A. Halimi and E. Ayday, "Profile Matching Across Unstructured Online Social Networks: Threats and Countermeasures *."
- [4] A. Rocha et al., "Authorship Attribution for Social Media Forensics," *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 1, pp. 5–33, 2017, doi: 10.1109/TIFS.2016.2603960.
- [5] S. Argamon, M. Koppel, J. W. Pennebaker, and J. Schler, "Automatically profiling the author of an anonymous text," *Commun. ACM*, vol. 52, no. 2, pp. 119–123, 2009, doi: 10.1145/1461928.1461959.
- [6] M. De-Arteaga, S. Jimenez, G. Dueñas, S. Mancera, and J. Baquero, "Author profiling using corpus statistics, lexicons and stylistic features: Notebook for PAN at CLEF-2013," *CEUR Workshop Proc.*, vol. 1179, 2013.
- [7] R. Bayot and T. Goncalves, "Multilingual author profiling using word embedding averages and SVMs," *Ski. 2016 - 2016 10th Int. Conf. Software, Knowledge, Inf. Manag. Appl.*, pp. 382–386, 2017, doi: 10.1109/SKIMA.2016.7916251.
- [8] M. Agrawal and T. Gonçalves, "Age and Gender Identification using Stacking for Classification★ Notebook for PAN at CLEF 2016," in *CEUR Workshop Proceedings*, 2016, vol. 18, no. 24, p. 28.
- [9] L. Miculicich Werlen, "Statistical Learning Methods for Profiling Analysis Notebook for PAN at CLEF 2015," *CLEF 2015 Labs Work. Noteb. Pap.*, 2015.
- [10] A. Vaswani et al., "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.
- [11] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019.
- [12] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," *ACL 2018 - 56th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (Long Pap.*, vol. 1, pp. 328–339, 2018, doi: 10.18653/v1/p18-1031.
- [13] F. Rangel, P. Rosso, M. Koppel, E. Stamatatos, and G. Inches, "Overview of the author profiling task at PAN 2013," *CEUR Workshop Proc.*, vol. 1179, pp. 8–11, 2013.
- [14] F. Rangel and I. Chugur, "Overview of the 2nd Author Profiling Task at PAN 2014," 2014.
- [15] F. Rangel, F. Celli, P. Rosso, M. Potthast, B. Stein, and W. Daelemans, "Overview of the 3rd Author Profiling Task at PAN 2015," in *CEUR Workshop Proceedings*, 2015, vol. 1179, [Online]. Available: <http://pan.webis.de>.
- [16] F. Rangel, P. Rosso, B. Verhoeven, W. Daelemans, M. Potthast, and B. Stein, "Overview of the 4th author profiling task at PAN 2016: Cross-genre evaluations," *CEUR Workshop Proc.*, vol. 1609, pp. 750–784, 2016.
- [17] F. Rangel, P. Rosso, M. Potthast, and B. Stein, "Overview of the 5th Author Profiling Task at PAN 2017: Gender and Language Variety Identification in Twitter," *CEUR Workshop Proc.*, vol. 2380, 2017.
- [18] H. Gómez-Adorno, I. Markov, G. Sidorov, J. P. Posadas-Durán, M. A. Sanchez-Perez, and L. Chanona-Hernandez, "Improving Feature Representation Based on a Neural Network for Author Profiling in Social Media Texts," *Comput. Intell. Neurosci.*, vol. 2016, 2016, doi: 10.1155/2016/1638936.
- [19] E. Lundqvist and M. Svensson, "Author profiling: A machine learning approach towards detecting gender, age, and native language of users in social media," no. 17013, p. 81, 2017.

- [20] S. Mamgain, R. C. Balabantaray, and A. K. Das, "Author profiling: Prediction of gender and language variety from document," *Proc. - 2019 Int. Conf. Inf. Technol. ICIT 2019*, pp. 473–477, 2019, doi: 10.1109/ICIT48102.2019.00089.
- [21] R. Bakkar Deyab, J. Duarte, and T. Gonçalves, "Author Profiling Using Support Vector Machines Notebook for PAN at CLEF 2016," in *CEUR Workshop Proceedings*, 2016, pp. 2–5.
- [22] G. Kheng, L. Laporte, and M. Granitzer, "INSA Lyon and UNI passau's participation at PAN@CLEF'17: Author Profiling task: Notebook for PAN at CLEF 2017," in *CEUR Workshop Proceedings*, 2017, vol. 1866.
- [23] M. Martinc, I. Skrjanec, K. Zupan, and S. Pollak, "PAN 2017: Author Profiling-Gender and Language Variety Prediction.," in *Working Notes of CLEF 2017-Conference and Labs of the Evaluation Forum, Ireland, 11-14 September, 2017*.
- [24] S. S. R. Seelam, S. Kumar, C. M. Gopi, and R. T. Raghunadha, "A New Term Weight Measure for Gender and Age Prediction of the Authors by analyzing their Written Texts," *Proc. 8th Int. Adv. Comput. Conf. IACC 2018*, pp. 150–156, 2018, doi: 10.1109/IADCC.2018.8692092.
- [25] F. Rangel, P. Rosso, M. Koppel, E. Stamatatos, and G. Inches, "Overview of the author profiling task at PAN 2013," *CEUR Workshop Proc.*, vol. 1179, 2013.
- [26] S. Panigrahi, A. Nanda, and T. Swarnkar, "A Survey on Transfer Learning," *Smart Innov. Syst. Technol.*, vol. 194, pp. 781–789, 2021, doi: 10.1007/978-981-15-5971-6_83.
- [27] Q. Yang, Y. Zhang, W. Dai, and S. J. Pan, "Transfer Learning in Natural Language Processing," *Transf. Learn.*, pp. 234–256, 2020, doi: 10.1017/9781139061773.020.
- [28] T. R. Reddy, B. V. Vardhan, and P. V. Reddy, "N-gram approach for gender prediction," *Proc. - 7th IEEE Int. Adv. Comput. Conf. IACC 2017*, pp. 860–865, 2017, doi: 10.1109/IACC.2017.0176.

Pedestrian Attribute Recognition using Gabor Wavelet Layers

Imran N. Junejo

Zayed University, Dubai, 19282, U.A.E.

Abstract. We address the problem of Pedestrian Attribute Recognition (PAR) in this paper. Owing to the presence of surveillance cameras in almost all outdoor and indoor public spaces, keeping an eye on pedestrian is a sought-after task with many useful applications. The problem entails recognizing attributes such as age-group, clothing style, accessories, footwear style etc. This is a multi-label problem and challenging even for human observers. We propose using a convolution neural network (CNN) with trainable Gabor wavelets (TGW) layers. The proposed layers are learnable and adapt to the dataset for a better recognition. The proposed multi-branch neural network is a mix of TGW and convolutional layers and we show its effectiveness on a public dataset.

Keywords: Gabor Wavelets, Convolutional Neural Networks, Pedestrian Attributes.

1 Introduction

Pedestrian attribute recognition is one of the active areas of research in the field of computer vision. The pedestrian attribute recognition deals with identifying a number of visual attributes from an image data. The identified attributes can belong to different classes, e.g. clothing style, footwear, gender, age group etc. A successful outcome of this research can be applied to various domains. It can be employed for motion analysis [20], where it can be used to identify crowd behavior attributes. Another important area of application is image-based surveillance or visual features extractions for person identification [18, 19]. Other applications include video analytics for business intelligence, or searching a criminal database for suspects using the identified visual attributes. Various factors make this a challenging problem. One of the main factors that makes this problem very difficult is the varying lighting conditions. Attributes of the same type of clothing can appear completely different under different lighting conditions. For example, distinguishing between black and dark blue colors is very difficult in certain weather conditions. Both colors will appear very similar to the camera in a darker environment. Occlusion also complicates the correct visual attribution identification and recognition. Occlusions can be either complete or partial and can result due to the camera orientation or from object self occlusions. For example, if a person is wearing a hat, it might appear partially in the image, or its shape might be completely different. Similarly, the orientation of a person or a camera can hide a backpack partially or completely from the view. These examples clearly show that settings of an acquisition environment for image or video capture result in a high intra-class variations for the same visual attributes.

The focus of this work is the identification of visual attributes from image and video data. The distance of an object from the camera affects how that object appears in the image. If an object is very far from the camera, or if the image resolution is very low, a visual attribute, e.g. dress, hat, backpack, scarf, shoes etc. will only occupy a few pixels in the image. The combination of low image resolution, in addition to the self-occlusions or view-oriented occlusions, makes visual attribute identification a very challenging problem. Many of these issues can be seen in the most widely used pedestrian dataset. Figure 1 shows some of the samples from the PEdesTrian Attribute (PETA) [8]. PETA is the largest benchmark dataset. It comprises of 19000 images of different resolution that cover more than 60 attributes. The dataset is acquired from real-world surveillance camera systems and includes images of 8,705 persons. It is a very challenging dataset because of the acquisition setup and scene settings. As can be seen in Figure 1, the quality of images is very low as well. This is due to a number of factors: images are very low resolution, acquisition problems result in a significant blur, many of the attributes are hidden due to severe occlusions. Moreover, due to the fast motion or acquisition problems some of the objects appear quite blurred thus making it a very challenging problem.

Visual attribute recognition problem can be solved in different ways, but the predominant solutions involve a two step process. In the first step, a feature extraction algorithm is employed to find a feature representation of the attributes. A number of feature extraction solutions are discussed in the computer vision literature. Most of these techniques require a very expert domain knowledge, and also needs a very high level of fine tuning for an accurate representation of visual attributes. For feature representation, methods like SIFT [16], HoG [7] or Haar-like features [25] have been employed in the field rigorously. Feature extraction is followed by the attributes classification step. For classification, Support Vector Machines (SVM) [8] has been the most widely used technique in the last decade.

In recent years, the convolutional neural networks (CNNs) have almost completely replaced SVMs for classification tasks. Compared to earlier attribute learning or image classification methods, CNNs are more effective and robust. Sarfraz et al. [23] proposed an end-to-end CNN-based network (VeSPA). This network had four parts, where each part corresponds to a specific pose category. Pose-specific attributes of each category are learned by each of these network parts. Their work demonstrated that coarse body pose information greatly influences the pedestrian attribute recognition. They extended their work in [21] and added a ternary view classifier in a modified approach that employed a global weighting solution. In this work, the global weighting solution for feature maps was employed before the final embedding. P-Net [2] employs a part-based approach. Based on GoogLeNet, the method guides the refined convolutional feature maps to capture different location information for the attributes related to different body parts. A joint person re-identification and attribute recognition approach (HydraPlus-Net) is presented by Liu et al. [15]. HydraPlus-Net is an Inception-based network and aggregates feature layers from multi-directional attention modules for the final feature representation. Sarafianos et al. [22] presented a multi-branch network that employed a simple weight scheme to address the class imbalance problem. They extracted visual attention masks to guide the network to crucial body parts. The masks are then fused at



Fig. 1: PETA [8] dataset Samples.

different scales to obtain a better feature representation. Another end-to-end method for person attribute recognition that uses Class Activation Map (CAM) network [27] to refine attention heat map is proposed by Guo et al [10]. The heat map identifies the areas of different image attributes. They use CAM network to refine the attention heat map for an improved recognition. A Harmonious Attention CNN (HA-CNN) based joint learning approach for person re-identification is presented in [14]. They used HA-CNN for the joint learning of hard regional attention and soft pixel attention. Feature representation is obtained by this simultaneous optimization. A Multi-Level Factorization Net (MLFN) that factors the visual appearance of a person into latent discriminative factors is proposed by [4]. The factorization is done without manual annotation at multiple semantic levels. A Transferable Joint Attribute-Identity Deep Learning (TJ-AIDL) model that allows for a simultaneous learning of an identity discriminative and

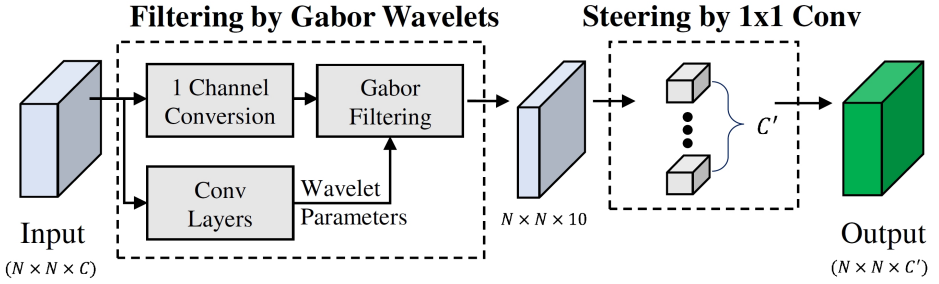


Fig. 2: Trainable Gabor Wavelet (TGW) layer [11]: Inputs and outputs are multichannel. A neural network is used to generate Gabor wavelet hyperparameters. These generated Gabor filters are then applied to the input. 1×1 convolution layer is added to enable the steerability of the Gabor wavelets.

attribute-semantic feature representation is proposed by [26]. Si et al. [24] proposed a Dual ATtention Matching network (DuATM), which is a joint learning end-to-end person re-identification framework. Their method simultaneously performs context-aware feature sequences learning and attentive sequence comparison in a joint learning mechanism for person re-identification.

Using Gabor wavelets with CNNs have received a tremendous attention as well [1, 3, 11, 17]. [1] use a Gabor filter bank as the first layer of a CNN and the bank gets updated using the standard back-propagation network leaning phase. [3] also use Gabor filters in the first layer of the network. While introducing lateral inhibition to enhance network performance, they use a n-fold cross validation to search for the best parameters. Authors in [17] introduce a Gabor Neural Network (GNN) where Gabor filters are incorporated into the convolution filter as a modulation process, in a spirit similar to the above mentioned works. In contrast to the above works where fixed Gabor filters are used, [11] introduce a trainable Gabor wavelets (TGW) layer. The authors present a method where the hyperparameters of the wavelets are learned from the input and a novel 1×1 convolution layers are employed to create steerable filters. In this paper, we propose using this TGW layer with our proposed CNN for a novel solution to the problem of PAR. We test on a challenging dataset and show a considerable improvement over state of the art.

2 Main Approach

In this section, we start with the description of the Gabor wavelet layer. Then we describe the architecture of our network in general.

2.1 Gabor Wavelet Layer

We make use of the Trainable Gabor wavelets (TGW) layer as proposed by Kwon et. al. [11] (see. Fig. 2). A neural network is used to generate the hyperparameters for the

Gabor wavelet and the generated Gabor filters are applied to filter inputs. In order to capture essential input features, a 1×1 convolution layer is added to the TGW layer to capture features at different orientations.

Hyperparameter estimation The 2D Gabor wavelet can be described as:

$$G(x, y) = \exp\left(-\frac{X^2 + \gamma Y^2}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda} X\right) \quad (1)$$

where γ represents aspect ratio, λ represents wavelength of the sinusoidal, σ represents width or the standard deviation, $X = x \cos(\theta) + y \sin(\theta)$, $Y = -x \sin(\theta) + y \cos(\theta)$, and θ is an angle in the range $[0, \pi]$. Thus in order to specify a continuous Gabor wavelet, we need to determine the set of hyperparameters $\{\gamma, \theta, \lambda, \sigma\}$. In order to convert the continuous filter to a discrete one, a sampling grids need to be defined, which is largely linked to σ . A new parameter is thus introduced to compute the discrete filter:

$$G[m, n] = g(u, v) = \left(\frac{m}{\lfloor \zeta \rfloor} \times \zeta, \frac{n}{\lfloor \zeta \rfloor} \times \zeta\right) \quad (2)$$

where m and n are in the interval $-\lfloor \zeta \rfloor, \lfloor \zeta \rfloor + 1, \dots, \lfloor \zeta \rfloor$, and by just varying $\lfloor \zeta \rfloor$, variety of sampling grids can be achieved [11]. For a loss function L , we need to compute $\frac{\partial L}{\partial \zeta}$ in order to train for the wavelet layer that is cascaded with our CNN. In order to train for the ζ , what remains is to compute $\frac{\partial G[m, n]}{\partial \zeta}$, as $\frac{\partial L}{\partial G[m, n]}$ is handled automatically by the deep learning libraries:

$$\frac{\partial G[m, n]}{\partial \zeta} = \frac{\delta g(u, v)}{\partial u} \frac{\partial u}{\partial \zeta} + \frac{\partial g(u, v)}{\partial v} \frac{\partial v}{\partial \zeta} \quad (3)$$

$$= \frac{\delta g(u, v)}{\partial u} \frac{u}{\zeta} + \frac{\partial g(u, v)}{\partial v} \frac{v}{\zeta} \quad (4)$$

as $\frac{d}{d\zeta} \lfloor \zeta \rfloor = 0$. The remaining parameters $\frac{\partial G[m, n]}{\partial \sigma}$, $\frac{\partial G[m, n]}{\partial \gamma}$, $\frac{\partial G[m, n]}{\partial \lambda}$ can be computed in a similar way and a similar parameterization can be adopted for the parameters σ, γ and λ .

A very significant parameter for the Gabor wavelet is the orientation (θ). These values are mostly chosen empirically. This parameter is also made trainable to better design orientations for the task at hand. To use the steering property, where a linear combination of finite set of responses can be used to represent convolution at any orientation, a 1×1 convolution layer, working as a linear combination layer, is added to the output of the generated filters. For this layer, ten equally spaced fixed orientations are selected, working as basis filters: $9^\circ, 27^\circ, 45^\circ, 63^\circ, 81^\circ, 99^\circ, 117^\circ, 135^\circ, 153^\circ$, and 171° [11].

2.2 Attribute Recognition Network

The above mentioned TGW layer can be thought of as a feature extracting layer. In addition to this, we also employ it as the key building block of our network. Thus, in

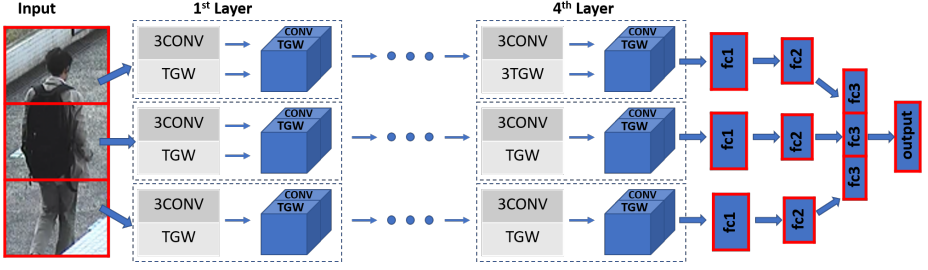


Fig. 3: Our Approach: The proposed method divides the input image into three parts. For each branch, the network contains 4 layers that are a mix between TGW and 3Conv layer (mixed-layers). The output of each branch is followed by three fc layers. Size of the last layer of the network matches the number of attributes of the dataset. Parameters of the network are mentioned in Table 1.

addition to functioning as the *lowest layer*, it also aids the network to learn high level features.

The proposed network is shown in Fig. 3. An input image is divided into three equal parts along on the vertical axis. Each part of the image passes through a separate branch of the network. As can be seen in the figure, each branch consists of 4 mixed-layers: combination of TGW layer and a 3×3 convolution layer. The input to the TGW layer starts with a 1-channel conversion, i.e. a multi-channel input is converted to a 1-channel, which is a summation over the channels operation for all layers except the first layer where we perform a simple color-to-gray image conversion. The parameters for these layers are given in Table 1.

Each mixed-layer (1 to 4) contains 256 channels from the TGW layer and 256 channels from a 3×3 convolution layer (denoted as 3Conv). Thus depth of each mixed-layer output is 512 (concatenation of TGW and 3Conv layer). The network thus contains blocks of layers stacked together. For each 3Conv layer, as the name suggest, the kernel size is 3×3 . The convolution is followed by ReLU activation function, max-pool layer (size 2×2), and Batch Normalization (BN) layer. The size of the input image to each of these stacked layers is, respectively: 48×48 , 24×24 , 12×12 , and 6×6 .

Output from each branch encounters three fully connected layers, i.e. fc1, fc2 and fc3, of size 512, 512 and 35, respectively. Each fc layer uses ReLU as the activation function, followed by a dropout layer ($p = 0.5$), to minimize the number of parameters of the network. fc3 from all branches are concatenated and the final output layer size matches the number of dataset attributes.

The method proposes using Gabor wavelets embedded with a deep neural network. Whereas other methods construct Gabor filters manually, the proposed network learns the wavelet parameters suitable to the dataset. Generated Gabor filters are stacked with convolution layers to build the overall network. As we shall show next, the proposed network is efficient and learns the dataset structure well to perform at par with state of the art.

Layer	γ_o	λ_o	σ_o	ζ_o	TGW Channels	Conv Channels
1	0.3	6.8	5.4	6	256	256
2	0.3	5.6	4.5	5	256	256
3	0.3	4.6	3.6	4	256	256
4	0.3	3.5	2.8	3	256	256

Table 1: Parameters used for the TGW layers.

3 Evaluation

Following channel conversion, the grayscale image is divided into three parts. Each part of the networks encounters 4 mixed-layers, consisting of equal number of channels from TGW and 3_{Conv} layer. Depth of each mixed-layer is 512. The mixed-layers are followed by a series of fully connected layers before the final output layer. ReLU is used as the activation function for all the layer. The output layer uses `sigmoid` as the activation function.

In order to evaluate our method quantitatively, we compute various measures and report the results below. Although mean accuracy has been widely used in the attribute recognition literature, it treats each attribute independent of the other attributes. This might not necessarily be the case and an inter-attribute correlation might exist. Therefore, researchers also report *example-based* evaluations, namely accuracy (Acc), precision ($Prec$), recall (Rec), and F1 score ($F1$) [13].

3.1 Dataset

PETA is one of the most widely used dataset for the problem of pattern attribute recognition. Collected from real-time surveillance cameras, the PETA dataset contains 19,000 images collected from 10 publicly available datasets. The resolution of the images ranges from 17×39 to 169×365 . Most of the previous works [12, 23] report results on the PETA dataset using only 35 attributes. Similarly, for a fair comparison, experiments are conducted on 5 random splits: we allocate 9,500 samples for training, 1,900 samples for validation, 7,600 samples for testing on the dataset.

Pre-processing: Before continuing to the next step, we perform **mean subtraction**: That is, we compute the mean for all the images for each color spaces and this value is subtracted from image data. Intuitively for each dimension, this step is equal to centering the data around its origin. Next step involves **normalization**: We compute the standard deviation separately for each color space and the image data is divided by this value.

3.2 Setup

For deep learning, we adopted the KERAS [6] library, which is based on the TensorFlow backend. All experiments were performed on a cluster node with 2 x Intel Xeon E5 CPU, 128GB Registered ECC DDR4 RAM, 32TB SAS Hard drive storage, and 8 x NVIDIA Tesla K80 GPUs.

	PETA [8]			
	<i>Acc</i>	<i>Prec</i>	<i>Rec</i>	<i>F1</i>
Chen et. al. [12]	75.07	83.68	83.14	83.41
Liu et. al. [28]	74.62	82.66	85.16	83.40
Sarfaraz et. al. [23]	77.73	86.18	84.81	85.49
ours	79.35	86.24	79.45	81.48

Table 2: Quantitative results (%) on PETA datasets. Results are compared with the other benchmark methods. As can be seen, we have comparable results, with considerable improved accuracy for the datasets.

3.3 Implementation Details

We train the network for 50 epochs. ReLU was used as the activation function for all layers of the network. We used the Adam for update optimizer using the parameters: learning rate = $1e^{-4}$, $\beta_1 = 0.9$ and $\beta_2 = 0.999$.

We added the dropout layers to the fc layers to prevent model over-fitting. We adopt weight decay by a factor of 0.1 after 15 epochs. The batch size was set to be 8. All weights in the network are initialized using He Normal initialization.

For the TGW layers with a steering block, we use the scheme suggested by [9]: we fix the parameters $\{\gamma, \sigma, \lambda\}$ as shown in Table 1 while training for ζ . This setup yields the best results in our experiments.

3.4 Results

We evaluate the effectiveness of the proposed method on PETA datasets. Table 2 shows a comparison of the proposed method with six current state of the art methods. For the PETA dataset, *Acc* obtained from our method is 79.35%. This is higher than all the other methods that we compare with. The obtained results for the other measures (*Pre*, *Rec* and *F1*) is 86.24%, 79.45%, and 81.48% respectively. Class-wise accuracy chart for the PETA dataset is shown in Fig. 4. Interestingly, the lowest accuracy is that for the class `upperBodyOther`. Considering the image resolutions in the dataset, this is indeed a very difficult class to accurately measure. On the other hand, the highest accuracy is that of the classes `upperBodyThinStripes` and `upperBodyVNeck`.

The proposed method makes a novel use of the Gabor wavelet layers. Instead of manually constructing Gabor filters, the layers are trainable and are able to correctly estimate model parameters. The method divides input image into three parts. For each part, we train four mixed-layers: combination of TGW and 3Conv layers. The output of these branches are concatenated and then followed by three fc layers. We have obtained very encouraging results for the key measures. The method is novel and unique in the sense that it does not resort to data augmentation or part-based computations, as employed by [13]. We also do not have to compute pose estimation [12], or construct any hand-crafted features [5]. Our results are an improvement over state of the art and clearly justifies the use of Gabor wavelet layers.

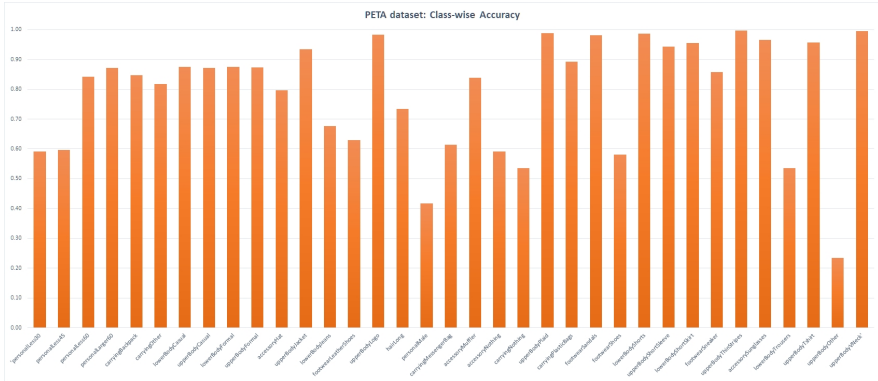


Fig. 4: Class-wise Accuracy - PETA dataset: the figure shows the obtained class-wise accuracy. The highest accuracy is for the class `upperBodyThinStripes`, `upperBodyVNeck`. The lowest accuracy is 23.4% for the class `upperBodyOther`.

4 Conclusion

This work proposes the idea of using trainable Gabor wavelets (TGW) for the task of pedestrian attribute recognition. We have proposed a multi-branch neural network. The input to the network is an image that is divided into three parts, each processed through a different branch of the network. Each branch contains mixed-layers that are capable of learning the Gabor wavelet parameters. The filters in each branch are learned from the data itself. We have tested the data on a challenging public dataset and are encouraged by the results. In future work, we aim to experiemnt with other publicly available datasets with possibly different network architectures.

References

1. Alekseev, A., Bobe, A.: Gabornet: Gabor filters with learnable parameters in deep convolutional neural network. In: 2019 International Conference on Engineering and Telecommunication (EnT). pp. 1–4 (2019). <https://doi.org/10.1109/EnT47717.2019.9030571>
2. An, H., Fan, H., Deng, K., Hu, H.M.: Part-guided network for pedestrian attribute recognition. 2019 IEEE Visual Communications and Image Processing (VCIP) pp. 1–4 (2019)
3. Bai, J., Zeng, Y., Zhao, Y., Zhao, F.: Training a v1 like layer using gabor filters in convolutional neural networks. In: 2019 International Joint Conference on Neural Networks (IJCNN). pp. 1–8 (2019)
4. Chang, X., Hospedales, T.M., Xiang, T.: Multi-level factorisation net for person re-identification. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
5. Chen, Y., Duffner, S., STOIAN, A., Dufour, J.Y., Baskurt, A.: Pedestrian attribute recognition with part-based CNN and combined feature representations. In: Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. pp. 114–122 (2018)

6. Chollet, F.: keras (2015), <https://github.com/fchollet/keras>
7. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). vol. 1, pp. 886–893 (2005)
8. DENG, Y., Luo, P., Loy, C.C., Tang, X.: Pedestrian attribute recognition at far distance. In: Proceedings of the 22nd ACM International Conference on Multimedia. pp. 789–792. MM '14 (2014)
9. Guo, G., Mu, G., Fu, Y., Huang, T.: Human age estimation using bio-inspired features. In: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009. pp. 112 – 119 (07 2009). <https://doi.org/10.1109/CVPR.2009.5206681>
10. Guo, H., Fan, X., Wang, S.: Human attribute recognition by refining attention heat map. Pattern Recognition Letters **94**(C), 38–45 (Jul 2017)
11. Kwon, H.J., Koo, H., Soh, J.W., Cho, N.I.: Age estimation using trainable gabor wavelet layers in a convolutional neural network. 2019 IEEE International Conference on Image Processing (ICIP) pp. 3626–3630 (2019)
12. Li, D., Chen, X., Zhang, Z., Huang, K.: Pose guided deep model for pedestrian attribute recognition in surveillance scenarios. In: 2018 IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6 (2018)
13. Li, D., Zhang, Z., Chen, X., Ling, H., Huang, K.: A richly annotated dataset for pedestrian attribute recognition. CoRR **abs/1603.07054** (2016)
14. Li, W., Zhu, X., Gong, S.: Harmonious attention network for person re-identification. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
15. Liu, X., Zhao, H., Tian, M., Sheng, L., Shao, J., Yan, J., Wang, X.: Hydraplus-net: Attentive deep features for pedestrian analysis. In: Proceedings of the IEEE international conference on computer vision. pp. 1–9 (2017)
16. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision. vol. 2, pp. 1150–1157 (1999)
17. Luan, S., Zhang, B., Zhou, S., Chen, C., Han, J., Yang, W., Liu, J.: Gabor convolutional networks. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1254–1262 (2018). <https://doi.org/10.1109/WACV.2018.00142>
18. Nanda, A., Chauhan, D.S., K. Sa, P., Bakshi, S.: Illumination and scale invariant relevant visual features with hypergraph-based learning for multi-shot person re-identification. Multimedia Tools and Applications **78**(4), 3885–3910 (Feb 2019)
19. Rahman, K., Abdul Ghani, N., Abdulbasah Kamil, A., Mustafa, A., Kabir Chowdhury, M.A.: Modelling pedestrian travel time and the design of facilities: A queuing approach. PLOS ONE **8**(5), 1–11 (05 2013). <https://doi.org/10.1371/journal.pone.0063503>
20. Raudies, F., Neumann, H.: A bio-inspired, motion-based analysis of crowd behavior attributes relevance to motion transparency, velocity gradients, and motion patterns. PLOS ONE **7**(12), 1–17 (12 2013). <https://doi.org/10.1371/journal.pone.0053456>
21. Saquib Sarfraz, M., Schumann, A., Eberle, A., Stiefelhagen, R.: A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
22. Sarafianos, N., Xu, X., Kakadiaris, I.A.: Deep imbalanced attribute classification using visual attention aggregation. In: Springer European Conference on Computer Vision. pp. 708–725 (2018)
23. Sarfraz, M., Schumann, A., Wang, Y., Stiefelhagen, R.: Deep view-sensitive pedestrian attribute inference in an end-to-end model. In: British Machine Vision Conference (BMVC) (09 2017)

24. Si, J., Zhang, H., Li, C.G., Kuen, J., Kong, X., Kot, A.C., Wang, G.: Dual attention matching network for context-aware feature sequence based person re-identification. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
25. Viola, P., Jones, M.: Robust real-time object detection. In: International Journal of Computer Vision (IJCV). vol. 57 (01 2001)
26. Wang, J., Zhu, X., Gong, S., Li, W.: Transferable joint attribute-identity deep learning for unsupervised person re-identification. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
27. Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A.: Learning deep features for scene recognition using places database. In: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1. pp. 487–495. NIPS'14, MIT Press, Cambridge, MA, USA (2014)
28. Zhou, Y., Yu, K., Leng, B., Zhang, Z., Li, D., Huang, K.: Weakly-supervised learning of mid-level features for pedestrian attribute recognition and localization. In: British Machine Vision Conference BMVC 4-7 (2017)

ONLINE OBSTRUCTIVE SLEEP APNEA DETECTION BASED ON HYBRID MACHINE LEARNING AND CLASSIFIER COMBINATION FOR HOME-BASED APPLICATIONS

Hosna Ghandeharioun

Department of Electrical and Biomedical Engineering,
Khorasan Institute of Higher Education, Mashhad, Iran

ABSTRACT

Automatic detection of obstructive sleep apnea (OSA) is in great demand. OSA is one of the most prevalent diseases of the current century and established comorbidity to Covid-19. OSA is characterized by complete or relative breathing pauses during sleep. According to medical observations, if OSA remained unrecognized and un-treated, it may lead to physical and mental complications. The gold standard of scoring OSA severity is the time-consuming and expensive method of polysomnography (PSG). The idea of online home-based surveillance of OSA is welcome. It serves as an effective way for spurred detection and reference of patients to sleep clinics. In addition, it can perform automatic control of the therapeutic/assistive devices. In this paper, several configurations for online OSA detection are proposed. The best configuration uses both ECG and SpO2 signals for feature extraction and MI analysis for feature reduction. Various methods of supervised machine learning are exploited for classification. Finally, to reach the best result, the most successful classifiers in sensitivity and specificity are combined in groups of three members with four different combination methods. The proposed method has advantages like limited use of biological signals, automatic detection, online working scheme, and uniform and acceptable performance (over 85%) in all the employed databases. These advantages have not been integrated in previous published methods.

KEYWORDS

Obstructive Sleep Apnea, Supervised Machine Learning, Feature Reduction, Classifier Combination, Biomedical Signal Processing.

1. INTRODUCTION

Obstructive sleep apnea (OSA) is the most prevalent sleep-related breathing disorder worldwide [1]. It has also established as a comorbidity to Covid-19 [2]. Intermittent episodes of airway subsidence during sleep characterizes OSA [3]. If OSA remains undetected and untreated, the resultant abrupt changes in sympathetic neural activity may cause severe cardiovascular side-effects [4], type 2 diabetes [5], impaired cognition, and psychiatric symptoms [6]. Hence the detection and immediate treatment of OSA is essential. The diagnosis of OSA requires the joint evaluation of related clinical features and the visible demonstrations of abnormal breathing during sleep. [7]. The gold standard for the detection of abnormal breathing during sleep is overnight polysomnography (PSG). The PSG-driven apnea-hypopnea index (AHI) characterizes the OSA severity [8, 9]. AHI derivation is currently performed visually according to the American Association of Sleep Medicine (AASM) guidelines [8]. This time-consuming and

expensive process imposes a heavy burden on the public health section [10]. Therefore, many automatic methods for pre-clinic detection and scoring of OSA have been developed in the literature [11-27, 31, 32, 34, 35]. These methods use analysis of a variety of biological signals and machine learning techniques. In some studies, electroencephalogram (EEG) is used for feature extraction based on occurred discrepancies between the right and left hemispheres [11] or tracking non-linear behavior of EEG due to fluctuations in sleep depth [12, 29, 30]. Single-channel ECG or combination of ECG and saturated oxygen level of the blood in peripheral veins (SpO₂) is also suggested in several studies due to easy and unobtrusive signal acquisition [13, 14, 16-19, 23].

In the most recent studies, OSA detection is accomplished based on ECG and the newly widespread deep learning techniques. In deep learning solutions, the feature extraction/selection is generally embedded in the learning algorithm, and no separate step is needed [32]. This advantage reduces the computational load. However, for deep learning training, high-performance computers are required [33], and the methodologies do not suit home-based and portable applications where the processing ability and data storage capacity are limited. Apnea is detected based on nasal pressure signals with the help of convolutional neural networks (CNN) in [34]. Several supervised machine learning methods are tested for OSA detection with a single channel ECG signal in [35]. The achieved results are promising, yet in a small database and with slightly less accuracy than our suggested strategy.

In this study, several configurations for online detection of OSA are suggested. Employing a limited number of biological signals, automatic and real-time detection, and uniform acceptable performance over several databases are the merits of our proposed method. To the knowledge of the author, these advantages are accumulated in none of the previous studies together.

2. MATERIAL AND METHOD

Automatic detection of respiratory events based on supervised machine learning is generally divided into several steps [28]. In the first step, the training set is made from signal records labeled as apnoeic and normal (by an expert clinician). In the second step, feature extraction is performed for each signal. The extracted features can be reduced to improve the performance of the next step. Finally, the last step is the classification of the test records. We will go through each step of our work in detail.

We conducted this study based on three databases. The first two databases are public and can be reached by anyone: St. Vincent, University College Dublin (UCD) database [36], eight subjects of Apnea-ECG database [37] whose data include more signals than one ECG channel. The third database is exclusively at our disposal. This database includes clinical records of the sleep laboratory of Ibn-e-Sina Hospital, Mashhad, Iran, from July 2012 to May 2014. The study was approved by the ethics committee overseeing the research proposal (permission no.92/620792, date 2014/03/07). We were allowed to use clinical data only, with no deviation from AASM protocol. The PSG (model: Alice LE, part no. 1002387, Philips Respironics) recordings were conducted in baseline montage with 16 channels on the 158 referred patients. Out of all participants, 134 subjects were diagnosed with OSA, and 24 healthy according to the International Classification of Sleep Disorders II (ICSD-II) [8]. We ascertained sleep apneas as ≥ 10 s of airflow pauses and hypopnea as a $\geq 3\%$ of oxygen desaturation/or arousal preceded by a 50% decrement in the amplitude of baseline airflow. From now on, we refer to this database as “the exclusive database”. Figure 1 shows a 1-minute frame of polysomnographic records of our exclusive database.

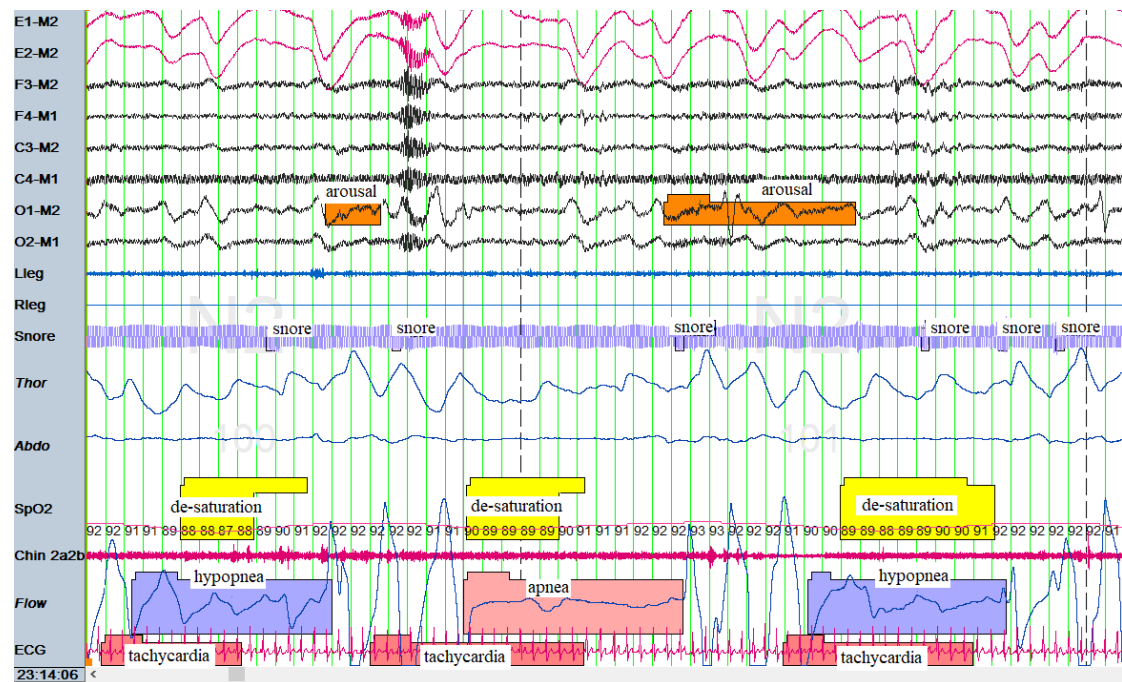


Figure 1. 1-minute frame of the polysomnographic records of a subject with severe OSA from the exclusive database

The three signals of EEG, SpO₂, and air pressure/flow have a central role in the clinical definition of apnea. We refer to them as “the main signals”. Other biological signals (such as ECG, voice, and actigraphy) have a supplementary role in the detection of OSA. We refer to this group as “the auxiliary signals”. Relying on the main signals for an OSA detection system is the first choice; however, the developed system must be more concise than PSG and perform a pre-clinic screening. Placing EEG electrodes on the scalp during sleep and pressure/flow sensors is rather obtrusive; besides, preparations and installation of electrodes and sensors are not straightforward for an ordinary user. For EEG acquisition and conditioning, a relatively expensive system is needed. The repeatability of the observed effects of OSA on EEG compared to SpO₂ signal is also on debate [38, 39]. That is why generally EEG and air pressure/flow signals are excluded.

Among the auxiliary signals, ECG is gained more attention in the OSA detection methods. The effects of the apneas on ECG signal are well understood [4]. The ECG electrodes are installed simpler than EEG and less obtrusive than those of air pressure/flow signal. The apparent effect of respiratory events on ECG is called Cyclical Variation of Heart Rate (CVHR) [4]. The challenge of ECG-based detection systems is their lower specificity since their modulating factor is not a respiratory event only. The presence of cardiovascular problems can also have considerable effects on ECG. In the absence of OSA, these effects can increase the false positive detection rate. In practice, the number of false-negative detections also increases, and the sensitivity of the OSA detection method drops. A decrease in sensitivity is because the database usually includes subjects with OSA whose problem has been un-diagnosed for years, and lack of treatment has led to cardiovascular complexities for them [4]. Up to 90% of subjects affected by OSA are not aware of their problem and have not been treated yet [1].

More successful results are reported for SpO₂-based detection methods compared to other single-channel detection systems. They have reasonable specificity and sensitivity, they can be

performed in real-time, and they have non-obtrusive sensors; additionally, some of them are realized in smartphones and can serve as useful home-based systems [27, 40].

In this study, we consider PPG (and SpO2) from “the main signals”, and ECG from “the auxiliary signals”. Parallel use of these signals, covers their deficiencies and increases the overall accuracy, sensitivity, and specificity of the detection system [16]. The OSA detection based on ECG and SpO2 is more popular than other multi-channel detection systems due to simple sensor installation and powerful representation of respiratory events [13, 14, 16-19, 23].

2.1. Pre-processing and Noise Rejection

Considering the ECG sampling frequency is essential. The insufficient sampling frequency may negatively affect the resolution and the signal-to-noise ratio of the R-R time series [41, 42]. The UCD and the Apnea-ECG databases have less sampling frequency than the specified 250Hz value of the American National Standard Institute (ANSI), yet they are good benchmarks for the evaluation of automatic OSA detection methods. We have assumed that their subjects are carefully selected so that exceptions, where their sampling frequencies are insufficient for representing ECG behavior, are deleted [42]. The ECG signals of the exclusive database are also down-sampled to 250Hz.

To avoid the aliasing effects of non-integer fractional down-sampling, equating the UCD and the Apnea-ECG sampling frequencies is avoided [43]. For de-trending and noise rejection, the decimated lifting wavelet transform (DWT) algorithm [44] is employed [13]. The Daubechies (D4) wavelet is used with seven levels of decomposition. The R-R time series is extracted by the famous and robust method of Hamilton-Tompkins [45, 46]. Impulses more or less than 20% distant to the last normal R-R interval, those with more than 30% values in the R-S difference or with the negative R-S difference values are assumed to be a sign of ectopic or abnormal beat and omitted; the resulting signal is called the R-R tachogram [38].

Table 1. The SpO2 features in each 1-minute frame: $\{spo2_i\}_{i=1}^{60}$

Name/ Definition
The minimum value of the frame
The average value of the frame
The standard deviation of the frame
Sequential correlation coefficients [20]
Sequential mutual information [52]
Average value crossing points
The absolute value of the slope of the line fitted over SpO2 [20]
y-Intercept value of the line fitted over SpO2 [20]
Approximate entropy [53]
Sample entropy [53]
Lempel-Zive complexity measure [54]
Central tendency measure (CTM_r) ($r=0.25, 0.75, 0.5, 1$) [54]
Delta measure (Δ) [30]
Baseline [22]
odi2, odi3, odi4: The number of 2%,3%, and 4% desaturations to the baseline [30]
ODI_{xy} : The number of desaturations more than or equal to $x\%$ lasting for y seconds [30]
$ODIS_x$: The number of desaturations more than or equal to $x\%$ [22]
Time elapsed under saturation level x ($\%tsax$); $x=70, 80, 85, 90, 95$) [30]

2.2. Feature Extraction

We consider values below 50% and fluctuations more than 40% in two consecutive samples of SpO2 signal (in the sampling period of 1s) artifacts [16, 19]. We eliminate these values and their corresponding values of other PSG signals from the records (2 minutes of the Apnea-ECG database, 37 minutes of the UCD database, and 78 minutes of the exclusive database, totally equal to 1.9% of available data). The resulting signal is divided into non-overlapping 1-minute frames and is used for feature extraction. Table 1 summarizes the SpO2 features.

We process the ECG signal in 1-minute time windows. The R-R tachogram is extracted from ECG. It is not a result of uniform ECG sampling. The points of this time series are scattered non-uniformly across the time axis based on the time interval of consecutive beats. In frequency analysis of ECG signal, this crucial fact is usually ignored. The pre-assumption of the fast Fourier transform (FFT) is the uniform sampling of the signal under analysis; hence the FFT-based frequency analysis of the R-R tachogram and its dependents like the ECG-derived respiration (EDR) are not appropriate. Frequency analysis tools needless of the uniform sampling assumption like the Lomb-Scargle periodogram are good candidates for calculating quantities related to the heart rate variability (HRV) [50].

The EDR is extracted by the T wave duration method [51, 52] in the UCD and ECG-Apnea databases. We calculate the EDR with the help of the area under the QRS graph [53] in our exclusive database.

We use the Lomb-Scargle periodogram and the DWT with Daubechies (D4) wavelet (with 18 levels of decomposition) to extract frequency-domain features of the R-R tachogram, and the EDR signals [44]. The ECG features are categorized as the time-domain, and the frequency-domain features in tables 2, 3 and 4.

Table 2. The time-domain ECG features

The R-R tachogram: $R(rr_{t_m}) = \{rr_i\}_{i=rr_{t_1}}^{rr_{t_m}}$, the EDR: $EDR(q) = \{edr_i\}_{i=1}^q$

Definition	Name
$\bar{rr}_t = \frac{1}{m} \sum_{i=1}^m rr_{t_i}$	Time window mid-time
M	length ECG
$\bar{rr} = \frac{1}{m} \sum_{i=1}^m rr_i$	Average beat [115]
$NN50v1 = \sum_{i=2}^m U(rr_i - rr_{i+1} - 50ms)$ U(.): step function	NN50-version 1 [115]
$NN50v2 = \sum_{i=1}^{m-1} U(rr_{i+1} - rr_i - 50ms)$ U(.): step function	NN50-version 2 [115]
$pNN50v1 = \frac{NN50v1}{m}$	pNN50-version 1 [115]
$pNN50v2 = \frac{NN50v2}{m}$	pNN50-version 2 [115]
$S_{rr} = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (rr_i - \bar{rr})^2}$	Tachogram standard deviation
$SDSD = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (rd_i - \bar{rd})^2}$ $rd_i = rr_{i+1} - rr_i$, $\bar{rd} = \frac{1}{m-1} \sum_{i=1}^{m-1} rd_i$	SDSD [115]

$RMSSD = \sqrt{\frac{1}{m-1} \sum_{i=1}^{m-1} rd_i^2}$	RMSSD [115]
$r_k = \frac{\sum_{i=1}^m (rr_i - \bar{rr})(rr_{i+k} - \bar{rr})}{\sum_{i=1}^m (rr_i - \bar{rr})^2}$	Sequential correlation coefficients [115]
$MI_k = \hat{I}(\{rr_i\}; \{rr_{i+k}\})$ $= \sum_{i=1}^m P_n(\{rr_i\}, \{rr_{i+k}\}) \log \frac{P_n(\{rr_i\}, \{rr_{i+k}\})}{P_n(\{rr_i\})P_n(\{rr_{i+k}\})}$ P_n : Probability distribution function	Sequential mutual information [319]
$AT_k = \frac{E((N_{i+1}(k) - N_i(k))^2)}{2E(N_{i+1}(k))}$ $N_i(k)$: Number of beats in the i^{th} section of a k -second signal	Allan Factor [124]
$NEP_k = \frac{1}{m-2} \sum_{i=2}^{m-1} (1 - U((rr_i - rr_{i-1})(rr_{i+1} - rr_i)))$	Number of Extreme Points [116]
$\overline{edr} = \frac{1}{q} \sum_{i=1}^q edr_i$	Average EDR
$S_{edr} = \sqrt{\frac{1}{q-1} \sum_{i=1}^q (edr_i - \overline{edr})^2}$	Standard Deviation EDR

Table 3. The frequency-domain features of the R-R tachogram: $R(rr_{t_m}) = \{rr_i\}_{i=rr_{t_1}}^{rr_{t_m}}$

Definition	Name
$S_{D_{rr}^k}^2 = \sum_{i=1}^{I_{rr,k}} (d_{rr,i}^k - \overline{d_{rr}^k})^2$ $\overline{d_{rr}^k} = \frac{1}{I_{rr,k}} \sum_{i=1}^{I_{rr,k}} d_{rr,i}^k$	Sample deviation of $\{D_{rr}^k\}_{k=2}^{17}$
$S_{D_{rr}^{LF}}^2 = \sum_{i=1}^{I_{rr,LF}} (d_{rr,i}^{LF} - \overline{d_{rr}^{LF}})^2$	Sample deviation of $\{D_{rr}^k\}_{k=2}^{17}$ (LF band)
$S_{D_{rr}^{HF}}^2 = \sum_{i=1}^{I_{rr,VLF}} (d_{rr,i}^{HF} - \overline{d_{rr}^{HF}})^2$	Sample deviation of $\{D_{rr}^k\}_{k=2}^{17}$ (HF band)
$P_{rr}^{VLF} = \int_{2\pi \times 0.04}^{2\pi \times 0.15} P_{rr}(\omega) d\omega$ $P_{rr}(\omega)$: Lomb-Scargel periodogram [348,86]	HRV Power spectrum (LF band)
$P_{rr}^{HF} = \int_{2\pi \times 0.15}^{2\pi \times 0.4} P_{rr}(\omega) d\omega$	HRV Power spectrum (HF band)
$LF/HF = P_{rr}^{LF} / P_{rr}^{HF}$	LF-HF power ratio in the HRV spectrum
$P_{rr}(\omega) _{2\pi \times 0.04}^{2\pi \times 0.4}$	Lomb-Scargel periodogram samples in LF-HF band
$\omega_{resp} = \operatorname{argmax}(P_{rr}(\omega) _{2\pi \times 0.15}^{2\pi \times 0.4})$	Estimated respiration frequency (Dominant HF-band frequency of HRV) [86]
$respMag = \max(P_{rr}(\omega) _{2\pi \times 0.15}^{2\pi \times 0.4}) = P_{rr}(\omega_{resp})$	Power at the dominant HF-band frequency of

	HRV
$respProb = Prob(P_{rr}(\omega_{resp}))$	Probability of estimated respiration frequency occurrence with power $P_{rr}(\omega_{resp})$
$\omega_{ProbMax} = argmax(Prob(P_{rr}(\omega) _{\frac{2\pi \times 0.4}{2\pi \times 0.04}}))$	Most probable frequency of the HRV spectrum
$ProbMax = max(Prob(P_{rr}(\omega) _{\frac{2\pi \times 0.4}{2\pi \times 0.04}}))$ $= Prob(P_{rr}(\omega_{ProbMax}))$	Probability of $\omega_{ProbMax}$ occurrence with power $P_{rr}(\omega_{ProbMax})$
$ProbMaxMag = P_{rr}(\omega_{ProbMax})$	Power of the HRV spectrum at $\omega_{ProbMax}$

Table 4. The frequency-domain features of the EDR: $EDR(q) = \{edr_i\}_{i=1}^q$

Definition	Name
$S^2_{D_{edr}^k} = \sum_{i=1}^{I_{edr,k}} (d_{edr,i}^k - \overline{d_{edr}^k})^2$ $\overline{d_{edr}^k} = \frac{1}{I_{edr,k}} \sum_{i=1}^{I_{edr,k}} d_{edr,i}^k$	Sample deviation of $\{D_{edr}^k\}_{k=2}^{17}$
$S^2_{D_{edr}^{LF}} = \sum_{i=1}^{I_{edr,LF}} (d_{edr,i}^{LF} - \overline{d_{edr}^{LF}})^2$	Sample deviation of $\{D_{edr}^k\}_{k=5}^{17}$ (LF band)
$S^2_{D_{edr}^{HF}} = \sum_{i=1}^{I_{edr,VLF}} (d_{edr,i}^{HF} - \overline{d_{edr}^{HF}})^2$	Sample deviation of $\{D_{edr}^k\}_{k=2}^4$ (HF band)
$P_{edr}^{VLF} = \int_{\frac{2\pi \times 0.04}{2\pi \times 0.15}}^{2\pi \times 0.15} P_{edr}(\omega) d\omega$	EDR Power spectrum (LF band)
$P_{edr}^{HF} = \int_{\frac{2\pi \times 0.15}{2\pi \times 0.4}}^{2\pi \times 0.4} P_{edr}(\omega) d\omega$	EDR Power spectrum (HF band)
$LF/HF_{edr} = P_{edr}^{LF} / P_{edr}^{HF}$	LF-HF power ratio in the EDR spectrum
$P_{edr}(\omega) _{\frac{2\pi \times 0.4}{2\pi \times 0.04}}$	Lomb-Scargel periodogram samples in LF-HF band
$\omega_{edr-resp} = argmax(P_{edr}(\omega) _{\frac{2\pi \times 0.4}{2\pi \times 0.15}})$	Dominant HF-band frequency of the EDR
$respMag_{edr} = max(P_{edr}(\omega) _{\frac{2\pi \times 0.4}{2\pi \times 0.15}})$ $= P_{edr}(\omega_{edr-resp})$	Power at the dominant HF-band frequency of the EDR
$respProb_{edr} = Prob(P_{edr}(\omega_{edr-resp}))$	Probability of $\omega_{edr-resp}$ occurrence with power $P_{edr}(\omega_{edr-resp})$
$\omega_{edr-ProbMax} = argmax(Prob(P_{edr}(\omega) _{\frac{2\pi \times 0.4}{2\pi \times 0.04}}))$	Most probable frequency of the EDR spectrum
$ProbMax_{edr} = max(Prob(P_{edr}(\omega) _{\frac{2\pi \times 0.4}{2\pi \times 0.04}}))$ $= Prob(P_{edr}(\omega_{edr-ProbMax}))$	Probability of $\omega_{edr-ProbMax}$ occurrence with power $P_{edr}(\omega_{edr-ProbMax})$
$ProbMaxMag_{edr} = P_{edr}(\omega_{edr-ProbMax})$	Power of the HRV spectrum at $\omega_{edr-ProbMax}$

2.3. Feature Reduction

Most automatic OSA detection methods [11-13, 16-19, 27, 29] use no feature reduction or employ linear dependency and correlation-based strategies or principal component analysis (PCA) for feature selection. Dependency and mutual information (MI) proved to outperform linear methods of feature selection, especially in respiratory event detection [14, 31]. Feature selection can be performed by individual analysis of each feature [13, 21, 26]. It is also possible to define a measure to evaluate a subset of features [14, 16]. The first method speculates the inter-relations among features but, the second method searches for features with both the tightest relations with the class label and the loosest interaction with each other. We use the second strategy for feature reduction.

To calculate the mutual interactions, we consider MI rather than a simple statistical correlation. We select the features which have the highest MI with the class label (normal or apnoeic) and the least MI with each other. The approach to search the feature space is forward feature selection. In this approach, the subset of selected features is gradually built by adding single features to an initial null set [14, 54].

2.4. Classification

We employ nine classifiers in this study; support vector machines (SVM) [55], K nearest neighbors (KNN) [60], decision table [56], C4.5 [57] decision tree, reduced-error pruning tree (REPT) [58], functional trees [59], the meta-algorithm of adaptive boosting accompanied with the simple classifier of decision stump [60], and the meta-algorithm of bagging along with the alternating decision tree (ADT) [61]. The meta-algorithms make a new data set out of the primary data set and devise a new classifier for each set in one trial. These trials are repeated T times, and eventually, the results of the T classifiers are combined to achieve a more accurate result.

In this study, four classifier combination methods are also performed on a group of three binary classifiers. Combination methods are max probability, average probability, the product of probability, and majority voting [16].

3. RESULTS

Table 5 demonstrates the selected features employing the MI measure. According to table 5, as the number of database subjects increases, the number of selected features also increases. There are several similarities between the selected measures; fewer ECG features are among the selected ones, mostly the time domain ECG features. This result is consistent with the previously published reports. Most of the selected features are based on the SpO2 signal, which indicates their power for the OSA detection. However, simultaneous use of the ECG and the SPO2 features enhances the performance of the OSA detection method [16].

Table 5. The selected features through forward feature selection based on the MI criterion. Name and definition of features stated in tables 1 to 4

Number	Selected features	Database
20	MI_3 , $spo2_{min}$, NEP_1 , S_{spo2} , $MI_{spo2,1}$, Δ , LZ_{down} , $odi4$, $CTM_{0.5}$, $ODI55$, $tsa80$, $tsa85$, $tsa90$, $S^2_{D_{rr}^4}$, P_{rr}^{HF} , $S^2_{D_{edr}^6}$, P_{edr}^{LF} , ‘samples 13 th and 55 th of $P_{rr}(\omega)$ sample 4 th of $P_{edr}(\omega)$	UCD
18	S_{spo2} , $MI_{spo2,1}$, Δ , LZC_{up} , $CTM_{0.25}$, $CTM_{0.5}$, $ODI55$, $tsa80$, $tsa85$, $tsa90$, $S^2_{D_{rr}^4}$, $S^2_{D_{edr}^6}$, P_{edr}^{LF} , samples 11 th , 18 th , 22 th and 55 th of $P_{rr}(\omega)$ and sample 4 th of $P_{edr}(\omega)$	Apnea-ECG
29	$Spo2_{min}$, $\overline{spo2}$, S_{spo2} , $r_{spo2,2}$, ZC , $ApEn$, $SpEn$, LZC_{up} , $MI_{spo2,3}$, $MI_{spo2,4}$, Δ , $ODIS4$, $ODI23$, $ODI25$, $ODI31$, $ODI35$, $ODI51$, $ODI53$, $ODI55$, $odi3$, $odi4$, $odi5$, $tsa95$, $tsa85$, $tsa80$, $CTM_{0.5}$, $CTM_{0.75}$, CTM_1 , ‘sample 4 th of $P_{edr}(\omega)$	Exclusive database

Table 6 illustrates the performance of our real-time detection method in each of the databases. We obtain the results from a system equipped with Windows 10 Pro, version 1511, the Intel processor Core i7CPU M640@2.8GHz and a RAM of 4GB. All the classifiers are realized in Java language. Evaluation is 10-fold cross-validation.

In some references, only the classifier's training time is reported [16]. This parameter is not enough to represent the total computational burden of the suggested method. In some previous works, the processing time is reported for a specified number of samples [14]. In our study, "the processing time for a fixed number of data samples" is not an accurate measure since several databases with different ECG sampling rates are observed.

Table 6. The performance of the suggested detection method in each of the databases: DT (Decision Table), REPT (Reduced-Error Pruning Tree), FT (Functional tree), AB+DS (Adaptive boosting + decision stump), B+ REPT (Bagging + REPT), B+ADT (bagging + alternating decision tree), AECG (Apnea-ECG database), EX (Exclusive database). Maximums in each column are shaded.

Processing time for 10 frames			Accuracy (%)			Specificity (%)			Sensitivity (%)			Classifier
EX	AECG	UCD	EX	AECG	UCD	EX	AECG	UCD	EX	AECG	UCD	
19	11.8	11.9	88.3	95.3	82	91	89.8	93	80.9	96.68	81.02	SVM
7	2.98	2.09	82.9	90.4	82	84.7	94	83	80.01	89	80.5	KNN
5.68	3.001	2.503	82	83.7	82	83	84.9	82	82.9	83	82.9	DT
4.001	1.45	1.076	82	85.6	81.7	86.1	89	85	73	82.1	72	C4.5
2.32	1.045	1.002	84.6	91.6	83.6	84.9	92.6	84	82.9	83.5	81.5	REPT
9.867	4.7	4.345	80	88.8	79.8	82	90.7	81.7	73	81.4	71.5	FT
2.383	1.32	1.205	92.6	87.3	79	93.3	79.3	78	89.9	92.6	88	AB+DS
4.794	2.97	2.164	88.5	91	85	89.9	92.2	86.3	82.1	89	81.03	B+REPT
29.9	15	13.98	85.6	95	84.5	85	95	83	86.78	89.9	85	B+ADT
18	9	8.99	55.1	63	57	55.8	57	54.3	55.01	65	59	SOM
10	5.7	4.897	35.6	38.6	34.5	33	37	33	38.4	40.1	37.3	K-means

Observing the processing time in table 6 reveals that the parameter value does not exceed 1s in the UCD and Apnea-ECG databases and 2s in our exclusive database. These margins are the minimum time needed for pre-processing and feature extraction at the specified sampling frequencies. Smaller values for processing times belong to the Apnea-ECG database with the lowest number of data points. The processing time for our exclusive database is the highest of all, nearly two times the minimum value. Regarding this quantity, two classifiers have the highest computational burden; the ADT and the SVM. The processing time for the SVM is more than two times higher than the others'. For the real-time OSA detection, these computationally intensive classifiers are not chosen despite their high classification ability.

Accuracy, sensitivity, and specificity in all the databases are satisfactory but, slightly better in the Apnea-ECG database compared to the others. The two unsupervised classifiers (the SOM and the K-means) do not exhibit acceptable results. Best sensitivity, but the worst specificity/accuracy belongs to adaptive boosting accompanied with the decision stump. On the other hand, bagging along with REPT achieves the best accuracy and specificity at the price of degrading sensitivity.

To reach a method with acceptable sensitivity and specificity, the combination routines declared in section 2.4 are used to fuse a group of three classifiers. Because the “boosting with the decision stump” and the “bagging along with the REPT” have better performances than others, they are the two fixed members of the group. The third member is chosen from the rest of the classifiers. We exclude the SVM and the “bagging with ADT” due to excessive computational load, so five options remain. These classifiers shape five different classifier groups to be fused. The classifier combination results are reported in tables 7 to 9.

According to tables 7 to 9, performance is nearly equal in all databases (slightly better performance for the Apnea-ECG database). Combining the classifiers, balances the performance measures in values around 80%. The most successful combination happened when the third group member is the KNN or the decision tree. In these cases, all the measures of performance, including sensitivity, specificity, and accuracy, have achieved values of more than 85%. These results outperform all the suggested methods to date [13, 14, 16, 19, 32, 35]. The principal difference between the KNN and the decision tree lies in their nature. KNN benefits from slow, moment-based training. It is appropriate for subject-dependant applications, in which models are built and tested with the same data. In subject-dependant applications each classifier model should be trained (i.e. updated) with the user data before utilization. On the other hand, the decision table is suitable for subject-independent applications where the classifier model is trained with a database of several subjects before being tested by the user.

Surveying the processing time shows that this quantity is approximately equal to the sum of the processing time needed for each classifier of the group. There is no distinguished difference between different combination routines. It is worth saying that combination methods based on probability need the sensitivity and the specificity of the classifier to weigh their decisions. This issue entails a more complex online realization than that of majority voting. Therefore, in online realization, the majority voting method will suffice.

Table 7. The performance of the suggested classifier combination detection method in the UCD database. Three classifiers are combined with four different methods (MP: Maximum probability, PP: Probability product, AP: Average probability, MV: Majority voting). Other abbreviations are similar to table 6. The two highest values in each column are shaded.

Processing time for 10 frames				Accuracy (%)				Specificity (%)				Sensitivity (%)				3 rd classifier
MV	AP	PP	M P	M V	AP	PP	M P	M V	AP	PP	M P	M V	AP	P P	M P	
4.36	4.68	4.47	4.4	85.28	86.12	86.2	86.12	85.25	86.03	86.16	86.07	87.55	87.41	87.19	85.87	KNN
4.55	4.65	4.76	4.869	85	85.68	85.7	85.64	84.16	85.35	85.42	85.47	87.61	86.68	86.57	86.14	DT
4	3.92	3.79	3.963	81.81	82.12	82.17	82.02	81.25	82.03	82.16	82.07	83.55	82.41	82.19	81.87	C4.5
3.65	3.56	3.39	3.245	81	81.68	81.70	81.64	80.16	81.35	81.42	81.47	83.61	82.68	82.57	82.14	REP T
5.21	5.34	5.63	5.56	81.03	80.95	80.98	80.96	80.43	80.48	80.57	80.69	82.9	82.41	82.25	81.82	FT

Table 8. The performance of the suggested classifier combination detection method in the Apnea-ECG database. Three classifiers are combined with four different methods (MP: Maximum probability, PP: Probability product, AP: Average probability, MV: Majority voting). Other abbreviations are similar to table 6. The two highest values in each column are shaded.

Processing time for 10 frames				Accuracy (%)				Specificity (%)				Sensitivity (%)				3 rd classifier
MV	AP	PP	MP	MV	AP	PP	MP	MV	AP	PP	MP	MV	AP	PP	MP	
5.68	5.634	5.555	5.29	85.38	86.2	86.27	86.15	85.34	86.61	86.23	86.17	87.6	87.5	87.2	86	KN
5.23	5.125	5.34	5.291	85.02	85.7	85.8	85.2	84.2	85.39	85.48	85.5	86.70	86.73	86.6	86.23	DT
3.99	3.7	3.65	3.49	82.4	82.25	82.2	82.23	82.33	82.1	82.2	82.1	83.6	82.5	82.2	81.94	C4.5
3.025	3.068	3.128	3.11	81	81.71	81.75	81.7	80.2	81.38	81.49	81.5	83.69	82.7	82.6	82.15	REP
5.969	5.79	5.87	5.9	81.1	81.2	81.01	81	80.57	80.6	80.7	81	83	82.5	82.31	81.91	FT

Table 9. The performance of the suggested classifier combination detection method in the exclusive database. Three classifiers are combined with four different methods (MP: Maximum probability, PP: Probability product, AP: Average probability, MV: Majority voting). Other abbreviations are similar to table 6. The two highest values in each column are shaded.

Processing time for 10 frames				Accuracy (%)				Specificity (%)				Sensitivity (%)				3 rd classifier
MV	AP	PP	MP	MV	AP	PP	MP	MV	AP	PP	MP	MV	AP	PP	MP	
12.05	11.81	11.6	12	85.32	86.03	86.13	86	85.24	86.6	86.11	86	87.23	87.35	87.1	85.67	KN
10.43	10.54	10.68	10.56	84.9	85.48	85.7	85.6	84.14	85.30	85.34	85.5	87.5	86.65	86.6	86	DT
9.34	9.47	9.24	9.04	81.78	82.10	82.13	81.95	81.20	82.2	82.14	82.01	83.51	82.13	82.17	81.85	C4.5
7.349	7.367	7.489	7.32	82.1	81.68	81.67	81.64	81.55	81.33	81.43	81.46	83.6	82.7	82.55	82.1	REP
15.004	14.96	14.62	14.86	81.01	80.87	80.93	80.92	80.32	80.36	80.44	80.59	82.8	82.3	82.15	81.72	FT

4. CONCLUSIONS

In this study, several configurations for online detection of the OSA are suggested. The advantages of the proposed method are: exploiting only two channels of biological signals, automatic and real-time detection, and uniform acceptable performance over several databases (over 85%). To date, no other study has achieved all these merits together. Acceptable performance in well-known databases is due to classifiers that do not possess database-related parameters (e.g. sampling frequency of signals). The classifiers have covered deficiencies of each other in a combinational configuration. To reach the best result, the most successful classifiers are combined in groups of three members with four different combination methods. The features are also calculated and selected considering generality; in frequency-domain analysis, the refined Lomb-Scargle periodogram is used to care for the inherent non-uniform sampling of the R-R tachograms and unequal sampling frequency of the ECG signal in different databases [50]. Feature selection is based on the MI. The MI measure considers non-linear correlations among features and selects effective features to decrease the computational burden of the classifiers and avoid over-fitting problems.

On the other hand, the MI feature reduction has an important impact on the family of decision tree classifiers. MI-based feature selection accompanied by decision tree classifiers, avoids the classifier sensitivity to MI-biased estimates. In other words, the decision-tree classifiers may be misled by a fake replica of a feature with more marginal samples and higher maximum entropy value [62]. Selection of the more appropriate feature with an entropy-normalised MI estimator is helpful [62, 63].

ACKNOWLEDGEMENTS

The author appreciates the cooperation of the Ibn-e-Sina Hospital sleep laboratory at Mashhad University of Medical Sciences.

REFERENCES

- [1] C. V. Senaratna, J. L. Perret et al, "Prevalence of obstructive sleep apnea in the general population: A systematic review", *Sleep Medicine Reviews*, vol. 34, pp. 70-81, 2017. ISSN 1087-0792, Available: <https://doi.org/10.1016/j.smr.2016.07.002>.
- [2] O. Berdina, I. Madaeva, and L. Rychkova, "Obstructive Sleep Apnea and COVID-19 Infection Comorbidity: Analysis of the Problem in the Age Aspect", *International Journal of Biomedicine*, vol. 10, no. 4, pp. 312-315, 2020. Available: 10.21103/article10(4)_ra1.
- [3] J. Remmers, W. deGroot, E. Sauerland and A. Anch, "Pathogenesis of upper airway occlusion during sleep", *Journal of Applied Physiology*, vol. 44, no. 6, pp. 931-938, 1978. Available: 10.1152/jappl.1978.44.6.931.
- [4] B. Dredla and P. Castillo, "Cardiovascular Consequences of Obstructive Sleep Apnea", *Current Cardiology Reports*, vol. 21, no. 11, 2019. Available: 10.1007/s11886-019-1228-3.
- [5] B. Phillips, "Association of Sleep Apnea and Type II Diabetes: A Population-Based Study", *Yearbook of Pulmonary Disease*, vol. 2007, pp. 249-251, 2007. Available: 10.1016/s8756-3452(08)70452-3.
- [6] W. Akberzie, S. Hesselbacher, I. Aiyer, S. Surani and Z. Surani, "The Prevalence of Anxiety and Depression Symptoms in Obstructive Sleep Apnea", *Cureus*, 2020. Available: 10.7759/cureus.11203.
- [7] W. McNicholas, "Diagnosis of Obstructive Sleep Apnea in Adults", *Proceedings of the American Thoracic Society*, vol. 5, no. 2, pp. 154-160, 2008.
- [8] International classification of sleep disorders. Darien, Ill.: American Acad. of Sleep Medicine, 4th ed. New York: Westchester, III; 2014.
- [9] T. Lee-Chiong, *Sleep*. Hoboken, N.J.: Wiley-Liss, 2006.

- [10] V. Kapur et al., "The Medical Cost of Undiagnosed Sleep Apnea", *Sleep*, vol. 22, no. 6, pp. 749-755, 1999. Available: 10.1093/sleep/22.6.749.
- [11] U. R. Abeyratne, V. Swarnkar, C. Hukins, and B. Duce, "Interhemispheric Asynchrony Correlates With Severity of Respiratory Disturbance Index in Patients With Sleep Apnea," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2947-2955, 2010.
- [12] Kim PY, McCarty DE, Wang L, Frilot C, Chesson AL, Marino AA. Two-group classification of patients with obstructive sleep apnea based on analysis of brain recurrence. *Clinical Neurophysiology*. 2014 Jun 30;125(6):1174-81.
- [13] Bsoul M, Minn H, Tamil L. Apnea MedAssist: real-time sleep apnea monitor using single-lead ECG. *IEEE Transactions on Information Technology in Biomedicine*. 2011 May;15(3):416-27.
- [14] Nguyen HD, Wilkins BA, Cheng Q, Benjamin BA. An online sleep apnea detection method based on recurrence quantification analysis. *IEEE Journal of Biomedical and Health Informatics*. 2014 Jul;18(4):1285-93.
- [15] H. M. Al-Angari, and Alan V. Sahakian, "Automated Recognition of Obstructive Sleep Apnea Syndrome Using Support Vector Machine Classifier," *IEEE Trans. Inf. Tech. Biomed.*, vol. 16, no. 3, pp. 463-468, 2012.
- [16] Xie B, Minn H. Real-time sleep apnea detection by classifier combination. *IEEE Transactions on Information Technology in Biomedicine*. 2012 May;16(3):469-77.
- [17] Sannino G, De Falco I, De Pietro G. Monitoring obstructive sleep apnea by means of a real-time mobile system based on the automatic extraction of sets of rules through differential evolution. *Journal of biomedical informatics*. 2014 Jun 30; 49:84-100.
- [18] J. V. Marcos et al, "Automated Prediction of the Apnea-Hypopnea Index from Nocturnal Oximetry Recordings," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 1, pp. 141-149, 2012.
- [19] Koley BL, Dey D. On-Line Detection of Apnea/Hypopnea Events Using SpO2 Signal: A Rule-Based Approach Employing Binary Classifier Models. *IEEE Journal of Biomedical and Health Informatics*. 2014 Jan;18(1):231-9.
- [20] Koley BL, Dey D. Automatic detection of sleep apnea and hypopnea events from single channel measurement of respiration signal employing ensemble binary SVM classifiers. *Measurement*. 2013 Aug 31;46(7):2082-92.
- [21] Koley BL, Dey D. Real-time adaptive apnea and hypopnea event detection methodology for portable sleep apnea monitoring devices. *IEEE Transactions on Biomedical Engineering*. 2013 Dec;60(12):3354-63.
- [22] S. I. Rathnayake, I. A. Wood, U. R. Abeyratne, and C. Hukins, "Nonlinear features for single-channel diagnosis of sleep-disordered breathing diseases," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 1973-1981, Aug. 2010.
- [23] Sanchez-Morillo D, Lopez-Gordo MA, Leon A. Novel multiclass classification for home-based diagnosis of sleep apnea/hypopnea syndrome. *Expert Systems with Applications*. 2014 Mar 31;41(4):1654-62.
- [24] Azarbarzin, Ali, and Zahra MK Moussavi. "Automatic and unsupervised snore sound extraction from respiratory sound signals." *IEEE Transactions on Biomedical Engineering* 58.5 (2011): 1156-1162.
- [25] E. Goldshtein, A. Tarasiuk, and Y. Zigel, "Automatic Detection of Obstructive Sleep Apnea Using Speech Signals," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 5, pp. 1373-1382, Feb. 2011.
- [26] Burgos A, Goni A, Illarramendi A, Bermudez J. Real-time detection of apneas on a PDA. *IEEE Transactions on Information Technology in Biomedicine*. 2010 Jul;14(4):995-1002.
- [27] Behar J, Roebuck A, Shahid M, Daly J, Hallack A, Palmius N, Stradling J, Clifford GD. SleepAp: An automated obstructive sleep apnoea screening application for smartphones. *IEEE journal of biomedical and health informatics*. 2015 Jan;19(1):325-31.
- [28] ZhaoD, WangY, WangQ, WangX. Comparative analysis of different characteristics of automatic sleep stages. *Comput Methods Programs Biomed* 2019; 175:53-72
- [29] Carrubba S, Kim PY, McCarty DE, Chesson Jr AL, Frilot C, Marino AA. Continuous EEG-based dynamic markers for sleep depth and phasic events. *J. Neurosci. Meth.* 2012; 208:1-9.
- [30] Ireneusz Jabłoński, "Modern Methods for the Description of Complex Couplings in the Neurophysiology of Respiration," *IEEE Sensors Journal*, vol. 13, no. 9, pp. 3182-3192, Sep 2013.
- [31] H. Ghandeharioun, F. Rezaeitalab, and R. Lotfi, "Analysis of respiratory events in obstructive sleep apnea syndrome: Inter-relations and association to simple nocturnal features", *Revista Portuguesa de Pneumologia (English Edition)*, vol. 22, no. 2, pp. 86-92, 2016. Available: 10.1016/j.rppnen.2015.09.008.

- [32] ErdenebayarU, KimYJ, ParkJ-U, JooEY, LeeK-J. Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram. *Comput. Methods Programs Biomed.* 2019; 180:105001.
- [33] DaldalN, CömertZ, PolatK. Automatic determination of digital modulation types with different noises using convolutional neural network based on time–frequency information. *Appl. Soft Comput. J* 2019.
- [34] ChoiSH, YoonH, KimHS, KimHB, KwonHB, OhSM, et al. Real-time apnea-hypopnea event detection during sleep by convolutional neural networks. *Comput. Biol. Med.* 2018; 100:123–31.
- [35] F. Bozkurt, M.K. Uçar, M.R. Bozkurt, C. Bilgin, Detection of Abnormal Respiratory Events with Single Channel ECG and Hybrid Machine Learning Model in Patients with Obstructive Sleep Apnea, *IRBM*, Volume 41, Issue 5, 2020, Pages 241-251.
- [36] St. Vincent's University Hospital/University College Dublin Sleep Apnea Database. (2008). Available: <http://www.physionet.org/pn3/ucddb/>
- [37] "CinC Challenge 2000 data sets: Data for development and evaluation of ECG-based apnea detectors," (2000). Available: <http://www.physionet.org/physiobank/database/apnea-ecg/>
- [38] A. Roebuck, V. Monasterio, E. Geder, M. Osipov, J. Behar, A. Malhotra, T. Penzel and G. Clifford, "A review of signals used in sleep analysis", *Physiological Measurement*, vol. 35, no. 1, pp. R1-R57, 2013.
- [39] Kuna S T, Benca R, Kushida C A, Walsh J, Younes M, Staley B, Hanlon A, Pack A I, Pien G W and Malhotra A 2013 Agreement in computer-assisted manual scoring of polysomnograms across sleep centres, *Sleep* 36 583–9
- [40] Behar J, Roebuck A, Shahid M, Daly J, Hallack A, Palmius N, Stradling JR, Clifford GD. SleepAp: An automated obstructive sleep apnoea screening application for smartphones, *Computing in Cardiology* 2013 2013 Sep 22 (pp. 257-260).
- [41] Malik M., Camm A.J. (eds.): *Heart Rate Variability*, Armonk, N.Y. Futura Pub. Co. Inc., 1995.
- [42] Abboud S., Barnea O.: Errors due to sampling frequency of electrocardiogram in spectral analysis of heart rate signals with low variability, *Computers in Cardiology*, pp 461-463, 1995.
- [43] Oppenheim, Alan V., Willsky, Alan S., Navab, S. Hamid. "Digital signal processing." Printice-Hall, London (2000). ISBN:0-13-651175-9.
- [44] Daubechies, Ingrid. "Ten lectures on wavelets, vol. 61 of CBMS-NSF Regional Conference Series in Applied Mathematics." (1992).
- [45] Hamilton P., Tompkins W.: Quantitative Investigation of QRS Detection Rules Using the MIT/BIH Arrhythmia Database, *IEEE Transactions on Biomedical Engineering*, vol. BME-33, NO. 12. 1986.
- [46] Pan J., Tompkins W.: A real-time QRS detection algorithm, *IEEE Transactions on Biomedical Engineering*, vol. BME-32 NO. 3. 1985.
- [47] Darbellay, Georges A., and Igor Vajda. "Estimation of the information by an adaptive partitioning of the observation space." *IEEE Transactions on Information Theory* 45.4 (1999): 1315-1321.
- [48] R. Hornero, D. Alvarez, D. Abasolo, F. del Campo, and C. Zamarron, "Utility of approximate entropy from overnight pulse oximetry data in the diagnosis of the obstructive sleep apnea syndrome," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 1, pp. 107-113, Jan. 2007.
- [49] D. Alvarez, R. Hornero, D. Ab'asolo, F. Campo, and C. Zamarr'on, "Nonlinear characteristics of blood oxygen saturation from nocturnal oximetry for obstructive sleep apnoea detection," *Physiol. Meas.*, vol. 27, pp. 399–412, 2006.
- [50] Clifford GD. Signal processing methods for heart rate variability (Doctoral dissertation, Department of Engineering Science, University of Oxford, September 2002).
- [51] G. D. Furman, Z. Shinar, A. Baharav, and S. Akselrod, "Electrocardiogram derived respiration during sleep," *Comput. Cardiol.*, vol. 32, pp. 351–354, 2005.
- [52] Raymond, B., Cayton, R. M., Bates, R. A., & Chappell, M. J., "Screening for Obstructive Sleep Apnoea Based on the Electrocardiogram—The Computers in Cardiology Challenge," *Proc. Computers in Cardiology*, Vol. 27, IEEE Press, 2000, pp. 267–270.
- [53] G. B. Moody, R. G. Mark, A. Zoccola, and S. Mantero, "Derivation of respiratory signals from multi-lead ECGs," *Comput. Cardiol.*, vol. 12, pp. 113–116, 1985.
- [54] P. A. Estevez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Trans. Neural Netw.*, vol. 20, no. 2, pp. 189–201, Feb. 2009.
- [55] Murphy KP. Machine learning: a probabilistic perspective. MIT Press; 2012 Sep 7.
- [56] N. Landwehr, M. Hall, and E. Frank. (2005, May). Logistic model trees. *Mach. Learn.* 59, pp. 161–205.

- [57] R. Quinlan, C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann, 1993.
- [58] Mingers, John. "An empirical comparison of pruning methods for decision tree induction." *Machine learning* 4.2 (1989): 227-243.
- [59] Gama, Joao. "Functional trees." *Machine Learning*, vol. 55, no. 3, pp. 219-250, 2004.
- [60] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [61] Y. Freund and L. Mason, "The alternating decision tree algorithm," in *Proc. 16th Int. Conf. Mach. Learning*, 1999, pp. 124–133.
- [62] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, pp. 81–106, 1986.
- [63] Paninski, Liam. "Estimation of entropy and mutual information." *Neural computation* 15.6 (2003): 1191-1253.

AUTHOR

Dr. Hosna Ghandeharioun received her B.S. degree in electrical engineering from Ferdowsi University of Mashhad (FUM), Iran, in 2003 and an M.S. degree in Biomedical Engineering from Iran University of Science & Technology, Iran in 2006. She received her Ph. D. degree in electrical engineering from FUM in 2016. Now She works as an assistant professor at the electrical and biomedical Engineering dept. of Khorasan Institute of Higher Educations. Her research interests are biological signal processing and mobile health.



ARTIFICIAL INTELLIGENCE & MACHINE LEARNING ROLE IN FINANCIAL SERVICES

Prudhvi Parne

Information Technology, Bank of Hope,
1655 W Redondo Beach Blvd, Gardena, CA, USA

ABSTRACT

Financial services are the economical backbone of any nation in the world. There are billions of financial transactions which are taking place and all this data is stored and can be considered as a gold mine of data for many different organizations. No human intelligence can dig in this amount of data to come up with something valuable. This is the reason financial organizations are employing artificial intelligence to come up with new algorithms which can change the way financial transactions are being carried out. Artificial Intelligence can complete the task in a very short period. Artificial intelligence can be used to detect frauds, identify possible attacks, and any other kind of anomalies that may be detrimental for the institution. This paper discusses the role of artificial intelligence and machine learning in the finance sector.

KEYWORDS

Artificial Intelligence, Machine Learning, Finance, Security

1. INTRODUCTION

The development of data analysis capabilities has allowed multiple sectors in the industries to identify various beneficial activities to improve efficiency, identify opportunities, develop better capabilities, improve reachability, develop better customer satisfaction, identify the products that can be developed and sustained in a longer run, improve the security measures, and many other benefits accurately implemented to support the business activities.

One of the primary beneficiaries of artificial intelligence and machine learning capabilities can be identified as the financial sectors that are one of the biggest repositories of data that can be explored to identify valuable insights that can be used by the sector to identify opportunities, improve the services, develop better products, provide better customer service capabilities, leverage the artificial intelligence and machine learning capabilities to identify risks, develop automated processes to improve the security measures on the infrastructure and information, and other activities implemented appropriately that can be effectively utilized to improve the business prospects of the financial institute.

2. ARTIFICIAL INTELLIGENCE IN FINANCE

BFSI as it is commonly known stands for Banking, Securities, Finance, and Insurance which forms the core of the financial sector. There are vast amounts of data generated by them because of which systems with high analytical capabilities are required to dig out the crucial information required to grow the business. This knowledge can be also used to make good decisions. Artificial intelligence helps the financial sector by reducing the number of manual errors that

were conducted earlier by consistently developing decision-making processes that verify every bit of the information available before the required decisions are established. Machine learning capabilities efficiently utilize artificial intelligence to identify future opportunities by understanding the communication and transactions to develop better strategies to improve the business prospects, identify opportunities, develop autonomous response capabilities to improve the communications with the customers, and many other beneficial activities implemented [1].

3. IMPORTANCE OF DATA

The sensitivity of the information available in the financial sector makes it one of the primary targets by attackers to gain access to the infrastructure and the information. Artificial intelligence capabilities established in financial institutes can efficiently identify the possible attacks by developing understanding based on signatures, patterns, anomalies identified, and many other identifiers used to detect abnormal activities in the network to alert the security team on possible intrusion attempts carried out. Mapping various aspects of the information with the historical information to identify the difference in the activities can efficiently identify anomalies in the transactions, analyze the activities of the users in the infrastructure to define activities beyond the roles and responsibilities which may develop into risks to the organizations and identify possible solutions that can be effectively utilized to reduce the security risks [2].

4. RISK MANAGEMENT

The financial sector is full of risks daily due to the nature of activities carried out. An organization can efficiently perform by identifying all the risks, identifying the new risks, and define the impact of identified risks in the earliest stage [3]. The business prospects in the Financial Industries are full of risks that need to be identified, protect the organization from higher exposure to risks, identify the risk tolerance and risk appetite of the organization, provide adequate alerts to the risk management team on identified possibilities of risks materializing in the environment, develop all at mechanisms that can appropriately report the risks which can be analyzed by the management to make appropriate decisions based on the data provided by the artificial intelligence and machine learning capabilities that are established to detect activities beyond the acceptable range of the organization to reduce exposure to higher risks which may impact the performance of the organization [5].

Risk management is one of the critical activities in the financial sector and identifying the triggers that may increase the risks for an organization is one of the key activities to be conducted to ensure the exposure in the financial sector by the organization is manageable and under control. The ability of the organization to inspect the live data, learn from the live feed, analyze the available information, identify and detect anomalies that can promote risks, and develop appropriate alert mechanisms and create preventive measures to contain the risk is the key to effectively manage the risks and develop better profitability [3].

5. TOOLS FOR THE FINANCIAL SECTOR

The financial sectors offer multiple products to the customers with different benefits. The capabilities of the marketing and sales team to promote their products improve the profitability of the organization [4]. Implementation of artificial intelligence and machine learning technologies can enable the organization to improve the business prospects by identifying the prospective customers worldwide, providing proposals for individual tailor-made products that can be provided by the organization, developing conversations and communication with the clients through automated processes, implementation of Chatbots in sales and services can drastically

improve the response of the organization consistently to the many queries of the customers with satisfactory responses developed based on the understanding created by the AI/ML process.

An organization can improve the efficiency of responses provided to the customers by implementing artificial intelligence, big data analytical capabilities, and machine learning capabilities to identify the historical information, analyze the information to develop a better understanding, identifying the probable responses that can be provided, implementing appropriate available solutions, identifying the possible combination of solutions that can satisfy the customer needs and also develop profitable business for the organization [2]. The new capabilities established can allow the banks to develop business solutions, standards of implementation, the power of artificial intelligence, the insides of historical information, the ability of the machine working to develop predictability, and other capabilities put together to achieve the best results for the instant services required by the customers to improve their activities.

6. BENEFACTORS OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

One of the primary beneficiaries of the advanced technologies used in the security measures on the information by implementing blockchain technologies can also use analytical capabilities on the secured information to efficiently conduct analytical capabilities with an accuracy of information, authenticity, and accuracy of the information without being adulterated. Blockchain has allowed higher security measures to be implemented on the information in transit and storage, data analytics capabilities have improved the decision-making capabilities, artificial intelligence improves the decision-making process by identifying media's insights from the historical information that can be applied, and the machine-learning capabilities evolve their understanding on the live information feed to develop the understanding instantly and identify the patents in the information that can be effectively utilized to develop accurate responses which can be accomplished automatically and with high accuracy [4].

On other important aspects in the financial industry is the regulator the aspect due to the sensitivity of the information and the high risks which requires the regulatory authorities to implement stringent rules and regulations to secure the interest of the investor, protect the information privacy, develop ethical practices by the organizations, and provide adequate suggestions on security implementation to the organizations implementing infrastructure and business activities. With the increased number of cyber incidents daily, the natural response from the regulatory authority to implement stricter measures to protect the interest of the citizens, to develop accountability and responsibilities assigned to promote ethical behavior, and provide judicial capabilities to resolve disputes, information protection, information privacy, and accountability of activity are mandated by regulatory authorities worldwide. An organization must abide by the rules and regulations and identify various breach incidents in the initial stages to improve the implementation of rules and regulations to protect the interest of the investor and protect the infrastructure from unauthorized activities [5].

7. EFFICIENT AND EFFECTIVE UTILIZATION OF TECHNOLOGY

Identifying the consistent utilization of artificial intelligence and machine-learning capabilities which may reduce the number of Manual jobs but provide consistency in the information and its authenticity, developing appropriate unbiased and decision-making capabilities, identifying the opportunities to diversify and improve business prospects, develop identification of future requirements of the customers in the financial sectors, understanding the needs of the customers, developing autonomous conversation capabilities by the organization with its existing and prospective customers can greatly improve the opportunities for the organization to improve the business prospects and develop better profitability [2].

Identifying the historical value of the information, developing analytical capabilities on the existing information, understanding the hidden value of the information available, using the value to develop better capabilities, improving the processes, discovering new processes, implementing improvements in customer services, improving the cybersecurity measures, enhancing the risk management process to detect risks at the earliest age, interacting with the customers to provide solutions available in the organization, and many other beneficial activities can be effectively implemented with the use of the data analytical capabilities, artificial intelligence implementation, and machine-learning processes established that can effectively identify, understand, develop responses accurately, improves the responses based on the feedback received to provide best solutions to the customers and improve the profitability of the organization consistently [3].

Firms can utilize AI to obtain more data from meager verifiable models or recognize non-direct connections in the request stream. Machine learning can be used to make 'trading robots' that then, at that point, show themselves how to respond to market changes. Market sway analysis includes assessing the impact of an association's trading on market costs. Since firms are worried about the effect of exchanges, particularly massive sales, on market costs, a more precise assessment of this effect is vital to timing exchanges and limiting trading execution costs. Firms are exploring utilizing AI tools to evaluate the market effect of a given business. The impact of an association's trading on market costs is famously tricky to show, particularly for less fluid protections, where information on similar past exchanges is scant. AI and machine learning can supplement traditional market sway models[2].

AI tools may help by enlarging models effectively or acquiring a machine learning approach to limit the trading sway on costs and liquidity. For the most dynamic efficient assets, as much as 66% of the gain on exchanges is assessed to be lost to market sway costs[6]. AI tools may help by enlarging models effectively being used, or by acquainting a machine learning approach with limit the trading sway on prices and liquidity for trading both into and out of enormous market positions, or as a piece of consistently trading strategies. Machine learning is frequently used to distinguish gatherings of bonds that act comparatively to one another. Like this, they can depend on more information focuses, giving better gauges of value developments when the market is dainty. The subsequent apparatus bunches bonds into wide, naturally comparative containers. Afterward, utilizing group analysis, gather the most tantamount items together in each pile to score the liquidity of individual bonds[5].

Additionally, AI can assist with distinguishing how the circumstance of exchanges can limit market sway. Market sway models can be fostered that depict how the impact of business relies upon past interactions as a beginning stage. The models endeavor not to plan trades too intently together to try not to have a market sway more prominent than the number of its parts[6]. These models can set out the ideal trading plans for a scope of situations and afterward change the timetable as the genuine exchange advances, using managed learning strategies to make the transient predictions deciding those changes.

8. CASE STUDY: AI AND ML BASED APPLICATIONS

In Insurance domain, AI and machine learning applications can significantly increase some protection area capacities, such as endorsing and preparing claims. In supporting, AI frameworks dependent on NLP can grow enormous business guaranteeing and life or inability endorsing. These applications can gain from training sets of past claims to feature critical contemplations for human decision-creators[4]. Machine learning procedures can decide repair costs and naturally sort the seriousness of vehicle mishap harm. Moreover, AI might help diminish claims preparing times and functional expenses. Insurance agencies are additionally investigating how AI and

machine learning and remote sensors (associated through the 'web of things') can distinguish, and sometimes forestall, insurable episodes before they happen, for example, compound spills or auto crashes[2].

Credit scoring tools/applications that utilize machine learning are intended to accelerate loaning decisions while possibly restricting gradual danger. Since a long time ago, Loan specialists have depended on credit scores to settle on loaning decisions for firms and retail customers. Information on exchange and installment history from monetary organizations generally became the establishment of most credit scoring models. These models use tools like relapse, decision trees, and measurable analysis to produce a credit score utilizing restricted measures of organized information. In any case, banks and different moneylenders are progressively going to extra, unstructured, and semi-organized information sources, including online media movement, cell phone use, and instant message action, to catch a more nuanced perspective on creditworthiness and further develop the rating exactness of advances. Applying machine learning algorithms to this star grouping of new information has empowered the evaluation of subjective factors like customer conduct and pay[1].

The capacity to use extra information on such measures considers a more prominent, quicker, and less expensive division of borrower quality and eventually prompts a fast credit decision. Be that as it may, the utilization of individual information raises other policy issues, including those identified with information security and information insurance[5]. As well as working with a conceivably more exact, divided evaluation of credit worthiness, the utilization of machine learning algorithms in credit scoring might empower more noteworthy admittance to credit[4]. In conventional credit scoring models utilized in specific markets, a potential borrower should have an adequate measure of recorded credit data available to be considered 'scorable.' without this data, a credit score can't be created, and a conceivably creditworthy borrower is frequently unfit to obtain credit and assemble a credit history. With the utilization of elective information sources and the use of machine learning algorithms to assist with fostering an evaluation of capacity and ability to reimburse, moneylenders might have the option to show up at credit decisions that beforehand would have been outlandish. While this pattern might profit economies with shallow credit markets, it could prompt non-sustainable expansions in credit exceptional in nations with profound credit markets. For the most part, it has not yet been demonstrated that machine learning-based credit scoring models beat customary ones for evaluating creditworthiness[5].

There are a few benefits and disservices to utilizing AI in credit scoring models. AI permits enormous measures of information to be dissected rapidly. Therefore, it could yield credit scoring arrangements that can deal with a more extensive scope of credit inputs, bringing down the expense of surveying credit hazards for specific people and expanding the number of people for whom firms can gauge credit hazards[5]. An illustration of the use of enormous information to credit scoring could incorporate the evaluation of non-credit charge installments, for example, the convenient installment of wireless and other service bills, in the mix with different information. Also, individuals without credit history or credit score might have the option to get an advance or a credit card because of AI, where an absence of credit history has customarily been a constraining element as option pointers of the probability to reimburse have been inadequate in ordinary credit scoring models[1].

Notwithstanding, the utilization of complex algorithms could bring about an absence of straightforwardness to shoppers. This 'discovery' part of machine learning algorithms may thus raise concerns. When utilizing machine learning to allot credit scores to settle on credit decisions, it is, for the most part, harder to give buyers, inspectors, and directors a clarification of a credit score and coming about credit decisions whenever tested [6]. Furthermore, some contend that using new elective information sources, like online conduct or non-conventional monetary data,

could bring predisposition into the credit decision. In particular, purchaser support bunches call attention to that machine learning tools can yield blends of borrower qualities that foresee race or sexual orientation, factors that fair loaning laws disallow considering in numerous locales. These algorithms may rate borrowers from an ethnic minority at a greater danger of default because comparable borrowers have customarily been given less ideal credit conditions. The availability of chronicled information across a scope of borrowers and credit items is critical to an exhibition of these tools. Moreover, the availability, quality, and dependability of information on borrower-item execution across a broad scope of monetary conditions are likewise crucial to display these danger models. Again, the absence of information on new AI and machine learning models, and the absence of data about the collection of these models in an assortment of monetary cycles, has been noted by certain specialists[3].

9. CONCLUSION

Artificial Intelligence and Machine Learning are groundbreaking technologies that are still in their primary stages of development and adoption. They have very high capabilities and if they are implemented correctly can change the very way the finance sector is currently operating. This is the reason, almost all the big financial organizations are investing heavily in these technologies because the return on investments is very high. There is a lot of hope riding on these technologies and the IT services companies are trying to deliver them with utmost accuracy. There is still some time required for these technologies to mature and recognize their utmost potential. Companies need to understand that these technologies must be used to make the lives of employees easy. They should not be a reason to replace the human workforce because nothing can beat the human instincts which are required in the financial sector. The management of these financial organizations needs to come up with the best combination of human and artificial intelligence for the development of the industry, mankind, and the world.

REFERENCES

- [1] Lee, J. (2020). Access to Finance for Artificial Intelligence Regulation in the Financial Services Industry. Springer Link Articles 731-740.
- [2] Joseph, B. & Collins, P. D. (2021). Adversary-Aware Learning Techniques and Trends in Cybersecurity. Springer Publishing 17-36.
- [3] John, G., & Mooney, M. C. (2018). Disrupting Finance - FinTech, and Strategy in the 21st Century. Springer Publishing, 33-50.
- [4] Itay, G. & Jagtiani, J. & Klein, A, (2018). Philadelphiafed 2018 fintech conference, Fintech and the new financial landscape , 4.
- [5] Raghad, G. L. (2019). The Application of Artificial Intelligence in Financial Compliance Management. ACM Article, 1-6.
- [6] FSB (2017), <https://www.fsb.org/2017/11/artificial-intelligence-and-machine-learning-in-financial-service/>, 5-20.

AUTHOR

Prudhvi Parne received the Master's (MS) degree in Computer Science from University of Louisiana, Lafayette, LA, USA. His expertise spans in the areas of Cloud Architecture, Software Development, Finance, Banking, Hybrid clouds, Product Management, and Product leadership.



OPEN LoRAWAN SENSOR NODE ARCHITECTURE FOR AGRICULTURE APPLICATIONS

Philipp Bolte¹, Ulf Witkowski¹ and Rolf Morgenstern²

¹Department of Electronics and Circuit Technology,
South Westphalia University of Applied Sciences, Soest, Germany

²Department of Agriculture, South Westphalia University
of Applied Sciences, Soest, Germany

ABSTRACT

In agriculture, it becomes more and more important to have detailed data, e.g. about weather and soil quality, not only in large scale classic crop farming applications but also for urban agriculture. This paper proposes a modular wireless sensor node that can be used in a centralized data acquisition scenario. A centralized approach, in this case multiple sensor nodes and a single gateway or a set of gateways, can be easily installed even without local infrastructure as mains supply. The sensor node integrates a LoRaWAN radio module that allows long-range wireless data transmission and low-power battery operation for several months at reasonable module costs. The developed wireless sensor node is an open system with focus on easy adaption to new sensors and applications. The proposed system is evaluated in terms of transmission range, battery runtime and sensor data accuracy.

KEYWORDS

Wireless Sensor Node, LoRa Communication, Real-Time Environmental Monitoring, Urban Agriculture.

1. INTRODUCTION

The success of crop farming is traditionally dependent on weather and climate patterns. Farmers have a long history of weather and climate observation as well as weather prediction, formerly based on local experience and intuition, later augmented with systematic weather data collection and forecasting by agricultural and national institutions. Research on low precipitation and drought of the recent years has revealed a strong spatial diversity [1]. This can mean that some plots of land received sufficient rainfall, but adjacent plots suffered from water shortage. Micro climates in cities can lead to similar effects. This situation generates the desire of field scale data acquisition for farmers to allow for adapted irrigation measures.

The coarse scale of data acquisition is being augmented with fine grained and more detailed data, not only concerning weather and climate parameters, but also soil and plant data, in the precision farming movement. This is even more important for Urban Agriculture (UA) where the plots are very small and scattered in the urban landscape. Partial sharing and the structure of surrounding buildings create microclimates that may differ widely between plots even though they are not very far apart.

The generally small total size of an UA operation requires an urban farmer to minimize waste and transportation and to maximize yield as well as farmer and staff productivity. Hyper local environmental, soil and plant data, possibly extended with presence and intrusion detection can facilitate the complex crop and harvest planning and general farm management task. Ideally plots are monitored on an individual basis.

1.1. Application Scope

Urban Farming oftentimes employs different production techniques, adapted to local conditions of the production locations. Market gardening, also known as Small Plot INTensive (SPIN) farming on open plots, is often enhanced with simple foil tunnels or small scale greenhouses. Increasingly Hydroponics and Aquaponics are utilized in order to produce in locations that do not offer arable soil [2]. These water based production methods require the monitoring of relevant water parameters like temperature, electronic conductivity (EC) and pH of the nutrient solution as well as the dissolved oxygen (DO) when fishes are involved. The welfare of the fishes in the aquaculture of such a system calls for near real-time monitoring of the mentioned parameters.

The collection of different environmental and production system parameters, that are relevant for such an operation, range from air temperatures and relative humidity over soil temperature and moisture, global radiation and daily light integral (DLI) to the mentioned process water parameters. The design of a hydroponic or aquaponic system might additionally require liquid flow and liquid level measurements. Finally, location detection of equipment and presence detection of staff as well as intrusion detection add one more dimension to be monitored.

Production plots are usually not all located in the direct vicinity of a building the farmer has authority over. On a case to case basis it might be possible to ask friendly neighbours for Wi-Fi connectivity to have wireless access to sensor devices. But this approach bears the risk of depending on a crucial part of the management infrastructure not being under control of the farmer. Therefore, alternative methods for data transportation from the field or greenhouse to the data management application are desirable.

1.2. Typical Requirements

The usage of data logging systems for agriculture application is associated with application specific requirements. The typical users of such systems do not have extensive technical expertise. Therefore, the deployment and particularly the maintenance of data logging systems and related sensors must be simple. A large battery lifetime is expected. The sensors are often placed in harsh conditions exposed to rain, condensing humidity and sun light exposure.

It must be distinguished between short term usage and continuous monitoring applications. In research the experiments are usually time limited and technical experienced staff is available. The overall requirements of the sensor system are not nearly as extensive compared to long-term usage. The availability of real-time data over long time periods promise benefits in the areas of food safety, cost reduction, operational efficiency and asset management [3]. Current research is furthermore utilizing machine learning to control the process aiming additional yield optimization using real-time data [4].

1.2.1. Sampling Frequency

Growing crops and fattening fish are rather slow processes that do not generally require real-time data acquisition or high sampling frequencies. Depending on the local context, sampling times between one minute and one hour should be sufficient for the bulk of applications. Suitable

transmission frequencies might even be lower than one per minute or one per hour if data is buffered in the sensor node. Two applications however benefit from near real-time sampling and transmission: vital water parameters for the aquaculture and intrusion detection. Low oxygen supply in the aquaculture requires an immediate action of the farmer as ensuring the welfare of animals in husbandry is not only necessary to mitigate the risk of losses, but also a legal requirement. The rationale for a timely reaction to an intruder is self-explanatory.

1.2.2. Transmission Distances

Sensor data needs to be transmitted over distances well beyond the range of conventional Wi-Fi networks. In agricultural settings fields are usually in a distance between 2 and 7.5 kilometres from the farm [5]. In urban agriculture scenarios production plots are typically between one to three kilometres apart [6]. Urban environments present an additional level of difficulty with buildings obstructing the line of sight between transmitter and receiver of a setup, lowering the signal quality and the maximum range of the chosen transmission technology [7].

1.2.3. Environmental Conditions

Sensors nodes and transmission equipment are exposed to outdoor conditions, with seasonally varying temperature ranges, rain and wind. Sensors placed in protected production facilities like greenhouses and foil tunnels can be exposed to elevated temperatures as well as to condensing humidity. Greenhouses and foil tunnels might additionally complicate the RF situation when the metal structure acts like a Faraday cage, dampening signal strength and distorting the signal.

1.2.4. Usability

Farmers and urban farmers are typically no experts in information technology. While both profession groups usually need to be able to adapt technology to their production intents, it is desirable for a sensor network setup to be as easy to deploy and to maintain as possible. The battery runtime of the sensor must exceed several months. Integrating additional sensors into a system should pose a low barrier. Urban Farming environments might require temporarily shifting sensors to new plots, helping the farmer to grasp the local conditions, allowing him to adapt the production concept accordingly.

1.3. Structure of the Paper

In Chapter 2 the current state of the art of data logging systems is presented. The features of the currently used technologies for data transmission of sensor nodes (SN) are compared. The potential for novel LoRaWAN based SN is highlighted. The overall system architecture is explained in Chapter 3. The focus of this chapter is on the data routing between the SN and the cloud application. The structure of proposed sensor node is explained in detail in Chapter 4. The used hardware and software components are introduced. A simplified device configuration approach and measurements for power consumption reduction are explained.

This work evaluates if the proposed sensor node is suitable for the use in UA applications in terms of provided range, battery runtime, and sensor accuracy. The range evaluation is performed in Chapter 5. Signal strength and quality parameters were recorded and evaluated for different locations. In Chapter 6 the battery runtime was analysed. A power interval analysis shows the current consumption for the different operating modes of the SN. The theoretical battery lifetime was estimated using the results from the power interval analysis. This estimation was validated and confirmed by an experiment. An accuracy evaluation of supported temperature and humidity sensors is performed in Chapter 7. Those physical quantities are essential in many UA

applications and are well assessable. The measured sensor values are compared to data from a professional weather station to gather the statistical parameters of the used sensors. A conclusion and suggestions for further research are given in Chapter 8.

2. STATE OF THE ART

A variety of different data logging systems are used in the agricultural sector. Different categories of devices are discussed in this section. Their usage depends mainly on the need for real-time data access, the requirements on simplicity and the duration of usage.

2.1. Offline Data Logging

Offline data loggers store measurement data locally. The data inventory needs to be read out manually. Those data loggers are available from 50\$ to 2000\$ with respect to functionality, the size of data storage and battery lifetime. The configuration and installation of these devices is usually simple compared to setup of wireless sensor networks (WSN). No communication network infrastructure or mains supply is required for operation. This type of data acquisition is typically used as a robust and yet simple solution if no real-time data is required. The missing capability of live data transmission and analysis therefore restricts the use cases substantially. Those systems achieve a typical battery lifetime of more than one year as a result of the lacking power intensive radio frequency (RF) transmissions [8, 9].

In other literature those devices are frequently used for time limited experiments, particularly due to the extensive effort of the manual data readout. Shaw et. al. are using an offline DL2e DeltaT data logger to estimate the spatial nitrogen variation within a grassland field [10]. Chatterjee, Dey and Sen developed a neural network based soil moisture quantity prediction model gathered from data using an offline HOBO U30 data logger [11].

2.2. Cellular Connected WSN

Data loggers with a cellular modem solve the problem of lacking online data. Those devices transmit the measurements periodically using a mobile radio. Common variants are using GSM, 2G or 3G cellular network technology which are not low-power optimized. Those systems have a fairly high power consumption when transmitting [12] and often require a mains power supply [13]. Currently new cellular technologies optimized for IoT applications are emerging. The LTE-Cat-NB1 and LTE-Cat-M1 extensions provide a narrow-band data transmission optimized for low-power and high range utilizing existing infrastructure [14]. Zhang et. al. proposed a sensor node equipped with an LTE NB-IoT modem for data transmission that transmits environmental parameters at an interval of one week with an estimated battery lifetime of 11 years [15]. Those systems still need a registered SIM card introducing recurring costs. As with all cellular networks the usage is limited to areas with actual network coverage. A basic LTE NB-IoT network coverage of Telekom in Germany at January of 2021 is given but especially rural parts are still lacking connectivity [16].

The proprietary SigFox network pursues a similar approach as the IoT optimized LTE protocols. A narrow-band technology is utilized for low bandwidth data transmission over long distances. The creator of the SigFox protocol acts as the only available provider of this technology at the same time. The costs of the service depend on the number of devices and the frequency of message transmission. The network coverage in Germany is especially problematic in rural areas [17]. The Thoreau project uses the SigFox network to transmit underground soil moisture and ambient temperature data from sensor nodes installed on multiple location on a campus into a

cloud application over long time periods [18]. Joris et. al. are using a solar powered SigFox sensor node to transmit temperature and humidity measurements on a vineyard to evaluate the weather influence on the yield [19]. The usage of cellular data loggers is introducing a dependency to 3rd party network providers. This should be specially considered for long term usage scenarios. The provider may increase the usage fees or even shut down the service in non-profitable situations.

2.3. UAV Supported Data Logging

A new approach is using autonomous drones for the readout of data loggers. A drone is used to temporarily activate a communication interface of the data logger using RF pulses [20]. The short distance between the data logger and the drone enables the usage of ultra-low power RF protocols for measurement data transmission providing only a short range [21]. Those systems could be extended by path planning techniques to automate the control of the drone that were designed for similar problems [22, 23]. Idbella et al. presented such a UAV based data logging approach for monitoring agro-ecological condition of vine plants [24]. In this experiment the placement of sensors and the control of the drone was performed manually. More research is necessary to adapt existing path planning techniques to this specific problem. This ambitious approach for sensor data collection is still in development. Particular aspects are already working but the challenge here is the integrating of sensor technology and complex automated control of the UAV into a usable product at reasonable costs. Real time data acquisition would still not be possible with this approach and the costs for the required infrastructure is fairly high compared to traditional WSN.

2.4. LoRaWAN connected data logging

A promising technique for low-power and long-range data transmission in agricultural applications is LoRaWAN. Data loggers equipped with a LoRaWAN modem were already used for experiments and show good results. Davcev et al. are utilizing a LoRaWAN technology from The Things Network to measure leaf wetness and soil moisture to control an irrigation system [25]. The focus of that research however was on the data analytic part of the system. Ibrahim et al. are using LoRaWAN development kits to measure and control the ambient humidity of a Shiitake fungi cultivation [26].

Those experiments are using LoRaWAN implementations that are not optimized for general agriculture applications. The used hardware was tailored to the specific experiments and does not provide a generalized interface for sensors. Furthermore, aspects as power consumption and range were not a focus of that work. The LoRaWAN technology itself is a promising approach if real-time data is required. The low-power narrow-band RF transmission enable small sensor nodes with a large battery lifetime while using a high sampling rate. The capability of self-deployment of gateways enable good network coverage even in rural areas and decouples the dependence of 3rd party providers.

3. SYSTEM ARCHITECTURE

The proposed sensing and data logging system is a wireless sensor network (WSN) using LoRaWAN technology as RF protocol to connect the sensors to a gateway, as shown in Figure 1. The architecture is made of three layers. The sensor data is gathered by the SN in the bottom layer. The network layer is using LoRaWAN technology for routing and aggregation of the sensor data into the application layer. Here, the gateway (GW) aggregates the data from the

sensor nodes. The data storage, analysis and representation is realized as a cloud service in the application layer. Those layers are described in more detail in the following subsections.

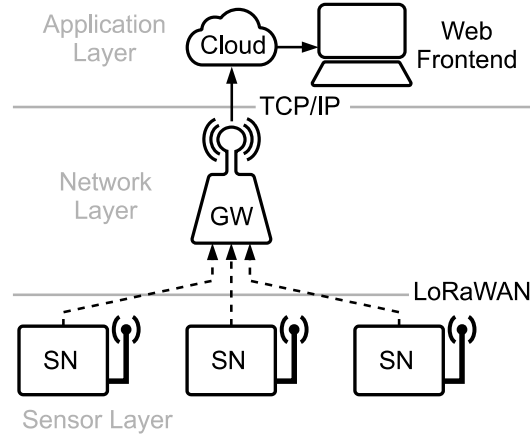


Figure 1. Overall system overview of the wireless sensor network with cloud integration

3.1. Data Routing and Aggregation

The measurement data is transmitted from the sensor to the application layer using LoRaWAN technology utilizing LoRa RF modulation. The physical layer of data transmission is using the proprietary narrow-band LoRa protocol from Semtech. The direct sequence spread spectrum (DSSS) is replacing each bit by a sequence of bits resulting in a signal with higher bandwidth that is less prone to narrow-band interference. Chirp spread spectrum (CSS) transmits each symbol using continuously varying frequency to eliminate the need for a precise reference clock [27]. The modulation provides a high range while using less energy though only achieving relatively low data rates. The specified LoRa modulation describes the raw RF transmission only. The adaption of parameters (e.g. spreading factor, bandwidth) enables a trade-off between range and data rate. More advanced features are implemented in the upper LoRaWAN layer.

The medium access control (MAC) and the aggregation and routing of messages is done using the LoRaWAN protocol extension. The data transfer is always initiated by the end devices (SN) followed by a receiving window for data uplink from the network [28]. The end devices stop listening after the receiving windows and enter a sleep state to save energy. Messages from a single or multiple gateways are aggregated by a central network server (NS) and from there redirected to specific application servers (AS) both using TCP/IP based protocols, as shown in Figure 2. The LoRaWAN infrastructure can be self-supplied by the operator of the network or a third party provider can be used. We are using the infrastructure from The Things Network (TTN) for the proposed WSN. TTN offers a free usage of their infrastructure. Therefore, the gateways need to be connected to their service. The gateway is then available for all registered TTN users. The motivation of TTN is to create a global LoRaWAN network only with community operated gateways. The separate encryption of payload and network metadata should prevent eavesdropping.

3.2. Application Layer

In the application layer all received sensor data is processed and stored depending on the specific application. In our setup the aggregated messages are fetched by a Node-RED instance from TTN by using the MQTT protocol, shown in Figure 2. Node-RED is a web-based tool for data

processing. The graphical flow-based approach allows a simple design of rules for data processing [29]. For the proposed architecture the received measurement data is validated and stored in an InfluxDB database. The Node-RED software provides interfaces for MQTT and InfluxDB.

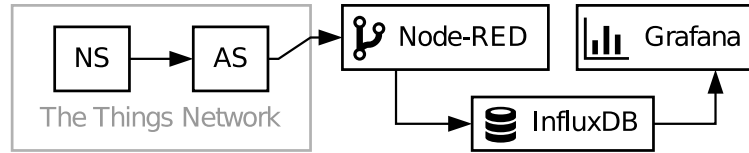


Figure 2. Application structure for data aggregated by a LoRaWAN network

The used InfluxDB time series database (TSDB) is specialized for storing periodical data. Compared to traditional relational database management systems (RDBMS, e.g. MySQL) the performance for storing single measurements is significant higher [30].

One crucial aspect of the overall system is the easy evaluation and processing of the measurements. The web-based Grafana frontend allows a simple query of the recorded measurements [31]. A visualisation using graphs (e.g. lines, bars, points), gauges, tables and integration of third party controls is supported. The user can set the time range of the output data. It is possible to show simple statistical data (min, max, average, sum) in a legend. Thresholds can be set to visualize critical periods and to send alerts via E-Mail. The graph raw-data can be exported to CSV files for further processing.

4. SENSOR NODE ARCHITECTURE

The major effort of the proposed system was put into the adaption of the developed IoTyze sensor node (SN) supporting LoRaWAN for agriculture applications. This SN architecture was originally developed by our faculty as a generic LoRaWAN sensor platform. This work optimizes the software stack in terms of easy deployment and simplified sensor inclusion. Figure 3 shows the logical structure of the proposed SN. Main processing component is the STM43L4 MCU. It supports multiple low-power states while providing high processing power if required. A trade-off between low power-consumption and computing power in active state is achieved by a flexible clock selection. The used sleep mode (Stop 2) consumes 2.4µA with enabled real-time clock (RTC) and backup memory.

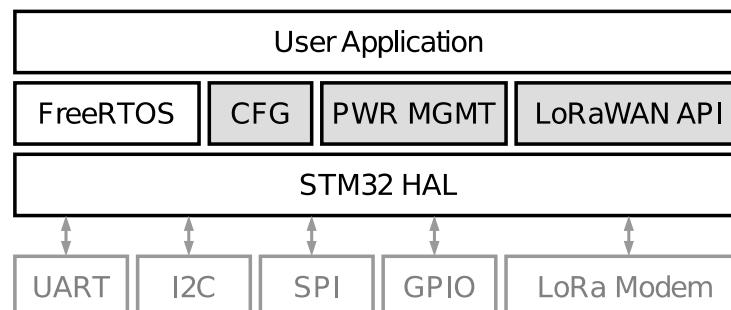


Figure 3. Sensor node stack of the IoTyze device

The sensors are physically connected to the various interfaces of the MCU. The STM32 HAL library from STMicroelectronics provides high-level hardware drivers for the MCU core and the peripherals. FreeRTOS is used as a real-time operating system for scheduling multiple tasks and synchronizing shared resources. The grey shaded boxes, these are CFG (configuration parser), PWR MGMT (power management unit), LoRaWAN API (LoRaWAN driver) are self-developed parts of the SN framework. These components are introduced in the following sections. Different parts of the developed software are implemented in dedicated tasks to improve the modularity of the software project and to ease code maintenance.

4.1. Program States

The software supports periodic data readings from the sensor devices via different interfaces, node integration into the sensor network, data transmission, and low-power sleep modes. The program flow of the SN is predefined by the developed software framework according to Figure 4.

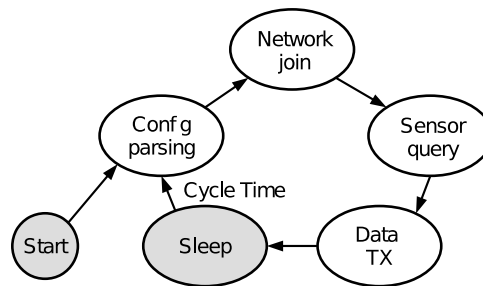


Figure 4. Program states of the sensor node

The system first parses the configuration file stored on the EEPROM. This step includes the recovery of persistent application data from the backup RTC RAM. This data is empty at first start. Afterwards the SN joins the LoRaWAN network, see 4.3 for detailed description. All enabled sensors are initialized and read out. The collected measurements are transmitted using the LoRaWAN modem. Finally, the device is set into a low-power sleep state to reduce the power consumption. The SN restarts after a configurable cycle time. The sleep state is also set on network errors. Unsuccessful measurements do not interrupt the program and only set a failure flag in the payload data to mark the certain measurement as invalid.

4.2. Configuration Management

The implemented configuration management provides a simple mechanism for storing application specific parameters. The configuration includes all connected sensors and their parameters (e.g. interface, slave address), the measurement interval and LoRaWAN related specifications (e.g. device EUI, application EUI, encryption keys, data rate). The configuration needs to be set during the deployment of the SN. Figure 5 shows the structure of the implemented configuration management.

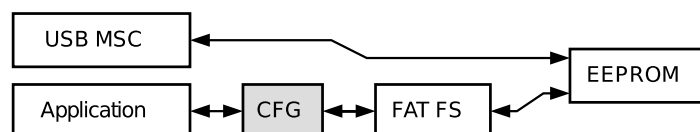


Figure 5. Configuration management of the sensor node

The configuration is stored in a text file on a FAT32 formatted I²C EEPROM. The file access on the MCU is provided by the used FAT FS library. The implemented configuration parser (CFG) retrieves the individual parameters to the application. The configuration file can be changed by connecting the SN to a computer using the provided USB interface. The SN implements the USB mass storage device class (MSC) that maps the EEPROM as a drive on the computer. Therefore, it is possible to change the configuration file with a regular text editor. Furthermore, a prepared file can be easily copied to the drive simplifying mass deployment. The utilization of the USB MSC standard provides OS independent compatibility and eliminates the need for a dedicated configuration software and hardware programmer.

4.3. LoRaWAN Driver

One module of the software stack is the LoRaWAN driver, cf. LoRaWAN API in Figure 3. This driver provides an API to the application that enables a simplified connection management and data transmission. The low-level driver manages the UART interface between the MCU and the LoRaWAN modem. The high-level driver transmits AT commands, parses the response and implements the state management. An API is declared to the application for simple usage of the LoRaWAN network.

LoRaWAN clients usually perform an over the air activation (OTAA) to join the network after each power-up. The clients send a join request and the gateway answers with a join accept response including a nonce for generation of the session keys for data encryption. The generated session keys are usually lost after power-down. The developed driver stores the session keys in the modem for reuse. The OTAA is replaced by activation by personalization (ABP) if session keys are present. The ABP approach does not require any join request, thus reducing the required duty cycle of the SN for subsequent measurements. An OTAA is only performed at the first start or if the ABP method fails (e.g. if the NS rejected the session keys after long inactivity). A cyclic regeneration of the session keys by performing OTAA can be optionally scheduled by the application developer to improve the security if required.

4.4. Power Management

A long battery lifetime is achieved by entering a low-power mode between the measurements. The software stack includes a dedicated power management (PWR MGMT) module to simplify the usage of low-power states for the application developer. All required setups are executed before entering the low-power state. This covers platform (e.g. power-down of LoRaWAN modem) and MCU (e.g. interrupt configuration) specific tasks and the configuration of the wake-up source. The Stop 2 state of the MCU with enabled RTC is entered between the measurements. The RAM is disabled when entering this state to reduce the power consumption to a minimum. The MCU is therefore rebooting after wake-up. The PWR MGMT module provides a mechanism to store application specific data into the backup memory of the RTC. The data is passed as a structure to the PWR MGMT module before entering the low-power state and is retrieved after the MCU is rebooted. This allows the application developer to store data between the measurement cycles.

4.5. Sensor Drivers

The modular system architecture allows a simple integration of sensor drivers using the C programming language. The included STM32 HAL library offers high-level access to all peripherals of the MCU. Platform specific example code for various interfaces (GPIO, ADC, I²C, SPI, UART) is available. The SN software framework includes drivers for various sensors that can be used in the application layer. Table 1 shows all provided sensor drivers.

Table 1. By SN supported sensor devices with related interfaces

Sensor	Physical quantities	Interface
DHT22 / AM2303	Temperature + r.H.	GPIO
Sensirion SHT21	Temperature + r.H.	I ² C
Sensirion SHT31	Temperature + r.H.	I ² C
Texas Instruments HDC1008	Temperature + r.H.	I ² C
Maxim DS18B20	Temperature	GPIO
Bosch BMP280	Pressure + temperature	I ² C
Bosch BME680	Press. + temp. + VOC	I ² C
TAOS TSL2561	Luminosity	I ² C
Capacitive Soil Moisture	Soil moisture	Analog
Sparkfun Soil Moisture	Soil moisture	Analog
Nova SDS010	Particulate matter	UART
Sensirion SPS30	Particulate matter	UART
Sensirion SCD30	CO ₂ + temp. + r.H.	I ² C

A template for developing a custom driver is additionally provided. A separation between the low-level hardware access and the device logic is introduced. Each part is implemented in a separate pair of .c/.h files. The programmer can utilize the synchronization functions and blocking delays from FreeRTOS eliminating the need of implementing own schedulers. The high-level functions are called from the application layer.

4.6. Reference Hardware

The described architecture is implemented in a reference hardware. Core component of the SN is the IoTyze LoRa board extended by various peripherals as shown in Figure 6.



Figure 6. Reference hardware of the sensor node

This credit card sized board integrates an STM32 host MCU, an RN2483 LoRaWAN modem and a power management system with lithium polymer (LiPo) battery charger [32]. The SN is supplied by one 3.7V LiPo cell with 2200mAh capacity. The system is mounted inside an IP65 classified case to provide a protection against external environmental influences. The USB interface used for device configuration and battery charging is realized using a robust aviation-grade GX12 connector. The sensors are connected using similar GX12 connectors. Those

measures prevent the entry of moisture. The SN can be mounted in environments facing splashing water (e.g. outdoor) or condensing humidity (e.g. green houses).

5. RANGE EVALUATION

The achieved range depends on the transmission power of the transmitter and used data rate. The data rate of a LoRa system depends on the used bandwidth and spreading factor (SF). The SF determines the required SNR at the receiver until demodulation becomes possible. The receiver sensitivity is given by Equation 1 [33].

$$S = -174\text{dB} + 10 \log BW + NF + SNR \quad (1)$$

where,

- S = Receiver sensitivity
- BW = Bandwidth
- NF = Noise figure
- SNR = Required signal-to-noise ratio

The bandwidth of LoRa modulation for Europe is fixed to 125kHz, while other parts of the world may use 250kHz [34]. The NF describes the inherent noise of the receiver. The only controllable factor is the used SF resulting in the minimum required SNR. The choice of SF influences the data rate. The use of high SF allows large coverage but reduces the possible data throughput. Table 2 shows minimum required SNR and achievable data rates (DR) for various SF.

Table 2. Resulting SNR and data rates for various SF using LoRa

SF	7	8	9	10	11	12
SNR (dB) [36]	-6	-9	-12	-15	-17.5	-20
DR (kb/s) [27]	5.47	3.13	1.76	0.976	0.537	0.293

5.1. Experimental Setup

For range evaluation the sensors were placed at different spatial conditions. Multiple LORIX One gateways are placed at a different position on the campus [35]. The measurements are only recorded from a single GW that was placed at a fixed position. A test software was developed to gather signal strength and transmission quality data. The received signal strength indicator (RSSI) and SNR from the GW is fetched for 100 sequent uplink packages. Each uplink package from the SN is confirmed by a downlink package from the GW. The test was done in a multi gateway environment. The downlink messages are sent by the GW with the best link to the SN. The SNR measurements by the SN were dropped in this evaluation, because the downlink messages are not sent by the same GW for each uplink package that affects the transmission quality. In total, four different spatial setups have been used:

Location 1 (indoor, 40m, NLOS):

The test SN was placed indoor in a distance of 40m to the GW. Several thick stone walls are located between the SN and the GW causing additional damping. The signal is partly reflected at the walls, introducing reflection of the transmitted signals.

Location 2 (outdoor, 200m, LOS):

The SN was placed outside on the campus in line of sight (LOS) condition to the GW. The distance between the SN and the GW was 200m. The SN was orientated almost straight to the window.

Location 3 (outdoor, 200m, NLOS):

In this scenario the SN was placed outdoor in a distance of 200m on the campus with one building between the GW. The signal has to pass multiple walls. Furthermore, the SN was placed angular to the window where the GW was located, requiring the signal to pass part of the facade.

Location 4 (outdoor, 300m, NLOS):

The SN was placed at the location on the campus with the distance of 300m to the GW. Three buildings are located between the SN and the GW.

5.2. Results

Figure 7 shows the results for location 1. The measured RSSI for all chosen SF are in the range between -84,79dB and -81,45dB. The RSSI measurements show a wider spread for SF of 10 and SF of 11. The average SNR values are in a range from 7,33dB for SF of 12 to 10,28dB for SF of 8.

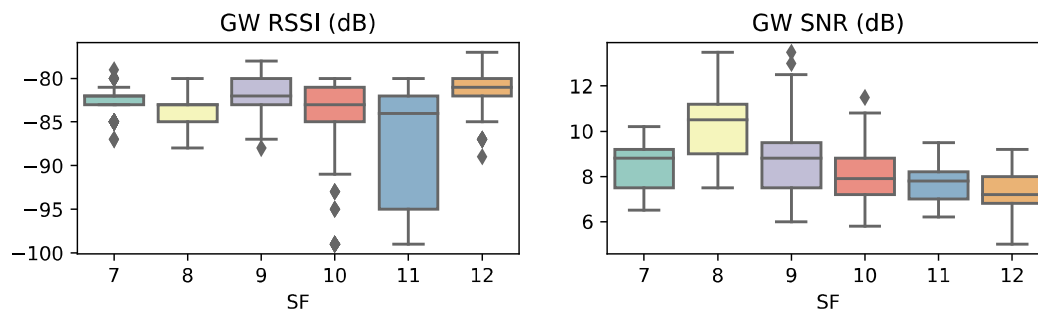


Figure 7. RSSI and SNR measurements at gateway for location 1

The results for location 2 are shown in Figure 8. The range of average RSSI values was between -109,65dB for SF of 11 and -107,92dB for SF of 12. The distribution of single measurements is, again, wider for SF of 10 and SF of 11. The lowest SNR average of 2,17dB was measured for SF of 12 and the highest SNR average of 4,05dB was reached for SF of 10. The SNR values are wider distributed for this location. As expected, the RSSI as well as the SNR are smaller compared to setup in location 1, because of the larger distance between SN and GW.

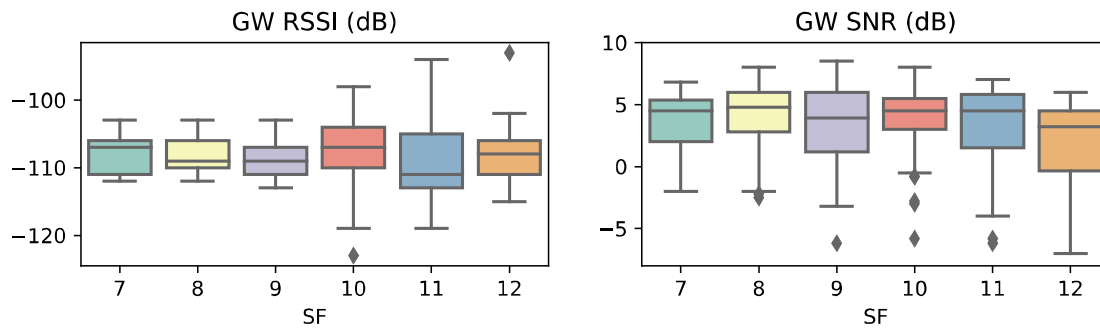


Figure 8. RSSI and SNR measurements at gateway for location 2

Figure 9 shows the result for location 3. The locations 2 and 3 are both located in a distance of 200m to the GW. While location 2 was in LOS to the GW, multiple walls causing a NLOS condition for location 3. For SF of 12 a loss of 7 packets was detected. The RSSI average values are closely around -112dB for all SF. The RSSI measurements are showing more outlier, compared to location 2. The lowest SNR average of -4,09dB was determined for SF12 and the highest average SNR was -2,28dB for SF9. The SNR values are wider distributed for location 3.

The average RSSI and SNR values have slightly deteriorated for NLOS conditions. The distribution of the single measurements is noticeably wider for NLOS conditions, compared to the values for LOS conditions of the same distance. Reflecting signals from the walls, resulting in a multipath effect, could be an explanation for this observation.

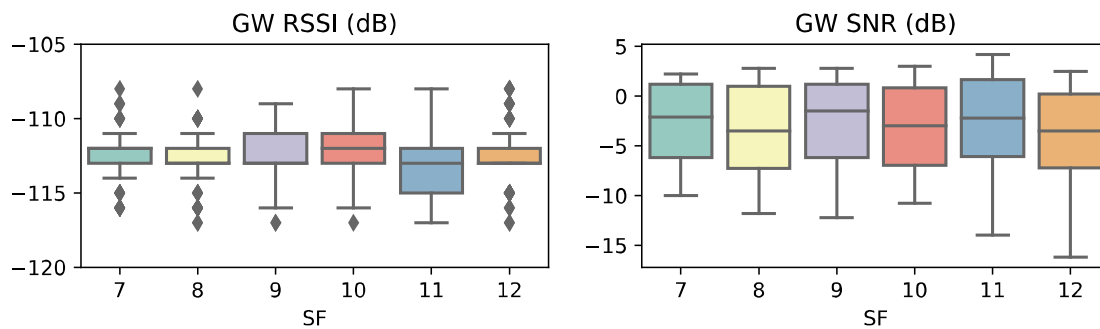


Figure 9. RSSI and SNR measurements at gateway for location 3

The results for the furthest location 3 are shown in Figure 10. For location 4 the distance was increased from 200m to 300m. The measurements for RSSI and SNR are similar to results in setup for location 3. At this location 2 packets were lost for SF of 8, while no packets were lost for all other SF.

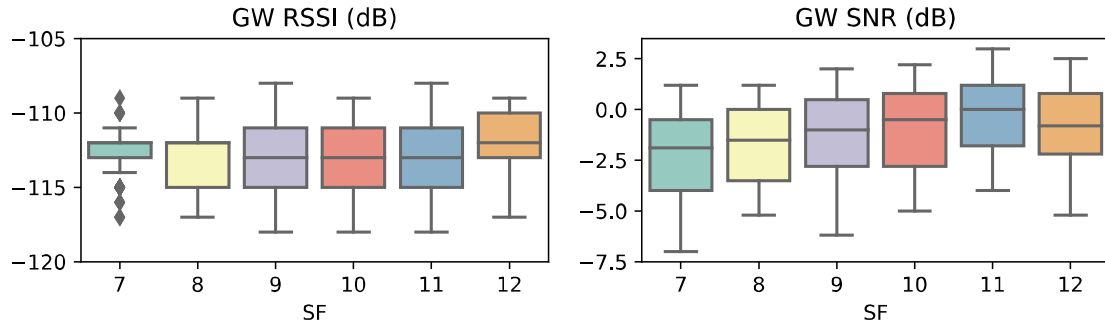


Figure 10. RSSI and SNR measurements at gateway for location 4

To sum up, the signal strength and signal to noise ratio depends on both distance between transmitter and receiver and presence of blocking objects, i.e. if we have LOS or NLOS condition. In NLOS condition the signal attenuation heavily depends of type of blocking object. Therefore, a general statement when a data transmission fails in case of NLOS can't be made. In our case, the presence of a building causing the NLOS condition as difference of scenarios location 2 and location 3 slightly lowers RSSI and SNR, but data transmission is still possible. The average RSSI over all measurements decreased by 4.2dB and the SNR by 6.49dB. This rate depends on type of object and has to be analysed for new setups. Between location 3 and 4, the distance was increased by 50%. The angle between the SN and the GW was changed, causing different objects as obstacles. The average RSSI over all measurements further declined by only 0.1dB. The average SNR even increased by 1.92dB. I.e., the increased distance does not have a significant effect. The measurements for location 3 and a SF of 12 results in a packet loss rate of 7%, while no packets were lost for smaller SF. A similar behaviour was also seen on location 4 and SF of 8. Particular spatial conditions can cause poor reception for certain SF. Then the usage of a lower SF can increase the receiving quality, although a better SNR is required for decoding. In total, transmission conditions and environments have to be analysed for new node sensor node application to select appropriate LoRaWAN transmission parameters and to ensure dependable data transmission.

6. BATTERY RUNTIME

The requirement for reduced maintenance demand long operating intervals between battery recharge. The proposed platform needs to compete against established products with a battery runtime of several months. Therefore, it features a low power consumption as a prerequisite. Space limitations and the used cell chemistry restrict the installable battery capacity. Nickel-metal hydride batteries show significant self-discharge over time. Lead-acid cells have a low energy density (Wh/cm³) compared to lithium-ion batteries [37]. The used lithium-ion battery combines a low self-discharge with a high energy density enabling small sensors with a long operating intervals. To optimize battery runtime a detailed power analysis of the sensor node components is performed.

6.1. Power Interval Analysis

The current consumption of the proposed sensor node depends on the operation state. The power consumption significantly changes between the states sensor readout, data transmission and sleep. The current consumption in all operation states was measured using a Keysight B2901A precision source measurement unit (SMU). For the measurements during the active states the supply voltage was set to 3.7V according to the nominal voltage of the used battery cell. The SF

of the SN was set to 11. The downlink messages were acknowledged by the gateway. The sampling frequency of the SMU was set to 10ms. The sensor node needs to perform an over-the-air activation (OTAA) when powering on for the first time to exchange the temporary session keys for encryption. This time-consuming procedure only needs to be repeated if the session keys are out of synchronization (e.g. the NS or AS dropped the session key). The required energy or respectively charge for the transmission of payload depends on TX power, selected SF and message type. In this example the worst case scenario with the largest possible SF, the highest allowed TX power and a confirmed uplink with acknowledge was used. The acquisition time and power consumption highly depends on the number and types of sensors that are connected. For this measurements a DHT21 temperature sensor and a BMP280 ambient pressure sensor were read out by the SN. The current consumption between the measurements is resulting from the real-time clock (RTC) of the MCU and leakage current of the circuit. The current consumption of the sensor in sleep mode is drastically reduced compared to the active states. Changes in the supply voltage show a non-negligible impact on the current consumption in sleep mode. Analysed states and related current consumptions are listed in Table 3.

Table 3. Current consumption profile

	Duration (s)	Charge (mAs)	Avg. current (mA)
OTAA Join	9.39	188.3	20.12
Transmit	2.54	67.9	26.73
Acquisition	0.69	10.63	15.41
Sleep	-	-	0.1524

6.2. Battery Runtime Estimation

The estimated battery runtime was estimated using the measurements from the power cycle analysis. The total charge for one cycle is calculated by Equation 2.

$$Q_{cycle} = \frac{1}{n_{rj}} Q_{join} + Q_{Acq} + Q_{TX} + I_{sleep} \cdot t_{sleep} \quad (2)$$

where,

Q_{cycle} = Charge required to perform one measurement cycle

n_{rj} = Average number of cycles until a re-join is required

Q_{join} = Charge required for joining the network

Q_{Acq} = Charge required for data acquisition

Q_{TX} = Charge required for data transmission

I_{sleep} = Sleep current

t_{sleep} = Time between two measurement cycles

The battery runtime estimation was calculated for multiple measurement cycle times. It was assumed that a re-join is required each 20 cycles. Table 4 shows the estimated battery runtimes for various cycle times.

Table 4. Battery lifetime estimation (in days) for various cycle times

t_{sleep} (m)	Q_{cycle} (mAs)	$t_{battery}$ (d)
5	133,7	243
10	179,4	362
30	362,3	538

A battery runtime of almost one year was estimated when using a cycle time of 10 minutes between the measurements. This setting strikes good balance between a long battery runtime and a high sampling rate. The real battery runtime as measured is based on this cycle time.

6.3. Experimental Battery Runtime

The battery runtime of a SN from the first revision was measured in a long-term experiment. The SN was mounted in a greenhouse. The SN was equipped with a DHT22 temperature and relative humidity sensor and a BMP280 ambient pressure sensor, as used for the estimation. The battery voltage curve is shown in Figure 11.

The sensor was installed on 15th of January 2020 and has sent data with a cycle time of 10 minutes until 8th of September 2020, which is 237 days. The battery had an open clamp voltage (OCV) of 3.622V at the end of the experiment. The full charge of the battery could not be used due to a non-optimal power supply circuit of the used prototype. The second revision of the SN has an optimized power supply circuit supporting operation down to 3.2V OCV.

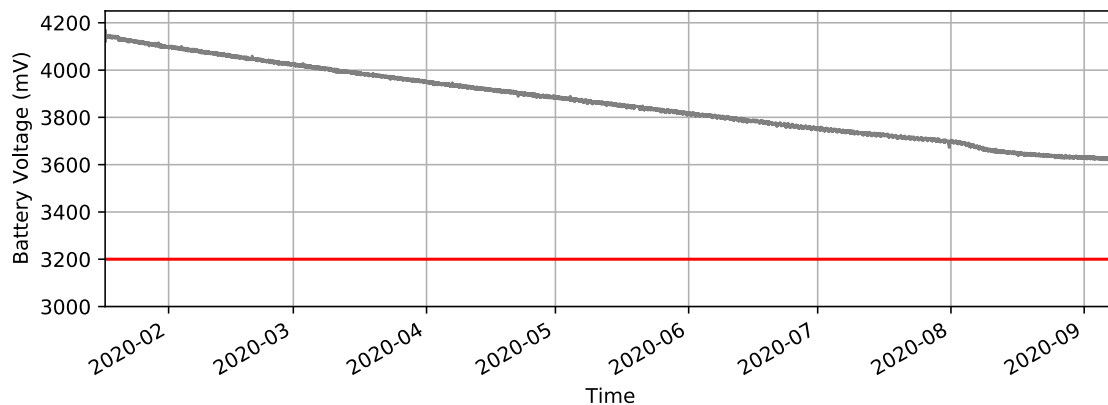


Figure 11. Battery runtime experiment for almost 8 months

7. SENSOR EVALUATION

For the productive use of the proposed SN it is essential that the achieved accuracy of the measurements is comparable to established products. A robust measurement of temperature and relative humidity is crucial for many UA applications. Furthermore, those measurements are well evaluable by comparison to a reference. Multiple supported sensors as listed in Table 1 were compared in terms of accuracy. The used SN was attached with two DHT22, a SHT21 and a SHT31 sensors. One of the DHT22 sensors was wrapped with a PTFE membrane as a vapour barrier to suppress condensed humidity inside the sensor. The SHT21 sensor was not equipped with a membrane and the SHT31 sensor had a factory mounted PTFE membrane. The SN was mounted outdoor next to a Pessl iMetos 3.3 weather station that was used as reference. The data was gathered over a period of 30 days. Data from the SN and the Pessl weather station was

individually transmitted with a cycle time of 10 minutes each. Both data sources are not synchronized causing a delay between both data sources.

The data was averaged over 30 minutes to compensate for time offset. The root-mean-square error (RMSE) was calculated based on the difference between the sensor readings and the reference. A linear regression was applied on the readings of each sensor and the reference. The correlation coefficient is a measure for strength of the linear correlation. The results for the temperature readings are shown in Table 5. The accuracy of the temperature sensors is quite good. Especially the cheap DHT22 sensors perform very well. The most expensive SHT31 has a small offset, but less outlier.

Table 5. Temperature measurements accuracy evaluation

Sensor	RMSE (°C)	Linearity	Offset (°C)	Corr. coeff.
DHT22	0.548	0.941	0.233	0.976
DHT22 PTFE	0.523	0.948	0.132	0.979
SHT21	0.457	0.974	0.308	0.988
SHT31 PTFE	1.134	0.962	1.119	0.978

The results for the humidity readings are represented in Table 6. The cheap DHT22 sensor with manually mounted PTFE membrane performs best. The most expensive SHT31 sensor has a poor performance.

Table 6. Humidity measurements accuracy evaluation

Sensor	RMSE (% r.H.)	Linearity	Offset (% r.H.)	Corr. coeff.
DHT22	3.612	1.382	-37.2	0.96
DHT22 PTFE	3.572	1.099	-7.67	0.933
SHT21	5.256	0.831	13.0	0.913
SHT31 PTFE	7.746	0.875	17.6	0.917

The majority of measurements is above 90% r.H. causing the relative high offset values. The linear regression would produce better results when being applied to more diversified data set. Furthermore, the data series of the sensors show a systematic error for conditions with strong solar radiation. The housing gets heated, resulting in a drop of relative humidity compared to the ambient humidity. A detailed evaluation of the humidity sensors is therefore only partly possible with the limited available data.

8. CONCLUSIONS

A sensor node has been developed that can be used in agriculture applications. Objectives were long range wireless communication, low-power design and modular structure to be able to easily support and to integrate different sensor devices. In this context, both, hardware and software have been optimized. The range of the developed sensor node is sufficient for urban farming applications. The distance of 300m could be bridged even with multiple buildings in between. The resulting signal quality has enough reserves for even larger distances. The results of the LOS data transmission test with a distance of 200m between sensor node and gateway show more than 20dB SNR margin, providing excellent performance application on open fields. For wireless communication a LoRaWAN modem with optimized parameter setup has been used.

The low-power feature of the sensor node has been implemented successfully, e.g. by supporting sleep modes of the processor. It was successfully verified that a battery runtime of about one year

is possible. The used battery still had some charge left after nine month of continuous operation with cycle time of ten minutes. For the first tests a non-optimal design of the power-supply circuit of the first hardware revision did not allow an operation below 3.6V battery voltage. This issue was fixed in a second hardware generation, now running down to 3.2V battery voltage. Based on the data of the first battery runtime evaluation, the revised hardware is able to operate one year on a single battery charge.

The evaluation of the supported temperature and humidity sensors show excellent results for the temperature measurements. All sensors except the SHT31 sensor showed a RMSE of around 0.5°C compared to the professional Pessl iMetos weather station. The evaluation of quality of the humidity measurements was partially possible only, because the humidity was almost above 90% r.H. during the whole test period. It was shown that the cheap DHT22 sensor with a manually attached PTFE membrane performs very good. In contrast, the expensive SHT31 showed a poor performance during the test period. The high accuracy and the reasonable costs of the DHT22 sensor supports the large-scale deployment of the proposed SN in UA applications.

The proposed sensor node is suitable for field soil measurements, aquaponics monitoring and urban farming applications. The long battery runtime and the continuous data transmission provide benefits compared to the usage of traditional offline or cellular data loggers. The extended range is sufficient for covering large areas under LOS conditions. A single gateway is able to cover a large field, eliminating e.g. the need to use UAV solutions to locally read sensor data.

8.1. Further Research

The collected data series of humidity measurements did not cover the full value range. The accuracy of the used sensor could be further evaluated using a more comprehensive data set. The construction of the sensor housing was not optimal. Solar radiation heated the housing causing a drop in relative humidity compared to ambient humidity. So far, only the accuracy of temperature and humidity measurements was evaluated. More sensors for measuring e.g. ambient pressure, light irradiation, and pH value need to be assessed in future research. The suitability of the proposed sensor node for application on large fields, e.g. soil measurements, could be further proved by experiments in large scale real-world scenarios. Further research could evaluate the usage of 5G technology as a replacement for the used LoRaWAN technology with the proposed sensor node architecture. This cellular technology would eliminate the need for a stationary gateway. The required power consumption needs to compete with the LoRaWAN solution to allow a comparable battery lifetime.

REFERENCES

- [1] S. M. Vicente-Serrano, F. Domínguez-Castro, C. Murphy, J. Hannaford, F. Reig, D. Peña-Angulo, Y. Trambay, R. M. Trigo, N. M. Donald, M. Y. Luna, M. M. Carthy, G. V. D. Schrier, M. Turco, D. Camuffo, I. Noguera, R. García-Herrera, F. Becherini, A. D. Valle, M. Tomas-Burguera, and A. E. Kenawy, "Long-term variability and trends in meteorological droughts in Western Europe (1851–2018)," *International Journal of Climatology*, vol. 41, no. S1, 2020.
- [2] R. Christensen, "SPIN-Farming: advancing urban agriculture from pipe dream to populist movement," *Sustainability: Science, Practice and Policy*, vol. 3, no. 2, pp. 57–60, 2007.
- [3] O. Elijah, T. A. Rahman, I. Orikumhi, C. Y. Leow, and M. N. Hindia, "An Overview of Internet of Things (IoT) and Data Analytics in Agriculture: Benefits and Challenges," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3758–3773, 2018.
- [4] A. Goldstein, L. Fink, A. Meitin, S. Bohadana, O. Lutenberg, and G. Ravid, "Applying machine learning on sensor data for irrigation recommendations: revealing the agronomist's tacit knowledge," *Precision Agriculture*, vol. 19, no. 3, pp. 421–444, 2017.

- [5] T. Machl and T. Kolbe, "Analyse landwirtschaftlicher Transportbeziehungen", Wege mit Zukunft, 2017.
- [6] K. Bradley, "Micro-farming on Rented Land: Curtis Stone Interview", Kewlona, 2017.
- [7] A. Farhad, D.-H. Kim, and J.-Y. Pyun, "Scalability of LoRaWAN in an Urban Environment: A Simulation Study," 2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN), 2019.
- [8] Onset, "U30 USB Weather Station," HOBO U30-NRC datasheet, 2021
- [9] Gemini, "Tinytag Plus 2 Dual Channel Temperature/Relative Humidity", TGP-4500 datasheet, Oct. 2014
- [10] R. Shaw, R. Lark, A. Williams, D. Chadwick, and D. Jones, "Characterising the within-field scale spatial variation of nitrogen in a grassland soil to inform the efficient design of in-situ nitrogen sensor networks for precision agriculture," Agriculture, Ecosystems & Environment, vol. 230, pp. 294–306, 2016.
- [11] S. Chatterjee, N. Dey, and S. Sen, "Soil moisture quantity prediction using optimized neural supported model for sustainable agricultural applications," Sustainable Computing: Informatics and Systems, 2018.
- [12] J. P. Becona, A. S. Pereira, C. Vazquez, and A. Arnaud, "A battery powered RTU: GPRS vs 3G comparison: IEEE Urucon 2017 paper 101," 2017 IEEE Urucon, 2017.
- [13] R. Yordanov, R. Miletiev, P. Kapanakov and E. Lontchev, "Design of a portable system for sensor data acquisition and transmission," 2017 XXVI International Scientific Conference Electronics (ET), Sozopolpp. 1-3, 2017.
- [14] A. D. Zayas and P. Merino, "The 3GPP NB-IoT system architecture for the Internet of Things," 2017 IEEE International Conference on Communications Workshops (ICC Workshops), 2017.
- [15] J. Zhang, P. Liu, W. Xue, and Z. Rui, "Farmland Intelligent Information Collection System Based on NB-IoT," Cloud Computing and Security Lecture Notes in Computer Science, pp. 331–343, 2018.
- [16] "Telekom LTE NB-IoT Coverage Map for Germany" [Online]. Available: <https://t-map.telekom.de/tmap2/nbiot>. [accessed: 23-Jan-2021].
- [17] "SigFox Coverage Map" [Online]. Available: <https://www.sigfox.com/en/coverage>. [accessed: 23-Jan-2021].
- [18] X. Zhang, A. Andreyev, C. Zumpf, M. C. Negri, S. Guha and M. Ghosh, "Thoreau: A subterranean wireless sensing network for agriculture and the environment," 2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Atlanta, GA, pp. 78-84 , 2017.
- [19] L. Joris, F. Dupont, P. Laurent, P. Bellier, S. Stoukatch, and J.-M. Redoute, "An Autonomous Sigfox Wireless Sensor Node for Environmental Monitoring," IEEE Sensors Letters, vol. 3, no. 7, pp. 01–04, 2019.
- [20] J. Chen, Z. Dai, and Z. Chen, "Development of Radio-Frequency Sensor Wake-Up with Unmanned Aerial Vehicles as an Aerial Gateway," Sensors, vol. 19, no. 5, p. 1047, Jan. 2019.
- [21] A. Rajakaruna, A. Manzoor, P. Porambage, M. Liyanage, M. Ylianttila, and A. V. Gurtov, "Lightweight Dew Computing Paradigm to Manage Heterogeneous Wireless Sensor Networks with UAVs", ArXiv, 2018.
- [22] A. Noriega and R. Anderson, "Linear-Optimization-Based Path Planning Algorithm for an Agricultural UAV," AIAA Infotech @ Aerospace, 2016.
- [23] L. H. Nam, L. Huang, X. J. Li and J. F. Xu, "An approach for coverage path planning for UAVs," 2016 IEEE 14th International Workshop on Advanced Motion Control (AMC), Auckland, pp. 411-416, 2016.
- [24] M. Idbella, M. Iadaresta, G. Gagliarde, A. Mennella, S. Mazzoleni, and G. Bonanomi, "AgriLogger: A New Wireless Sensor for Monitoring Agrometeorological Data in Areas Lacking Communication Networks," Sensors, vol. 20, no. 6, p. 1589, 2020.
- [25] N. H. N. Ibrahim, A. R. Ibrahim, I. Mat, A. N. Harun, and G. Witjaksono, "LoRaWAN in Climate Monitoring in Advance Precision Agriculture System," 2018 International Conference on Intelligent and Advanced System (ICIAS), 2018.
- [26] D. Davcev, K. Mitreski, S. Trajkovic, V. Nikolovski, and N. Koteli, "IoT agriculture system based on LoRaWAN," 2018 14th IEEE International Workshop on Factory Communication Systems (WFCS), 2018.
- [27] Semtech, "AN1200.22", LoRa Modulation Basics, 2015.
- [28] LoRa Alliance, "LoRaWAN 1.1 Specification", 2017.
- [29] Node-RED, "About", [Online] Available: <https://nodered.org/about/>. [accessed: 29-Jan-2021].

- [30] D. Arnst, V. Plenk, A. Woeltche., "Comparative Evaluation of Database Performance in an Internet of Things Context", 2018.
- [31] Grafana Labs, "Grafana Features", [Online] Available: <https://grafana.com/grafana/>. [accessed: 29-Jan-2021].
- [32] P. Bolte and U. Witkowski, "Energy self-sufficient sensor node for long range wireless networks," IOP Conference Series: Earth and Environmental Science, vol. 431, 2020.
- [33] Semtech, "AN1200.13", SX1272/3/6/7/8 LoRa Modem Designer's Guide, 2013.
- [34] Semtech, "RP002-1.0.0", LoRaWAN Regional Parameters, 2019.
- [35] Wifx, "Compact and Robust Professional Grade LoRaWAN Gateway", LORIX One, 2020.
- [36] T. Elshabrawy and J. Robert, "Analysis of BER and Coverage Performance of LoRa Modulation under Same Spreading Factor Interference", 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2018.
- [37] S. Anuphappharadorn, S. Sukchai, C. Sirisamphanwong, and N. Ketjoy, "Comparison the Economic Analysis of the Battery between Lithium-ion and Lead-acid in PV Stand-alone Application," Energy Procedia, vol. 56, pp. 352–358, 2014.

AUTHORS

Philipp Bolte is research assistant at the Department of Electronics and Circuit Technology of the South Westphalia University of Applied Sciences since 2016. He received his master's degree in systems engineering in 2017. Currently, he is working on his PhD research with focus on industrial IoT sensor nodes.



Dr. Ulf Witkowski heads the Electronics and Circuit Technology research group at the South Westphalia University of Applied Sciences in Soest, Germany. He has been an active researcher for about 20 years in the area of wireless networking, sensor networks, cognitive systems, and mini-robotics. He has established his research group at the South Westphalia University as a professor in 2009. His research areas include wireless communication involving mobile ad-hoc networking, radio-based node localization, sensor networks, and embedded systems. He received the diploma degree in electrical engineering in 1995 from the Technical University of Hamburg-Harburg, Germany and in 2003 the Dr.-Ing. degree from the University of Paderborn. U. Witkowski has published more than 80 scientific articles.



Chemical engineer **Rolf Morgenstern** (51), started researching sustainable food production, focussing on Urban Agriculture and Aquaponics, at the department of agriculture of SWUAS in 2015.



DEEP LEARNING FOR IDENTIFYING MALICIOUS FIRMWARE

David Noever and Samantha E. Miller Noever

PeopleTec, Inc., 4901 Corporate Drive. NW, Huntsville, AL, USA

ABSTRACT

A malicious firmware update may prove devastating to the embedded devices both that make up the Internet of Things (IoT) and also that typically lack the same security verifications now applied to full operating systems. This work converts the binary headers of 40,000 firmware examples from bytes into 1024-pixel thumbnail images to train a deep neural network. The aim is to distinguish benign and malicious variants using modern deep learning methods without needing detailed functional or forensic analysis tools. One outcome of this image conversion enables contact with the vast machine learning literature already applied to handle digit recognition (MNIST). Another result indicates that greater than 90% accurate classifications prove possible using image-based convolutional neural networks (CNN) when combined with transfer learning methods. The envisioned CNN application would intercept firmware updates before their distribution to IoT networks and score their likelihood of containing malicious variants.

KEYWORDS

Neural Networks, Internet of Things, Image Classification, Firmware, MNIST Benchmark.

1. INTRODUCTION

One classic benchmark for machine learning is handwriting digit recognition (Modified National Institute of Standards and Technology database, or MNIST) [1-12]. The original digit recognition challenge has since seen widespread generalization to include alphabetic versions [2] in multiple languages [10-12] and multiple unrelated topic areas [13-19] ranging across medical [13], fashion [14], and satellite imagery [17]. A common element of these generalizations has been that small images (either 28x28 or 32x32) [1,7,16] can be addressed with both statistical machine learning (e.g. tree-based algorithms) or deep learning (multi-layer neural networks) [7]. We have recently built many cyber-security challenge datasets for malware and intrusion detection by first assembling the dataset in formats compatible with previous MNIST solutions [17-19], but also adding to the conversation begun by Intel and Microsoft Research to go beyond the signature-based methods of identifying viruses in their STAMINA initiative [20-21]. Our datasets for malware (V-MNIST) [18] and image-based intrusion detection [19] are starting points for motivating the current approach to map firmware updates [22-23] that are either malicious, hacks, or benign into a similar format. The approach builds on the extensive publication history of mapping integer datasets to images, then applying the power of convolutional neural networks (CNNs) along with other algorithms to compare their ability to detect malicious or rogue firmware updates [24-25]. One motivation for converting the malware to imagery stems from the advanced feature extractions available for performing convolutions on pixel maps. The core mathematical transformation applied in two-dimensional convolutions includes sliding a small weight matrix over the image, performing elementwise multiplication within that particular sliding window, then finally summing up the results to generate new output pixel layers.

Successive layers involving convolutions automate feature extraction and hierarchies of related image parts. A second investigative motivation behind this approach follows from the success already demonstrated by STAMINA for other categories of malware [20-21], but extended here for firmware rather than traditional malware.

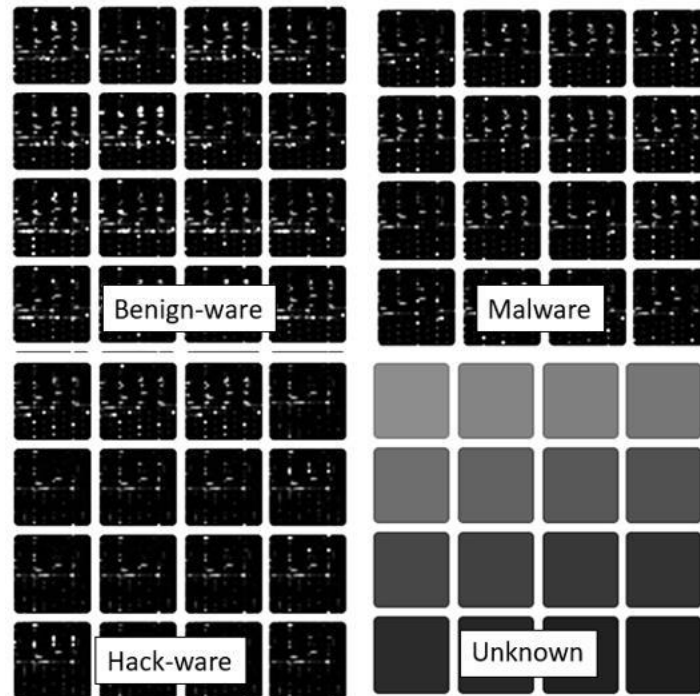


Figure 1. Firmware ELF binaries as Thumbnail Images

The future of embedded and Internet of Things (IoT) infrastructure depends on updates that users and industry can trust. What's unclear presently however is whether these updates will prove equally trustworthy given the lackadaisical approach to decent password protection or verifiable software integrity [26-27]. In 2020, 50 billion IoT devices worldwide are specifically designed to attach to a network with little or no administrative management or oversight [27]. While advanced persistent threats (APTs) have previously exploited weak passwords for devices like thermostats, home appliances, and personal assistants, the infection of firmware updates represents a larger attack surface to exploit. Anecdotal reports from the 2018 Olympics noted that hacked remote printers were unable to issue gate tickets for the opening ceremonies [27-28]. Ideally, a simple image classifier that quickly identifies and isolates rogue firmware might prove useful in the same way that program hashes and signatures defined a previous generation of malware protection layers. The original contribution of this work is to 1) map the firmware labeled dataset to a representative image and 2) solve the classification problem as a proof of principle for future development.

2. METHODS

This research extends the labeled ELF-binary dataset [22] to image classification. We accept the multi-class labels for malware, hack-ware, and benign-ware, which include over 40,000 examples of small compiled binaries. We add class specific to image classification which is grayscale "unknown" and bears no family resemblance to compiled software. The unknowns are just a spectrum of flattened backgrounds shades. The original dataset encodes the binary files using the following annotation and naming scheme:

{Architecture}__{Bit width}__{Endianess}__{ABI}__{Compiler used to compile the exe}__{Optimization level}__{Whether obfuscation was applied}__{Is the file stripped of debug symbols}__{Package name}__{Program name}.

2.1. Dataset Preparation

Employing the methods of Project STAMINA from Intel and Microsoft [21], we similarly convert the first 1024 bytes of each firmware binary to its decimal equivalent then scale those integers (0-15) to span the full 0-255 interval to create small images as JPEGs. Because the class imbalances include dominant benign firmware (75% of examples), we produced an alternative public dataset (published on Kaggle [29]) that includes both a long and a short-form version. The short-form version includes 3,000 examples of benign-ware, 714 examples of malware, and 100+ examples of hack-ware. While not balanced, it matches with the presentation of a basic confusion matrix of train-valid-test split. The choice for 1024 bytes as a small thumbnail (32x32 pixels in grayscale) derives from matching this complex problem to previous MNIST approaches but with attention to the stride-length (powers of 2) preferred by some modern deep learning frameworks like Keras. The area of the sliding weight matrix or kernel in 2D convolution determines the number of input features from the firmware that get passed to generate new output features in the deeper layers of the neural network.

2.2. Model Parameters and Quantitative Metrics

As an example of applying deep learning, we solve the firmware-image classifier problem using transfer learning from MobileNetV2 starting networks [30]. This network provides an optimized algorithm for feature hierarchies but efficiently extends to new areas beyond its original training datasets. We have previously found this approach useful to understand the image classification for both malware (V-MNIST) [18] and intrusion detection [19]. We use transfer learning over 50 epochs, with a 0.001 learning rate, and report four firmware classes: “malware”, “hack-ware”, “benign-ware”, [22] and the new class labeled “unknown”. The unknowns were to handle images outside of the patterns of ELF headers, such as flat grayscale backgrounds. We generate all the images using the ImageMagick tool suite [31] after binary-to-scaled decimal conversions of 1024 pixels, which subsequently rescale to meet the 32x32 requirement. The accuracy and misclassification (via error matrix) provide a score to assess effectiveness. We assess the learning parameters and sample sizes [32] using error and accuracy values per training epoch for both validation and training subsets.

2.3. Traditional Statistical Machine Learning Approaches

To compare the effectiveness of deep learning, we solve the tabular equivalent of the firmware in pixel format but applying tree-based methods [33]. These methods such as decision trees and random forests offer robust interpretability for why they may assign a class label to the malicious firmware. The choice between accuracy, speed, and explainability thus provides additional model tradeoffs and focuses future avenues for investigation. For example, particularly appealing output from tree-based methods includes the assignment of variable or feature importance in an automated way; among the 1024 bytes in the firmware’s header, the method can extract the key positional bytes that signal a possible malicious operation.

	benignware	hackware	malware
pixel265	10.07	11.52	12.81
pixel633	10.48	9.01	12.78
pixel725	7.45	10.43	12.37
pixel685	8.69	9.67	11.47
pixel597	8.19	7.94	11.13
pixel781	6.09	7.21	10.94
pixel329	11.01	11.21	10.93
pixel275	8.26	7.97	10.61
pixel653	7.52	9.30	10.39
pixel621	8.34	8.80	10.30

Figure 2. Most Determinant Byte (or Pixel) Positions for Firmware Class Assignment using Random Forest

3. RESULTS

3.1. Transfer Deep Learning

Table 1 shows the accuracy for class determinations for the small (96x96) and large (224x224) images when custom training the MobileNetV2 architecture. The choice of small images (which are rescaled from the original 32x32) accommodates cameras for embedded systems such as Arduino BLE Sense micro-controllers. The accuracy for a class decision approaches 100% for the larger images and suggests the ELF headers provide a sufficiently rich pattern in the first 1024 bytes to assign a risk factor to each firmware binary.

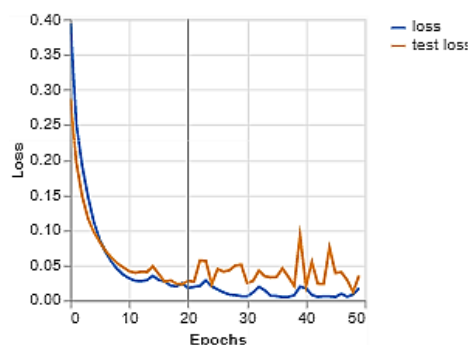


Figure 3. Learning Loss Rates over Time in Epochs

The training time versus accuracy (entropy loss) is shown in Figure 3. After 10 epochs, the network has effectively reached its plateau both for training and validation subsets. The execution times for this style of MobileNetV2 approaches real-time (equivalent to 30 frames per second), such that the overall processing for validating firmware might be limited only by the time to read the first 1024 bytes and flatten them to a decimal equivalent in pictures.

Table 1. Accuracy Results for Four Class MobileNetV2		
Class	Lg. Accuracy (Test Samples)	Sm. Accuracy (Test Samples)
Benign-ware	0.98 (451)	0.98 (451)
Malware	1.00 (107)	0.94 (107)
Hack-ware	1.00 (16)	0.81 (16)
Unknown	1.00 (101)	1.00 (101)

3.2. Single Decision Tree

Figure 4 shows a single decision tree based on considering all 1024 pixel values but splitting firmware class determinations based on ranges of grayscale (or decimal-byte conversions) in the ELF header. Using a subset (4%) of the full training dataset and further holding out a 15% test dataset for evaluation, the decision tree method achieves 95.9% accuracy (4.1% error) across all three classes (benignware, hackware, and malware). This result is competitive with the deep learning approach (99+% accuracy, Table 1). Single trees offer the additional advantage of easier interpretability. One can, for instance, envision a simple algorithm for detecting malicious firmware by examining the decimal conversion of selected key binary bytes in the ELF header. Figure 4 shows the most important 10 bytes as positions at 1285, 377, 298, and so forth. A shortcoming of this approach for single decision trees, however, stems from their brittleness, particularly when applied to test data outside of the narrow training threshold. If an attacker discovers the key 10 bytes for this method to assign a malware or hackware class to the binary, then the decision tree suffers from the same fragility as hash-based or signature methods. A single-byte change can render the detector ineffective.

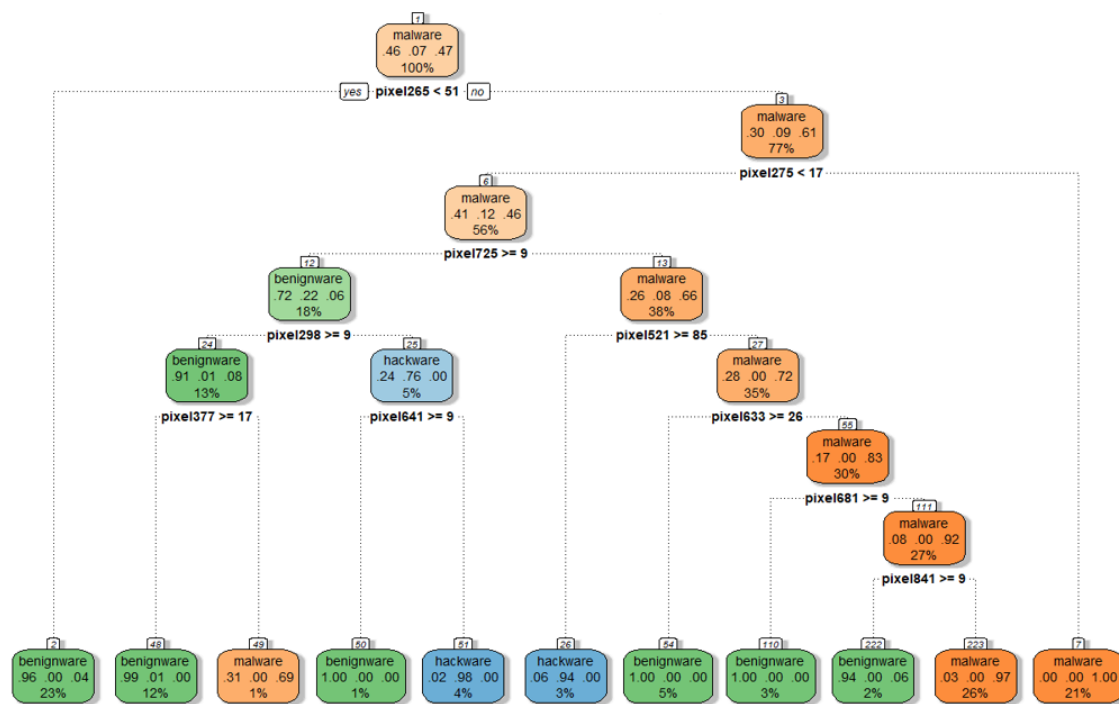


Figure 4. Single Decision Tree Applied to Firmware ELF Bytes

3.3. Multiple Decision Trees, or Random Forest

To investigate the robustness of statistical methods compared to deep learning, Figure 5 illustrates the application of a random forest [33]. Compared to Table 1 for CNNs, the random forest achieves 100% class accuracy. The circular plot in Figure 5 is much denser with decision branches than the single tree shown in Figure 4. Starting in the center of the plot, decision branches for (yes-no) choices span out until a labeled class can be identified by the outer (colored) tags. The resulting high accuracy model combines an ensemble of 500 such trees to render a perfect classification for withheld testing data. The approach of combining many (often weaker) learners to render an ensembled strong learner is well-known for its enhanced robustness

and ability to generalize better than single trees when confronted with out-of-band or under-represented data.

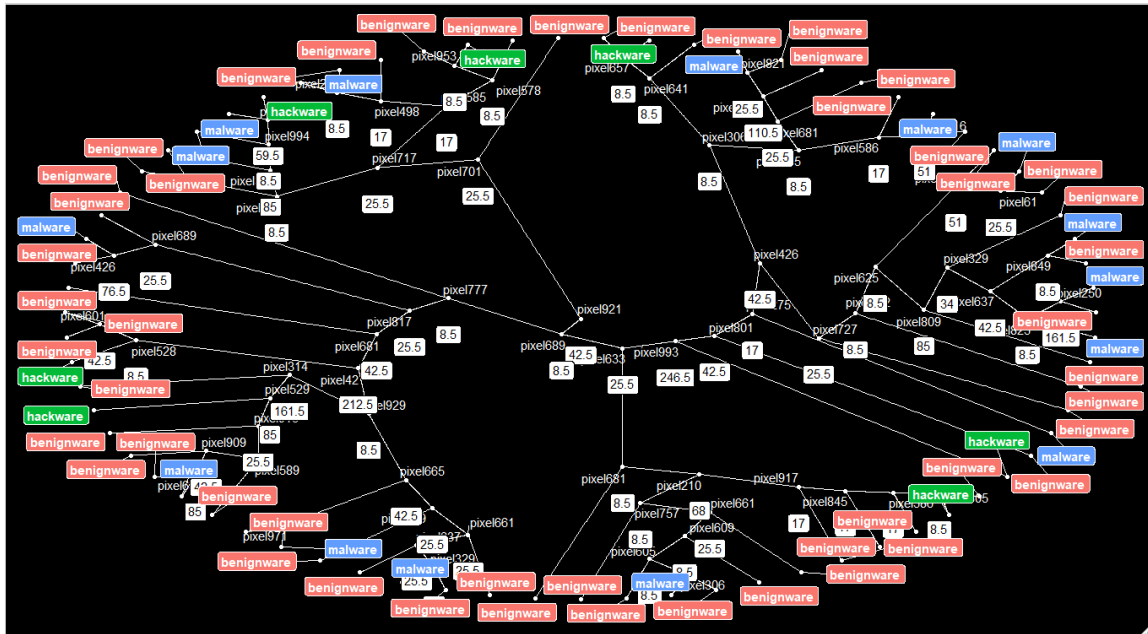


Figure 5. Random Forest (Tree 1) For Firmware Class

4. DISCUSSION AND CONCLUSIONS

By applying deep (transfer) learning to converted images of firmware headers, an optimized neural network can classify malicious Executable and Linkable Files (ELF). The small (32x32) grayscale images match with a decimal conversion (0-15) of the raw binary and then are scaled to a wider (0-255) pixel value range. Each pixel represents a byte in order and the underlying pattern of malicious behavior appears across the file and image nomenclature [22] for architecture, compiler, program name, etc. A procedure to under-sample the benign firmware better rebalances the dataset but leaves between 100-3000 images per class. This number of representative samples has previously been shown to be sufficient, particularly when not training the network from scratch but inherited the weighted features from a previous run on unrelated classes (transfer-learning) [32]. Future work can apply the large research efforts of MNIST derivatives to this firmware classification including simpler or more easily explainable algorithms that are tree-based methods. The research highlights an accurate tree-based method that offers additional interpretability advantages and suggests new ways to apply “if-then” filtering to ELF binaries before firmware updates.

ACKNOWLEDGMENTS

The authors would like to thank the PeopleTec Technical Fellows program for its encouragement and project assistance.

REFERENCES

- [1] LeCun, Yann, Corinna Cortes, and C. J. Burges. "MNIST handwritten digit database." (2010): 18.<http://yann.lecun.com/exdb/mnist/> and Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner.

- "Gradient-based learning applied to document recognition." *Proceedings of the IEEE*, 86(11):2278-2324, November 1998
- [2] Cohen, Gregory, Saeed Afshar, Jonathan Tapson, and Andre Van Schaik. "EMNIST: Extending MNIST to handwritten letters." In *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 2921-2926. IEEE, 2017.
 - [3] CV Online (accessed 01/2021) <http://homepages.inf.ed.ac.uk/rbf/CVonline/Imagedbase.htm>
 - [4] Google Scholar search (accessed 01/2021), <https://scholar.google.com/scholar?q=mnist> and https://trends.google.com/trends/explore?date=all&q=mnist,ImageNet,%2Fg%2F11gfhw_78y
 - [5] Chen, Li, Song Wang, Wei Fan, Jun Sun, and Satoshi Naoi. "Beyond human recognition: A CNN-based framework for handwritten character recognition." In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 695-699. IEEE, 2015.
 - [6] Image Classification on MNIST, (accessed 01/2021), <https://paperswithcode.com/sota/image-classification-on-mnist>
 - [7] Grim, Jiri, and Petr Somol. "A Statistical Review of the MNIST Benchmark Data Problem." <http://library.utia.cas.cz/separaty/2018/RO/grim-0497831.pdf>
 - [8] Schott, Lukas, Jonas Rauber, Matthias Bethge, and Wieland Brendel. "Towards the first adversarially robust neural network model on MNIST." *arXiv preprint arXiv:1805.09190* (2018).
 - [9] Cheng, Keyang, Rabia Tahir, LubambaKasangu Eric, and Maozhen Li. "An analysis of generative adversarial networks and variants for image synthesis on MNIST dataset." *Multimedia Tools and Applications* 79, no. 19 (2020): 13725-13752.
 - [10] Preda, Gabriel, Chinese MNIST: Chinese Numbers Handwritten Characters Images, (accessed 01/2021) <https://www.kaggle.com/gpreda/chinese-mnist>
 - [11] CoMNIST: Cyrillic-oriented MNIST, A Dataset of Latin and Cyrillic Letters, (accessed 01/2021) <https://www.kaggle.com/gregvial/comnist>
 - [12] Prabhu, Vinay Uday. "Kannada-MNIST: A new handwritten digits dataset for the Kannada language." *arXiv preprint arXiv:1908.01242* (2019). <https://www.kaggle.com/higgstachyon/kannada-mnist>
 - [13] Noever, David, Noever, Sam E.M. "Expressive Multimodal Integrated Learning (EMIL): A New Dataset for Multi-Sense Integration and Training", 2020 Southern Data Science Conference, August 12-14 2020, Atlanta, GA (poster) and Sign Language MNIST: Drop-In Replacement for MNIST for Hand Gesture Recognition Tasks, <https://www.kaggle.com/datamunge/sign-language-mnist>
 - [14] Mader, K Scott, Skin Cancer MNIST: HAM 10000, A Large Collection of Multi-Source Dermatoscopic Images of Pigmented Lesions, (accessed 01/2021) <https://www.kaggle.com/kmader/skin-cancer-mnist-ham10000>
 - [15] Xiao, Han, Kashif Rasul, and Roland Vollgraf. "Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms." *arXiv preprint arXiv:1708.07747* (2017). <https://www.kaggle.com/zalando-research/fashionmnist> See also Fashion-MNIST, (accessed 01/2021), <https://paperswithcode.com/sota/image-classification-on-fashion-mnist> and <https://github.com/zalando-research/fashion-mnist>
 - [16] Lu, Arlen, "Convert-own-data-to-MNIST-format" (accessed 01/2021) <https://github.com/Arlen0615/Convert-own-data-to-MNIST-format>
 - [17] Noever, D., & Noever, S. E. M. (2021). Overhead MNIST: A benchmark satellite dataset. *arXiv preprint arXiv:2102.04266*. <https://www.kaggle.com/datamunge/overheadmnist> and Github, <https://github.com/reveondivad/ov-mnist>
 - [18] Noever, D., & Noever, S. E. M. (2021). Virus-MNIST: A Benchmark Malware Dataset. *arXiv preprint arXiv:2103.00602*.
 - [19] Noever, D. A., & Noever, S. E. M. (2021). Image Classifiers for Network Intrusions. *arXiv preprint arXiv:2103.07765*.
 - [20] Freitas, S., Duggal, R., & Chau, D. H. (2021). MalNet: A Large-Scale Cybersecurity Image Database of Malicious Software. *arXiv preprint arXiv:2102.01072*
 - [21] Chen, L., Sahita, R., Parikh, J., Marino, M. (2020). STAMINA: Scalable Deep Learning Approach for Malware Classification, <https://www.intel.com/content/dam/www/public/us/en/ai/documents/stamina-scalable-deep-learning-whitepaper.pdf>
 - [22] Partush, N. (2021). Labeled-Elfs, <https://github.com/nimrodpar/Labeled-Elfs>

- [23] Kairajärvi, S., Costin, A., &Hämäläinen, T. (2019). Towards usable automated detection of CPUarchitecture and endianness for arbitrary binary files and object code sequences. arXiv preprint arXiv:1908.05459.
- [24] Clemens, J. (2015). Automatic classification of object code using machine learning. *Digital Investigation*, 14, S156-S162.
- [25] Xie, H., Abdullah, A., &Sulaiman, R. (2013). Byte frequency analysis descriptor with spatial information for file fragment classification. In *Proceeding of the International Conference on Artificial Intelligence in Computer Science and ICT (AICS 2013)*.
- [26] Constantin, L. (2015). Cisco warns customers about attacks installing rogue firmware on networking gear, Network World. Aug 10, 2015. <https://www.networkworld.com/article/2970954/cisco-warns-customers-about-attacks-installing-rogue-firmware-on-networking-gear.html>
- [27] Microsoft Threat Intelligence Center(2019). Corporate IoT – a path to intrusion. <https://msrc-blog.microsoft.com/2019/08/05/corporate-iot-a-path-to-intrusion/>
- [28] Greenberg, A. (2019). The Untold Story of the 2018 Olympics Cyberattack, the Most Deceptive Hack in History. Wired Magazine. <https://www.wired.com/story/untold-story-2018-olympics-destroyer-cyberattack/>
- [29] Noever, D. (2021). IoT Firmware Image Classifier: Rendered ELF Binaries by Class as Malware, Kaggle. <https://www.kaggle.com/datamunge/iot-firmware-image-classification>
- [30] Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. "Mobilenetv2: Inverted residuals and linear bottlenecks." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510-4520. 2018.
- [31] Salehi, Sohail. *ImageMagick Tricks*. Packt publishing ltd, 2006.
- [32] Warden, P. "How many images do you need to train a neural network?" (2017).<https://petewarden.com/2017/12/14/how-many-images-do-you-need-to-train-a-neural-network/>
- [33] Morales-Molina, C. D., Santamaria-Guerrero, D., Sanchez-Perez, G., Perez-Meana, H., & Hernandez-Suarez, A. (2018, November). Methodology for malware classification using a random forest classifier. In 2018 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC) (pp. 1-6). IEEE.

LIMITING FACTORS IN WIDESPREAD ADOPTION OF ACTIVE QUEUE MANAGEMENT IN THE PHILIPPINES' CONSUMER ELECTRONICS SPACE

Min Guk I. Chi

Bachelor of Business Administration, S P Jain School of Global Management

ABSTRACT

The premise that Active Queue Management (AQM) is effective in both quantitative and qualitative settings in residential and enterprise networks has repeatedly been established in multiple papers from academic journals along with private studies in addressing bufferbloat, characterized as excessive latency because of heavy network utilization. However, the presence and understanding of bufferbloat mitigation is absent and not well-known in the Philippine Internet of Things space except enthusiasts, willing to take the time to examine the concept along with its benefits. Hence, this paper examines possible reasons as to why AQM is not widely adopted by Philippine consumers and industries in increasing productivity considering the COVID-19 Pandemic: a lack of basic understanding of bufferbloat and its implications, the complexity of the concept, the know-how required to execute its implementation being far too high, and the lack of perceived benefit by existing telecommunications players in the country.

KEYWORDS

Active Queue Management, Consumer Adoption, COVID-19, Bufferbloat.

1. INTRODUCTION

In an increasingly digital world, a strong and robust internet infrastructure is paramount; this is more so considering the context in which this paper was made: during the Severe Acute Respiratory Syndrome — Coronavirus 2 pandemic, colloquially known as COVID-19. With major events around the world being moved to a virtual medium considering the virus spreading through respiratory droplets, the internet is increasingly utilized to compensate for productivity in many fields, including but not limited to the academe and commercial — events that generally can be held from the comfort of an individual's home. Hence, the need for a robust internet is essential since any further disruptions will increase the losses of productivity that have been incurred due to the global pandemic.

This premise is given weight thanks to the medium of these events: video conferencing applications such as Zoom have risen to prominence thanks to the need for virtually distant conferences. Considering this, video conferencing is a latency-sensitive application which requires that the latency of the internet is kept at a minimum to avoid video and audio degradation. Additionally, latency-sensitive activities such as Voice over IP (VoIP), Video Streaming, and Low Latency Online Gaming are some of the other examples where sudden increases in latency prove significantly detrimental. This phenomenon in internet networks is

known as bufferbloat; according to DSLReports, this is characterized as “the undesirable latency caused by routers and cable/DSL modems buffering more data than necessary.” [5]

One of the mitigations that is present thanks to the Institute of Electrical and Electronics Engineers (IEEE) is Active Queue Management (AQM), characterized as the management of data packets via proactively dropping packets before it exceeds the buffer, preventing excessive latency thanks to heavy load. Therefore, this study seeks to examine the reasons as to why AQM is noticeably absent in the Internet of Things: consumer electronics space despite the clear benefits of its applications in existing network infrastructure.

With that said, the primary contributions of this paper are as stated:

- i. First, it provides an analysis of consumer preferences in the Philippines in terms of price, purchase behavior, and factors that may consider a purchase of a third-party router in order to solve their problem of bufferbloat.
- ii. Second, it shows the current level of understanding of Filipinos on what bufferbloat is and what are the effective mitigations in solving such a problem, or lack thereof.
- iii. Third, it provides possible opportunities into hastening the adoption of Active Queue Management strictly in terms of a business perspective along with causing disruption in the Philippine space.
 - a. By extension, it provides possible product offerings of third-party routers at reasonable price points in the context of the Philippines.
- iv. Lastly, the paper is an attempt into bridging the concept of Active Queue Management into the applied space. More specifically, Active Queue Management being given in the hands of consumers and allowing for a better internet experience overall, thus consumers staying relevant in the Fourth Industrial Revolution.

2. REVIEW OF RELATED WORK

2.1. Robust Telecommunications Industry

According to Chi (2020) [1], the paper asserts that modern telecommunications are essential in today’s technologically adept population, along with the premise that higher internet speeds are positively correlated with the economic status of a country. With these two statements, this gives a clear foundation for the basis of the introduction — a robust infrastructure being a must. Additionally, on the same paper, it asserts that the improvement of telecommunications is a must to provide better opportunities for all Filipinos who increasingly rely on access to the internet in order to have a chance at improving their quality of life; with this, further pretext is given to the necessity of a strong base for the telecommunications industry. Without a strong industry, the mitigations given to solve bufferbloat will have minimal impact if the internet constantly goes out thanks to poor reliability from internet providers. Furthermore, the impact of these anti-bufferbloat mitigations is also lost if the power grid infrastructure is not fully developed. Power outages will cause a loss of internet given the logical premise that these are connected to electricity. Hence, in conclusion, this source proves relevant given the foundation it provides for the requirements needed to give AQM a consistently meaningful impact.

2.2. Active Queue Management

The concept of Active Queue Management in and of itself must be well-defined given that said concept is the core foundation on which this research lies. For instance, the concept was formalized into actual use by the Institute of Electronics and Electrical Engineers in 1993

according to Adams (2013)[12]; from this, conclusions can be inferred. One, with respect to the context in which AQM was appearing in research papers, it can be opined that signs for the internet were already on the point for civilian use. Coincidentally, this was also the time where Berners-Lee introduced the concept of the World Wide Web (WWW): a way for the internet to be bridged from military use to general civilian use. Hence, with the increase in data packets moving because of the increased reach of civilians using the internet, the problem of buffers being full quickly due to said increase in data packets leading to internet degradation would arise. Therefore, because of this, AQM models have been examined by multiple organizations seeking to solve the problem.

Initial frameworks that involved the AQM system during the 1990s-2000s were the Tail Drop and the Random Early Detection algorithms. These are characterized as two methods in which packets that are incoming whilst the buffer is full is effectively dropped. This means that the data incoming will not be accepted into the current buffer until the buffer can allocate can process the existing data packets. The tail-drop method has its limitations: for instance, this method is a passive method of AQM. Based on a previous lecture on internet protocols from NC State University in 2014, "Tail drop is a passive queue management algorithm. In this algorithm, the traffic is not differentiated, and each packet is at the same priority. Also, the main consideration here is the maximum queue length at each router and its services using the first in first out algorithm." This effectively does not differentiate between latency-sensitive packets and only adjusts the buffer once it's full. Hence, it has a significant shortcoming since it does not factor in differing types of packets that require a higher priority than others [9]. According to a paper published by the IEEE in 2002 [3], Random Early detection is inherently limited considering that RED is inflexible with respect to the paper mentioning "A drawback in deploying RED stems from its apparent tuning difficulties. As we now show, we believe this difficulty stems in large part to RED's use of average queue length." Since it uses an average, limitations inherent to using an average are applicable such as susceptibility to extremes and inaccuracies in data throughput to make the RED algorithm tailor-made to the concerned network.

More developed frameworks have been developed to compensate for the weaknesses of the initial algorithms as mentioned above. For instance, the development of Fair Queue — Controlled Delay (FQ — CoDel) by Jacobson and Nichols in 2012 non-exhaustively improves on the initial frameworks of Tail-Drop and RED on the following [10]:

1. It is parameter-deficient, meaning that it is inherently easier to configure considering networks with dynamic throughput.
2. CoDel also can determine certain types of traffic: those that cause bufferbloat, and those that don't. a. Those that don't are effectively ignored by the AQM algorithm, whilst those that are will be subjected to the algorithm to minimize delay as much as possible.
3. Implementation is simple, so it can be utilized in consumer-grade products and in high-end networking hardware.

Given the effectiveness of FQ — CoDel in reducing bufferbloat, a study has been made to examine the feasibility of implementation into network equipment for commercial use. A study conducted by White and Rice in 2013 [16] has shown significant benefits when correctly implementing the use of CoDEL and other AQM implementations such as Proportional-Integral Controller, that was mentioned above, and Stochastic Fair Queuing with CoDel. Surprisingly, AQM is already present in existing cable modems as of the time of this research paper and was just not implemented by operators; when configured optimally, it has been quantitatively and qualitatively shown that improvements in latency are significant. This, however, is only limited to upload speeds and does not examine the download side, what many households use. Hence,

this paper is valuable in proving that the theoretical algorithms do make a meaningful difference in internet experience.

Over the years, research has been made to see which AQM algorithm is the most optimized for network traffic. With the current AQM implementations Stochastic Fair Queuing was the best algorithm for implementation of reducing excessive delay when it comes to heavy network utilization [7]. This is relevant considering that there have been many AQM implementations that have been proposed by researchers across the world, seeking to solve the issue of bufferbloat.

However, while attempting to find the best AQM algorithm available, there has been a clear lack of standardization of metrics to quantitatively measure the results of researchers' algorithms to determine what algorithm is best all-around in order to see which implementation of AQM should be actively used in the real-world. The general metrics that developers of future AQM algorithms should use to benchmark performance. Specifically, they recommend merging the processes of Analytic Hierarchy Process (AHP) along with Technique in Order of Preference by Similarity to the Ideal Solution (TOPSIS) to create a new benchmarking standard that is ideal for new AQM algorithms to be tested against: promoting robustness and quality (Khatari et.al., 2019) [7].

Perhaps another clear application of the effectiveness of AQM is in big data: significant amounts of data running through a server at once. Given that big data requires quick execution of data processing with mass inputs into meaningful outputs to be of use to experts, it also requires that the connection to the internet for this is responsive (i.e., latency-free). Hence, an examination was made to see whether AQM would provide a tangible benefit in Hadoop clusters and in the MapReduce Programming Model. Given the analysis of the research that has been done, it has also shown that bufferbloat can be reduced significantly, by 85% whilst only increasing the Hadoop execution time using the MapReduce Model by 5%. However, it strongly cautions that getting the configuration right is the only way to achieve optimal results, as poor configurations lead to increases in Hadoop executions and non-ideal reductions in bufferbloat. Hence, it can be concluded that AQM is clearly scalable in many settings, from the confines of one's home to the large data processing units that are used by technology companies to process large amounts of information. [13]

As of today, in open-source software, AQM has more developments. In OpenWrt, an open-source Linux base for networking hardware such as routers and wireless extenders, the implementation of Cake (Common Applications Kept Enhanced) — from the Bufferbloat community [2] — has provided for further improvements in solving this problem. This discipline for queuing network packets considers AQM as only one of the measures necessary to address bufferbloat. Cake includes the following:

- 1) Traffic Shaper
- 2) Priority Queue
- 3) Flow Isolation
- 4) AQM
- 5) Packet Management

All these factors lead to reduced latency in general internet use, especially in latency-sensitive applications and has factored in additional information that makes the difference between a low-latency internet and one that is significantly crippled by bufferbloat. Additionally, cake is superior to CoDel, non-exhaustively, in a couple of ways:

- 1) Command Line interface is simpler than CoDel.
- 2) Reduced CPU Load thanks to an integral shaper.
- 3) Explicit Congestion Notification (ECN) is always on, avoiding false positives.
- 4) Ease of availability; a. Since it is on Linux, an open-source platform, it is easily available with commands understood on the platform.

With all of this, it shows the progression of AQM into actual practice and is relevant to gain a better understanding of the merits of applying this into the consumer market, especially considering the COVID-19 pandemic, and its impacts still forcing many events and industries to go virtual, as any sign of latency means missed productivity which often proves to be major.

2.3. Consumer Behavior

For instance, Consumer Behavior: 11th Edition [14] has relevant concepts that are of value. When it comes to effectively appealing to consumers, there are four elements to properly understand consumer behavior: Motives, Cues, Responses, and Reinforcement. These are characterized as:

1. Motives: The incentive behind doing something.
2. Cues: The mechanism in which consumers will know that a certain product/service is what they need (e.g., Marketing, Advertising).
3. Responses: How the consumer reacts to the former two factors.
4. Reinforcement: The way in which consumers are solidified towards believing a certain view.

In terms of how to create effective advertising, there are many effective avenues: comparative advertising, appeal to humor, fear, and sexual appeal. These avenues are defined as follows:

1. Comparative Advertising: By making claims that one's product is superior to the competition, this appeals to consumers since they only want what's best for them.
2. Appeal to Humor: Using advertising to induce laughter, it promotes a positive brand image and leaves a long-term impact.
3. Appeal to Fear: By presenting a clear threat that consumers should be fearful of, and providing a solution to said fears, it promotes consumers to buy your products for a sense of safety.
4. Sexual Appeal: Through this method, consumers are captivated by an ideal human image. This, however, requires careful execution. Else, consumers will only be captivated on the model and not the product.

Extending on consumer behavior, another research paper proves its value given the recency of the paper. A paper made by Moon, Choe, and Song in 2021 [8] describes consumer behavior in South Korea considering the COVID-19 pandemic; most respondents would prefer to acquire their goods online rather than getting goods in person considering personal safety. With this, it can be argued via the use of the Protection-Motivation Theory; this is simply explained as protection being the first reason for their actions and with that framework in mind, people will gravitate their decisions for the sake of protection.

Another relevant part of a business is pricing and how deciding how much or little a product or service is priced has a meaningful impact on whether a consumer would purchase said product or service. Given a research paper by Quan, Quan, and Wang in 2019 [11], consumers' expectations play a factor as to influencing pricing. On one hand, if consumers' expectations are very high for the product in question and prioritize psychological satisfaction, the price is effectively damaged

and the same goes with profits. This is because a higher standard is required — and this compels the seller to reduce prices to not disappoint them. On the other hand, if a consumer is only expecting the bare minimum and does not prioritize their happiness, prices can rise and profits by extension. This is due to their perspective on only focusing with the product in question getting the job done. Given that one of the goals of this paper is to create a potentially disruptive business, the goal is to reach as many customers as possible.

Lastly, of relevance is consideration for how consumers are influenced in terms of their purchases specifically with Internet of Things devices. There are certain factors that play a larger role in a consumer as to whether they will buy a product under this category i.e., most important is trust. When a consumer has confidence in a brand's products and services through superior customer experience, perceived ease of use, and a proven track record, adoption is significantly hastened. On the other, factors such as social influence do play a part though to a lesser extent against the factors (Tsourela & Nerantzaki, 2020). [15]

3. RESEARCH FRAMEWORK

With the available literature, the concept of Active Queue Management itself is only available with no research papers on examining the actual limitations of full market adoption; hence, this warrants the use of an individual framework unique to this research to fully ascertain the reasons for the lack of market adoption. Given that the market is the body that is directly concerned with the apparent lack of adoption, marketing fundamentals make the most sense to apply as a theoretical framework.

Regarding the use of marketing fundamentals, parts of a market plan — according to Pearson's 17th Edition of Principles of Marketing [6] — make most sense in applying here; for instance, parts of this paper can already be used for the current marketing situation such as the market description, existing products, and competition. Marketing actions will be manifested as recommendations. Essentially speaking, this is a market reach problem, basics in marketing are used for most of the paper to solve the question of the lack of market adoption on the internet of things — consumer electronics space.

In this case, the foundational framework that is most applicable will be thematic analysis. Considering the nature of the research question which requires that factors be fleshed out to sufficiently answer the question, dividing the independent variables (factors) into themes will be of use to the fulfillment of this research paper. Hence, there will be no use of hypothesis testing when it comes to the information that has been obtained along with corresponding analysis and discussion as a result.

Consequently, given no hypothesis testing, what the data will be used for is to determine the extent of the veracity of the hypotheses that will be mentioned in the latter part of this paper. In other words, the goal is to determine to what extent the hypotheses are true in the context of the data that has been obtained, and whether said data confirms or rejects the possible factors that answer the research question at hand.

4. METHODOLOGY

The primary mechanism to successfully answer the research question will primarily be done through data collection; through the input of respondents, information relevant to the research objectives will be revealed via a survey questionnaire. A secondary method will be via the use of existing literature from relevant and quality sources.

Respondent Profile & Sample

Information obtained from respondents such as age, employment, income bracket, will provide relevant information as to how they are influenced in their purchase decisions and their preferences in the context of this study.

For the purposes of this study, a sample of 200 respondents will be obtained from multiple areas of the country (in Luzon, Visayas, and Mindanao) to provide a diverse picture of their perceived wants to promote an ideal internet experience.

Data Collection Method

The variables in this paper will be ascertained via an online survey which is distributed to respondents. A prototype of the survey has already been made to ensure that the base version is fully functional; hence, a final version was made with questions that have been tweaked to be more relevant to the study. Consequently, the sampling technique that was used for this paper is non-probability based; in particular, convenience and snowball sampling are the prevalent mechanisms used.

Questionnaire& Tools

Survey questions are segmented into five parts which contain questions that ask for certain information; the Likert scale was often used to determine the respondents' inclinations to agreeing or disagreeing to certain statements. Additionally, open-ended questions were used to factor in a variety of responses that respondents may give to certain questions which requires to be flexible about the context of the respondent.

Data that has been obtained through the questionnaire will be analyzed via IBM's SPSS and Microsoft Excel with PH Stat 4.1 in order to tabulate, organize, and analyze the data; the primary mechanism of analysis is through the cluster method: done in order to create a customer profile; three segments have been made for the sake of the objectives mentioned above. Standard statistical treatment will also be utilized: Sample Mean, Median, Mode, Standard Deviation; this will be relevant in analysis of the data that has been given by respondents and will promote a representative picture of the market to see whether or not a business undertaking is feasible. For the purposes of keeping the paper concise, only the customer profiles from the cross-tabulation will be shown. All other data will be available for viewing in the original paper.

5. RESULTS

Segments	Characteristics &Needs	Solution
Budget Consumer	a. A Monthly Income of Minimum Wage Up Until Less Than ₱22,000. b. Spending ₱500 - ₱2,000 for Internet Usage Monthly. c. Knows Little to Moderate Knowledge on Bufferbloat, And. d. Prefers Buying Technology Products	A Budget Router with a Low Production Cost and Low Sale Cost is Attractive. <i>Must be Cheap but Robust in Terms of Build Quality.</i> Low Specifications: Basic Single-Core, Low Frequency CPU 8 MB of Flash 64 MB of RAM

	Physically.	Providing Basic Wi-Fi Standard (802.11ac); Wireless Protected Access 2 (WPA 2) Low-Powered Antennas: Maximum of 20 decibel-milliwatts (dBm)
Mid-Range Consumer	<ul style="list-style-type: none"> a. A Monthly Income of ₱22,000 Up Until Less Than ₱132,000. b. Spending ₱2,000 - ₱5,000+ for Internet Usage Monthly. c. Knows Little to Moderate Knowledge on Bufferbloat, And. d. Prefers Buying Technology Products Physically. 	<p>A Mid-Range Router with Slightly Higher Production Cost and Increased Price is Attractive: <i>An Increased Budget for Specifications and Better Build Design.</i></p> <p>Medium Specifications: Single Core, Higher Frequency CPU 8-16 MB Flash 64-128 MB RAM</p> <p>Providing Slightly Better Wi-Fi Standard (802.11ac): Wireless Protected Access 2 (WPA 2) Medium Powered Antennas Maximum of 25 decibel-milliwatts (dBm)</p>
High-End Consumer	<ul style="list-style-type: none"> a. A Monthly Income of ₱132,000 Up Until ₱220,000+. b. Spending ₱2,000 - ₱5,000+ for Internet Usage Monthly. c. Knows Little to Moderate Knowledge on Bufferbloat, And. d. Prefers Buying Technology Products Physically. 	<p>A Flagship Router with the Best Specifications that are Currently Available: <i>Significantly Increased Specifications and Build Quality:</i></p> <p>Single Core, Even Higher Frequency to Dual Core CPU. 16-32 MB Flash 128 MB RAM +</p> <p>Latest Wi-Fi Standard (802.11ax): Wireless Protected Access 3 (WPA 3) High Powered Antennas: Maximum of 30 decibel-milliwatts (dBm)</p>

6. CONCLUSIONS

Given the objectives of the research paper at hand, most respondents are unable to find a meaningful answer to bufferbloat and thus provides significant demand for a networking equipment solution at the consumer level. Though reading this paper, one can understand from a sample of respondents in the Philippines that Filipinos are clearly frustrated with their experience when it comes to their internet and there is clearly a prospective market given the significant perceived value for security, latency management, service, customizability, and differentiation.

A guideline has been made for the Philippine context when it comes to the specifications given their income levels and cost of internet use monthly. As mentioned in the literature review, there are substantial quantitative and qualitative benefits when it comes to an optimal implementation

of Active Queue Management. Should this be implemented on a commercial scale for consumers, this would be of major benefit to consumers which will increase productivity during this pandemic — especially when productivity is already compromised.

Consequently, when the internet infrastructure proves to be robust and extremely responsive, this further allows for the exchange of information on a faster rate. Although the effects of anti-bufferbloat measures may be not apparent from an individual standpoint immediately, the benefits accumulate over time in the form of saved time, increased productivity, and increased research into the latest standards of networking given the widespread implementation and its clear merits. This improves the sharing of information on a great scale, improving the well-being of a country. By extension, keeping said country competitive in the Fourth Industrial Revolution.

REFERENCES

- [1] Arora, N., & Singh, G. (2015). Practical Appraisal of Distinguish Active Queue Management Algorithms. *International Journal of Computer Science and Mobile Computing*, 496-505.
- [2] Bufferbloat.net. (n.d.). Cake - Common Applications Kept Enhanced. Retrieved March 8, 2021, from Bufferbloat.net: <https://www.bufferbloat.net/projects/codel/wiki/Cake/>
- [3] C. V. Hollot, V. Misra, D. Towsley and Weibo Gong, "Analysis and design of controllers for AQM routers supporting TCP flows," in *IEEE Transactions on Automatic Control*, vol. 47, no. 6, pp. 945-959, June 2002, doi: 10.1109/TAC.2002.1008360.
- [4] Chi, M G. I. (2020). A Study on the Factors Inhibiting High Speed Internet in the Philippines. Retrieved March 8, 2021, from Academia: https://www.academia.edu/44296610/A_Study_on_the_Factors_Inhibiting_High_Speed_Internet_in_the_Philippines
- [5] DSLReports. (n.d.). Bufferbloat. DSLReports. Retrieved April 14, 2021, from <http://www.dslreports.com/faq/17883>
- [6] Kotler, P., & Armstrong, G. (2018). *Principles of Marketing 17th Edition*. In P. Kotler, & G. Armstrong, *Principles of Marketing 17th Edition* (pp. 603-612). New York: Pearson.
- [7] Khatari, M., Zaidan, A., Zaidan, B., Albahri, O., & Alsalem, M. (2019). Multi-Criteria Evaluation and Benchmarking for Active Queue Management Methods: Open Issues, Challenges and Recommended Pathway Solutions. *International Journal of Information Technology and Decision Making*, 1187-1242.
- [8] Moon, J.; Choe, Y.; Song, H. Determinants of Consumers' Online/Offline Shopping Behaviours during the COVID-19 Pandemic. *Int.J. Environ. Res. Public Health* 2021,18,1593. <https://doi.org/10.3390/ijerph18041593>
- [9] Mulky, E., Jain, P., Bhatia, S., Dash, S., & Dutta, R. (2012). Tail Drop Algorithm. Retrieved April 19, 2021, from <https://sites.google.com/a/ncsu.edu/tail-drop-vs-red/plan-of-work/tail-drop-algorithm>
- [10] Nichols, K., Jacobson, V. (6 May 2012). "Controlling Queue Delay". *ACM Queue*. ACM Publishing. doi:10.1145/2209249.2209264. Retrieved March 8, 2021.
- [11] Quan, J., Wang, X., & Quan, Y. (2019). Effects of Consumers' Strategic Behavior and Psychological Satisfaction on the Retailer's Pricing and Inventory Decisions. *IEEE Access*, Access, IEEE, 7, 178779–178787. <https://doi.org/10.1109/ACCESS.2019.2958685>
- [12] R. Adams, "Active Queue Management: A Survey," in *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1425-1476, Third Quarter 2013, doi: 10.1109/SURV.2012.082212.00018.
- [13] R. F. E. Silva and P. M. Carpenter, "Controlling Network Latency in Mixed Hadoop Clusters: Do We Need Active Queue Management?" 2016 IEEE 41st Conference on Local Computer Networks (LCN), Dubai, United Arab Emirates, 2016, pp. 415-423, doi: 10.1109/LCN.2016.70.
- [14] Schiffman, L. G., & Wisenblit, J. (2015). *Consumer Behavior: 11th Edition (Eleventh ed.)* (pp. 122-123, 170-183). Pearson.
- [15] Tsourela, M., & Nerantzaki, D.-M. (2020). An Internet of Things (IoT) Acceptance Model. Assessing Consumer's Behavior toward IoT Products and Applications. *Future Internet*, 12(11), 1. <https://doi.org/10.3390/fi12110191>

- [16] White, G., & Rice, D. (2013, April). Cablelabs Access Network Technologies. Retrieved March 5, 2021, from Cablelabs: https://www-res.cablelabs.com/wp-content/uploads/2019/02/28094033/Active_Queue_Management_Algorithms_DOCSIS_3_0.pdf

Authors

Min Guk I. Chi was born in 2001 and graduated High School from Xavier School San Juan in the Philippines. Possessing a love for research, he has successfully completed major research papers for academic requirements independently. Currently, he is taking up 3rd Year Business Administration majoring in entrepreneurship at S P Jain School of Global Management.



© 2021 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.

DEVELOPMENT OF AN AUTISM SCREENING CLASSIFICATION MODEL FOR TODDLERS

Afef Saihi and Hussam Alshraideh

Department of Industrial Engineering,
American University of Sharjah, Sharjah, UAE

ABSTRACT

Autism spectrum disorder ASD is a neurodevelopmental disorder associated with challenges in communication, social interaction, and repetitive behaviors. Getting a clear diagnosis for a child is necessary for starting early intervention and having access to therapy services. However, there are many barriers that hinder the screening of these kids for autism at an early stage which might delay further the access to therapeutic interventions. One promising direction for improving the efficiency and accuracy of ASD detection in toddlers is the use of machine learning techniques to build classifiers that serve the purpose. This paper contributes to this area and uses the data developed by Dr. Fadi Fayed Thabtah to train and test various machine learning classifiers for the early ASD screening. Based on various attributes, three models have been trained and compared which are Decision tree C4.5, Random Forest, and Neural Network. The three models provided very good accuracies based on testing data, however, it is the Neural Network that outperformed the other two models. This work contributes to the early screening of toddlers by helping identify those who have ASD traits and should pursue formal clinical diagnosis.

KEYWORDS

Autism Spectrum Disorder, Screening, Machine Learning, Decision Tree, Random Forest, Neural Network, Classifier, Accuracy.

1. INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental condition that impacts the way a person perceives others and socializes with them. It affects three main developmental areas which are communication, social interaction, and repetitive patterns of behavior. As shown in Figure 1, Autism rates continue to rise dramatically, and according to Center for Disease Control (CDC) reports [1], prevalence rate increases with 1 in 54 children diagnosed with autism in 2020 compared to 1 in 59 in 2018, 1 in 110 in 2006, and 1 in 150 in 2000.

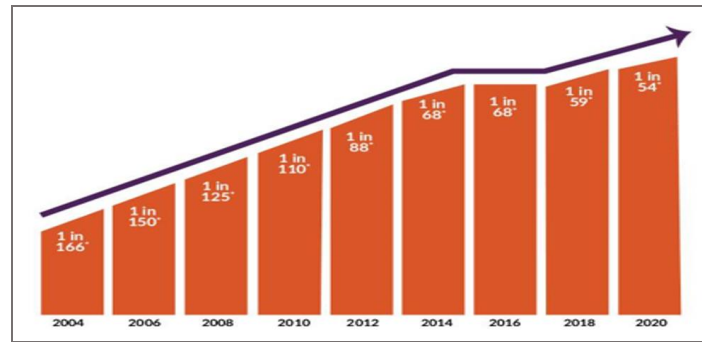


Figure 1. Autism prevalence estimates (Source: CDC reports [1])

Research has shown that there is no cure for ASD, however, early diagnosis and intensive intervention can make a big difference in the lives of many children and their families. Early diagnosis is very important for children on the spectrum because it allows to teach them the skills and behaviors that they lack at an early age when they still have good brain plasticity and therefore, the impact of intervention can be maximized and helps them reach their full potential. Interventions such as special education, behavior modification techniques, speech and occupational therapies help to bridge the gap that ASD kids have compared to their peers and speed up their development [2]. Therefore, early identification and diagnosis of children with ASD is very beneficial for the kids and their families and early screening is always recommended by experts because it allows to detect at an early stage the kids with more risk to be on the autism spectrum disorder, and thus, pushes their families to take the required measures to get the kids access the early intervention services.

Early screening of ASD has an important role in improving prognosis via early diagnosis and intervention [3]. For these kids with autism, both research and practice have shown that there is no magic pill for cure and early intervention through intensive education and behavior modification is the key due to its capability of changing the quality of life of these kids and their families and improving long-term outcomes [4]. However, the standardized tests for diagnosis such as Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview-Revised (ADI-R), among others are time consuming, very costly and can be run only by trained clinicians. Thus, there are many barriers of getting these kids screened for autism at an early stage which contribute to delaying the therapies and interventions that they require. Nowadays, due to the advancement made in artificial intelligence and machine learning, autism can be predicted at an incredibly early stage [5]. Having access to an accurate, automated, cost effective and fast instrument for ASD screening at an early age will be greatly beneficial for detecting autism traits in the kids and securing a much better future for them.

In this work, our aim is to contribute to the early screening of toddlers to help identify the kids who have ASD traits and should pursue formal clinical diagnosis. This is done through the development of a classification model that, based on some defined attributes, will identify the kids who have ASD symptoms and need to undergo further advanced assessments with professionals.

The remainder of this paper contains a literature review section that discusses the related works, a methodology section that describes the data and details the steps to be followed for the analysis, and a model building section, in which, the selected classifiers have been trained and their respective performances have been analyzed and compared.

2. LITERATURE REVIEW

Although concerns about children development milestones are mostly reported by parents of children with ASD within their first year, these children are rarely diagnosed before the age of 4 years [3, 6]. Early identification is key as, if followed by intensive early intervention, will enable these kids to acquire the necessary skills, improve the core behavioral symptoms, and there is a chance that they will grow out of the diagnosis or at least lose many of the autistic traits [7]. A significant number of studies investigated the potential for early intervention in helping kids with this neurodevelopmental condition. In this context, [8] conducted a systematic review that highlighted the large number of studies, which are over 83% of the published literature since 2010, reflecting the increased interest among researchers in this area.

Autism cannot be identified using conventional clinical methods such as blood tests. Instead, ASD screening is a crucial phase, and it is the process of determining the autistic symptoms of an individual. Many screening tools have emerged overtime, they include direct observations, questionnaires and interviews and they should be performed by specialists. However, these tools are lengthy, costly, and low-and-middle-income countries have shortage of mental health clinicians. This constitutes a major barrier for obtaining an early diagnosis and accessing intervention services [9]. Therefore, a viable screening instrument that is cost effective, available and less time consuming to identify the risk of ASD at a preliminary stage is highly needed.

There is a growing interest among researchers in the early screening and intervention for young children at risk of ASD [8] and in the use of machine learning and intelligent methods for autism screening and detection [9, 10]. In this context, [11] investigated fuzzy data mining models to detect autism features for both cases and control groups between 4 and 11 years. [5] developed an effective prediction model by merging Random Forest-CART and Random Forest-Id3 and complemented their work by developing a mobile application based on the proposed model. [12] developed and tested an Artificial Neural Network (ANN) for diagnosing ASD, and the test data evaluation showed that ANN model was able to diagnose ASD with 100% accuracy. [13] used machine learning to investigate the accuracy and reliability of the Q-CHAT method in classifying autistic kids. The authors used three different ML algorithms and found that the Support Vector Machine was the most effective. Similarly, [10] proposed a new machine learning method which is the “Rules-Machine Learning”. This method detects the ASD traits in cases and offers rules that can be used by domain experts to understand the reasons behind the classification. The authors compared this method with other classifiers and found that it leads to higher predictive accuracy, sensitivity, and specificity than those of other models such as bagging, decision trees and rule induction. In line with these studies, this paper contributes to this area by proposing a toddler screening prediction model for ASD traits.

3. METHODOLOGY AND DESCRIPTION OF DATA

The dataset used in this study is obtained from Kaggle website, and it was developed by Dr Fadi Fayed Thabtah [14] to screen autism in toddlers. In this dataset, 10 behavioral features have been recorded in addition to some individual characteristics that are effective in detecting ASD cases. Table 1 summarizes the various attributes that the dataset includes.

Table 1. Dataset Description (source: [14])

Variable	Type	Description
A1: Response_to_name	Binary (0, 1)	Does your child look at you when you call his/her name?
A2: Eye_contact	Binary (0, 1)	How easy is it for you to get eye contact with your child?
A3: Point_to_objects	Binary (0, 1)	Does your child point to indicate that s/he wants something?
A4: Sharing_interest	Binary (0, 1)	Does your child point to share interest with you?
A5: Pretend_play	Binary (0, 1)	Does your child pretend?
A6: Follow_looking	Binary (0, 1)	Does your child follow where you are looking?
A7: Confort_someone	Binary (0, 1)	does your child show signs of wanting to confort someone upset?
A8: First_words	Binary (0, 1)	Description of child first words
A9: Simple_gesture	Binary (0, 1)	Does your child use simple gestures?
A10: Stare_at_nothing	Binary (0, 1)	Does your child stare at nothing with no apparent purpose?
Age	Number	Toddlers (months)
Score by Q-chat-10	Number	1-10 (Less than or equal 3 no ASD traits; > 3 ASD traits)
Sex	Character	Male or Female
Ethnicity	String	List of common ethnicities in text format
Born with jaundice	Boolean (Y/N)	Whether the case was born with jaundice
Family member with ASD history	Boolean (Y/N)	Whether any immediate family member has a PDD
Who is completing the test	String	Parent, self, caregiver, medical staff, clinician, etc.
Class variable (ASD)	String	ASD traits or No ASD traits (Yes / No)

The items A1 to A10 are answers to questions, the possible answers are Always, Usually, Sometimes, Rarely and Never. For the questions A1 to A9, 1 is assigned if the response was Sometimes or Rarely or Never, and 0 is assigned if the response was Always or Usually. However, for question 10, if the response was Always or Usually or Sometimes, 1 is assigned, otherwise 0 is assigned.

The proposed plan for analysis comprises the following steps:

1. Examining the data and its different attributes to see if any preprocessing steps are required before using the data with the prediction models.
2. Performing an exploratory data analysis to generate insights and hidden patterns from the data.
3. Providing some classification models such as Decision Trees, Ensemble methods, Neural networks that predict the ASD screening status based on the other provided attributes.
4. Comparing the various models based on the prediction accuracy and selecting the model that provides the best accuracy based on the validation data.
5. Evaluating the model using the testing dataset.

Throughout the analysis steps, R-Studio which is a development environment for R is used. Moreover, the dataset is divided into training dataset that is used for training and validation and testing data set that is used only at the very end to evaluate the model performance. The expected outcome of this study is to have an accurate prediction model that will help the autism community by contributing to screening the toddlers at an early stage.

4. ANALYSIS AND RESULTS

4.1. Exploratory Data Analysis

An initial investigation has been performed on the data to discover if any patterns exist and to generate insights. Table 2 summarizes the proportions of female vs male toddlers screened for ASD symptoms. Of the identified cases as having ASD traits, 26% are female and 74% are male which is in line with the research of [15] that analyzed fifty-four studies and found that, among children meeting criteria for ASD, the male-to-female ratio is close to three to one.

Table 2. Male vs female kids screened for ASD

	ASD traits	ASD traits	No ASD traits	Totals
Sex				
Female		26.6%	38.3%	30.3%
Male		73.4%	61.7%	69.7%
Totals		100%	100%	100%

The collected data contains cases from different ethnicities, and we are interested in investigating the distribution of the cases by ethnicity. Figure 2 provides a summary about this and shows that the majority of the toddlers having ASD traits are White European followed by Asian then Middle Eastern.

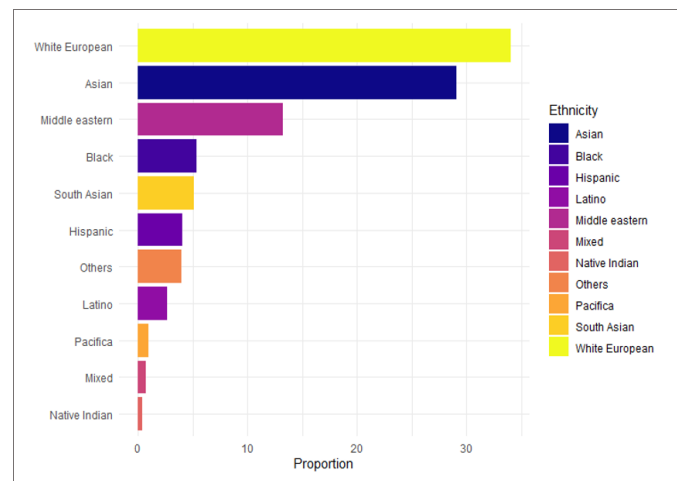


Figure 2. Proportion of ASD cases by ethnicity

4.2. Model Building

The ASD screening process is a binary classification problem as it aims to classify toddlers to either having or not having Autism traits [10]. We intend here to use various classification models, assess and compare their performances to see which model provides better performance measures and is therefore more suitable for this purpose. The evaluated models include Decision trees CART and C45, Ensemble methods Bagged cart and Random Forest, and Neural Networks. The outcome variable is the ASD.Traits (Yes/No) which indicates if the child is classified as having autism symptoms or no. This outcome is predicted based on the rest of the attributes presented in Table 1 except the score by Q-Chat-10 test, based on it, the data has been labelled. This variable is considered as an outcome variable and should not be included as predictor because this leads to model overfitting.

The classification models were built by performing a 5-fold cross-validation, using 70% of the dataset (739 observations) for training and the remaining 30% (315 observations) to test the accuracy of the various classifiers. Regarding the cross-validation, the training data set is partitioned into 5 subsets, and the algorithm randomly uses 4 data subsets to train the model and then tests it on the remaining subset.

4.2.1. Decision Tree Models

CART and C4.5 decision trees were used. Based on the validation dataset, CART tree gave an accuracy of 88% and C4.5 gave an accuracy of 92%. Table 3 presents the confusion matrices of both models. The plots of the two trees as well as the ROC curve of the C4.5 are in the Appendix. Given its better accuracy based on validation dataset, the C4.5 decision tree is a candidate for this category.

Table 3. Confusion matrices based on validation dataset

CART Tree			C4.5 Tree		
	Reference			Reference	
Prediction	No	Yes	Prediction	No	Yes
No	25	5.4	No	27	3.9
Yes	6	63.7	Yes	3.9	65.2
Accuracy	0.8861		Accuracy	0.9222	

The C4.5 model was applied to the test dataset (315 observations) to predict the outcome variable, and Table 4 summarizes its performance.

Table 4. C4.5 performance on testing data

	Reference			
Prediction	No	Yes	Sensitivity	0.9908
No	25	5.4	Specificity	0.9794
Yes	6	63.7	Pos Pred Value	0.9908
Accuracy	0.9873		Neg Pred Value	0.9794
95% CI	(0.9678, 0.9965)		Prevalence	0.6921
No Information Rate	0.6921		Detection Rate	0.6857
P-Value	<2e-16		Detection Prevalence	0.6921
Kappa	0.9702		Balanced Accuracy	0.9851
			'Positive' Class	Yes

4.2.2. Ensemble Methods

From the Ensemble methods, Bagged CART and Random Forest models were used. Based on the validation dataset, Bagged CART gave an accuracy of 94% and Random Forest gave an accuracy of 95%. Table 5 presents the confusion matrices of both models. Therefore, the Random Forest is candidate for this category, and its ROC curve is presented in the Appendix.

Table 5. Confusion matrices based on validation dataset

Bagged CART			Random Forest		
	Reference			Reference	
Prediction	No	Yes	Prediction	No	Yes
No	28.1	2.6	No	27.6	0.8
Yes	2.8	66.4	Yes	3.4	68.2
Accuracy	0.9459		Accuracy	0.9581	

The Random Forest model was applied to the test dataset to predict the outcome variable, and Table 6 summarizes its performance on this data.

Table 6. Random Forest performance on testing data

	Reference			
Prediction	No	Yes	Sensitivity	1
No	85	0	Specificity	0.8763
Yes	12	218	Pos Pred Value	0.9478
Accuracy	0.9619		Neg Pred Value	1
95% CI	(0.9344, 0.9802)		Prevalence	0.6921
No Information Rate	0.6921		Detection Rate	0.6921
P-Value	<2.2e-16		Detection Prevalence	0.7302
Kappa	0.9074		Balanced Accuracy	0.9381
			'Positive' Class	Yes

4.2.3. Neural Networks

Artificial Neural Network (ANN) model was built using R Neuralnet package. The model has two hidden layers; the first one with 5 hidden nodes and the second one with 3 hidden nodes. The plot of the model is presented in the Appendix and Table 7 summarizes its performance on the testing data.

Table 7. ANN performance on testing data

	Reference			
Prediction	No	Yes	Sensitivity	0.9794
No	95	0	Specificity	1
Yes	2	218	Pos Pred Value	1
Accuracy	0.9937		Neg Pred Value	0.9909
95% CI	(0.9773, 0.9992)		Prevalence	0.3079
No Information Rate	0.6921		Detection Rate	0.3016
P-Value	<2e-16		Detection Prevalence	0.3016
Kappa	0.985		Balanced Accuracy	0.9897
			'Positive' Class	No

4.2.4. Models Comparison

The three candidate models are the C4.5, the Random Forest and the Artificial Neural Network. Table 8 summarizes the prediction accuracy, sensitivity, and specificity of these three models. All

the candidate models have an extremely good performance and are expected to have an excellent predictive power in detecting autistic traits in toddlers.

Table 8. Models comparison

Model	Accuracy	Sensitivity	Specificity
C4.5 Tree	0.98	0.99	0.97
Random Forest	0.96	1	0.87
Neural Network	0.99	0.97	1

5. CONCLUSION

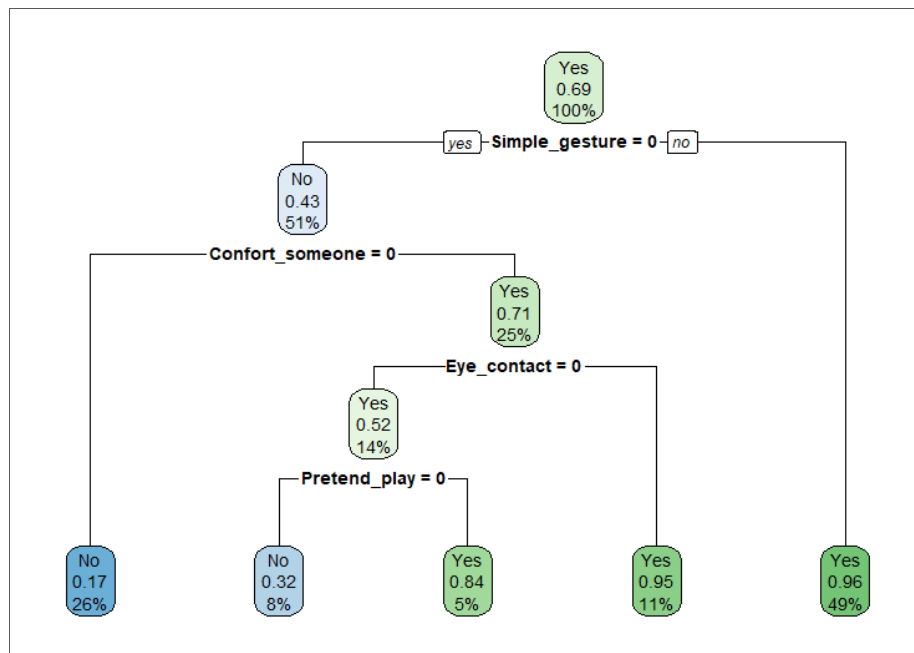
Due to the increasing rate of autism diagnosis cases and the multiple barriers that hinder getting a timely screening and therefore a diagnosis which allows the kids to get access to the early intervention services, there is an urgent need to think about developing and implementing a fast, accessible, and cost-effective autism screening tool. This was the aim of this work that showed the predictive power of the machine learning techniques in detecting autistic traits. The dataset used for training the model recorded ten behavioral features that indicate the development milestones and are concerned with communication and social behaviors, in addition to the gender, age, ethnicity, some attributes for the family history and the relationship to the person who completed the screening. After training multiple classifiers, three models were shortlisted as candidates for providing exceptionally good accuracy, sensitivity, and specificity. Although the three models' performances are excellent, still the Neural Network is outperforming the other two models, and therefore we propose to implement it. This study showed promising results for ASD screening, and the proposed model can be complemented by developing a mobile application that will add to the convenience and accessibility of the screening method. Furthermore, this research can be improved further and extended by performing feature selection and finding which features are most significant for ASD screening prior to building the models through dimensionality reduction.

REFERENCES

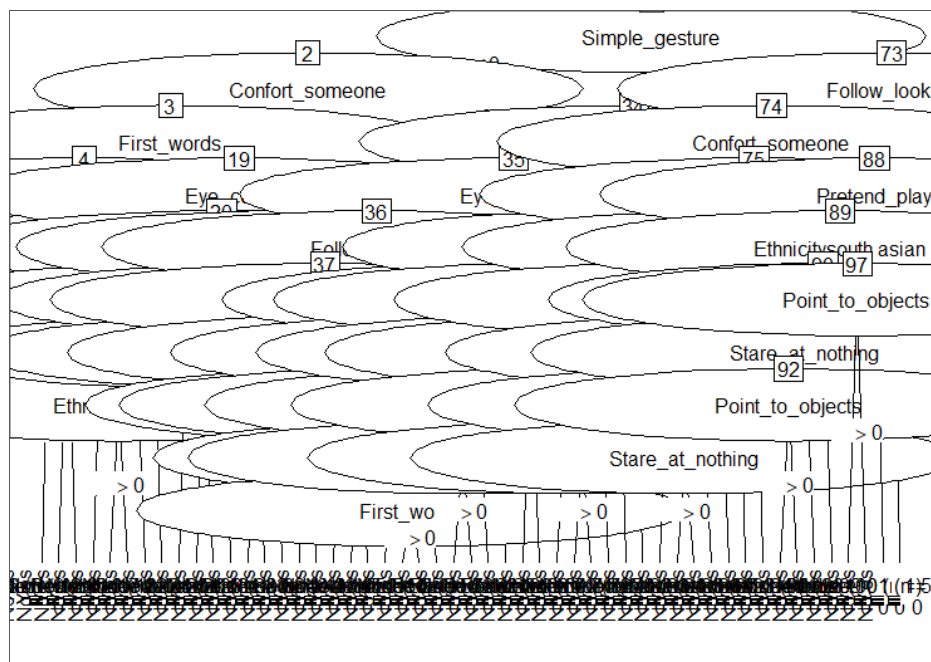
- [1] CDC. Reports, (2020) "AUTISM AND DEVELOPMENTAL DISABILITIES MONITORING (ADDM) NETWORK," ed.
- [2] V. Jagan and A. Sathiyaseelan, (2016) "Early intervention and diagnosis of autism," *Indian Journal of Health and Wellbeing*, vol. 7, no. 12, pp. 1144-1148.
- [3] L. E. K. Achenie, A. Scarpa, R. S. Factor, T. Wang, D. L. Robins, and D. S. McCrickard, (2019) "A Machine Learning Strategy for Autism Screening in Toddlers," *Journal of Developmental and Behavioral Pediatrics*, vol. 40, no. 5, p. 369, doi: 10.1097/DBP.0000000000000668.
- [4] P. S. Carbone et al., (2020) "Primary Care Autism Screening and Later Autism Diagnosis," *Pediatrics*, vol. 146, no. 2, doi: 10.1542/peds.2019-2314.
- [5] K. S. Omar, P. Mondal, N. S. Khan, M. R. K. Rizvi and M. N. Islam, (2019) "A Machine Learning Approach to Predict Autism Spectrum Disorder," in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE): IEEE*, pp. 1-6.
- [6] S. Broder Fingert et al., (2019) "Implementing systems-based innovations to improve access to early screening, diagnosis, and treatment services for children with autism spectrum disorder: An Autism Spectrum Disorder Pediatric, Early Detection, Engagement, and Services network study," *Autism : the international journal of research and practice*, vol. 23, no. 3, pp. 653-664, doi: 10.1177/1362361318766238.
- [7] C. Kamuk, C. Cantio, and N. Bilenberg, (2017) "Early screening for autism spectrum disorder," *European Psychiatry*, vol. 41, no. Supplement, pp. S131-S132, doi: 10.1016/j.eurpsy.2017.01.1948.

- [8] L. French and E. M. M. Kennedy, (2018) "Annual Research Review: Early intervention for infants and young children with, or at-risk of, autism spectrum disorder: a systematic review," *Journal of Child Psychology and Psychiatry*, vol. 59, no. 4, pp. 444-456, doi: 10.1111/jcpp.12828.
- [9] B. Wingfield et al., (2020) "A predictive model for paediatric autism screening," *Health informatics journal*, p. 1460458219887823, doi: 10.1177/1460458219887823.
- [10] F. Thabtah and D. Peebles, (2020) "A new machine learning model based on induction of rules for autism detection," *Health informatics journal*, vol. 26, no. 1, pp. 264-286, doi: 10.1177/1460458218824711.
- [11] M. Al-diabat, (2018) "Fuzzy data mining for autism classification of children," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 7, pp. 11-17, doi: 10.14569/IJACSA.2018.090702.
- [12] I. M. Nasser, M. O. Al-Shawwa, and S. S. Abu-Naser, (2019) "Artificial Neural Network for Diagnose Autism Spectrum Disorder " *International Journal of Academic Information Systems Research (IJASIR)*, vol. 3, no. 2, pp. 27-32.
- [13] T. Gennaro, C. Giovanni, D. P. Davide, and A. Stefania, (2020) "Use of Machine Learning to Investigate The Quantitative Checklist For Autism in Toddlers (QCHAT) Towards Early Autism Screening," ed.
- [14] F. Thabtah, (2020) "Autism screening data for toddlers." <https://www.kaggle.com/fabdelja/autism-screening-for-toddlers> (accessed).
- [15] R. Loomes, L. Hull, and W. Polmear, (2017) "What is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis " *Journal of the American Academy of Child & Adolescent Psychiatry*.

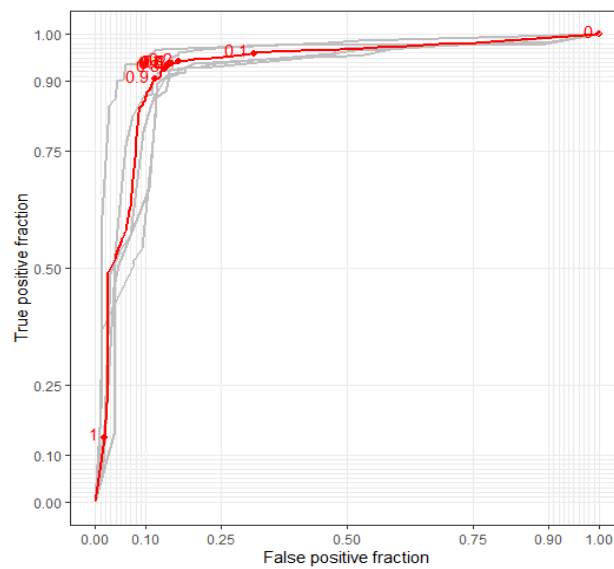
APPENDIX



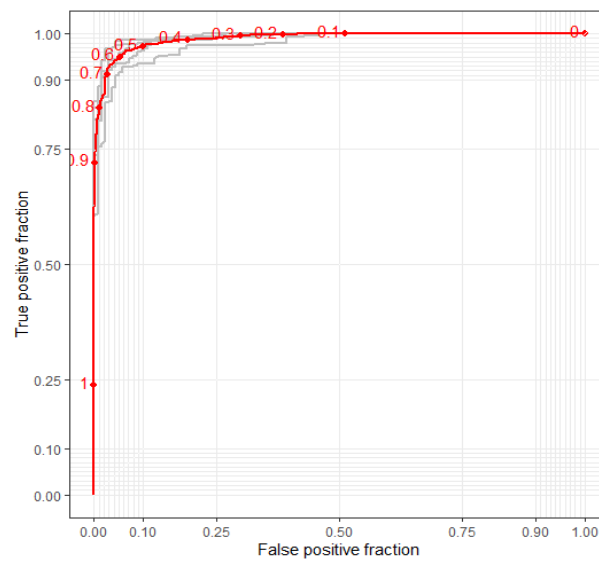
CART Tree



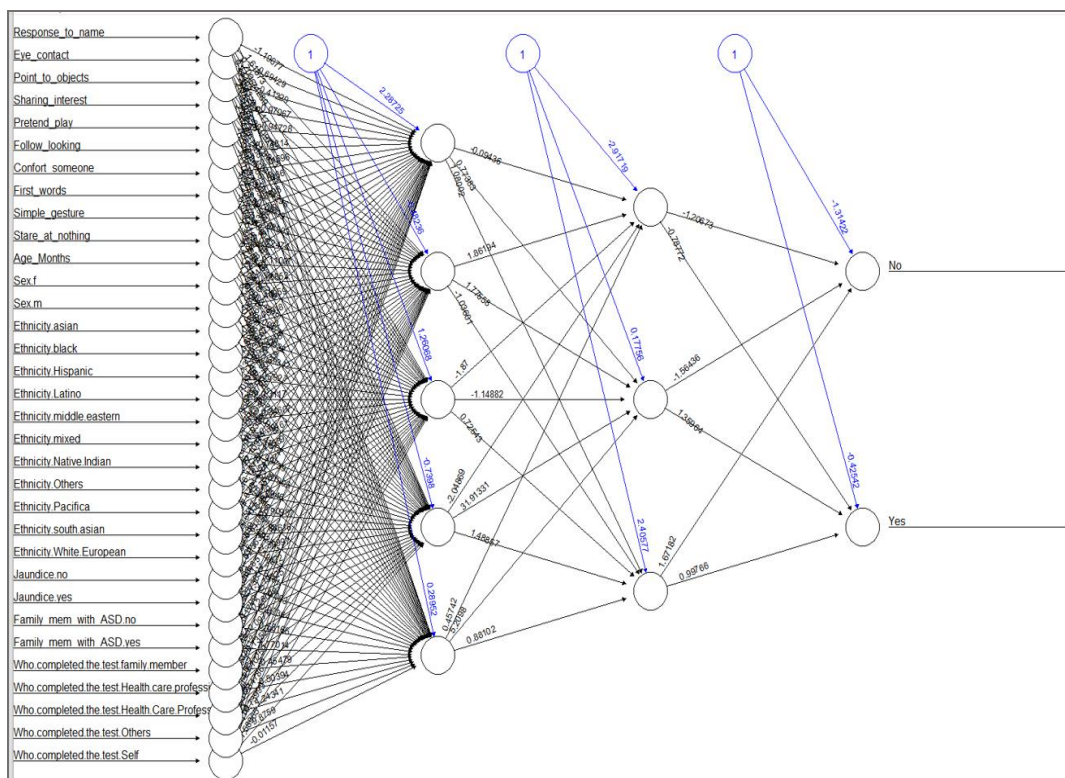
C4.5 Tree



C4.5 ROC curve



Random Forest ROC curve



Neural Network

AUTHORS

Ms Afef Saihi is currently working as graduate research and teaching assistant and pursuing her Ph.D. in Engineering Systems management at the American University of Sharjah. Her research interests are in the fields of supply chain management, maintenance planning and optimization, digital transformation and innovation management.



Dr Hussam Alshraideh is an Associate Professor of Operations Research and Statistics at the Industrial Engineering Department at the American University of Sharjah (AUS). He holds a dual Ph.D. degree in Industrial Engineering and Operations Research with a minor in Statistics from The Pennsylvania State University. His current research interests include statistical process optimization and smart data analytics applications in healthcare related fields. He has published more than forty papers in highly reputable journals on health informatics and process control.



REVISITING MOBILE CROWDSENSING: AN OPEN CHALLENGE

Vittalis Ayu

Department of Informatics, Sanata Dharma University, Indonesia

ABSTRACT

Mobile crowdsensing has become a new paradigm that enables citizens to participate in the sensing process by voluntarily gathering data from their smartphones to accomplish some given task. However, performing the sensing task generate lots of data resulting in various quality of the sensed data and high sensing cost in term of resource consumption. This matter became a significant concern in mobile crowdsensing as the mobile nodes which act as crowd sensors have limited resources. Moreover, an opportunistic mobile crowdsensing mechanism does not require user involvement, so the data collection process must be autonomous and intelligent to sense the data in the proper context. That is why context-awareness is also essential in opportunistic crowdsensing to maintain the sensed data quality. In this mini-review, we revisit the possibility of enhancing the mobile crowdsensing mechanism. We argue that improving the data collection process, including context-awareness, can optimize in-node data availability and sensed data quality. Besides, we also argue that finding optimization on inter-node data exchange mechanisms will increase the quality of the in-node data. Furthermore, smartphones that are related to humans as their owners reflect humans' physical and social behavior. We believe that considering contexts such as human social relationships and human mobility patterns can benefit the optimization strategies.

KEYWORDS

Mobile Crowdsensing, Data Quality, Context-awareness, Social Relation, Mobility Pattern.

1. INTRODUCTION

The emerging technology has enabled smartphones to enhance their connectivity and computing capabilities. Different sensors like gyroscope, accelerometer, camera, and microphone can be integrated easily into a single smartphone. This development empowers mobile devices to be able to sense, collect, and process information. Furthermore, sharing the collected information with nearby devices will form a collective knowledge about a particular phenomenon. This mechanism of sensing, collecting, processing, and sharing is called Mobile Crowdsensing (MCS). This term was first introduced by Ganti [1] as the collective ability to sense a kind of phenomenon from different devices.

A sensor-rich smartphone is a good upgrade for the existing environment-sensing mechanism. Furthermore, fast connectivity and various kind of sensors which already embedded in a mobile device can lead to lower deployment costs than those of the infrastructure-based sensing system. Ganti [1] observe the possibility of using sensing and connectivity capable devices, especially those that participate in the Internet of Things, to report their sensing result and form a collective knowledge. From then onwards, mobile crowdsensing has attracted many researchers to deepen their knowledge in this area.

In this mini-review, we intend to present the following aspects of MCS:

1. We provide an overview of MCS, existing implementation, and related research.
2. We investigate the existing implementation and research whether it has already incorporated social aspect and mobility model
3. We introduce potential challenge in MCS related to minimizing sensing and communication cost while maximizing data quality and a possibility to enhance it with the collaboration of social aspect and mobility model

2. OVERVIEW OF MOBILE CROWDSENSING

The context scope in the sensing process is a crucial part of Mobile Crowdsensing. Lane [2] categorized mobile crowdsensing into two: personal and community sensing based on the scope. In personal sensing applications as depicted in Figure 1, the device user is the point of interest. For instance, monitoring and recognizing user-related posture and movement patterns for personal fitness logs or health care reasons. Community sensing as illustrated in Figure 2 monitors a particular phenomenon such as traffic congestion, pollution level or noise level that cannot be measured only by a single device. This phenomenon can be accurately measured when many observers such as stationary sensor, embedded sensors in mobile phone, car or bus sense the same phenomenon to acquire complete information.

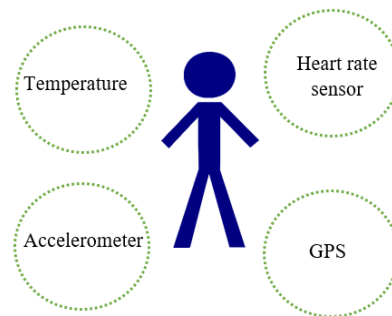


Figure 1. Personal sensing

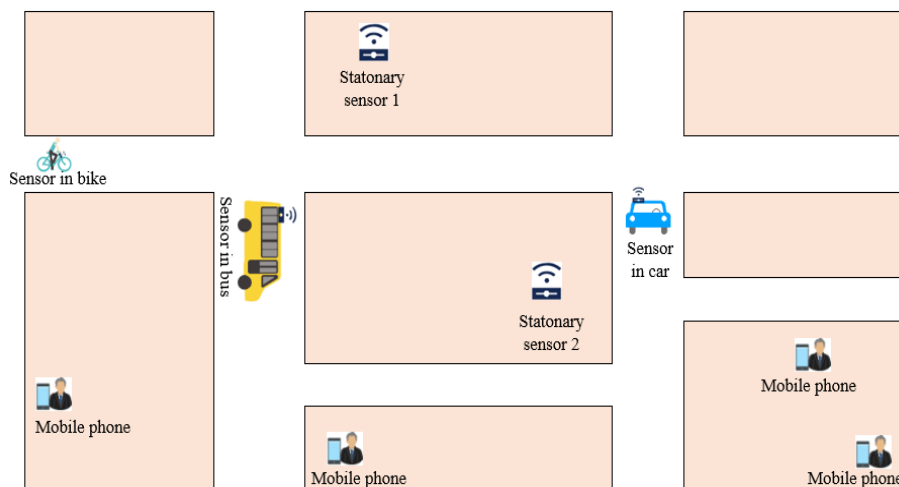


Figure 2. Community Sensing

Due to the large-scale observation scope, there are two approaches to implement community sensing: participatory sensing and opportunistic sensing. The former required the user to register and thus be actively involved in the data sensing and collecting process, whereas the latter only needed passive participation. The participatory sensing process required awareness from the contributed citizen. The citizen as smartphone's owner has to define what, when and where to sense a specific phenomenon. On the other hand, opportunistic sensing does not require user intervention as the data was collected automatically.

Consequently, opportunistic mobile crowdsensing relies heavily on the application to do the sensing task. Hence, this sensing mechanism needs to be autonomous and more intelligent since it has to recognize the sensing context on its own without the assistance of a human. Moreover, opportunistic sensing needs to sample data more frequently because of the user's passive participation and likely lower data quality due to context inaccuracy than participatory sensing. However, these data collection processes raise concerns about resources such as storage and energy. Meanwhile, as mobile devices' resources are limited, we must optimize resources while maximizing data quality and availability.

Crowdsensing management as depicted in Figure 3 includes three processes: data collection, data forwarding, and data analysis. The first process is data collection which aims to minimize the sensing cost of crowdsensors while maximizing the sensing coverage and quality. This process includes user recruitment and task allocation. The user recruitment concern is selecting the suitable crowd sensors for the sensing tasks, while task allocation assigns suitable tasks for crowdsensors. The second process is the data forwarding process which includes the communication aspect of the crowdsensing process. Lastly, the data analysis process aims to aggregate and extract the information from collected raw sensed data.

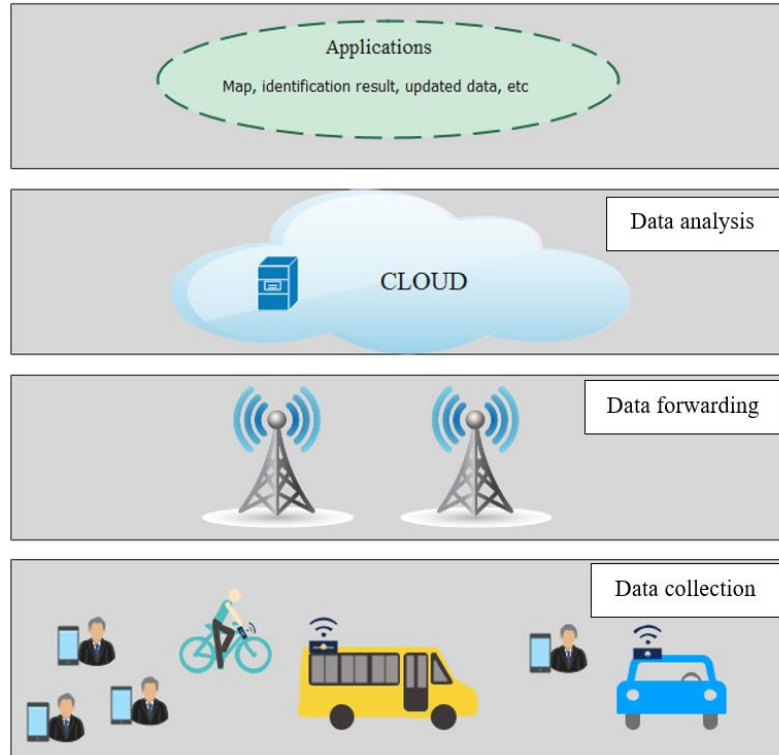


Figure 3. Crowdsensing management layer

Crowdsensing applications are already implemented to build collective knowledge in some domains of interest, such as environment monitoring, e-commerce, and healthcare. Several studies [3][4][5][6] had already implemented the crowdsensing process to observe some aspects of the environment. Yan's research [3] relies on a driver's smartphone to monitor a traffic congestion situation and report it to the cloud-based server. After the cloud-based server receives the report, it aggregates the data and subsequently distributed the overall traffic condition to the other drivers who use the same application to be informed of the event and thus avoid the congested road. FixMyStreet [4] collects reports from citizens regarding some trouble related to the street and its supporting public facilities. Research conducted by El-Wakeel [5] monitors road conditions and road hazards by analyzing vehicle movement recorded by sensors both in the land vehicle and driver's smartphone. Air quality monitoring using SecondNose [6] had been done in Trento, Northern Italy, with 80 participants on the data collection process. These participants collected air condition with sensors and aggregated the data. The smartphone's owner gets the complete information of the air quality in its surroundings (measured by the smartphone's sensors) and the overall air quality map generated by the aggregated data. In e-commerce, LiveCompare [7] uses participatory sensing to compare different prices from a specific product in different grocery stores. Abu-Elkheir [8] introduces crowdsensing as a means to get collective knowledge about an emergency. Healthcare framework AllergyMap [9] collect data from patients and can further visualize the distribution of different allergens and irritants concerning the spatial and temporal distribution.

These crowdsensing applications, as mentioned earlier, can be deployed with various kinds of underlying technologies. Capponi [10] differentiates the underlying of mobile crowdsensing technologies into two significant groups: infrastructure and infrastructure-less. Infrastructure technology refers to the need for data delivery infrastructures such as cellular or vast area networks. In contrast, the infrastructure-less technology depends on device-to-device (D2D) communication such as WiFi Direct and Bluetooth.

The availability of infrastructure will be a crucial point to define proper crowdsensing management. In infrastructure-based crowdsensing, crowdsensors upload the sensed data via infrastructure such as cellular networks to the remote server. Subsequently, data analysis will be performed in the remote server then a global view of data is achieved. Whereas in infrastructure-less-based crowdsensing, the data aggregation will be done in a distributed environment within the crowd. Consequently, data analysis is performed locally and the global view of the sensed data can not be achieved,

3. ISSUES IN MOBILE CROWDSENSING

Generally, the more data can be collected in MCS; the more complete the knowledge will be. However, having a sufficient number of users to perform the sensing task can be a challenging problem. In the data collection process, some aspects concerned are data quality, sensing cost, and privacy guarantee. Even though a large crowd of participants is helpful to do the data collecting process because of its extensive sensing coverage, not all of the collected data have the same quality, yet the amount of generated data is enormous. Therefore, it is necessary to find optimization between the number of recruited participants and sensed data quality.

On the other hand, from the user's perspective, doing the sensing task can deplete its resources such as energy and storage. Consequently, users will not always be willing to cooperate in the sensing process. Hence, there must be some incentive mechanism to compensate the willingness of the user to contribute to the sensing process. Privacy also becomes a concern in the data collection process. The anonymity of user identity has to be guaranteed in the sensing process while still maintain user verification to avoid the threat from malicious users.

Meanwhile, there are more to concern with the different perspectives of how users collect the data by participatory and opportunistic sensing. In participatory sensing, to get involved in the sensing process, the user's decision on how, what, when and where to sense is purely human's who act as the smartphone owner, thus indirectly including human's intelligence in the process. On the other hand, in opportunistic crowdsensing, the burden shifted to the application, which obliges intelligently aware of the sensing context hence makes context-awareness become an addition of data collection aspects to be concerned in opportunistic crowdsensing.

Data collection management in crowdsensing includes two processes: user recruitment and task allocation. The user recruitment process addresses the necessity in mobile crowdsensing to minimize the number of contributed users required in the sensing process while maintaining the sensing coverage and sensed data quality. On the other hand, the task allocation process concerns how to assign the right task to the suitable recruited user. Furthermore, those four aspects mentioned earlier (data quality, sensing cost, privacy, and context awareness) are partially addressed in the research in user recruitment and task allocation to improve the data collection process in mobile crowdsensing.

Research in [11][12] has been conducted to improve the user recruitment process. Wang [11] introduced PURE-DF to recruit participants in participatory crowdsensing. PURE calculate the estimation of the user's arrival in PoI and the user's subscription plan to minimize recruitment cost. Furthermore, Delegation Forwarding (DF) is used to optimize the trade-off between the delivery ratio of the sensed data and the number of required participants. Although this method can successfully minimize the recruitment cost and the number of recruited participants, this study assumes there is always sufficient bandwidth and contact duration to successfully deliver the sensing task and exchange sensing data with other nodes. The Secure User Recruitment (SUR) protocol proposed by Xiao [12] uses a Greedy algorithm to recruit the nearly minimum user while ensuring the data quality and privacy using the semi-honest model. However, this mechanism assumes that the user mobility model is independent while the human mobility model is not fully independent. Besides smartphones, a vehicle with embedded sensors can act as crowd sensors. However, not all vehicles will voluntarily contribute to the sensing process. Research conducted by Liu [13] introduces an incentive mechanism to increase the participation of vehicles in the crowdsensing process.

To address task allocation, Capponi et al. [14] introduce an energy-efficient data collection framework that can reduce the sensing cost and maximize the data quality considering the user's level of remaining energy and matching it with a suitable sensing task. The approach use context in the allocation process called Context-Aware Task Allocation (CATA) has been introduced by Hassani [15]. CATA measures similarity between task and user's context to compute the probability of successfully executing the task for a given user. When the probability of success is high, then the task will be allocated to the user.

4. RESEARCH CHALLENGES IN MOBILE CROWDSENSING

From those issues mentioned above, we found two challenges that can still be further studied: maximize data quality; minimize sensing and communication cost.

Maximize data quality. To maximize data quality, we need to find a suitable number of recruited users to perform the sensing task, hence covering a certain area. Furthermore, because we focus on opportunistic mobile crowdsensing, we need to improve the sensing process to be more intelligent.

We argue that applying context-awareness will improve the data quality in the data collection process. In mobile crowdsensing, each node can do the local computation. Hence, the extraction of contextual information from raw sensed data can be performed inside the node itself. Hassani [15] implement context-awareness in the crowdsensing process by matching the task context with the participant context to allocate the suitable task to a certain participant. However, although the data were collected by smartphones, the collected data are then processed in a centralized server. The implementation on device to device communication network (D2D) and further use of distributed processing is possible. Artificial intelligence also can be utilized to derive the semantic characteristic of collected raw data. El-Wakeel [5] incorporate Support Vector Machine (SVM) to classify the collected data into eight different road anomaly contexts. Although the classification is done successfully, the diversity of sensors, vehicle's condition, and weather condition influence the classification process.

We also argue that involving the influence of mobility pattern of mobile nodes is advantageous to define which node to recruit based on spatial-temporal dimension because some people may regularly visit someplace after some interval. Moreover, we expect that considering the mobility pattern of the mobile node will benefit the context inference of the crowdsensing process. The mobility models presented in Table 1 already captured some mobility models used in earlier experiments. There is much diversity of mobility models such as synthetic mobility, real-world, and random mobility. However, none of these models are used to define the overlay social characteristics derived from the participant's movement.

Table 1. Summary of Mobile Crowdsensing Existing Solutions

Ref	Participation		Underlying technologies		Data quality	Sensing cost	Social aspect	Mobility model and geographical scope
	Participatory	Opportunistic	Infrastructure	Infrastructure-less				
[3]	✓	-	✓	-	-	✓	-	real vehicle movement of Guangzhou City
[4]	✓	-	✓	-	-	-	-	geographical region of Brussel
[5]	-	✓	-	✓	✓	-	-	vehicle movement in Kingston, Canada
[6]	-	✓	-	✓	-	-	-	human real movement in Trento
[7]	✓	-	✓	-	✓	-	-	Geographical region of Durham, North Carolina
[8]	✓	-	-	✓	✓	-	✓	drones movement based on roads
[9]	✓	-	✓	-	-	✓	-	geographical region of Greece
[11]	✓	-	✓	-	-	✓	-	roma/taxi trace set, epfl trace set, geolife trace set
[12]	✓	-	-	✓	✓	-	-	synthetic mobility model
[13]	-	✓	✓	-	-	✓	-	cawdad dataset of taxis in San Francisco
[14]	-	✓	✓	-	✓	✓	-	pedestrian mobility in Luxemburg
[15]	-	✓	✓	-	✓	✓	-	synthetic mobility model

Besides, we believe that data quality can be further refined by the data aggregation process facilitated by an inter-node data exchange mechanism. While inspecting other node's data, we may found similar data with different quality. Hence, we can compare these data and choose which data to keep.

In addition, different mobility characteristics of the nodes can be considered in the sensing process related to the coverage and density of the sensing area. For example, people who tend to travel in the more extensive coverage area can most likely sense a lot of different data. However,

the communication cost will be huge as the assigned task will rarely match with the in-node available data.

Also, people whose mobility patterns covered a dense area tend to have more similar data with other people compared to people whose mobility patterns covered a sparse area. When people arrive in a dense area, she/he can drop the data more frequently as another node may have similar data.

Minimize sensing and communication cost. Other than minimizing sensing and communication costs, there are factors to consider, such as the underlying communication technology and resource allocation. We focus on inspecting the scenario of infrastructure-less scenario. D2D communication over Mobile Adhoc Network (MANET) and Opportunistic Network (OppNet) are the two candidates for the underlying technology of the crowdsensing process in the D2D scenario. Although these two underlying communication technologies are the basis of mobile ad-hoc communication, they differ in some aspects. Even though node mobility in MANET is dynamic, there is always an end-to-end path between these nodes, so the global knowledge is available in MANET. In contrast, in the opportunistic network, there are not.

On the other hand, in OppNet, there are intermittent connectivity and opportunistic encounter between nodes. Thus, the end-to-end path does not exist. Moreover, the contact duration between nodes and encounter probability of nodes varies. Consequently, data aggregation and data forwarding in this network have to be distributed because of the absence of global knowledge of the network. However, the opportunistic network will benefit the crowdsensing process because of the availability of bundle layers inside the nodes, which enable the store-carry-forward mechanism. In an opportunistic network, data can be carried around even when the end-to-end path does not exist. This behavior will reduce the communication cost as the data can be carried around between the nodes themselves without infrastructure involvement. Researchers in [5][6][8][12] are already conducted in infrastructure-less settings. However, these four researchs still incorporate edge cloud in the data aggregation process.

Because of the limitation of mobile node resources such as storage and energy, we argue that resource-aware data collection and forwarding are essential to reduce the sensing cost. The willingness to participate in the crowdsensing process must be compensated with some reward from the incentive mechanism. In future research, a game theory-based incentive mechanism can be utilized to optimize the resource and the willingness to cooperate. The reputation-based reward can benefit the mechanism as the incentive scheme does not have to be in monetary forms. Although research in [9][11][13][14][15] implemented an incentive mechanism to encourage participation and minimize resource consumption, to the best of our knowledge, there are none that incorporate the derivation of the social context of the collaborating nodes in the opportunistic mobile crowdsensing. Elkheir's research [8] infuses social media to collect emergency data. However, the overlay social characteristics such as community and social ties of the collaborating nodes have not been explored yet. There are some efforts in [16][17][18] to incorporate social aspects of the routing process in the opportunistic network, such as similarity and centrality.

5. CONCLUSION

The emerging technology has enabled the mobile crowdsensing paradigm to form collective knowledge within the crowd, with some aspects to be concerned, such as sensed data quality and sensing cost. Furthermore, in opportunistic mobile crowdsensing, context-awareness is essential because the sensing mechanism needs to be intelligent as the sensing process does not include active human participation. As the opportunistic network will serve as the underlying

communication network, the intermittent connectivity, and various contact opportunity, the crowdsensing mechanism should employ a distributed algorithm which can be later implemented in this network. Therefore, mobile crowdsensing still faces open challenges to maximize the sensed data quality while minimizing the sensing cost. Moreover, incorporating social aspects and mobility patterns as additional aspects is a possible thing to be done in future studies.

REFERENCES

- [1] R. K. Ganti, F. Ye and H. Lei, (2011) "Mobile crowdsensing: current state and future challenges", *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32-39.
- [2] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury & A. T. Campbell, (2010) "A survey of mobile phone sensing", *IEEE Communications Magazine*, vol. 48, no. 9, pp. 140-150.
- [3] Hehua Yan, Qingsong Hua, Daqiang Zhang, Jiafu Wan, Seungmin Rho, & Houbing Song (2017) "Cloud-Assisted Mobile Crowd Sensing for Traffic Congestion Control", *Mobile Netw Appl* 22, 1212–1218
- [4] Pak, Burak, Alvin Chua & Andrew Vande Moere. (2017) "FixMyStreet Brussels: Socio-Demographic Inequality in Crowdsourced Civic Participation", *Journal of Urban Technology*, 24:2, 65-87
- [5] A. S. El-Wakeel, J. Li, A. Nouredin, H. S. Hassanein & N. Zorba, (2018) "Towards a Practical Crowdsensing System for Road Surface Conditions Monitoring", *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4672-4685.
- [6] Chiara Leonardi, Andrea Cappellotto, Michele Caraviello, Bruno Lepri, & Fabrizio Antonelli, (2014) "SecondNose: an air quality mobile crowdsensing system", *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (Nordi CHI '14)*, Association for Computing Machinery, New York, NY, USA, 1051–1054.
- [7] Linda Deng & Landon P. Cox, (2009) "LiveCompare: grocery bargain hunting through participatory sensing", *Proceedings of the 10th workshop on Mobile Computing Systems and Applications (HotMobile '09)*. Association for Computing Machinery, New York, NY, USA, Article 4, 1–6.
- [8] M. Abu-Elkheir, H. S. Hassanein & S. M. A. Oteafy, (2016) "Enhancing emergency response systems through leveraging crowdsensing and heterogeneous data," *International Wireless Communications and Mobile Computing Conference (IWCMC)*, Paphos, Cyprus, pp. 188-193,
- [9] L. A. Kalogiros, K. Lagouvardos, S. Nikolettas, N. Papadopoulos & P. Tzamalīs, (2018) "Allergymap: A Hybrid mHealth Mobile Crowdsensing System for Allergic Diseases Epidemiology : a multidisciplinary case study," *IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Athens, Greece, pp. 597-602.
- [10] A. Capponi, C. Fiandrino, B. Kantarci, L. Foschini, D. Kliazovich & P. Bouvry, (2019) "A Survey on Mobile Crowdsensing Systems: Challenges, Solutions, and Opportunities", *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2419-2465.
- [11] E. Wang, Y. Yang, J. Wu, W. Liu & X. Wang, (2018) "An Efficient Prediction-Based User Recruitment for Mobile Crowdsensing", *IEEE Transactions on Mobile Computing*, vol. 17, no. 1, pp. 16-28.
- [12] M. Xiao, J. Wu, S. Zhang & J. Yu, (2017) "Secret-sharing-based secure user recruitment protocol for mobile crowdsensing", *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, Atlanta, GA, USA, pp. 1-9.
- [13] L. Liu, X. Wen, L. Wang, Z. Lu, W. Jing & Y. Chen, (2020) "Incentive-Aware Recruitment of Intelligent Vehicles for Edge-Assisted Mobile Crowdsensing", *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 12085-12097.
- [14] A. Capponi, C. Fiandrino, D. Kliazovich & P. Bouvry, (2017) "Energy efficient data collection in opportunistic mobile crowdsensing architectures for smart cities," *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Atlanta, GA, pp. 307-312.
- [15] A. Hassani, P. D. Haghighi & P. P. Jayaraman, (2015) "Context-Aware Recruitment Scheme for Opportunistic Mobile Crowdsensing", *IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS)*, Melbourne, VIC, Australia, pp. 266-273.
- [16] P. Hui, J. Crowcroft & E. Yoneki, (2011) "BUBBLE Rap: Social-Based Forwarding in Delay-Tolerant Networks", *IEEE Transactions on Mobile Computing*, vol. 10, no. 11, pp. 1576-1589.

- [17] E. M. Daly & M. Haahr, (2009) "Social Network Analysis for Information Flow in Disconnected Delay-Tolerant MANETs", *IEEE Transactions on Mobile Computing*, vol. 8, no. 5, pp. 606-621.
- [18] D. Rothfus, C. Dunning & X. Chen, (2013) "Social-similarity-based routing algorithm in Delay Tolerant Networks", *IEEE International Conference on Communications (ICC)*, Budapest, Hungary, 2013, pp. 1862-1866.

AUTHORS

Vittalis Ayu received the Bachelor degree in Computer Engineering from Telkom University and Master degree in Computer Science from Gadjah Mada University. She is currently a Lecturer in Department of informatics, Sanata Dharma University, Yogyakarta, Indonesia. Her research interest includes computer network, distributed network, mobility model, Internet of Things and mobile crowdsensing.



PREDICTING CONSUMER PURCHASING DECISIONS IN THE ONLINE FOOD DELIVERY INDUSTRY

Batool Madani and Hussam Alshraideh

Department of Industrial Engineering,
American University of Sharjah, Sharjah, UAE

ABSTRACT

This transformation of food delivery businesses to online platforms has gained high attention in recent years. This due to the availability of customizing ordering experiences, easy payment methods, fast delivery, and others. The competition between online food delivery providers has intensified to attain a wider range of customers. Hence, they should have a better understanding of their customers' needs and predict their purchasing decisions. Machine learning has a significant impact on companies' bottom line. They are used to construct models and strategies in industries that rely on big data and need a system to evaluate it fast and effectively. Predictive modeling is a type of machine learning that uses various regression algorithms, analytics, and statistics to estimate the probability of an occurrence. The incorporation of predictive models helpsonline food delivery providers to understand their customers. In this study, a dataset collected from 388 consumers in Bangalore, India was provided to predict their purchasing decisions. Four prediction models are considered: CART and C4.5 decision trees, random forest, and rule-based classifiers, and their accuracies in providing the correct class label are evaluated. The findings show that all models perform similarly, but the C4.5 outperforms them all with an accuracy of 91.67%

KEYWORDS

Food Delivery Industry, Purchasing Prediction, Machine Learning, Decision Trees, Random Forest, Rule-Based Classifier.

1. INTRODUCTION

The internet's fast evolving technology has infiltrated practically every part of our lives, giving limitless opportunity for businesses and customer relationships. It has boosted online food services by allowing consumers to search, compare the provided services, and select companies with high service performance. The food delivery industry has experienced a lot of transformations, moving from restaurant-to-consumer delivery to platform-to-consumer delivery [1]. At first, restaurants had dedicated websites or phone numbers to allow customers to place their orders as well as have a dedicated delivery team. Then, in 2013, the business of platform-to-consumer came, in which specialized delivery players provide logistic support for restaurants. This new business model, which is referred to as online food delivery, provides customers with the abilityto compare menus, check reviews, and place many orders from different restaurants at the same time. Online food delivery has grown by 25% between 2015 and 2018 and it is expected to grow at a rate of 10.7% by 2023 [2]. Online food delivery is defined as the process of food online ordering, order process and preparation, and delivery. The platforms of online food

delivery such as Uber Eats, Deliveroo, and Zomato provide a variety of functions, such as providing customers with a wide variety of food choices, taking the order and transferring it to the food provider, monitoring the payment and managing the delivery of the food [1]. Worldwide, China leads the way in the market share of this industry, followed by the US and then India. Between 2020 and 2024, the market's revenue is expected to grow at a 7.5% annual rate, resulting in a market volume of US \$182,327 million by 2024 [3]. This indicates the rapid increase in the online food delivery market, which intensifies the competition between companies to gain dominance among others and increases the need to determine the key success factors that are critical to online food delivery providers. Online food delivery providers need to gain insights and reviews from customers to capture a larger segment of the market share. Besides, the decision of whether or not to make a purchase is a complex process, impacted by many aspects.

Due to the major dependence of consumers on online services, the online food delivery industry has been proactive in a way that creates a highly competitive market, which makes companies more susceptible to losing their marketplace. In order for companies to gain a competitive edge over others, they should understand their customers' needs, expectations, and requirements. Otherwise, the misunderstanding of customers' expectations can lead to the loss of customers' purchases and commitment. The reasoning behind that requires the evaluation of multiple measures, such as the timing of deliveries, the performance of tracking systems, quality of packaging, food temperature, food freshness, etc. Because obtaining customers' purchasing decisions is critical, determining the right reasoning can increase customer satisfaction and thus improve the company's market position. This requires the collection of customer data about their evaluation of the performance of online food delivery companies in terms of different aspects. While purchase prediction has been discussed in consumer research for a long time, the emergence of customer analytics has reignited such issues recently. One possible way of exploiting the data of customers' purchasing decisions is via machine learning techniques to construct accurate prediction models. Machine learning is a highly advanced, rapid, and accurate technology [4]. In the customer relationship management domain, the use of machine learning techniques for predictive purposes on a customer base is frequently investigated, with customer churn prediction being the most prominent goal. For maintaining customer relationships, accurate prediction of a customer's activity state and future purchasing propensities are critical [5]. Predicting purchasing decisions is a time series forecasting task that can be solved using traditional statistical techniques such as autoregressive moving average [6]. However, machine learning techniques are often more powerful and versatile, when dealing with time series forecasting. This is because they enable the employment of cutting-edge supervised learning algorithms like regression support vector machines and model trees.

In this study, machine learning techniques are used to anticipate customer purchasing decisions in the context of online food delivery. This is accomplished through the use of a dataset about customer purchasing experiences, which covers a variety of characteristics related to online food delivery providers. A comparison of three prediction models will be provided in order to determine which model is the most suitable and provides the best performance in terms of accuracy. The remainder of the paper is structured as follows. In section 2, the literature review is presented, while section 3 describes the used dataset. This is followed by the analysis and the results in section 4. Finally, section 5 provides the conclusion.

2. LITERATURE REVIEW

The rapid increase in demand for online services has motivated practitioners and academicians to seek a better understanding of customers' purchasing decisions and behaviors. An increasing number of studies are adopting prediction models to forecast purchasing decisions under different problem settings and varied inputs [7-16]. Van Den Poel and Buckinx [7], investigated the impact

of different sets of predictors on online purchasing behavior using logit modeling. The logit modeling method is used to answer the question of whether or not a purchase will be made during the next visit using a set of predictors: general clickstream behavior, detailed clickstream behavior, customer demographics, and historical purchase behavior. Using the same prediction model, Yilmaz and Belbag [8] predicted consumer behavior regarding purchasing remanufactured products, which indicated that low prices, product reliability, and product promotions affect positively the purchasing decisions of consumers. In the context of product promotions, Ling et al. [9] proposed a feature-combined deep learning framework for predicting consumers' purchase intent during promotions across multiple online channels. The study also suggested that including demographics information enhances the prediction performance, but, increases the methodological challenge. The importance of demographic information has also been considered in [10], which emphasized the importance of defining the demographic of people in a certain region to the marketing of the automobile industry in order to define the target group and integrate marketing strategies to enhance the purchase decision of a car. Therefore, they have investigated the prediction of consumer purchase decisions using the demographic structure of premium car owners using the logistic regression classification model. Due to the complexity of online markets and the diversity of their consumers, prediction models with powerful self-learning capabilities, such as artificial neural networks, decision trees, and random forest, to name a few, are increasingly relied on. Gupta and Pathak [11] applied different classification algorithms, such as decision trees, support vector machine, and rule-based method, to predict customers' purchase decisions, whether a user will be interested in buying a certain set of products that are placed in the online shopping cart or not. Similarly, Tang et al. [12] developed a hybrid model that is based on the technique of support vector machine and the firefly algorithm, for predicting online-purchasing behavior to forecast whether or not a customer will purchase during the next visit. Martínez et al. [13] developed an advanced analytics technique for non-contractual customer behavior prediction by establishing a dynamic and data-driven machine learning framework. Among the state-of-the-art machine learning algorithms, the gradient tree boosting method has outperformed the other methods and provided a prediction accuracy of 89%. Liao and Tsai [14] proposed a multimodel fusion B2C online marketing algorithm based on the least squares-support vector machine method, which has proved to have a high prediction accuracy compared to traditional prediction single-model.

Wang and Xu [15] examined the Chinese government's introduction of a 7-day unreasonable return policy to boost customer trust in e-commerce companies. The ease of return has a direct impact on customer purchase decisions, which is investigated in this study. An ensemble learning method based on a fuzzy support vector machine is used to predict customers' purchasing intentions. The proposed method outperformed a set of several classifiers such as logistic regression, support vector machine, and random forest in terms of prediction accuracy. Ghosh and Banerjee [16] proposed a modified random forest algorithm-based predictive analytic methodology. Using five parameters (previous purchasing habits, a sequence of online advertisements viewed, customer location, number of clicks, and last used service), the model seeks to predict purchasing decisions in cloud services. The model also had high forecast accuracy, with online advertisements being the most important component in making a purchase decision.

In the context of online food delivery, Natarajan et al. [17] investigated the impact of online food delivery service providers such as Swiggy, Foodpanda, and Zomato on Indian consumer preferences in the setting of online food delivery. According to the study's findings, consumers favor originality in terms of pricing, quality, and delivery. The online food delivery market in India is one of the world's largest markets. According to a study conducted in the years, 2019-2020 [18], the Indian online food delivery market was estimated to be valued at \$4.35 billion in 2020. This was a significant gain over the previous year, when the market was estimated to be

worth roughly 2.9 billion dollars. In addition, the food delivery sector is predicted to reach about 13 billion dollars in value by 2025. According to Anusha and Panda [19], “young India’s appetite is one of the key drivers for demand in the food and beverage industry on the whole”. As a result, the analysis provided in this study is centered on the consumers of India. Furthermore, compared to other fields of prediction research, existing research on online purchasing choice prediction is limited, particularly in the application of online food delivery. Therefore, the purchasing decisions in an online food delivery segment will be investigated here, which will help decision makers anticipate their customers’ buying intentions and determine the most influential factors in purchasing decisions.

3. DATA DESCRIPTION

To predict whether the consumer will buy again or not, a dataset obtained from the open-source database Kaggle is used [20]. The obtained dataset was collected from 388 consumers in Bangalore, India, and it has 55 variables consisting of the consumers’ demographics and consumers’ inputs about the delivery service, including the time, packaging, delivery person, and many others. There are 25 variables with a 5-point Likert-type scale (1 = Strongly disagree, 5 = Strongly agree), 8 variables about the level of importance of certain aspects, 10 demographic variables, 2 categorical variables with three levels, about the influence of delivery timing and the rating of restaurants, and a combination of categorical and numerical variables. Finally, the response variable is a categorical variable with two classes: “will purchase (yes)” and “will not purchase (no)”.

4. RESULTS AND ANALYSIS

As a first step, data pre-processing will be carried out, in which some of the input variables will be eliminated. Secondly, an exploratory data analysis will be performed to summarize the data, obtain insights and understanding of the demographics of the consumers, and investigate the relationship between the purchase decision and the other attributes. Due to a large number of input attributes, feature selection and elimination are considered to reduce the number of inputs and determine the significant ones. Table 1 provides a detailed description of the attributes and the variables of interest. Finally, using the significant attributes, different classification methods, which are decision-tree, random forest, and rule-based classifier, are used to predict the purchase decision. A comparison between them will be made based on the accuracy and the significance of the difference between them. The classification models will be first exposed to model parameter tuning using cross-validation to enhance their performance, and then will be tested on a new dataset to evaluate their performance.

4.1. Exploratory Data Analysis

Prior to this analysis, data preprocessing is performed, in which some of the variables are removed due to their irrelevancy to the problem, such as latitude, longitude, pin code, reviews. Afterward, an analysis of the demographics and the preferences of participants is presented. Table 2 presents the demographics summary of the participants.

As it can be observed, the mean average age of respondents is 25 years. The mix of respondents is fairly balanced, with males contributing to 57.2%. In terms of marital status, singles (69.1%) have a comparatively large presence, followed by married. Most of the respondents were students (53.3%), followed by employees (30.4%). For educational qualifications, graduates (45.6%) followed by postgraduates (44.9%) represent the majority of the respondents. Additionally, the majority of the respondents (46.3%) live with 3-4 members.

Table 1. Dataset description

	Attribute	Type	Description
1	Age	Integer	Age of participants
2	Gender	Character	Gender of participants
3	Marital Status	Character	Marital status of participants
4	Occupation	Character	Job occupation of participants
5	Monthly income	Character	Monthly income of participants
6	Educational Qualifications	Character	Educational qualification of participants
7	Family size	Integer	Number of family members/ friends living with
8	Ordering medium preference 1	Character	Through which medium participants are ordering
9	Ordering medium preference 2	Character	Through which medium participants are ordering
10	Meal preference 1	Character	What type of meal participants are ordering
11	Meal preference 2	Character	What type of meal participants are ordering
12	Ordering ease and convenience	Character	Ease and convenience of online ordering
13	Time saving	Character	Does it save time?
14	Restaurant choices	Character	More restaurant choice influence
15	Easy payment option	Character	Payment option influence
16	More offers and discounts	Character	Offers and discount influence
17	Good food quality	Character	Food quality influence
18	Good tracking system	Character	Tracking system influence
19	Self-cooking	Character	Self-cooking causes not purchasing
20	Health concern	Character	Health concern causes not purchasing
21	Late delivery	Character	Later Delivery causes not purchasing
22	Poor hygiene	Character	Poor Hygiene causes not purchasing
23	Bad experience	Character	Past experiences cause not purchasing
24	Unavailability	Character	Unavailability causes not purchasing
25	Unaffordable	Character	Un-affordability causes not purchasing
26	Long delivery time	Character	Long delivery causes cancellation
27	Delay of delivery person	Character	Delay of delivery person assigned causes cancellation
28	Delay of picking up food	Character	Delay of delivery person picking up food causes cancellation
29	Wrong order delivered	Character	Previous wrong order causes cancellation
30	Missing item	Character	Missing item in order causes cancellation
31	Order placed by mistake	Character	Placed order by mistake causes cancellation
32	Influence of delivery time	Character	Time of delivery influencing purchasing decision
33	Order time	Character	When do you order?
34	Maximum waiting time	Character	How long can you wait?
35	Residence in busy locations	Character	Residence in busy location
36	Google maps accuracy	Character	My location in google maps is accurate
37	Good road conditions	Character	My residence area road condition is good
38	Low quantity	Character	low quantity low delivery time
39	Delivery person ability	Character	Delivery person ability depends on time of delivery
40	Influence of restaurant rating	Character	Rating of restaurant influencing purchasing decision
41	Less delivery time	Character	Importance of Less delivery time
42	High quality of package	Character	Importance of Quality of package
43	Number of calls	Character	Importance of Number of calls made by delivery captain
44	Politeness	Character	Importance of Politeness of delivery captain
45	Freshness	Character	Importance of Freshness of food
46	Temperature	Character	Importance of Temperature of food
47	Good taste	Character	Importance of taste
48	Good quantity	Character	Importance of Quantity in food
49	Purchasing decision	Character	Will the customer purchase again (output variable)

Table 2. Demographics Summary

Category	Subcategory	Value
Age	Mean	24.6 years
Gender	Female	42.8%
	Male	57.2%
Marital Status	Married	27.8%
	Prefer not to say	3.1%
	Single	69.1%
Occupation	Employee	30.4%
	Housewife	2.3%
	Student	53.3%
	Self Employed	13.9%
Educational qualifications	Graduate	45.6%
	Ph.D	5.9%
	Postgraduate	44.9%
	School	3.1%
	Uneducated	0.5%
Family size	Less than 3	32.2%
	3-4	46.3%
	5-6	21.4%

Table 3 shows that customers prefer to use food delivery applications the most, ordering mostly food for snacks (32.0%) and dinner (80.4%).

Table 3. Preference Summary

Category	Subcategory	Percentage
Ordering medium preference 1	Direct call	1.3%
	Food delivery apps	92.3%
	Walk-in	5.7%
	Web browser	0.8%
Ordering medium preference 2	Direct call	53.6%
	Walk-in	26.8%
	Web browser	19.6%
Meal preference 1	Breakfast	13.7%
	Dinner	23.4%
	Lunch	30.9%
	Snacks	32.0%
Meal preference 2	Dinner	80.4%
	Lunch	7.2%
	Snacks	12.4%
Cuisine preference 1	Bakery items (snacks)	0.3%
	Non-Veg foods (Lunch / Dinner)	81.2%
	Sweets	0.8%
	Veg foods (Breakfast / Lunch)	17.8%
Cuisine preference 2	Bakery items (snacks)	3.4%
	Ice cream / Cool drinks	9.0%
	Sweets	11.9%
	Veg foods (Breakfast / Lunch)	75.8%

Figure 1 summarizes the relationship between purchase decisions and gender and marital status, respectively. Single customers are more likely to use online food delivery services than customers who are not single. Males and females use online services in comparable amounts, with males having a higher proclivity to buy from online food providers.

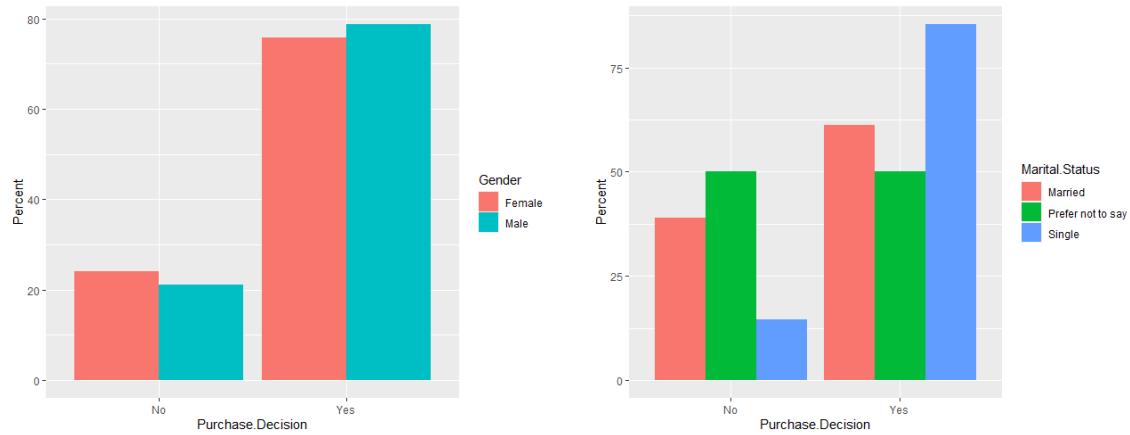


Figure 1. Relationship between purchasing decision and gender and marital status

4.2. Prediction Models

Three prediction models: decision tree, random forest, and rule-based classifier, will be compared based on their performances. In all three prediction models, the attributes of marital status, occupation, educational qualifications, family size, ordering medium preferences, meal preferences, and cuisine preferences are eliminated. In all three models, the training data covers 75% of the data, while the remaining 25% is assigned to test data. The method for data training is cross-validation with 10 folds. Finally, they are compared based on prediction accuracy. The prediction accuracy is derived from the confusion matrix which summarizes a classifier's classification performance in relation to some test data. It's a two-dimensional matrix with the true class of an object in one dimension and the class that the classifier assigns in the other [21]. The confusion matrix is frequently used with two classes, one of which is labeled as positive and the other as negative. True positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) are the four cells of the matrix in this context (FN). The following presents a description of each parameter [4].

- True positive (TP): A positive sample predicted by the model.
- False positive (FP): A negative sample predicted by the model as a positive example.
- False negative (FN): The positive sample predicted by the model is used as a negative sample.
- True negative (TN): A sample predicted to be negative by the model.

The prediction accuracy is defined as the number of correct predictions divided by the total number of input samples. It is calculated as the following:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

In this study, all the calculations of confusion matrices, prediction accuracies, and other parameters are performed using R software.

4.2.1. Decision tree

Classification and regression tree (CART): The CART decision tree is used for regression predictive modeling problems. It is a binary recursive partitioning tree, where each parent node in the tree is split into two child nodes [22]. Further, CART is known for its simple interpretation and inherent logic. Here, CART is used to predict the purchasing decisions of online food consumers. From Figure 2, we can obtain the following conclusions:

- The probability of a customer purchasing the next time, who evaluated the “ease and convenience of ordering” and “ordering saving time” elements with more than 3 is $(1-0.06) = 0.94$, and this node covers 77% of the dataset.
- The probability of a customer who will not purchase the next time, who evaluated the ease and convenience element with less than 3 is $(1-0.85) = 0.15$, and this node covers 18% of the dataset.

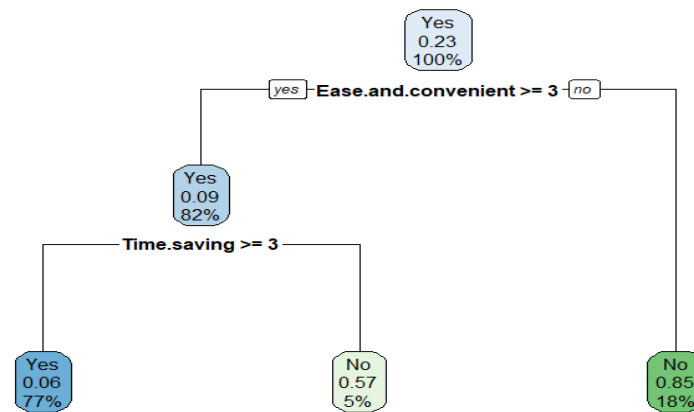


Figure 2. CART decision tree

For training data, the confusion matrix shows that 71.6% of those who will purchase again are classified correctly and 15.4% of those who will not purchase again are classified correctly. On the other hand, 5.8% of consumers are classified as not purchasing, while 7.2% of consumers are classified as purchasing wrongly. The accuracy of the CART tree is 87%. On other hand, the accuracy of this tree on the test data is 84.38%, which implies that CART is performing well.

C4.5 decision tree: In Data Mining, the C4.5 algorithm is utilized as a decision tree classifier, which can be used to make a decision based on a sample of data [23]. It is known for its ability to work with discrete and continuous data as well as handling incomplete data. After implementing the C4.5 tree, its accuracy on both training and testing data outperforms the CART decision tree, as seen in Figures 3 and 4.

Reference			
Prediction	Yes	No	
Yes	71.6	7.2	
No	5.8	15.4	
Accuracy (average) : 0.8699			

Reference			
Prediction	Yes	No	
Yes	74.7	6.2	
No	2.8	16.2	
Accuracy (average) : 0.9098			

Figure 3. Confusion matrices based on training data (CART – C4.5 decision trees)

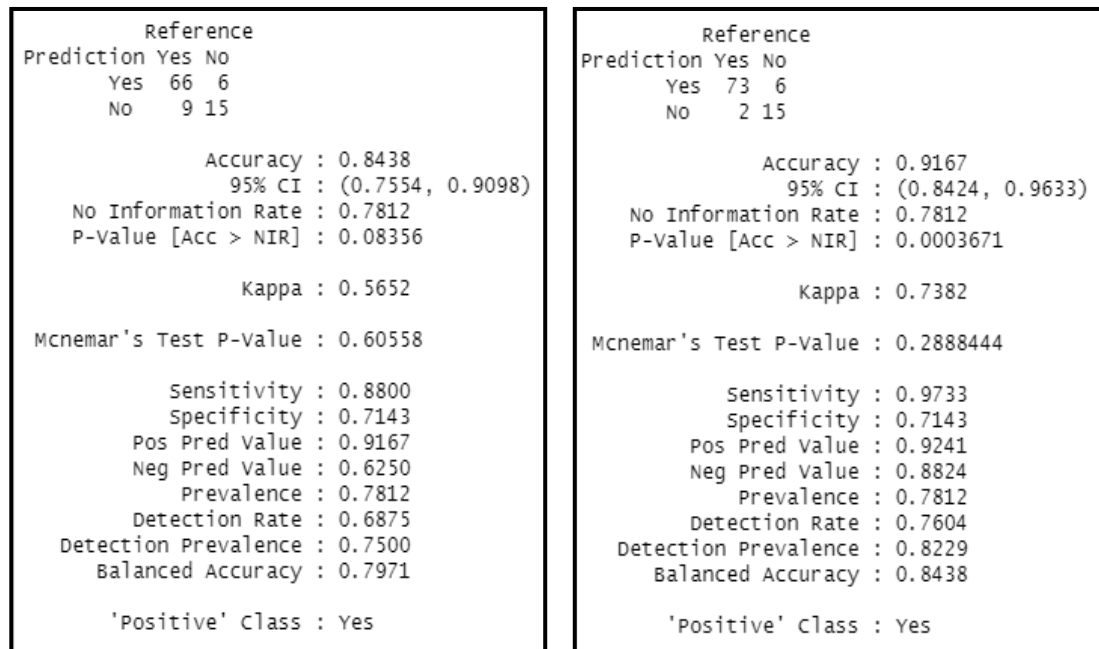


Figure 4. Confusion matrices based on testing data (CART – C4.5 decision trees)

The resulted C4.5 tree is illustrated in Figure 5, and the following conclusions are obtained.

- If the “ease and convenient” and “good taste” are rated with less than or equal to 2, the customer will purchase again with a probability of 100%.
- If the “ease and convenient” is rated with less than or equal to 2 and “good taste” was given a rate of greater than 2, the customer will not purchase again with a probability of 90%.
- If the “ease and convenient” and “time saving ” were given a rate of greater than 2, the customer will purchase again with a probability of 95%.
- If the “ease and convenient” was given a rate of more than 2, “time saving ” was given a rate of less or equal to 2, and “more offers and discounts” was rated greater than 4, the customer will purchase again with a probability of 100%.
- If the “ease and convenient” was given a rate of more than 2, “time saving ” was given a rate of less or equal to 2, “more offers and discounts” was rated less or equal to 4, and the age of the consumer is greater than 25, the customer will not purchase again with a probability of 100%.
- If the “ease and convenient” was given a rate of more than 2, “time saving ” was given a rate of less or equal to 2, “more offers and discounts” was rated less or equal to 4, age of the consumer is less than/equal to 25, and there is no influence of restaurant rating, the customer will not purchase again with a probability of 100%.
- If the “ease and convenient” was given a rate of more than 2, “time saving ” was given a rate of less or equal to 2, “more offers and discounts” was rated less or equal to 4, age of the consumer is less than/equal to 25, there is an influence of restaurant rating, and the rate of “good road condition” is greater than 2, the customer will purchase again with a probability of 100%.
- If the “ease and convenient” was given a rate of more than 2, “time saving ” was given a rate of less or equal to 2, “more offers and discounts” was rated less or equal to 4, age of the consumer is less than/equal to 25, there is no influence of restaurant rating, and the rate of

good road condition is less than/equal to 2, the customer will not purchase again with a probability of 100%.

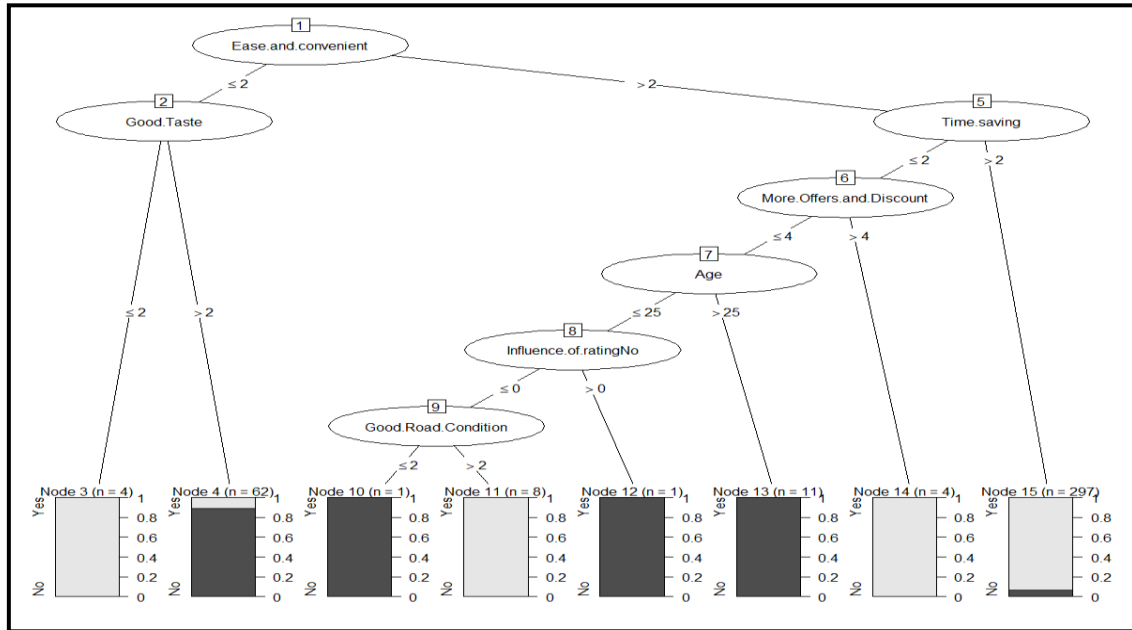


Figure 5. C4.5 decision tree

In other words, the predictors selected by the C4.5 decision tree are the importance of good taste, age, the influence of restaurant rating, ordering ease and convenience, the goodness of road conditions, time-saving, and availability of offers and discounts.

4.2.2. Random forest

A random forest is made up of many separate decision trees that work together to form an ensemble. Each tree in the random forest produces a class prediction, and the class performing the best becomes the prediction of the model [24]. When applied to our problem, it outperforms the C4.5 decision tree on training, with a 94.18% accuracy rate. On the other hand, its performance on the test is comparable to that of the C4.5 decision tree (90.62%).

4.2.3. Rule-based classifier

The rule-based classifier is employed in the class prediction method to give the rules a rating, which is then used to predict the class of future cases [25]. When compared with the other prediction models used, it performs less than the random forest model, in which its accuracy on the training data is 91.44% while on the testing data is 87.5%. Figures 6 and 7 show the comparison between the random forest and the rule-based classifier. The resulted rule-based classifier, shown in Figure 8, draws the following findings.

- If the “ease and convenience” and “time saving” are rated above 2 and “unaffordable” is rated less or equal to 3, the consumer will decide to purchase.
- If the “ease and convenience” is rated less or equal to 3 and “low quantity-time” are rated above 1, and female, the consumer will not decide to purchase.
- If the “ease and convenience” is given a rate of greater than 3. “More restaurant choices” is rated above 2, “good tracking system” is rated less than/equal to 3, a consumer is a female,

and the “delay of the delivery person assigned” was rated less than/equal to 4, the consumer will decide to purchase.

- If the consumer is a female with age less than/equal to 30, who gave a rate above 2 for “more restaurant choices”, he/she will purchase again.
- If the rate of “low quantity-low time” is above 1, “time saving” is less than/equal to 4, “ease and convenient” is greater than 1. And “wrong order delivered” is less than/equal to 4, the consumer will not decide to purchase again.
- If there is an influence of the time, and order time is not on Saturday or Sunday, and “late delivery” was given a rate of less than/equal to 4, the consumer will purchase again.
- If “self-cooking” is given a rate above 3, the consumer will not purchase again.
- If the age of the consumer is less or equal to 25 years, then he/she is expected to purchase again.
- If everything other than the aforementioned is not satisfied, the consumer will not purchase again.

Reference		
Prediction	Yes	No
Yes	76.4	4.8
No	1.0	17.8
Accuracy (average) : 0.9418		

Reference		
Prediction	Yes	No
Yes	73.3	4.5
No	4.1	18.2
Accuracy (average) : 0.9144		

Figure 6. Confusion matrices based on training data (random forest – rule-based classifier)

Reference		
Prediction	Yes	No
Yes	72	6
No	3	15
Accuracy : 0.9062		
95% CI : (0.8295, 0.9562)		
No Information Rate : 0.7812		
P-value [Acc > NIR] : 0.001067		
Kappa : 0.7108		
McNemar's Test P-value : 0.504985		
Sensitivity : 0.9600		
Specificity : 0.7143		
Pos Pred Value : 0.9231		
Neg Pred Value : 0.8333		
Prevalence : 0.7812		
Detection Rate : 0.7500		
Detection Prevalence : 0.8125		
Balanced Accuracy : 0.8371		
'Positive' Class : Yes		

Reference		
Prediction	Yes	No
Yes	69	6
No	6	15
Accuracy : 0.875		
95% CI : (0.7918, 0.9337)		
No Information Rate : 0.7812		
P-value [Acc > NIR] : 0.01391		
Kappa : 0.6343		
McNemar's Test P-value : 1.00000		
Sensitivity : 0.9200		
Specificity : 0.7143		
Pos Pred Value : 0.9200		
Neg Pred Value : 0.7143		
Prevalence : 0.7812		
Detection Rate : 0.7188		
Detection Prevalence : 0.7812		
Balanced Accuracy : 0.8171		
'Positive' Class : Yes		

Figure 7. Confusion matrices based on testing data (random forest – rule-based classifier)

4.2.4. Comparison

As seen in the previous section, there is a difference in the accuracies of the four classification models. However, this difference should be evaluated based on the p-value criterion to determine if the difference is significant. Figure 9 shows that based on accuracy, there is a significant difference between the CART decision tree and the rule-based classifier, and the random forest.

Add to that, based on the conclusions obtained previously, the random forest and C4.5 models perform comparably. However, the accuracy provided by the C4.5 decision is better.

```

PART decision list
-----

Ease.and.convenient > 2 AND
Time.saving > 2 AND
Unaffordable <= 3: Yes (185.0/4.0)

Ease.and.convenient <= 3 AND
Low.quantity.low.time > 1 AND
Ease.and.convenient > 1 AND
GenderMale <= 0: No (26.0)

Ease.and.convenient > 3 AND
More.restaurant.choices > 2 AND
Good.Tracking.system <= 4 AND
GenderMale > 0 AND
Delay.of.delivery.person.getting.assigned <= 4: Yes (13.0)

GenderMale <= 0 AND
Age <= 30 AND
More.restaurant.choices > 2: Yes (17.0)

Low.quantity.low.time > 1 AND
Time.saving <= 4 AND
Ease.and.convenient > 1 AND
Wrong.order.delivered <= 4: No (30.0/1.0)

Influence.of.timeYes > 0 AND
Order.Timeweekend (Sat & Sun) <= 0 AND
Late.Delivery <= 4: Yes (12.0)

Self.Cooking > 3: No (5.0)

Age <= 25: Yes (2.0)

: No (2.0)

Number of Rules :      9

```

Figure 8. Rule-based classifier

Accuracy				
	PART	RandomForest	CART	C4.5
PART		-0.027135	0.045304	0.004845
RandomForest	0.77657		0.072438	0.031980
CART	0.03673	0.02140		-0.040459
C4.5	1.00000	0.37712	0.42723	
Kappa				
	PART	RandomForest	CART	C4.5
PART		-0.06748	0.14927	0.02931
RandomForest	1.00000		0.21675	0.09679
CART	0.11409	0.07706		-0.11997
C4.5	1.00000	0.42192	0.79425	

Figure 9. Comparison between the classification models based on p-value

5. CONCLUSION

In this study, we used several prediction models to determine whether a customer would purchase again from the online food delivery platforms. The ability to do so provides a strong predictive tool for online food delivery providers to have a better understanding of their customers, and to improve their services accordingly. Building the right prediction mode, which combines high

prediction accuracy with sound reasoning, can assist decision-makers in reaching accurate conclusions about the major determinants of customer satisfaction, hence increasing the likelihood of repeat purchases. Past research has considered the implementation of prediction models on purchasing decisions. However, a limited number of studies have incorporated their use into the industry of online food delivery. In this study, we used CART and C4.5 decision trees, a random forest, and a rule-based classifier. The four models performed outstandingly in predicting the purchasing decision, but the C4.5 decision tree performed the best, by providing an accuracy of 91.67%.

Among other algorithms, the C4.5 algorithm is a decision tree algorithm that can be used to build rules that are easy to understand and fast. The approach can also provide a basic model subsystem that can be utilized to support a decision-making system. The C4.5 decision tree has an improved tree pruning strategy that lowers misclassification errors in the training data set owing to noise and too much information. Add to that, it can handle missing attribute values as well as handling different types of data. However, it is only used for small datasets where all or a fraction of the entire dataset must be kept in memory permanently. As a result, its suitability for mining massive databases must be examined. In addition, an improved version of the traditional prediction models must be developed to enhance their accuracy and the time taken to derive the tree. The pruning strategy of C4.5 may allow the trimming of nodes with high value information. Thus, adding enhancements and treatments to the selection of the nodes to be trimmed can increase the output accuracy.

REFERENCES

- [1] C. Li, M. Miroso, and P. Bremer, "Review of Online Food Delivery Platforms and their Impacts on Sustainability," *Sustainability*, vol. 12, no. 14, p. 5528, 2020.
- [2] A. C. Dave and R. Trivedi, "Predicting Youngster's Attitude towards Online Food Delivery," *International Research Journal of Business Studies*, vol. 12, no. 3, pp. 289-299, 2019.
- [3] "Online Food Delivery." <https://archive.is/e7OK5>.
- [4] C.-J. Liu, T.-S. Huang, P.-T. Ho, J.-C. Huang, and C.-T. Hsieh, "Machine learning-based e-commerce platform repurchase customer prediction model," *Plos one*, vol. 15, no. 12, p. e0243105, 2020.
- [5] M. Platzer and T. Reutterer, "Ticking away the moments: Timing regularity helps to better predict customer activity," *Marketing Science*, vol. 35, no. 5, pp. 779-799, 2016.
- [6] G. Tsoumakas, "A survey of machine learning techniques for food sales prediction," *Artificial Intelligence Review*, vol. 52, no. 1, pp. 441-447, 2019.
- [7] D. Van den Poel and W. Buckinx, "Predicting online-purchasing behaviour," *European journal of operational research*, vol. 166, no. 2, pp. 557-575, 2005.
- [8] K. G. Yilmaz and S. Belbag, "Prediction of consumer behavior regarding purchasing remanufactured products: a logistics regression model," *International Journal of Business and Social Research*, vol. 6, no. 2, pp. 01-10, 2016.
- [9] C. Ling, T. Zhang, and Y. Chen, "Customer purchase intent prediction under online multi-channel promotion: A feature-combined deep learning framework," *IEEE Access*, vol. 7, pp. 112963-112976, 2019.
- [10] S. P. Kumar, "Prediction of consumer purchase decision using demographic variables: A study with reference to premium car," *IOSR Journal of Business and Management*, vol. 12, no. 5, pp. 117-120, 2013.
- [11] R. Gupta and C. Pathak, "A machine learning framework for predicting purchase by online customers based on dynamic pricing," *Procedia Computer Science*, vol. 36, pp. 599-605, 2014.
- [12] L. Tang, A. Wang, Z. Xu, and J. Li, "Online-purchasing behavior forecasting with a firefly algorithm-based SVM model considering shopping cart use," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 12, pp. 7967-7983, 2017.
- [13] A. Martínez, C. Schmuck, S. Pereverzyev Jr, C. Pirker, and M. Haltmeier, "A machine learning framework for customer purchase prediction in the non-contractual setting," *European Journal of Operational Research*, vol. 281, no. 3, pp. 588-596, 2020.

- [14] H. Liao and S.-B. Tsai, "Research on the B2C Online Marketing Effect Based on the LS-SVM Algorithm and Multimodel Fusion," *Mathematical Problems in Engineering*, vol. 2021, 2021.
- [15] P. Wang and Z. Xu, "A Novel Consumer Purchase Behavior Recognition Method Using Ensemble Learning Algorithm," *Mathematical Problems in Engineering*, 2020.
- [16] S. Ghosh and C. Banerjee, "A Predictive Analysis Model of Customer Purchase Behavior using Modified Random Forest Algorithm in Cloud Environment," in *2020 IEEE 1st International Conference for Convergence in Engineering (ICCE)*, 2020: IEEE, pp. 239-244.
- [17] N. Chandrasekhar, S. Gupta, and N. Nanda, "Food Delivery Services and Customer Preference: A Comparative Analysis," *Journal of Foodservice Business Research*, vol. 22, no. 4, pp. 375-386, 2019.
- [18] "Size of the online food delivery market across India from 2019 to 2020, with estimates until 2025." <https://www.statista.com/statistics/744350/online-food-delivery-market-size-india/> (accessed August 14, 2021).
- [19] M. D. Anusha and M. P. R. Panda, "A Study On Analysis of Consumer Decision Making Variable on Food Products."
- [20] "Online Food Delivery Preferences-Bangalore Region." <https://www.kaggle.com/benroshan/online-food-delivery-preferencesbangalore-region>.
- [21] K. M. Ting, "Confusion Matrix," in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb Eds. Boston, MA: Springer US, 2010.
- [22] R. J. Lewis, "An introduction to classification and regression tree (CART) analysis," in *Annual meeting of the society for academic emergency medicine in San Francisco, California*, vol. 14, 2000.
- [23] B. Hssina, A. Merbouha, H. Ezzikouri, and M. Erritali, "A comparative study of decision tree ID3 and C4. 5," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 2, pp. 13-19, 2014.
- [24] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R news*, vol. 2, no. 3, pp. 18-22, 2002.
- [25] A. K. H. Tung, "Rule-based Classification," in *Encyclopedia of Database Systems*, L. Liu and M. T. Özsu Eds. Boston, MA: Springer US, pp. 2459-2462, 2009.

AUTHORS

Batool Madani holds a B.Sc. in Nuclear Engineering. She received her M.Sc. in Engineering Systems Management from the American University of Sharjah, U.A.E, in 2019. She is currently a PhD candidate and a Graduate Teaching Assistant at the American University of Sharjah. Her research is oriented around the integration of technologies in the Last Mile delivery problem. Her research interests include wireless technologies, machine learning, decision making, drones, logistics, and optimization.



Hussam Alshraideh is an Associate Professor of Operations Research and Statistics at the Industrial Engineering Department at the American University of Sharjah (AUS). He holds a dual Ph.D. degree in Industrial Engineering and Operations Research with a minor in Statistics from The Pennsylvania State University. He also holds a master's degree in Industrial Engineering/Quality Engineering from Arizona State University. His current research interests include statistical process optimization and smart data analytics applications in healthcare related fields. He has published more than forty papers in highly reputable journals on health informatics and process control.



SHAPEIOT: SECURE HANDSHAKE PROTOCOL FOR AUTONOMOUS IoT DEVICE DISCOVERY AND BLACKLISTING USING PHYSICAL UNCLONABLE FUNCTIONS AND MACHINE LEARNING

Cem Ata Baykara¹, Ilgın Şafak² and Kübra Kalkan¹

¹Department of Computer Science, Ozyegin University, Istanbul, Turkey

²Fibabanka R&D Center, Istanbul, Turkey

ABSTRACT

This paper proposes a new lightweight handshake protocol implemented on top of the Constrained Application Protocol (CoAP) that can be used in device discovery and ensuring the IoT network security by autonomously managing devices of any computational complexity using whitelisting and blacklisting. A Physical Unclonable Function (PUF) is utilized for the session key generation in the proposed handshake protocol. The CoAP server performs real-time device discovery using the proposed handshake protocol, and anomaly detection using machine-learning algorithms to ensure the security of the IoT network. To the best of our knowledge, the presented PUF-based handshake protocol is the first to performs blacklisting and whitelisting. Whitelisted IoT devices not displaying anomalous behavior can join and remain in the IoT network. IoT devices that display anomalous behavior are autonomously blacklisted by the CoAP server and are either disallowed from joining the IoT network or are removed from the IoT network. Simulation results show that amongst the five machine learning algorithms studied, the stacking classifier displays the highest overall anomaly detection accuracy of 99.98%. Based on the results of the network simulation performed, the CoAP server is capable of blacklisting malicious IoT devices within the network with perfect accuracy.

KEYWORDS

IoT Networks, Network Security, Handshake Protocols, Anomaly Detection, Machine Learning

1. INTRODUCTION

IoT has gained immense mind share in both academic and industry alike over the past several years. In our everyday lives, IoT enables devices to be aware of their surroundings, efficiently communicate, and ultimately create a better environment for the people [1]. Devices from the same owner effectively forms a smart environment for that owner where each device can communicate and effectively combine their strengths to overcome their weaknesses [2]. To be able to recognize and utilize the potential of IoT, secure discovery and access control is essential [2]. However, these devices have differing computational complexities, and not all of them are equipped with means to prevent themselves from malicious malware. Many IoT devices are manufactured with inherent security vulnerabilities [3]. These vulnerabilities often pose a risk for the security of an entire IoT network that includes vulnerable devices[4]. In a centralized IoT

network, it is often the responsibility of the central entity to ensure the availability and security of the network[5].

This paper proposes a lightweight and secure handshake protocol that is computationally inexpensive and suitable for detection of IoT devices of any computational complexity for the purpose of device discovery and device management in an IoT network. A Physical Unclonable Function (PUF) is used in securely generating a session key by the IoT device. An autonomous anomaly detection, whitelisting and blacklisting approach is proposed to prevent unwanted or malicious devices from joining, re-joining, or remaining in the IoT network. IoT network simulation results are provided of the proposed handshake protocol, where the CoAP server in the network performs real-time autonomous anomaly detection using pre-trained machine learning (ML) algorithms. To the best of our knowledge, the handshake protocol proposed in this paper is the first to address both the computational complexity and security challenges of IoT devices in the device discovery utilizing PUFs and a whitelisting and blacklisting approach in autonomously ensuring IoT network security. The contributions of this paper are as follows.

1. A new lightweight handshake protocol implemented on top of CoAP that utilizes PUFs to generate the secure session key, in addition to a whitelisting and blacklisting approach in ensuring IoT network security is proposed. To the best of our knowledge, this is the first PUF-based handshake protocol that performs blacklisting and whitelisting. The computationally inexpensive property of PUFs allows even the most basic IoT devices to access an IoT network by using the handshake protocol.
2. The proposed protocol allows central or distributed authorities of the IoT network to establish negotiated communications, provide novel services, and autonomously prevent undesired devices from joining, re-joining, or remaining in the IoT network using whitelisting and blacklisting.
3. This paper proves that the proposed handshake protocol is secure against our threat model, including Man in the Middle (MITM), forgery and replay based attacks, with a security analysis.
4. The paper presents an IoT network simulation to test the effectiveness of the proposed handshake protocol. The simulation includes a CoAP server that performs real-time anomaly detection using pre-trained machine learning algorithms. The paper also provides different machine learning classifiers that can be used and presents a detailed discussion and analysis of which classifier is preferable.
5. Based on the analysis and network simulation performed, the handshake protocol proposed in this paper can easily be adapted to real-life IoT networks.

The remainder of this work is structured as follows. Section 2 presents the existing approaches and related work. Section 3 presents a description and the security analysis of the proposed handshake protocol in detail. Section 4 presents the experimental work performed, namely the network simulation and its results. Finally, Section 5 discusses the conclusions of this work.

2. RELATED WORK

This section provides a summary of related work and discusses the differences between the proposed handshake protocol.

In the Transport Layer Security (TLS) protocol, session keys are randomly generated by the client [6]. This approach is prone to replay and forgery attacks, which is dependent on how the client generates the random string. Compared to the TLS protocol, the proposed handshake protocol in this work offers a more lightweight and secure approach to mutual authentication by

generating session keys using a PUF and utilizing a whitelisting and blacklisting approach in autonomously ensuring IoT network security using machine learning.

An approach to using PUFs to address the problems of low computational power of IoT devices is proposed in [7]. This work shows in detail the benefit of using PUFs to establish a secure session with IoT devices by comparing their proposed protocol with other existing solutions such as Datagram Transport Layer Security (DTLS) handshake protocol and User Datagram Protocol (UDP). The results indicate that with the use of PUFs, the authentication process results in a reduction in power of up to 45% by also using 12% less memory, compared with the existing solutions listed before. The authors propose a method of detecting artificially generated challenge-response pairs (CRPs) by using neural networks that their proposed verifier (server) uses to authenticate the nodes. By employing this technique, the server can detect malicious nodes that try to authenticate with the server by using replayed CRPs. However, the protocol can only authenticate existing and trusted nodes in the network; it does not utilize whitelisting or machine learning techniques for autonomous blacklisting of suspicious devices as proposed in this paper. Thus, it is not suitable for general and public use since it constrains the system from allowing new nodes to join the network. In contrast, the protocol presented in this paper, while keeping a record of previously explored nodes, allows new nodes to join the network and uses whitelisting and blacklisting techniques in ensuring the security of the IoT network.

Lightweight key exchange protocols that use pre-shared secret symmetric keys are proposed in [8]-[9]. These protocols assume that one or more symmetric keys are readily available to the client and the server before the handshake protocol begins. However, this is not a secure or scalable approach since it requires every new client to obtain the shared keys a forehand and is prone to forgery and brute force attacks. In comparison, the proposed handshake protocol generates a secure session key on-the-fly without requiring pre-shared symmetric keys, as well as utilizing whitelisting and blacklisting in ensuring the network security.

Lightweight key exchange protocols using mutual authentication are proposed in [10]-[11]. These protocols use PUFs not only for generating session keys, but also for registration and authentication purposes. However, these protocols do not study autonomous blacklisting techniques in ensuring the network security. The protocols store long-term keys to authenticate the server or a device, a process which requires extensive computation and data storage. Given the limited computational capability of IoT devices, the proposed protocol in this paper tries to minimize the computation required from the IoT nodes for server authentication. Thus, the proposed protocol in this paper authenticates the server with the help of a Certificate Authority (CA). However, using CA has its downsides as it requires the assumption that the CA will always be available since the protocol cannot be completed otherwise. Using CA for authentication also increases the communication flow required to perform the handshake. The proposed protocol in this paper performs the key exchange with 7 communication steps, whereas the protocol proposed in [11] only needs 3 communication steps for a registered user. However, unlike these PUF-based authentication and key exchange protocols, the protocol presented in this paper lets any user to initiate the key exchange without client registration or client verification processes.

Similar handshake protocols are proposed in [12]. These protocols also utilize the low computational complexity of PUFs for IoT devices and provide detailed analysis in terms of computation, memory, and communication overhead. The results in [12] show that using PUFs for secure session generation is a suitable solution for IoT devices. Another protocol proposed in [13] can generate a secure session between IoT devices. This protocol uses a trusted and pre-authenticated server for secure session generation, similar to the handshake protocol proposed in this paper, which uses a CA for server authentication. Results in [13] are also in agreement with the previous related work on showing the effectiveness of using PUFs instead of using existing

solutions. However, even though authors of [12] state that their proposed protocols can be adapted to real-life use cases, they do not provide a real experimental setup or a real time simulation to prove this statement. In contrast, this paper provides the results and details of a real time network simulation using the handshake protocol proposed to show that it can easily be adapted to real life use cases.

[14] proposes a secure PUF based authentication and identity-based key exchange protocol suitable for a distributed IoT network. It differs from this work in that a certificate-less identity based key exchange approach is used, and that it does not study whitelisting and blacklisting in ensuring the network security.

[15] proposes a lightweight mutual authentication protocol based on a new public key encryption scheme that uses the encryption scheme to transmit challenges and check whether the recipient can respond accordingly. The protocol is shown to have a performance significantly better than existing RSA and ECC based protocols. It does not require the use of a CA for authentication, as proposed in this paper. Additionally, whitelisting and blacklisting using ML techniques in ensuring the network security are not studied.

In [16], a new lightweight mutual two-factor authentication mechanism is proposed, where an IoT device and server authenticate each other and establish a key exchange using PUFs and a hashing algorithm. The main difference between this paper is that a CA is not utilized in mutual authentication, and whitelisting and blacklisting in ensuring the network security is not considered.

[17] and [18] propose variations of the Datagram Transport Layer Security (DTLS) handshake protocols for IoT networks. [17] proposes a simplified version of the Datagram Transport Layer Security (DTLS) handshake protocol suitable for IoT devices for a general scenario of end-to-end communications based on software-defined networking (SDN). A controller is utilized in generating a symmetric key dynamically, then encrypting and distributing the key to two communicating IoT devices. Certificate verification is shifted from the IoT device to the more powerful controller, where the controller replaces the DTLS server to make a cookie exchange with the DTLS client. The computational overhead and the energy consumption in the IoT devices and the overall duration of the handshake protocol are shown to reduce. [18] separates the DTLS protocol into the handshake phase and the encryption phase, which is shown to enhance the performance in both the device and the network by using a way to delegate the DTLS handshake phase. The proposed scheme supports secure end-to-end communication despite using delegation. However, neither paper utilizes whitelisting or blacklisting methods in ensuring the network security, as proposed in this paper.

[19] proposes a different fingerprinting approach for keeping a log of previously known clients and detect malware and other malicious processes trying to initiate a secure connection using the TLS handshake protocol before the secure session is established. The fingerprinting technique proposed in [19] uses the initial unencrypted hello message sent by the clients after the TCP connection is established. Since this message is unencrypted, TLS fingerprinting extracts metadata presented in the message and generates a fingerprint string using a pre-defined schema. After generating a fingerprint, the server then maps it to the client by keeping a dictionary of known fingerprint to client mappings. By doing so, the server can detect previously known or blacklisted clients trying to initiate a connection and take actions accordingly. However, it is stated that the TLS fingerprinting technique is not sufficient per se to profile clients effectively. A single fingerprint may map to tens or hundreds of unique clients. Thus, the TLS fingerprinting itself is often a poor indicator and additional information is required to increase its performance. The handshake protocol presented in this paper uses PUFs for both client fingerprinting and

session key generation. In order to verify the identity of a client, the server may log the challenges and the session keys used to create sessions and authenticate clients by sending the logged challenges. Since the clients in the proposed handshake protocol use their PUFs to generate the session keys, the client is expected to generate the same session key given the same challenge where only a specific client can create that unique session key due to how PUFs work. This approach is both more precise compared to [19], and it does not require additional information about the client for fingerprinting. Hence, it is more lightweight than TLS fingerprinting technique in [19] which puts additional computational overhead on the server.

3. HANDSHAKE PROTOCOL

CoAP is a specialized internet application layer protocol that allows constrained nodes to communicate with the Internet or with each other in a lightweight way suitable for IoT devices [20],[21], [22], [23]. CoAP provides a request, and response-based interaction model between nodes, while supporting built-in discovery of services and resources. It is designed to easily interface with HTTP to be used on the Web. The key features of CoAP that encouraged us to implement the proposed handshake protocol on top of it includes, but are not limited to [23]:

1. Web protocol fulfilling machine-to-machine (M2M) requirements in constrained environments.
2. Asynchronous message exchanges.
3. Low header overhead and parsing complexity.

PUF is a physical object that for a given input (challenge) produces an unclonable output (response) that serves as a unique identifier for that specific object [24]. The response generated by the PUF can also be called as the digital fingerprint of that object. This concept is often achieved by a semiconductor device such as a microprocessor.

The inimitable feature gives PUFs an advantage over other hardware-based security concepts as even if the attacker has physical access to the device, they cannot clone its intrinsic properties [24]. Thus, PUFs enable us to perform device identification and authentication in a secure manner. It is also stated in [24] that PUFs provide a low-cost alternative when compared with the conventional methods for cryptographic key generation, making them a suitable solution for low complexity IoT devices. The handshake protocol presented in this paper utilizes PUFs to generate the secure session key.

3.1. System and Adversary Model

3.1.1. System Model

The ideal model proposed in this paper assumes that the protocol is used to serve all devices ranging from high complexity power to lightweight IoT devices. The proposed ideal model scheme is depicted in Figure 1. The IoT devices generate the network traffic with the CoAP server after establishing the secure session using the proposed handshake protocol. The malicious devices are assumed to be able to listen every message within the network traffic and can also connect to the main server like other devices. All devices perform the key exchange protocol to achieve secure communication with the server. After concluding the handshake protocol, devices proceed to use the CoAP for secure communication. The proposed model does not assume initially trusted devices, so any device that can provide the necessary information during the key exchange protocol obtains a secure key to communicate. On the other hand, devices must verify the legitimacy of the server by sending the server's certificate to the CA during the handshake

protocol. This verification process is assumed to be secure and encrypted between the device and the CA and its security is not within the scope of this paper.

This delegated approach for IoT devices provides efficiency and consistency, which is crucial for use cases that require both communicational integrity and low computational power such as secure payment from IoT devices. The server adds to the IoT network the devices that successfully complete the handshake protocol. The server also performs autonomous real time anomaly detection on the IoT network traffic. Anomalous network traffic, which may pose a threat to the network, is detected by the server and the devices that produce the anomalous traffic are removed from the IoT network and recorded into the blacklist. Further network traffic from the devices within the blacklist is declined by the server.

The proposed model provides a secure and lightweight solution for IoT device communication on CoAP. The key exchange message sizes are kept within the boundaries of the CoAP protocol. The model uses asymmetric encryption and decryption once to establish a secure communication channel with symmetric keys.

3.1.2. Adversary Model

The attacker model provided in [25] is assumed in this paper. The attacker can listen, replay, and create messages in the network, where the goals of the attacker are as follows:

- To generate or obtain the key used in the session.
- To acquire device information.
- To obtain the confidential information shared within the communication.

The aim is to provide a secure connection to any device that successfully performs the key exchange protocol. Therefore, the denial of service and jamming attacks are out of the scope for the key exchange protocol analysis.

3.2. Protocol Description

This section presents the proposed handshake protocol in detail. The protocol allows IoT devices to establish a secure session for communication by utilizing asymmetric encryption and physical unclonable functions. Clients and servers that follow the protocol can securely generate a one-time session key to achieve end-to-end encryption for secure communication. Definition of acronyms used within this section are provided in Table 1.

The server makes a lookup on the blacklist for the C_{ID} received from the client. If the C_{ID} is not blacklisted, the server generates a random nonce SN and XOR's this with the CN_0 to generate ch . The server then sends a response to the client's initial hello message by a hello message of its own including SN , ch , and its sc . Upon receiving the server's hello message, the client first XORs

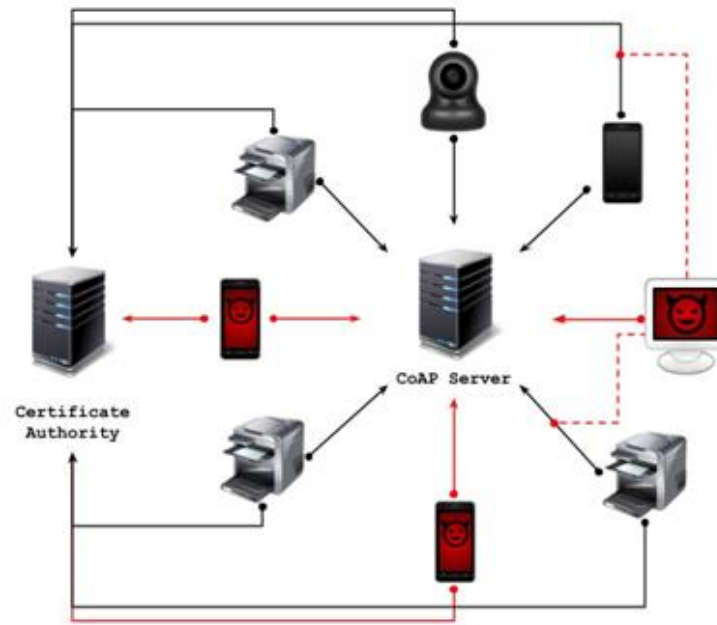


Figure 1. The proposed network model.

the NS and the ch to reconstruct its CN_0 and check if the server received and used CN_0 when generating ch . If the client cannot reproduce its own CN_0 from the ch and NS , the server is not trusted and the client terminates the protocol immediately, otherwise the protocol proceeds as normal. After obtaining the sc , the client checks its legitimacy through the CA. The client sends out a message containing the sc to the certificate authority along with NC_I , to show that the communication between the client and the CA is live. The CA responds to the client by sending out a message saying either OK or NOK. If the client receives NOK from the CA, the server is not trusted and the client terminates the protocol immediately, otherwise if the client receives an OK, this means that the server is trusted, and the protocol may proceed to the session key generation phase. The communication between the client and the CA is assumed to be secure (e.g., TLS, pre shared key), and the security between the CA and the client is not within the scope of this paper. After the client confirms the identity of the server through CA, it will now generate the K_S by using the ch . The client uses its PUF to generate the K_S . The client directly uses the ch as an input for the PUF, generating a one-time and non-replicable session key K_S .

Table 1. Definitions and acronyms.

Acronym	Definition
CA	Certificate Authority
C_{ID}	Client ID
CN_0	First Client Nonce
CN_I	Second Client Nonce
SN	Server Nonce
ch	Challenge
sc	Server Certificate
S_{ID}	Session ID
K_M	Private Key
K_P	Public Key
K_S	Session Key
N_8	8 Bytes Nonce
D_i	Device Information

The client then extracts the K_P from the sc and sends the K_S by encrypting it with the K_P of the server. To prevent replay attacks, the client also includes the ch inside this message. The server then decrypts the session key by using its K_M . At this point the server also generates a S_{ID} , by feeding the sum of C_{ID} , ch and K_S to the SHA256 function. To ensure that the S_{ID} is unique every time, the server also inserts a random numeric value with a length of 8 bytes to the end of the output given by the SHA256 function. After obtaining the K_S the server sends out a finished message to the client, encrypted with the K_S . This finished message includes the newly generated S_{ID} , to prevent replay attacks. This point is very important as the client will decrypt the finished message of the server to confirm that the server received the K_S , but more importantly that the server decrypted the message 5. If message 6 cannot be decrypted by the client using the K_S , this may signal that the server either cannot decrypt the message 5 and is an untrusted server who does not have access to the K_M , or either the message 5 or 6 has been tampered with. In either of these cases, the protocol is terminated and must be restarted from the beginning. If, however, message 6 can be decrypted by the client using the K_S it sends out a final finished message back to the server which includes its device type and Operating System (OS), and the secure session is generated, and secure symmetric encryption is achieved. The ensuing session is continued using the K_S as the symmetric key.

3.3. Security Analysis

Based on the model described, the security analysis of the proposed protocol is provided in the Section. The likelihood of an attacker breaking the security guarantees of the protocol to achieve his malicious goals is examined, where the analysis is based on the following assumptions:

1. The signature scheme used by the protocol participants (server and client) and the CA is secure (it is impossible for the attacker to forge a signature without the private key).
2. Both the server and client nonce is only picked twice with inconsequential probability.
3. It is not possible to clone or copy the session key generation function used by an arbitrary client.
4. The communication between the client and the CA is secure.

Based on these assumptions, the proposed handshake protocol provides five guarantees.

3.3.1. Guarantee 1

If the protocol is completed successfully, a private session key is generated which is only known by the server and the client. For an attacker to gain access to the session key, the message 5 which is carrying the encrypted session key created by the client must be decrypted. This message is encrypted by the server's public key and can only be decrypted using the server's private key. The only way for the attacker to decrypt this message is by creating the server's private key. An attacker can never access the private key by assumption 1, so assuming the attacker cannot retrieve the physical storage of the server where the private key is stored, the attacker cannot decrypt the session key.

3.3.2. Guarantee 2

For both the client and the server, the protocol guarantees that mutual authentication is achieved. If an attacker creates a fake server, to proceed with the handshake protocol, he is required to provide the clients trying to connect his server with a signed certificate. Since a client confirms the identity of the server through the certificate authority, the two ways an attacker can pose as the real server is by creating or replaying an OK message in message 4 after the client presents the attacker's certificate in message 3. The assumption 4 states that the session between client

and the certificate authority is secured. The attacker cannot create message 4 for an illegitimate certificate because message 4 is encrypted with unique session keys only known by the real server and the client. If the attacker replays the message 4, an arbitrary client receiving the message 4 will not obtain an OK message after decrypting it with its unique session key, because the encrypted OK/NOK message is unique for every client.

3.3.3. Guarantee 3

The protocol guarantees that a secure and unique session key is generated for each session. The attacker has two options for obtaining the session key: creating the message 5 or replaying the message 5. To replicate the session key for creating the message, the attacker needs the same key generation function and its unique parameters. These parameters are client nonce, server nonce and the timestamp information. Client nonce information is shared in messages 1 and 2. In the case of an attacker capturing both messages and obtaining client and server nonce, the attacker still cannot retrieve any information regarding the timestamp since the timestamp information is never shared in the protocol. By assumption 3, even when the attacker has the client nonce, server nonce and timestamp information, the attacker cannot replicate the key generation process because a PUF is used by the client to generate the session key. For a replay attack to work, the adversary must force both the client and the server to choose nonces to generate the same server challenge. The assumption 2 states that the probability of getting the same nonce values twice is negligible. Hence, the replayed challenge value cannot match with the current session challenge. Therefore, by replaying message 5, the attacker cannot get the expected response in message 6.

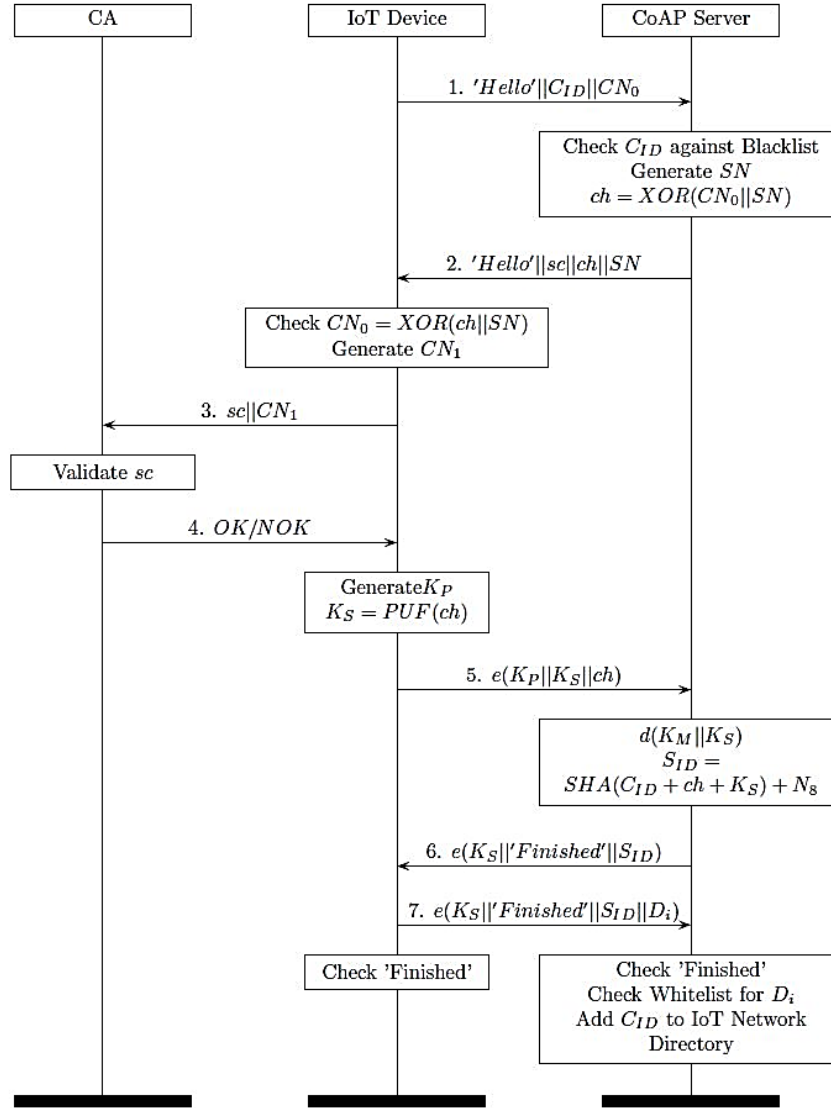


Figure 2. Secure Session Key Generation Protocol using PUF

3.3.4. Guarantee 4

For the client, the protocol guarantees that the generated session is with the trusted server. After sending the encrypted session key to the server in message 5, client now listens for a message that is encrypted with the session key. If this message results in the string “finished” after decryption, client confirms that the server has the session key. An attacker trying to act as the server has two options: replaying the message or creating the message and sending it to the client. Each session has a unique session key that is sent encrypted through message 6. If the attacker replays the captured message 6, an arbitrary client will not obtain the “finished” message after decryption for the similar reasons explained in proof 3. Therefore, the Attacker cannot complete the key exchange protocol using a replay attack. To create message 6 the attacker needs the session key. The attacker cannot recreate the session because of assumption 3, the client generates the session key using a PUF. PUF is a physical entity, which resides within the client and its behavior can never be replicated by the attacker. Therefore, if the client receives message 6, then the sender is guaranteed to be the trusted server and the real sender of the message.

3.3.5. Guarantee 5

If the server receives the encrypted finished message, the client proves that it can use the session key it sent in the message 5. After sending the message 6, the server listens to an encrypted message from the client that reads as “finished” after decryption. An attacker trying to pose as the client has 2 options: replaying the or creating the message 7. Each session has a unique session key encrypted message 6. If the attacker replays the captured message 7, the server will not obtain the “finished” message after decryption for the same reasons explained in proof 3. Therefore, the Attacker cannot complete the key exchange protocol using a replay attack. To create the message 7 the attacker needs the session key. By assumption 3, the attacker can never recreate the session key. Thus, the attacker cannot create the encrypted “finished” message. If the server successfully gets this message, it guarantees that the client which sent the session key can also use it.

4. EXPERIMENTAL WORK

This section will present information and discussion of the IoT network simulation performed to measure the effectiveness of the proposed handshake protocol in detail. The network simulation includes a central CoAP server and 30 IoT nodes. By utilizing the device information obtained through the handshake protocol, the server rejects certain devices and only allows devices which satisfy certain specifications to enter to the network. By performing autonomous real-time anomaly detection using machine-learning, the server aims to detect anomalous or malicious network traffic and block further network traffic generated by suspected malicious nodes. This section will provide the dataset used to train the classifiers, while also comparing five different machine learning classifiers, which were considered based on their performance and suitability for real-time anomaly detection. This section will also discuss the simulation environment, results obtained from the simulation, as well as how the preferred classifier performed on the simulation.

4.1. The Dataset Description

The IoTID20 dataset is used as the main source of data [26]. The data is generated from a common smart home setup where victim and attacking devices are present. The IoTID20 data contains over 625,000 instances which consists of 80 network features and 3 label features which can be seen below as:

- Binary: Normal, Anomaly
- Category: Normal, DoS (Denial of Service), Mirai, MITM (Man in the Middle), Scan
- Subcategory: Normal, Syn Flooding, Brute Force, HTTP Flooding, UDP Flooding, ARP Spoofing, Scan Host Port, Scan OS

[26] states that the most important benefit of the IoTID20 data is that it replicates a modern approach of IoT device communication, and it is among the few publicly available IoT intrusion detection datasets. The features present in the data is ranked using the Shapira-Wilk algorithm to measure the regularity of the distribution of instances with respect to the feature. They argue that more than 70% of the features ranked above 0.50 and state that these high ranked features will improve the classification capability of detection algorithms and techniques. To generate the data to be used for the network simulation, 20,000 instances were randomly chosen from each of the five main categories. To use the data within the network simulation, additional features were added which are not used by the classifiers during the training phase, and the dataset is revised to be compatible with the CoAP protocol.

4.2. Machine Learning Techniques for Autonomous Anomaly Detection

Machine learning is used for autonomous real-time anomaly detection in this work. Usage of machine learning classifiers for intrusion and anomaly detection for network security is thoroughly researched in the past decade [27], [28]. The classifiers studied in this paper include:

- Random Forest (RF)
- Decision Tree (DT)
- Stacking
- K-Nearest Neighbors (KNN)
- Gaussian Naïve Bayes (GNB)

DT, KNN and GNB classifiers are preferred due to their suitability for intrusion and anomaly detection [28], [29]. RF and Stacking classifiers are also included in this paper because they perform relatively well on the IoTID20 data as presented in [26]. The stacking classifier is used in this paper as the ensemble classifier.

4.2. Performance Evaluation of the Classifiers

The performance of each classifier on the test data can be seen from Table 2 and Table 3. It can clearly be seen that from the five different classifiers trained, the random forest, decision tree and the stacking classifiers performs the best on the IoTID20 data. The performance of the KNN classifier is similar with the RF and DT classifiers. However, it is stated in [30] that the KNN is not applicable for critical real time systems where high amounts of training samples are present. Furthermore, since the classification of a new instance x requires the calculation of all the distances between x and the training data in the KNN algorithm, it comes with a significant computational cost. Finally, the GNB classifier performed the worst considering the performances of other classifiers, as can be observed from Figure 3, Table 2 and Table 3.

Considering only the best performing three algorithms, namely the RF, DT and stacking classifiers, for the network simulation purposes, the stacking algorithm was preferred in this paper. Although the stacking algorithm is an ensemble classifier and it requires more computational power and resources for training, the difference in performance between the stacking algorithm and other better performing classifiers cannot be neglected. Moreover, from Figure 3, it is observed that the stacking algorithm is by far the best classifier for detecting scan port OS attacks on the IoTID20 data, outperforming other classifiers drastically.

Table 2. Macro avg. performances of the classifiers trained on category as target label.

Classifier	Accuracy	Precision	Recall	F-1 Score
Random Forest (RF)	0.90	0.93	0.90	0.90
Decision Tree (DT)	0.90	0.93	0.91	0.91
Stacking	0.91	0.93	0.92	0.92
K-Nearest Neighbor (KNN)	0.87	0.91	0.87	0.86
Naïve Bayes (GNB)	0.56	0.66	0.56	0.56

Table 3. Macro avg. performances of the classifiers trained on subcategory as target label.

Classifier	Accuracy	Precision	Recall	F-1 Score
Random Forest (RF)	0.82	0.73	0.74	0.71
Decision Tree (DT)	0.85	0.78	0.77	0.74
Stacking	0.89	0.86	0.85	0.84
K-N Neighbor (KNN)	0.78	0.71	0.72	0.69
Naïve Bayes (GNB)	0.48	0.46	0.43	0.40

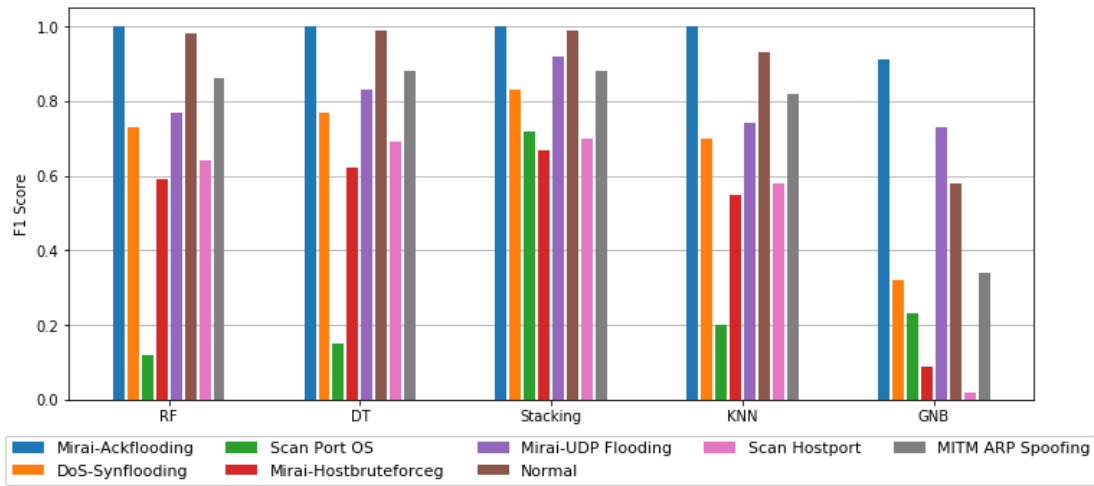


Figure 3. F1 Scores of the classifiers on the test data for subcategory label.

The classification duration is also considered when choosing the classifier to use during the network simulation. The stacking algorithm takes 0.06 ms to classify a single instance, whereas the decision tree classifier takes only 0.01 ms. This is to be expected due to the stacking classifier being computationally more expensive, but the difference between the two is quite negligible for the network simulation performed in this paper. However, if the number of IoT devices were to increase considerably in the network, the CoAP server could experience problems detecting anomalous network traffic using the stacking algorithm in real time. In such a case decision tree classifier should be preferred to stacking classifier.

4.3. Network Simulation

The network simulation which uses the proposed handshake protocol presented in this paper is implemented on Python 3.7. To simulate the distributed network a single central CoAP server, another server to act as the CA, and 30 IoT devices as the client nodes were used. The IoT devices range from simple and lower complexity devices such as smart cameras, controller and hubs, smoke alarms and printers, to more complex devices such as laptops, PCs, phones, and tablets. Each of these devices are given a unique client ID as required within the handshake protocol. The simulation assumes that none of these devices are previously known by the CoAP server, and all of them are required to perform the handshake protocol with the server to join the IoT network securely.

A total of 30 IoT devices are used to simulate the IoT network. Information and the assigned roles about these devices (see Table 4 and Table 5).

Table 4. Low complexity IoT devices used within the network simulation.

Device Name	Type	OS	ID	Role
Amazon Echo	Controller/Hubs	Nan	1	Victim
Belkin Motion Sensor	Energy Management	Nan	2	Victim
LIFX Light Bulb	Energy Management	Nan	3	Victim
iHome Power Plug	Energy Management	Nan	4	Victim
Belkin Switch	Energy Management	Nan	5	Victim
TP Link Power Plug	Energy Management	Nan	6	Victim
Netatmo Camera	Cameras	Nan	7	Victim
Nest Drop Camera	Cameras	Nan	8	Victim
Samsung Smart Camera	Cameras	Nan	9	Victim
TP Link Camera	Cameras	Nan	10	Victim
HP Envy Printer	Appliances	Nan	11	Victim
Pixstar Photo Frame	Appliances	Nan	12	Victim
Tribby Speaker	Appliances	Nan	13	Victim
Withthings Sleep Sensor	Health-Monitor	Nan	14	Victim
Blipcare BP Meter	Health-Monitor	Nan	15	Victim
Netatmo Weather Station	Health-Monitor	Nan	16	Victim
Nest Smoke Alarm	Health-Monitor	Nan	17	Victim
Withthings Scale	Health-Monitor	Nan	18	Victim

Table 5. High complexity IoT devices used within the network simulation.

Device Name	Type	OS	ID	Role
Samsung Galaxy Tablet	Tablet	Android	19	Victim
Android Phone	Phone	Android	20	Victim
Laptop	PC	Ubuntu	21	Victim
MacBook	PC	Mac OS	22	Victim
Android Phone 2	Phone	Android	23	Victim
iPhone	Phone	iOS	24	Victim
MacBook 2	PC	Mac OS	25	Victim
iPhone 2	Phone	iOS	26	Malicious
iPhone 3	Phone	iOS	27	Malicious
MacBook 3	PC	Mac OS	28	Malicious
Laptop 2	PC	Windows	29	Malicious
Laptop 3	PC	Windows	30	Malicious

After generating a secure session with the CoAP server using the proposed handshake protocol, each device starts generating a network traffic. The devices were split into malicious/anomalous and benign devices as stated before. Each packet sent by the IoT devices were chosen from within the test data as discussed within the dataset section. Each device picks a random packet information to send from within its assigned instances in the test data. For ease of implementation and simulation, the information about the packets picked and sent by the devices are known by the CoAP server. This information is used by the server to perform real time anomaly detection using the trained stacking classifier as discussed before.

The CoAP server adopts a whitelisting approach. In the presented setup, higher complexity devices such as PCs, phones, laptops, and tablets are the whitelisted devices. As such, after the CoAP server receives the type and OS information of each device during the handshake protocol. The connection of devices which does not satisfy the whitelist policy described before are declined. Thus, the devices which satisfy the above conditions can generate a secure session with the CoAP server and are registered to the IoT network. Currently the type of a device and its

respective OS is enough to be added into the IoT network. However, more specific whitelist policies can easily be added to tend to different scenarios if needed.

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

The network traffic each individual IoT device generates within the simulation follows a Poisson distribution. The Poisson distribution expresses the probability of a given number of events occurring in a fixed amount of time. It assumes that these events occur independently with a constant average rate. It is widely used for network and communication traffic simulations [31]. Poisson distribution can be expressed using the equation above where x denotes the number of occurrences and λ denotes the expected number of occurrences. A Poisson distribution with $\lambda=3$ packets sent per second by each client is used for the network simulation in this paper. It is important to mention that the λ can be any value as long as the CoAP server can handle the generated network traffic.

It is to be noted that the packet collision, loss, and faulty packets are omitted during the network simulation for convenience. The number of packets received by the CoAP server at each second throughout a 3-minute simulation can be seen from Fig. 4. The initial increase in the number of packets seen at the start of the simulation is due to each device performing the handshake protocol before being accepted to the IoT network. Due to the whitelisting policy, the connection of devices with IDs 1-18 are declined by the server during the handshake protocol. Thus, only the other 12 devices can join the IoT network and begin generating the network traffic. The number of packets at each second converges to around 36 as expected from 12 devices generating a network traffic based on a Poisson distribution with $\lambda=3$ packets sent per second.

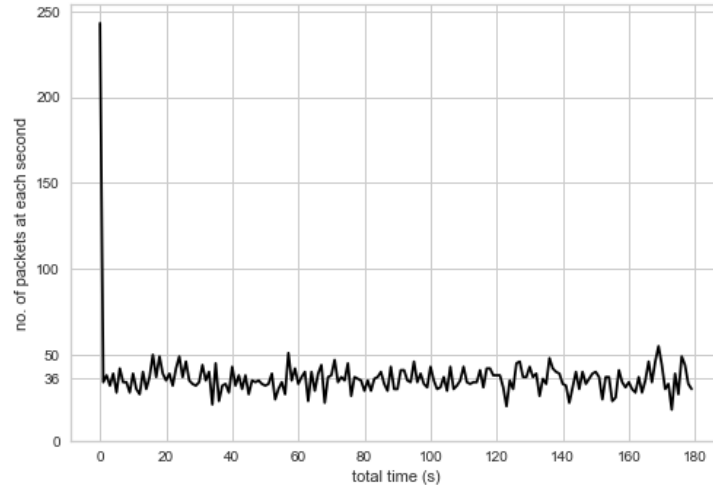


Figure 4. Number of packets received by the CoAP server at each second.

The CoAP server feeds the information about each packet within the IoT network traffic into the classifier for autonomous anomaly detection in real time. The binary performance of the classifier, which can be seen from Table 6, is measured as the classifier's ability to detect whether a packet is an anomaly or not. From Table 6, it is shown that the classifier performs almost perfectly when detecting anomalous packets, with a false negative rate of only 0.038%. The classifier classified some packets in the benign network traffic as anomalous, which can be seen from the misclassified instances detected from victim devices, yielding a false positive rate of 1.702%. The overall accuracy of the classifier in the binary setting is 99.98%, misclassifying only

67 packets out of 6591 packets in the IoT network simulation. The CoAP server inside the proposed network simulation setup was able to blacklist all the malicious devices with perfect accuracy whereas none of the benign devices were blacklisted by accident.

Table 6. Performance of the classifiers on detecting anomalous packets correctly from each device.

Device ID	Total Packets	Misclassified Packets	Accuracy
19	562	10	98.22%
20	558	15	97.31%
21	573	10	98.25%
22	556	9	98.38%
23	513	7	98.63%
24	557	8	98.56%
25	552	7	98.73%
26	553	0	100.00%
27	537	0	100.00%
28	549	0	100.00%
29	540	1	99.81%
30	541	0	100.00%

Table 7. Performance of the classifiers on correctly classifying packets from each device.

Device ID	Total Packets	Misclassified Packets	Accuracy
19	562	10	98.22%
20	558	15	97.31%
21	573	10	98.25%
22	556	9	98.38%
23	513	7	98.63%
24	557	8	98.56%
25	552	7	98.73%
26	553	47	91.50%
27	537	42	92.17%
28	549	59	89.25%
29	540	51	90.55%
30	541	55	90.57%

The classifiers performance on correctly classifying each packet can be seen from Table 7. Comparing both Table 6 and Table 7, since there is no additional classification of normal packets, the number of misclassified instances for victim devices remain the same. However, there are significantly more misclassified instances of the malicious devices. However, this does not mean that the classifier fails to detect anomalous packets, but rather fails to classify the type of the anomalies.

Considering both Table 6 and Table 7, the ability to detect the anomalous packets in the network traffic is enough for the CoAP server to maintain the security of the IoT network, and the classification of anomalies is not a high priority. Therefore, it can be said that the stacking classifier performs quite remarkably with an overall accuracy of 99.98%. Since the accuracy of the classifier is very high, a blacklisting approach can be used by the CoAP server to prevent malicious devices from generating network traffic within the IoT network.

Current limitations of the proposed system model include the assumption of secure verification between the IoT devices and the CA for server authentication. The model assumes a secure communication between the two which may be achieved through other handshake protocols like

TLS or a long term pre-shared secret which is inserted into the IoT devices and the CA. Both assumptions have their downsides in the proposed model. TLS uses traditional session key generation and public key encryption, which requires high computational power and hence not suitable for all IoT devices. Similarly, using a long term pre-shared secret is not suitable for the proposed use case and system model. Since any device capable of completing the proposed handshake protocol may enter the IoT network, distributing the pre-shared secret to the new devices poses a challenge.

The deficiencies of the proposed network simulation setup include the behavior of the malicious IoT nodes. The attacker IoT nodes within the network simulation generate high amounts of malicious network traffic which makes it relatively easy for the server to detect the malicious nodes since the classifier used for autonomous anomaly detection has a high performance as presented before. This may not be the case in a real-life scenario where the malicious devices actively try to impersonate benign devices by generating small amounts of malicious network traffic to avoid detection. In such a case, the proposed network simulation setup may be improved to include a more comprehensive trust-score based approach adopted by the CoAP server.

5. CONCLUSIONS

In this work, a secure, lightweight handshake protocol, designed to work on top of CoAP for the autonomous device discovery and management of IoT devices of any complexity is proposed. The protocol makes use of a PUF for session key generation, as well as whitelisting and blacklisting methods for ensuring the network security. To the best of our knowledge, the presented PUF-based handshake protocol is the first to perform blacklisting and whitelisting. Anomaly detection using machine-learning techniques is utilized for blacklisting. An in-depth security analysis of the proposed handshake protocol is provided, where the resilience of the protocol against forgery, replay and MITM attacks is shown with simulation results. The paper also provides and discusses on a network traffic simulation carried out using the proposed handshake protocol. By gathering the specifications of IoT nodes, the CoAP server can decide to only allow nodes join or remain in the IoT network that satisfy certain specifications using whitelisting and blacklisting methods. Additionally, a comparison of various machine learning algorithms performances and their fitness for real-time anomaly detection is provided. Amongst the five machine learning algorithms studied, the stacking algorithm is shown to display the highest level of accuracy of 99.98% in detecting anomalous data in the network. Based on the results of the network simulation performed, the CoAP server inside the proposed setup was able to blacklist all the malicious devices within the IoT clients with perfect accuracy.

Future work includes a comprehensive trust-score based blacklisting approach to further improve the performance of the proposed protocol. This improvement can make it easier for the server to detect malicious devices that impersonate benign devices by generating smaller amounts of malicious network traffic.

REFERENCES

- [1] A. Rayes and S. Salam, *Internet of Things From Hype to Reality*, Springer International Publishing, 2019.
- [2] A. Aktypi, K. Kalkan and K. B. Rasmussen, "SeCaS: Secure Capability Sharing Framework for IoT Devices in a Structured P2P Network," in *Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy*, New Orleans, Association for Computing Machinery, 2020, p. 271–282.

- [3] T. D. Nguyen, S. Marchal, M. Miettinen, H. Fereidooni, N. Asokan and A.-R. Sadeghi, "D²IoT: A Federated Self-learning Anomaly Detection System for IoT," in 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), 2019, pp. 756-767.
- [4] G. Fortino, F. Messina, D. Rosaci and G. M. Sarne, "ResIoT: An IoT social framework resilient to malicious activities," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 5, pp. 1263-1278, 2020.
- [5] G. Fortino, L. Fotia, F. Messina, D. Rosaci and G. M. Sarné, "Trust and Reputation in the Internet of Things: State-of-the-Art and Research Challenges," *IEEE Access*, vol. 8, pp. 60117-60125, 2020.
- [6] G. Fortino and Mozilla, "The Transport Layer Security (TLS) Protocol Version 1.3.," 2018.
- [7] Y. Yildiran, G. S. R and H. Basel, "Lightweight PUF-Based Authentication Protocol for IoT Devices," in 2018 IEEE 3rd International Verification and Security Workshop (IVSW), 2018, pp. 38-43.
- [8] A. Bhattacharyya, T. Bose, S. Bandyopadhyay, A. Ukil and A. Pal, "LESS: Lightweight Establishment of Secure Session: A Cross-Layer Approach Using CoAP and DTLS-PSK Channel Encryption," in 2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops, 2015, pp. 682-687.
- [9] A. Bin Rabbiah, K. Ramakrishnan, E. Liri and K. Kar, "A Lightweight Authentication and Key Exchange Protocol for IoT," 2018.
- [10] J. W. Byun, "End-to-End Authenticated Key Exchange Based on Different Physical Unclonable Functions," *IEEE Access*, vol. 7, pp. 102951-102965, 2019.
- [11] B. Genge and J. W. Byun, "A Generic Multifactor Authenticated Key Exchange with Physical Unclonable Function," 2019.
- [12] M. N. Aman, K. C. Chua and B. Sikdar, "Mutual Authentication in IoT Systems Using Physical Unclonable Functions," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1327-1340, 2017.
- [13] A. Braeken, "PUF Based Authentication Protocol for IoT," *Symmetry*, vol. 10, no. 8, 2018.
- [14] U. Chatterjee, V. Govindan, R. Sadhukhan, D. Mukhopadhyay, R. S. Chakraborty, D. Mahata and M. P. Mukesh, "PUF + IBE: Blending Physically Unclonable Functions with Identity Based Encryption for Authentication and Key Exchange in IoTs.," 2017.
- [15] N. Li, D. Liu and S. Nepal, "Lightweight Mutual Authentication for IoT and Its Applications," *IEEE Transactions on Sustainable Computing*, vol. 2, no. 4, pp. 359-370, 2017.
- [16] A. Mostafa, S. J. Lee and Y. K. Peker, "Physical Unclonable Function and Hashing Are All You Need to Mutually Authenticate IoT Devices," *Sensors*, vol. 20, no. 16, 2020.
- [17] Y. Ma, L. Yan, X. Huang, M. Ma and D. Li, "DTLShtps: SDN-Based DTLS Handshake Protocol Simplification for IoT," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3349-3362, 2020.
- [18] J. Park and N. Kang, "Lightweight secure communication for CoAP-enabled Internet of Things using delegated DTLS handshake," in 2014 International Conference on Information and Communication Technology Convergence (ICTC), 2014, pp. 28-33.
- [19] B. Anderson and D. McGrew, "Accurate TLS Fingerprinting using Destination Context and Knowledge Bases," 2020.
- [20] S. F. Danilo, A. O. Hygo and P. Angelo, "A personal connected health system for the Internet of Things based on the Constrained Application Protocol," *Computers & Electrical Engineering*, vol. 44, pp. 122-136, 2015.
- [21] T. A. Alghamdi, A. Lasebae and M. Aiash, "Security analysis of the constrained application protocol in the Internet of Things," in Second International Conference on Future Generation Communication Technologies (FGCT 2013), 2013, pp. 163-168.
- [22] D. Rathod, "Security Analysis of Constrained Application Protocol (CoAP): IoT Protocol," vol. 6, p. 37, 2017.
- [23] Z. Shelby, K. A. Hartke and C. Bormann, "The Constrained Application Protocol (CoAP)," Universitaet Bremen TZI, 2014.
- [24] A. Shamsoshoara, A. Korenda, F. Afgah and S. Zeadally, "A survey on physical unclonable function (PUF)-based security solutions for Internet of Things," *Computer Networks*, vol. 183, p. 107593, 2020.
- [25] D. Dolev and A. Yao, "On the Security of Public Key Protocols," *IEEE Transactions on Information Theory*, 1983.
- [26] I. Ullah and Q. H. Mahmoud, "A Scheme for Generating a Dataset for Anomalous Activity Detection in IoT Networks," in *Advances in Artificial Intelligence*, Cham, *Advances in Artificial Intelligence*, 2020, pp. 508-520.

- [27] T. Mehmood, R. Md and B. Helmi, "Machine learning algorithms in context of intrusion detection," in 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), 2016, pp. 369-373.
- [28] P. Illavarason and B. Kamachi Illavarason, "A Study of Intrusion Detection System using Machine Learning Classification Algorithm based on different feature selection approach," in 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2019, pp. 295-299.
- [29] A. Srivastava, A. Agarwal and G. Kaur, "2019 4th International Conference on Information Systems and Computer Networks (ISCON)," in Novel Machine Learning Technique for Intrusion Detection in Recent Network-based Attacks, 2019, pp. 524-528.
- [30] A. Starzacher and B. Rinner, "Evaluating KNN, LDA and QDA classification for embedded online feature fusion," 2009.
- [31] B. Chandrasekaran, "Survey of Network Traffic Models," Washington University in St. Louis - Computer Science & Engineering at WashU, 1006.

ARABIC POEMS GENERATION USING LSTM, MARKOV-LSTM AND PRE-TRAINED GPT-2 MODELS

Asmaa Hakami, Raneem Alqarni, Mahila Almutairi and Areej Alhothali

Department of Computer Science,
King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT

Nowadays, artificial intelligence applications are increasingly integrated into every aspect of our lives. One of the newest applications in artificial intelligence and natural language is text generation, which has received considerable attention in recent years due to the advancements in deep learning and language modeling techniques. Text generation has been investigated in different domains to generate essays and books. Writing poetry is a highly complex intellectual process for humans that requires creativity and high linguistic capability. Several researchers have examined automatic poem generation using deep learning techniques, but only a few attempts have looked into Arabic poetry. Attempts to evaluate the generated pomes coherence in terms of meaning and themes still require further investigation. In this paper, we examined character-based LSTM, Markov-LSTM, and pre-trained GPT-2 models in generating Arabic praise poems. The results of all models were evaluated using BLEU scores and human evaluation. The results of both BLEU scores and human evaluation show that the Markov-LSTM has outperformed both LSTM and GPT-2, where the character-based LSTM model gave the lowest yields in terms of meaning due to its tendency to create unknown words.

KEYWORDS

Arabic Poems, Markov, GPT-2, Deep Neural Networks, & Natural Language Processing.

1. INTRODUCTION

The developments of artificial intelligence have made it possible to compare the capabilities of machines with human abilities, such as the ability to generate texts of various forms [1]. The developments of artificial intelligence have made it possible to compare the capabilities of machines with human abilities, such as the ability to generate texts of various forms [1]. One of these texts is poetry, which is artistic literature that uses aesthetic and rhythmic language style to convey meanings or evoke emotions that affect the person who reads or hears [2]. Poetry can be used to express a specific feeling, situation, or scene or describe qualities of a character or place. It is one of the essential aspects of language in the world. Moreover, it is important to introduce the history and culture of the people, especially the Arab community. Their history, customs, and social principles are held over in this art. It also indicates the strength and durability of their language [3].

Poetry generation is one of the most interesting yet challenging Natural Language Processing (NLP) tasks. Several researchers seek to build models to generate textual data in different domains. One of the domains that were recently examined is poem generation. Several attempts were made in the NLP to generate poems in different languages, but many challenges were

encountered due to the meaning of the generated poem being unclear and understandable [4] or the structure of the poem being so chaotic and not thematic. Furthermore, other researchers focused on a specific type of poem or the style of specific writers [5] (see Section 2 for more details).

This work aims to take up this challenge and develop three different models to generate Arabic praise poems. LSTM, Markov-LSTM, and Pre-trained GPT-2 were chosen in this research due to their promising performance in other text generation tasks. We also evaluate the performance of the models against each other. This paper is organized as follows: Section 2 gives details about related work. Section 3 presents the used dataset. Section 4 describes the methodology for Arabic poem generation, which includes the pre-processing, and the proposed approaches to be compared. Section 5 presents the used evaluation criteria and analysis of the obtained results. Finally, Section 6 shows conclusions and future work.

2. RELATED WORK

Several studies that looked into generating poems and stories were considered in this research to benefit from the previous experiences in the same field. A number of these studies are represented in this section.

Authors in [4] built a model to generate coherent Chinese poetry in meaning with a flexible clear description of the topic for which the poem was created. The model was evaluated on three poetry domains which are quatrain, iambic, and chinoiserie lyric. Working Memory Model has been used as a method to create a poetry line by taking the previous line into account. The previous line is stored in local memory to be combined with the following line. A Topic Trace (TT) mechanism has been created to record the topics in a more explicit way. The poetry experts compared the results of this model with other approaches. The model received higher scores which indicates that the model generated poems with better quality and cohesion. Moreover, this model can create different types of poetry. The mechanism of TT helped increasing performance; however, there is still a gap between the generated poetry and human poetry.

Talafha and Rekabdar's [6] work was the first to propose an Arabic poem generation model using deep learning algorithms. The model contains two parts: The first part is a Bi-directional Gated Recurrent Unit (Bi-GRU) to generate the first line. The second part is a modified Bi-GRU encoder-decode. The proposed approach uses a hybrid model that combines a TextRank algorithm and a word embedding technique to extract keywords. FastText approach used to build the word-level embedded model. The dataset contains 80,506 verses from 20,106 Arabic poems that expressed love and religion. Quantitative evaluation using BLEU scores shows that the proposed model outperforms other deep learning approaches. On the other hand, the results of qualitative evaluation by humans show that the proposed model gives higher scores in terms of Coherence, Meaning, and Poeticness compared to other approaches in the same area.

Another study [5] developed a model called "poet without emotions" using deep learning algorithms for generating Arabic poems simulating the poems of the poet Nizar Qabbani. The model has built using Long Short-Term Memory (LSTM) networks that are a modified Recurrent Neural Network (RNN) version, making it more convenient to remember memory data. The model trained on 10,000 verses of Nizar Qabbani poems that are text sequences with the same length. So, the most generated poetic text by this model contains one verse of poetry. The accuracy of the generated poetic text reached 93%, which is a fairly good result. As Gharbi [5] states that, "it is an acceptable result if we take into account the simplicity of the used structure, the training and data formatting processes, in addition to the volume of the training text, compared to traditional text generation methods.

While [7] proposed a story generation model called “Story Scrambler” using RNN and LSTM. The model has given a sequence of input, which is considered a window. After that, the model had to predict the following word using the SoftMax activation function then the window updated. There were two types of input the first one is two stories with different content, and the second input is two stories with the same content but different narration. The model tested on various numbers of RNN layers, batch size, and input sequence length. It obtained minimal train loss of 0.01 when the size of RNN was 512 with three layers, the batch size was 100, and the input sequence length was 50. Moreover, it was discovered through experiments when increasing the number of layers beyond three and the batch size beyond 100 results in overfitting. The model evaluated by humans, and the accuracy of it was 63%.

Astigarraga et al. [8] proposed a model to generate poetry in the Basque language using two Markov chains. Poetry is one of the exhibitions of traditional Basque culture, especially in events and competitions. The model generates poems in the style of an existing author in less than a minute. Two different datasets were used to train the model. The first one was the Txirrita dataset, which has 2127 verses of poetry by a famous Txirrita. The second one was the Mixed dataset, which has 18913 lines compilation of sentences collected from Basque newspaper and poetry sung. Besides, some linguistic tools have been used to generate verse, such as rhyme search to find words that rhyme with the given word. Also, Latent Semantic Analysis (LSA) method has been used to measure the semantic relationship between pairs of words and sentences. The poem result will consist of four lines, each has thirteen syllables long, and all of them sharing a rhyme. The evaluation metric was only 2-gram because the system goal was to produce a poem somewhat different from the dataset. However, human evaluation is needed to assess poems. Therefore, four peoples familiar with the poems evaluated the generated poems. Each of them analyzed twenty poems, ten from each dataset. Their impression was positive, stating that they were well-formed poems, although not of human-produced quality. But they also found that the internal coherence of the whole poem was pretty poor. Furthermore, they stated that poems created from the Txirrita dataset seemed more natural and closer to the style of the Basque poems compared with the Mixed dataset.

From the previous studies, we found that most of the researches that are concerned with building a model to generate poems is limited to both English and Chinese languages. There is a lack of research that generates Arabic poems. Also, none of the previous studies were specialized in generating praise poems in the Arabic language.

3. DATASET

The dataset in this research was collected from praise poems written by different Muslim poets who were born and lived during different eras and in several countries, such as AL-Arjani, who lived in the Andalusian era, Ibn Al-Khayyat, who lived in the Mamluk era. The dataset consisted of 34,466 verses that have been collected manually from the AL-diwan website [9], which includes a large number of poems in various fields such as Ghazal poetry. There were also many different fields/aspects of praise poetry, including praising people such as Prophet Muhammad peace be upon him, tribes such as Quraysh. Also, countries such as Iraq and Egypt. This is an example of a poem that praise Prophet Muhammad peace be upon him:

أَجْدَ مَدَحَ خَيْرِ الْخَلْقِ ذَاتًا وَجُودَةً،
وَجَدَ عَنْ سُبُوحِ مَا سَنَّهُ لَكَ حَبِيدَةً
وَأَنْشِدْ هَوًى فِيهِ إِكْتَفَى وَمَوَدَّةً
مَنْحَتْ رَسُولَ اللَّهِ بَدَأَ وَعُودَةً
”وَمِقْدَارُهُ فِي الْبَدَاءِ وَالْعُودِ أَعْظَمَ

4. METHODOLOGY

As shown in Figure 1, the followed methodology to generate praise poems consists of four phases. In this section, these phases are discussed in more detail.

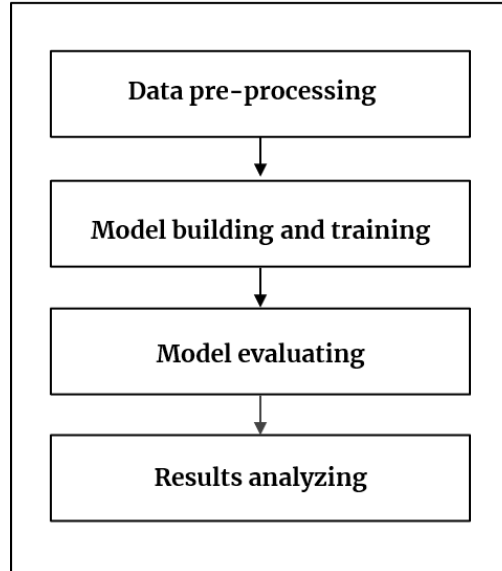


Figure 1. The phases of the methodology.

4.1. Data pre-processing

In this phase, we performed dataset pre-processing by removing or modifying data that is incorrect, incomplete, irrelevant, or duplicate. We removed unrelated characters or symbols such as punctuation marks, line space, \$, #, brackets, parentheses, and other unrelated characters from the dataset. Table 1 represents an example from the input before and after processing.

Table 1. Example of the pre-processing stage

Before processing	After processing
! غير الرياح التي في التيه تنزلق	غير الرياح التي في التيه تنزلق
هنا مدائح (حسنان) على شفتي	هنا مدائح حسنان على شفتي
من ذا يطيب في الإنسان جوهرة؟	من ذا يطيب في الإنسان جوهرة

4.2. Proposed Methods

In this work, three approaches for text generation (praise poems generation in particular) are proposed, which are the character-based LSTM model, Markov-LSTM model, and GPT-2 pre-trained model. In this section, the models are discussed in more detail.

4.3. Character-based LSTM model

The first proposed model is the character-based LSTM model. It is implemented by using the Keras library ¹ and composed of three layers. The first layer is the embedding layer that takes integers indices (which stand for specific characters) and turns them into dense vectors of 256 dimensions. Before inputting the data to this layer, we map all existing characters in the dataset to a numerical representation. The second layer is an LSTM layer with 1024 units. LSTM is a type of RNN that is capable of handling long-term dependency and vanishing gradient problem. RNN models can be useful to model time series data or sequential data such as natural language text [10]. The third layer is a dense layer with size outputs equals to vocab size. This model takes the input as a sequence of characters, and the length of this sequence is 200 characters, and tries to predict the next character at each time step. For example, if the input is the sequence shown in Figure 2 (A), and in this case, the model expects the letter "ر" and the output to be as shown in Figure 2 (B). The problem can be regarded as a classification issue at this point. The model at this time step will predict the class of the next character based on the previous LSTM state and the current input. Therefore, the categorical cross-entropy loss function is used, and it is employed across the last dimension of the predictions. This model is compiled with Adam optimizer. The string length of the input through the training is 200 characters, but the model can be run on start strings of any length. An example of the generated verses by using this model is shown and analysed in Section 5.

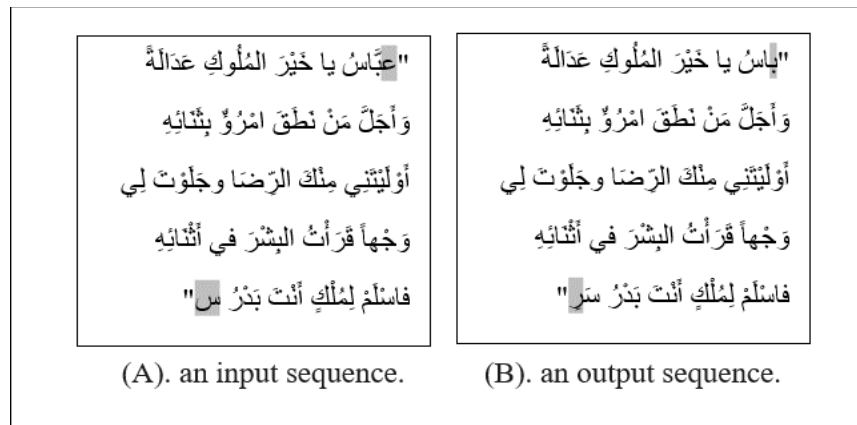


Figure 2. An example of the input-output of character-based LSTM model.

4.4. Markov-LSTM Model

The second model in this study generates Arabic poems using the Markov-LSTM model. Markov models were used in several fields. One of them is in text generation [11] and shows promising results in short text generation. Markov-LSTM depends on the probability of the next word based on the current [11]. This model uses Markovify functions to generate verses based on every word probability. So, if the current word is "وإني", and the probability of the word "لما" to come after it is 20%, where the word "حين" has the probability of 25%, Then the word "حين" will be chosen and so on. The first word in each verse was chosen randomly. Any verses result from the Markovify functions that contain the word "الله" (Allah) was removed to avoid inappropriate use of this word. Furthermore, the result verses were encoded and used as input to the LSTM model. The LSTM model is composed of four layers to generate a poem. LSTM has long-term memory,

¹<https://keras.io/>

so that it is better to predict the properties of the next verse, such as rhyme scheme. So, the appropriate new verse of the poem can be selected each time by the LSTM method from all previous verses generated by Markovify. An example of the generated verses by using this model is shown and analysed in Section 5.

4.5. Pre-trained GPT-2 Model

The third model focuses on fine-tuning a pre-trained GPT-2 model. GPT-2 is a large-scale unsupervised language model produced by Open-AI. It was trained on 40GB of Internet text to predict the next word. Due to the concern of Open-AI about malicious technical applications, Open-AI is not releasing the trained model. Instead, Open-AI launches a much smaller model as an experiment in responsible disclosure [12]. There are three versions of GPT-2 are released, which are the small version (124 Million Parameters), the medium version (355 Million Parameters), and the large version (774 Million Parameters). Moreover, larger models are more knowledgeable, but these models take a longer time for fine-tuning and text generation [12]. For this work, the small version of the GPT-2 model is fine-tuned on the dataset to generate praise poems. Because our dataset with a size of 2.01 BM and that less than the minimum recommended size (10 MB) to use the medium version of GPT-2. We fine-tune this model by using the gpt-2-simple package that is a Python package, and it wraps existing model fine-tuning. Also, it makes text generation easier and allowing for prefixes to force the text to start with a given phrase. The input data to this model is a single text file as the model requires. We use the Finetune function to fine-tune the pre-trained GPT-2 model on our datasets. We set the parameters for the Finetune function as following: The steps parameter was set to 2000. The restore From parameter set to "fresh"; to begin training from the base GPT-2. The learning rate parameter for the training is set to $1e-4$ by default. Also, we use the Generate function to generate text after fine-tuning this model on our datasets. This function has an important parameter which is the temperature. The higher the temperature, the syntactically incorrect and the unique the text. We set it to 0.7 as the minimum recommended value (recommended to keep the temperature value between 0.7 and 1.0) [13]. An example of the generated verses by using this model is shown and analysed in Section 5.

5. RESULT AND DISCUSSION

In this section, the results and evaluation of the three models that were used to generate Arabic poems will be represented.

5.1. Quantitative Evaluation

The BLEU scores were used to evaluate models, which indicate the identity of the generated text and the reference text. It has many ways of measuring, such as 1-gram, 2-gram, 3-gram, and 4-gram. Each gram represents the number of words that will be taken from both texts and compared to each other. The value of BLEU ranges from 0 to 1. The higher the BLEU value, the higher the similarity between the generated sentence and the reference sentence. The BLEU scores are calculated by the following equation [14]:

$$BLEU = BP.exp\left(\sum_{n=1}^N w_n \log p_n\right) \quad (5.1)$$

Table 2 represents a sample of the outputs for all models. The verses of the Character-based LSTM model were inconsistent and had no meaning because the model predicted a letter by letter. It creates new words that are not present in the dataset or the Arabic language dictionary.

Also, we noticed that sometimes the generated text reproduces the exact verses in the dataset. The Markov-LSTM model verses have common characteristics with the previous model, except that sometimes it changes one word when taking a verse from the dataset. The generated verse is better than the Character-based LSTM model because it does not create new unknown words. The verse of the Pre-trained GPT-2 model was close to the Markov-LSTM model in terms of relational coherence and logic.

Table 2. Models' output

Model name	The generated verses in Arabic	The generated verses in English
Character-based LSTM	وَأَحْسَنُ مِنْكَ لَمْ تَرَ قَطُّ عَيْنِي وَأَعْظَمُ غُرَّةً يُعَازِلُ فِي حَقِّي إِنْخَدَتَ بِهَا وَفِي الْعِرَاقِ إِذَا أَضَلَّتْ حَبًّا مَعْفُونَتَيْنِ بِهِ مَكَانَ عَلَى أَيْامِنِ الْجَنْزَلِيسِ وَهُوَ سُورِي ذُرَيْغٌ كَمَا غَدَّ الْعُمَامُ عَلَى الدُّجَى	My eyes did not see better than you (seed text), nor did they see greater than your face. And in Iraq if she astray because the huge love (metaphor). He is dismissing in the right way that you take it. They are chaste by it (word of chaste written with misspelling). Place on Ayamen (Ayamen is the plural of word of right in Arabic), jeans, and it is Syrian (these unrelated words to each other). It is fast as the spread of a thin white cloud over the darkness of the night.
Markov-LSTM	إِذَا قِيلَ هَلْ سَارَ فَاغْلِقْنِي أَنْشَدْتُ مَدْحِي فِيكَ مِنْ فَنُونِ الْمَعَانِي عَجَائِبُ لَا تَنْفَكُ عَيْنِي إِلَى جَدِّ بَعْدَ الْحَيَاةِ إِذَا التَّذْكَارُ أَحْيَانِي وَلَا تَعْتَذِرْ غَيْرَ مَا يَعْنِي أَنْتَ إِلَّا فِي مَقَامِ الشُّكْرِ يَا رَبِّ اشْفَعْ عَنِّي أَوْتِرْنِي	If it is said he walked, worry me I sang praise in you from the arts of meanings Wonders do not open my eyes To a grave after life, if the souvenir revives me And do not apologize except for what is meant You are only in the place of thanks giving, O Lord, preemptive me or see me.
Pre-trained GPT-2	وَأَحْسَنُ مِنْكَ لَمْ تَرَ قَطُّ عَيْنِي وَلَمْ أَجِرْ قَوْمَ يَخَافُ الذُّبَابَ بِهَا وَنَزَرَهُ مِنْ حَيْثُ لَا تَنْتَازِرُ أَخْرَفُ فَكَمْ بُعِدَ غَيْرَ الْقَوْلِ فِيمَا خُلِقُونِي تَوَارَى أَنَّهُ عَنِ الْأَعْمَارِ وَحَلَهُ وَمَا إِذَا حَلَّ قَلْبُ الْمُقَلِّدِ فِي كَرَمِ	And better than you, my eyes did not see(seed text) And I did not reward a people that feared flies for it And he pure from where Lantizar is senile How much a dimension of unspoken from what they created me He hid as from the ages and resolved it And whether the heart is resolved of the imitator in the vineyard

Table 3 shows the average BLEU-1 scores of all models. From the Character-based LSTM model and Pre-trained GPT-2 model scores, they noticed that several new words were created not from the dataset since the ratio of 1-gram is not high. The Markov-LSTM model had a high score on the 1-gram which means the model does not generate new words that are not in the dataset, but it used some consecutive words as in the dataset.

Table 3. Models' BLEU Score

GRAM	CHARACTER BASED LSTM	MARKOV-LSTM	PRE-TRAINED GPT-2
1-GRAM	0.552026	0.760297	0.560736

5.2. Qualitative Evaluation

The BLEU scores do not evaluate the quality of generated poems, such as meaning and coherence. So, it is required to get a human evaluation in this field of research. Besides, human evaluation is hard because they have different opinions and tastes in poems. In this work, three poems were chosen randomly from each model to evaluate them by two experts. The evaluation was based on four criteria which are: meaning, coherence, rhyme, and rhythm. Each criterion takes a score from zero to five (zero is the worst). Table 4 shows the result of human evaluation. As shown in the result, the Markov-LSTM model got higher scores in terms of meaning, coherence, and rhyme compared with the other two models.

Table 4. Human Evaluation

Criteria Model name	Meaning	Coherence	Rhyme	Rhythm
Character-based LSTM	0.5	0.5	1.5	1.5
Markov-LSTM	1	0.75	2	1.5
Pre-trained GPT-2	0.5	0.5	1.5	1.5

6. CONCLUSION AND FUTURE WORK

In this paper, three models are presented for generating Arabic poems in the field of praise. Three models were built, which are Character-based LSTM, Markov-LSTM, and pre-trained GPT-2. The results of the Markov-LSTM model were better than the other two models based on the BLEU-1 score. As for the consistency of the verses, and the clarity of their meaning there is no wide difference between it and the pre-trained GPT-2 model. The generated poems still lacked some grammatical rules and logical sequences of words and their interconnection with each other. In the future, we aim to reduce the runtime and increase the number of verses in the dataset. Finally, improve the grammatical rules of the lines that the models generate.

REFERENCES

- [1] Huang, M. H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, 21(2), 155-172.
- [2] Ibrahim Mahmud Ahmad, Aabd Alrahim (2018). Alqusu abalaghatuh fy alshier alerby alqdyd [The storytelling and its rhetoric in ancient Arabic poetry]. *Majalat albahth alelmy fy aladab [Journal of Scientific Research in Literature]*, 211-232.
- [3] Abdul-Sahib, Ali (2011). Fi mafhum alshier wlgth: khasayis alnas alshaerii[On the concept of poetry and its language: characteristics of the poetic text] *University of Sharjah Journal for Humanities and Social Sciences*, 111(460), 1-17.
- [4] Yi, X., Sun, M., Li, R., & Yang, Z. (2018, July). Chinese poetry generation with a working memory model. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence* (pp. 4553-4559).
- [5] Gharbi, G. (2019). Poetry Without Emotion: An experiment in generating Arabic poetry by using deep learning. In *Artificial Intelligence Applications In The Service of The Arabic Language*

- (1st ed., pp. 174-186). Riyadh, KSA: Wojooh Publishing & Distribution House. Retrieved October 02, 2020, from <https://kaica.org.sa/site/page/89>.
- [6] Talafha, S., & Rekadbar, B. (2019, January). Arabic Poem Generation with Hierarchical Recurrent Attentional Network. In 2019 IEEE 13th International Conference on Semantic Computing (ICSC) (pp. 316-323). IEEE.
 - [7] Pawade, D., Sakhapara, A., Jain, M., Jain, N., & Gada, K. (2018). Story scrambler-automatic text generation using word level rnn-lstm. *International Journal of Information Technology and Computer Science (IJITCS)*, 10(6), 44-53.
 - [8] Astigarraga, A., Martínez-Otzeta, J. M., Rodríguez, I., Sierra, B., & Lazkano, E. (2017, August). Markov Text Generator for Basque Poetry. In *International Conference on Text, Speech, and Dialogue* (pp. 228-236). Springer, Cham.
 - [9] Aldiwan: mawsueat alshier alearabi [Aldiwan: An Encyclopedia of Arabic Poetry] [Online] Available at: <<https://www.aldiwan.net>>.
 - [10] Shukla, N., & Fricklas, K. (2018). *Machine learning with TensorFlow*. Greenwich: Manning.
 - [11] Szymanski G., & Ciota, Z. (2004). On-line text generation using Markov models, *Proceedings of the International Conference Modern Problems of Radio Engineering, Telecommunications and Computer Science*, pp. 339-341.
 - [12] Radford, A., Wu, J., Amodei, D., Amodei, D., Clark, J., Brundage, M., & Sutskever, I. (2019). Better language models and their implications. OpenAI Blog <https://openai.com/blog/better-language-models>.
 - [13] Utane, N. (2020, April 17). Complete guide to build and deploy a tweet generator app into production. Retrieved December 22, 2020, from <https://towardsdatascience.com/complete-guide-to-build-and-deploy-a-tweet-generator-app-into-production-5006729e583c>.
 - [14] Zhukov, V., Golikov, E., & Kretov, M. (2017). Differentiable lower bound for expected BLEU score. arXiv preprint arXiv:1712.04708.

AUTHORS

Asmaa Hakami is a senior computer science student at King Abdul-Aziz University. Her research interest lies in the areas of machine learning, deep learning, natural language processing, and computer version.

Raneem Alqarni is a senior computer science student at King Abdul-Aziz University. Her research interest lies in the areas of machine learning, deep learning, natural language processing and computer version.

Mahila Almutairi is a computer science student at King Abdul-Aziz University. Her research interest lies in the areas of machine learning, deep learning, and natural language processing.

Areej Alhothali is an assistant professor in the faculty of computer science and information technology at King Abdul-Aziz University. She earned her master's and Ph.D. degrees in computer science (artificial intelligence) from the University of Waterloo, Canada. Her research interest lies in the areas of machine learning, deep learning, natural language processing, intelligent agent systems, affective computing, and sentiment analysis.

ALIASING FREE FOR MIXED SPECTRA FOR STABLE PROCESSES

Rachid Sabre

Biogeosciences (UMR CNRS/uB 6282), University of Burgundy, 26, Bd
Docteur Petitjean, Dijon, France

ABSTRACT

This work focuses on the symmetric alpha stable processes with continuous time frequently used in modeling the signal with indefinitely growing variance when the spectral measure is mixed: sum of a continuous measure and discrete measure. The objective of this paper is to estimate the spectral density of the continuous part from discrete observations of the signal. For that, we propose a method based on a sample of the signal at a periodic instant. The Jackson polynomial kernel is used for construct a periodogram. We smooth this periodogram by two spectral windows taking into account the width of the interval where the spectral density is non-zero. This technique allows to circumvent the phenomenon of aliasing often encountered in the estimation from the discrete observations of a process with a continuous time.

KEYWORDS

Spectral density, stable processes, periodogram, smoothing estimate, aliasing

1. INTRODUCTION

Stable alpha processes have been of interest to several research authors for their multiple applications when we have random signals with variance indefinitely increase. The harmonizable process is an important example of a symmetric α -stable process, and its proprieties have been considered by numerous authors like [1]-[10] to name a few.

In particular, stable symmetric processes find their place in various applications and in various fields such as: physics, biology, electronics and electricity, hydrology, economies, communications and radar applications, ...ect. See: [11]-[22]. This work considers a symmetric alpha stable harmonizable process $X = \{X(t) : t \in R\}$. Alternatively X has the integral representation:

$$X(t) = \int \exp[i(t\lambda)] d\xi(\lambda) \quad (1)$$

where $1 < \alpha < 2$ and ξ is a complex valued symmetric α -stable random measure on R with independent and isotropic increments. The measure defined by $m(A) = |\xi(A)|_\alpha^\alpha$ (see [4]) is called "control" measure or spectral measure." The spectral density function was already estimated in different cases: by E.Masry and S.Combanis [4] when the time of the process is continuous, by Sabre [23] when the time of the process is discrete and by R. Sabre [24]-[25] when the time of the process is p-adic.

This work considers a general case where we suppose that the spectral measure is the sum of an absolutely continuous measure with respect to Lebesgue measure and a discrete measure:

$$d\mu(\lambda) = \phi(x)dx + \sum_{i=1}^q c_i \delta_{w_i}$$

where δ is a Dirac measure, ϕ is nonnegative integrable and bounded function. c_i is an unknown positive real number and w_i is an unknown real number. Assume that $w_i \neq 0$. The function ϕ is called the spectral density. Discrete measure is due to random repeated value jumps during experimental measurements. The spectral density ϕ represents the distribution of the energy carried by the signal.

Our goal is to establish a non-parametric estimate of the spectral density ϕ from discrete observations of $X(t)$. This is motivated by the fact that, in practice, it is not obvious to observe the process on continuous interval of time. Indeed, we sampled the process at instants t_n , equally distant, i.e., $t_n = n\tau$, $\tau > 0$. It is known that aliasing of ϕ occurs. For more details about aliasing phenomenon, see [26]. To avoid this difficulty, we suppose that the spectral density ϕ is vanishing for $|\lambda| > \Omega$ where Ω is a nonnegative real number. From some smoothing, we construct an estimate depending on Ω and we show that it is asymptotically unbiased and consistent.

Briefly, we indicate the organisation of this paper: the outline in this paper is as follows: we present in the second section two technical lemmas, the preiodogram and show that this periodogram is asymptotically unbiased estimated but not consistent. In the third section, we smooth this periodogram by two chosen spectral windows to estimate the spectral density at jump points. We show that the smoothing periodogram is a consistent estimator.

2. THE PERIOGRAM AND ITS PROPRIETIES

First, we introduce some basic notations and properties of the Jackson's polynomial kernel. Let N is the size of sample of X . Let k and n are the numbers satisfying:

$$N-1 = 2k(n-1) \quad \text{with} \quad n \in N \quad k \in N \cup \left\{ \frac{1}{2} \right\} \quad \text{if} \quad k = \frac{1}{2} \quad \text{then} \quad n = 2n_1 - 1, \quad n_1 \in N.$$

The Jackson's polynomial kernel is defined by: $|H_N(\lambda)|^\alpha = |A_N H^{(N)}(\lambda)|^\alpha$ where

$$H^{(N)}(\lambda) = \frac{1}{q_{k,n}} \left(\frac{\sin \frac{n\lambda}{2}}{\sin \frac{\lambda}{2}} \right)^{2k} \quad \text{with} \quad q_{k,n} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{\sin \frac{n\lambda}{2}}{\sin \frac{\lambda}{2}} \right)^{2k} d\lambda.$$

where $A_N = (B_{\alpha,N})^{\frac{-1}{\alpha}}$ with $B_{\alpha,N} = \int_{-\pi}^{\pi} |H^{(N)}(\lambda)|^{\alpha} d\lambda$.

We give the following lemmas which are used in the rest of this paper. Their proof are given in [23].

Lemma 2.1 *There is a non negative function h_k such as:*

$$H^{(N)}(\lambda) = \sum_{m=-k(n-1)}^{k(n-1)} h_k\left(\frac{m}{n}\right) \cos(m\lambda)$$

$$B'_{\alpha,N} = \int_{-\pi}^{\pi} \left| \frac{\sin \frac{n\lambda}{2}}{\sin \frac{\lambda}{2}} \right|^{2k\alpha} d\lambda \quad \text{and} \quad J_{N,\alpha} = \int_{-\pi}^{\pi} |u|^{\gamma} |H_N(u)|^{\alpha} du,$$

Lemma 2.2 *Let*

where

$\gamma \in]0, 2]$, *then*

$$B'_{\alpha,N} \begin{cases} \geq 2\pi \left(\frac{2}{\pi}\right)^{2k\alpha} n^{2k\alpha-1} & \text{if } 0 < \alpha < 2 \\ \leq \frac{4\pi k\alpha}{2k\alpha-1} n^{2k\alpha-1} & \text{if } \frac{1}{2k} < \alpha < 2 \end{cases}$$

And

$$J_{N,\alpha} \leq \begin{cases} \frac{\pi^{\gamma+2k\alpha}}{2^{2k\alpha}(\gamma-2k\alpha+1)} \frac{1}{n^{2k\alpha-1}} & \text{if } \frac{1}{2k} < \alpha < \frac{\gamma+1}{2k} \\ \frac{2k\alpha\pi^{\gamma+2k\alpha}}{2^{2k\alpha}(\gamma+1)(2k\alpha-\gamma-1)} \frac{1}{n^{\gamma}} & \text{if } \frac{\gamma+1}{2k} < \alpha < 2 \end{cases}$$

In this section, we give a periodogram and we develop its proprieties. Assume that the process

$X(t)$, defined in (1), is observed at instants $t_j = j\tau$, $j = 1, 2, \dots, N$ and $\tau = \frac{2\pi}{\omega}$, where ω is a real number strictly greater than 2Ω . We define the periodogram \hat{I}_N on $]-\Omega, \Omega[$ as follows:

$$\hat{I}_N(\lambda) = C_{p,\alpha} |I_N(\lambda)|^p, \quad 0 < p < \frac{\alpha}{2}$$

where

$$I_N(\lambda) = [\tau]^{\frac{1}{\alpha}} A_N \operatorname{Re} \left[\sum_{n'=-k(n-1)}^{n'=k(n-1)} h_k\left(\frac{n'}{n}\right) \exp\{-i(n'\tau\lambda)\} X(n'\tau + k(n-1)\tau) \right],$$

and the normalisation constant $C_{p,\alpha}$ is given by $C_{p,\alpha} = \frac{D_p}{F_{p,\alpha}[C_\alpha]^{p/\alpha}}$, with

$$D_p = \int_{-\infty}^{\infty} \frac{1 - \cos(u)}{|u|^{1+p}} du \quad \text{and} \quad F_{p,\alpha} = \int_{-\infty}^{\infty} \frac{1 - e^{-|u|^\alpha}}{|u|^{1+p}} du.$$

Lemma 2.3

The characteristic function of $I_N(\lambda)$, $E \exp[irI_N(\lambda)]$, converges to $\exp[-C_\alpha |r|^\alpha (\psi_{N,1}(\lambda) + \psi_{N,2}(\lambda))]$, where

$$\psi_{N,1}(\lambda) = \int_{-\pi}^{\pi} |H_N(y - \tau\lambda)|^\alpha \phi\left(\frac{y}{\tau}\right) dy \quad \text{and} \quad \psi_{N,2}(\lambda) = \sum_{i=1}^q c_i |H_N(w_i - \tau\lambda)|^\alpha$$

Proof

By substituting (1) in the expression of I_N , we have:

$$I_N(\lambda) = [\tau]^\alpha A_N \operatorname{Re} \int_{R'} \sum_{n=-k(n-1)}^{n'=k(n-1)} h_k\left(\frac{n'}{n}\right) \exp\{i[n'\tau(\lambda - u)]\} \exp\{i[\tau uk(n-1)]\} d\xi(u).$$

It follows from [1] and the definition of the Jackson polynomial kernel that the characteristic function is the form:

$$E \exp[irI_N(\lambda)] = \exp[-C_\alpha |r|^\alpha \psi_N(\lambda)]. \quad (2)$$

where $\psi_N(\lambda) = \psi_{N,1}(\lambda) + \psi_{N,2}(\lambda)$ with

$$\psi_{N,1}(\lambda) = \int_R |H_N(v - \tau\lambda)|^\alpha \phi\left(\frac{v}{\tau}\right) dv \quad \text{and} \quad \psi_{N,2}(\lambda) = \sum_{i=1}^q c_i |H_N(w_i - \tau\lambda)|^\alpha$$

$$\psi_{N,1}(\lambda) = \int_R |H_N(v - \tau\lambda)|^\alpha \phi\left(\frac{v}{\tau}\right) dv = \sum_{j \in \mathbb{Z}} \int_{(2j-1)\pi}^{(2j+1)\pi} |H_N(v - \tau\lambda)|^\alpha \phi\left(\frac{v}{\tau}\right) dv.$$

We can write

Putting $v = y - 2\pi j$ and using the fact that H_N is 2π -periodic, we obtain

$$\psi_N(\lambda) = \sum_{j \in \mathbb{Z}} \int_{-\pi}^{\pi} |H_N(y - \tau\lambda)|^\alpha \phi_j(y) dy, \quad \text{where} \quad \phi_j(y) = \phi\left(\frac{y}{\tau} - \frac{2\pi}{\tau} j\right). \quad \text{Let } j \text{ be an integer}$$

such that $-\Omega < \frac{y-2\pi j}{\tau} < \Omega$. Using the fact that $\tau\Omega < \pi$ and $|y| < \pi$, we get $|j| < \frac{\tau\Omega}{2\pi} + \frac{1}{2} < 1$ and then $j = 0$. Consequently:

$$\psi_{N,1}(\lambda) = \int_{-\pi}^{\pi} |H_N(y - \tau\lambda)|^\alpha \phi\left(\frac{y}{\tau}\right) dy. \quad (3)$$

Theorem 2.4 Let $-\Omega < \lambda < \Omega$ then $E[\hat{I}_N(\lambda)] = [\psi_N(\lambda)]^{\frac{p}{\alpha}}$,

Proof

As in [4], we use the following equality: for all real x and $0 < p < 2$,

$$|x|^p = D_p^{-1} \int_{-\infty}^{\infty} \frac{1 - \cos(xu)}{|u|^{1+p}} du = D_p^{-1} \operatorname{Re} \int_{-\infty}^{\infty} \frac{1 - e^{ixu}}{|u|^{1+p}} du, \quad (4)$$

replacing x by I_N , we obtain

$$\hat{I}_N(\lambda) = \frac{1}{F_{p,\alpha}[C_a]^{p/\alpha}} \operatorname{Re} \int_{-\infty}^{\infty} \frac{1 - \exp\{iuI_N(\lambda)\}}{|u|^{1+p}} du, \quad (5)$$

Using (6) and the definition of the $F_{p,\alpha}$, we get

$$\begin{aligned} E\hat{I}_N(\lambda) &= \frac{1}{F_{p,\alpha}[C_a]^{p/\alpha}} \int_R \frac{1 - \exp\{-C_a |u|^\alpha \psi_N(\lambda)\}}{|u|^{1+p}} du. \\ &= [\psi_N(\lambda)]^{p/\alpha}. \end{aligned}$$

3. SMOOTHING PERIODOGRAM

In order to obtain a consistent estimate of $[\phi(\lambda)]^{\frac{p}{\alpha}}$, we smooth the periodogram via spectral windows depending on whether $\tau\lambda$ is a jump point or not ($\tau\lambda \neq w_i$).

$$f_N(\lambda) = \begin{cases} f_N^{(1)}(\lambda) & \text{if } \tau\lambda \notin \{w_1, w_2, \dots, w_q\} \\ \frac{f_N^{(2)}(\lambda) - cf_N^{(1)}(\lambda)}{1-c} & \text{else} \end{cases}$$

where $f_N^{(1)}(\lambda) = \int_{-\pi}^{\pi} W_N^{(1)}(\lambda - u) \hat{I}_N(u) du$ and $f_N^{(2)}(\lambda) = \int_{-\pi}^{\pi} W_N^{(2)}(\lambda - u) \hat{I}_N(u) du$.

The spectral windows $W_N^{(1)}$ and $W_N^{(2)}$ are defined by: $W_N^{(1)}(x) = M_N^{(1)}W(M_N^{(1)}x)$ and $W_N^{(2)}(x) = M_N^{(2)}W(M_N^{(2)}x)$ with W is an even nonnegative, continuous function, vanishing for

$$|\lambda| > 1 \text{ such that } \int_{-1}^1 W(u)du = 1. \text{ The bandwidths } M_N^{(1)} \text{ and } M_N^{(2)} \text{ satisfying: } c = \frac{M_N^{(2)}}{M_N^{(1)}} \\ \lim_{N \rightarrow \infty} M_N^{(i)} = +\infty, \quad \lim_{N \rightarrow \infty} \frac{M_N^{(i)}}{N} = 0 \text{ for } i=1,2, \quad \lim_{N \rightarrow \infty} \frac{M_N^{(2)}}{M_N^{(1)}} = 0 \text{ and such that} \\ W(M_N^{(2)}\theta) = W(M_N^{(1)}\theta) \quad \forall \theta \in \left[-\frac{1}{M_N^{(1)}}, \frac{1}{M_N^{(1)}}\right].$$

We first show that $f_N(\lambda)$ is an asymptotically unbiased estimator of $[\phi(\lambda)]^{\frac{p}{\alpha}}$ for $-\Omega < \lambda < \Omega$ and $\tau\lambda \notin \{w_1, w_2, \dots, w_q\}$.

Theorem 3.1

Let $-\Omega < \lambda < \Omega$, such that $\tau\lambda \notin \{w_1, w_2, \dots, w_q\}$. Then, $E[f_N(\lambda)] - [\phi(\lambda)]^{\frac{p}{\alpha}} = o(1)$.

If ϕ satisfies the hypothesis $|\phi(x) - \phi(y)| \leq cste |x - y|^{-\gamma}$, with $\gamma < 2k\alpha - 1$, then,

$$E[f_N(\lambda)] - [\phi(\lambda)]^{\frac{p}{\alpha}} = \begin{cases} O\left(\frac{1}{n^{(2k\alpha-1)}} + \frac{1}{M_N^{(1)\gamma}}\right) & \text{if } \lambda \neq 0 \\ O\left(\frac{1}{M_N^{(1)}n^{(2k\alpha-1)}} + \frac{1}{M_N^{(1)\gamma}} + \frac{1}{n^{2k\alpha-1}}\right) & \text{if } \lambda = 0 \end{cases}.$$

Proof

By the definition of the spectral window, we have:

$$E[f_N(\lambda)] = \int_R M_N^{(1)}W[M_N^{(1)}(\lambda - u)]E[\hat{I}_N(u)]du.$$

Let $M_N^{(1)}(\lambda - u) = v$ and from (11), we obtain:

$$E[f_N(\lambda)] = \int_{-1}^1 W(v) \left[\psi_N \left(\lambda - \frac{v}{M_N^{(1)}} \right) \right]^{\frac{p}{\alpha}} dv. \quad (7)$$

Using the fact that $\int_{-1}^1 W(u)du = 1$ and the inequality (4), we get:

$$\left| E[f_N(\lambda)] - [\phi(\lambda)]^\alpha \right| \leq \int_{-1}^1 W(v) \left| \psi_N \left(\lambda - \frac{v}{M_N^{(1)}} \right) - \phi(\lambda) \right|^\alpha dv.$$

Since $\frac{p}{\alpha} < 1$, we obtain

$$\left| \psi_N \left(\lambda - \frac{v}{M_N^{(1)}} \right) - \phi(\lambda) \right|^\alpha \leq \left| \psi_{N,1} \left(\lambda - \frac{v}{M_N^{(1)}} \right) - \phi(\lambda) \right|^\alpha + \left| \psi_{N,2} \left(\lambda - \frac{v}{M_N^{(1)}} \right) \right|^\alpha$$

We now examine the limit of $\psi_{N,1} \left(\lambda - \frac{v}{M_N^{(1)}} \right)$. From (3) we get:

$$\psi_{N,1} \left(\lambda - \frac{v}{M_N^{(1)}} \right) = \int_{-\infty}^{\infty} \left| H_N \left(u - \tau \left(\lambda - \frac{v}{M_N^{(1)}} \right) \right) \right|^\alpha \phi \left(\frac{u}{\tau} \right) du.$$

Let $u - \tau \left(\lambda - \frac{v}{M_N^{(1)}} \right) = y$, we obtain:

$$\begin{aligned} \psi_{N,1} \left(\lambda - \frac{v}{M_N^{(1)}} \right) &= \int_{\mathbb{R}} |H_N(y)|^\alpha \phi \left(\lambda - \frac{v}{M_N^{(1)}} + \frac{y}{\tau} \right) dy \\ &= \sum_{j \in \mathbb{Z}} \int_{(2j-1)\pi}^{(2j+1)\pi} |H_N(y)|^\alpha \phi \left(\lambda - \frac{v}{M_N^{(1)}} + \frac{y}{\tau} \right) dy. \end{aligned}$$

(8)

Let $y - 2j\pi = s$. Since $|H_N(\cdot)|^\alpha$ is 2π -periodic function, we get

$$\psi_{N,1} \left(\lambda - \frac{v}{M_N^{(1)}} \right) = \sum_{j \in \mathbb{Z}} \int_{-\pi}^{+\pi} |H_N(s)|^\alpha \phi \left(\lambda - \frac{v}{M_N^{(1)}} + \frac{s}{\tau} + \frac{2\pi}{\tau} j \right) ds.$$

Since the function ϕ is uniformly continuous on $[-\Omega, \Omega]$ and the fact that $|H_N|^\alpha$ is a kernel,

the right hand side of the last equality converges to $\sum_{j \in \mathbb{Z}} \phi \left(\lambda + \frac{2\pi j}{\tau} \right)$. Let j be an integer

such that $-\Omega < \frac{\tau\lambda + 2\pi j}{\tau} < \Omega$. The definition of τ implies that $|\tau\lambda| < \tau\Omega < \pi$. It is easy to

see that $|j| < 1$ and then $j = 0$. Since $H_N^{(1)}$ is a kernel, we obtain that $\psi_{N,1}\left(\lambda - \frac{v}{M_N^{(1)}}\right)$ converges to $\phi(\lambda)$. On the other hand,

Since w_i is different from $\tau\lambda$ and from the lemma 2.2, we get

$$\psi_{N,2}\left(\lambda - \frac{v}{M_N^{(1)}}\right) \leq \left[2\pi\left(\frac{2}{\pi}\right)^{2k\alpha} n^{2k\alpha-1}\right]^{-1} \frac{1}{cte} \sum_{i=1}^q c_i, \text{ where } cte = \inf \left| \sin \left(\frac{w_i - \tau\left(\lambda - \frac{v}{M_N^{(1)}}\right)}{2} \right) \right|^\alpha.$$

Therefore, $\psi_{N,2}\left(\lambda - \frac{v}{M_N^{(1)}}\right) = O\left(\frac{1}{n^{2k\alpha-1}}\right)$. Thus, we have $E[f_N(\lambda)] - [\phi(\lambda)]^\frac{p}{\alpha} = o(1)$.

The rate of convergence: We assume that the spectral density ϕ satisfies the hypothesis H . We denote by $F = |Bias(f_N(\lambda))| = |E[f_N(\lambda)] - [\phi(\lambda)]^\frac{p}{\alpha}|$. It follows that

$$F \leq \frac{p}{2\alpha} \int_{-1}^1 W(v) \left[\left| \psi_N\left(\lambda - \frac{v}{M_N^{(1)}}\right) \right|^\frac{p-1}{\alpha} + [\phi(\lambda)]^\frac{p-1}{\alpha} \right] \left| \psi_N\left(\lambda - \frac{v}{M_N^{(1)}}\right) - \phi(\lambda) \right| dv.$$

Since $\psi_N\left(\lambda - \frac{v}{M_N^{(1)}}\right)$ converges to $\phi(\lambda)$, getting the rate of the convergence for F requires to

examine the rate of convergence of $\int_{-1}^1 W(v) \left| \psi_N\left(\lambda - \frac{v}{M_N^{(1)}}\right) - \phi(\lambda) \right| dv$. Indeed, from (3), we

obtain $\psi_N\left(\lambda - \frac{v}{M_N^{(1)}}\right) = \int_{-\pi}^{\pi} |H_N\left(y - \tau\lambda + \frac{\tau v}{M_N^{(1)}}\right)|^\alpha \phi\left(\frac{y}{\tau}\right) dy$.

Denote by $\Delta(\psi_N, \phi) = \psi_N\left(\lambda - \frac{v}{M_N^{(1)}}\right) - \phi(\lambda)$. Putting $t = -\left(y - \tau\lambda + \frac{\tau v}{M_N^{(1)}}\right)$ and using the

$$|\Delta(\psi_N, \phi)| \leq C_1 \int_{\tau\lambda - \frac{\tau v}{M_N^{(1)}} - \pi}^{\tau\lambda - \frac{\tau v}{M_N^{(1)}} + \pi} |H_N(t)|^\alpha \left| \frac{v}{M_N^{(1)}} + \frac{t}{\tau} \right|^\gamma dt.$$

condition H , we get

It is easy to show that

$$\begin{aligned}
\int_{-1}^1 W(v) |\Delta(\psi_N, \phi)| dv &\leq 2^\gamma C_1 \left| \frac{1}{M_N^{(1)}} \right|^\gamma \int_{-1}^1 W(v) |v|^\gamma dv \\
&+ 2^\gamma \frac{C_1}{\tau^\gamma} \int_{-1}^1 W(v) \int_{\tau_1 \lambda - \frac{\tau v}{M_N^{(1)}} - \pi}^{\tau \lambda - \frac{\tau_1 v}{M_N^{(1)}} + \pi} |H_N(t)|^\alpha |t|^\gamma dt dv
\end{aligned}$$

The second integral of the right hand side is bounded as follows:

$$\begin{aligned}
\int_{\tau \lambda - \frac{\tau v}{M_N^{(1)}} - \pi}^{\tau \lambda - \frac{\tau_1 v}{M_N^{(1)}} + \pi} |H_N(t)|^\alpha |t|^\gamma dt &\leq \int_{-|\tau \lambda| - \frac{\tau v}{M_N^{(1)}} - \pi}^{-\pi} |H_N(t)|^\alpha |t|^\gamma dt \\
&+ \int_{-\pi}^{\pi} |H_N(t)|^\alpha |t|^\gamma dt \\
&+ \int_{\pi}^{|\tau \lambda| + \frac{\tau v}{M_N^{(1)}} + \pi} |H_N(t)|^\alpha |t|^\gamma dt.
\end{aligned} \tag{9}$$

The function $|H_N(\cdot)|$ is even, then the first and the last integrals in the right hand side of (9) are

equal. Since $\frac{\tau v}{M_N}$ converges to zero and $\tau \lambda < \pi < \tau \lambda + \frac{\tau v}{M_N}$, for a large N we have:

$$\begin{aligned}
\int_{\pi}^{|\tau \lambda| + \frac{\tau v}{M_N^{(1)}} + \pi} |H_N(t)|^\alpha |t|^\gamma dt &\leq (2\pi)^\gamma \int_{\pi}^{|\tau \lambda| + \frac{\tau v}{M_N^{(1)}} + \pi} |H_N(t)|^\alpha dt \\
&\leq \frac{(2\pi)^\gamma}{B_{\alpha, N}} \frac{|\tau \lambda| + \frac{\tau}{M_N^{(1)}}}{\left| \sin \left(\pi + |\tau \lambda| + \frac{\tau}{M_N^{(1)}} \right) \right|^{2k\alpha}}
\end{aligned}$$

From the lemma 2.1, we obtain,

$$\int_{\pi}^{|\tau \lambda| + \frac{\tau v}{M_N^{(1)}} + \pi} |H_N(t)|^\alpha |t|^\gamma dt = \begin{cases} O\left(\frac{1}{n^{2k\alpha-1}}\right) & \text{if } \lambda \neq 0, \\ O\left(\frac{1}{M_N^{(1)} n^{2k\alpha-1}}\right) & \text{if } \lambda = 0, \end{cases}$$

Thus, the result follows.

Theorem 3.2. Let λ a real number belonging to $] -\Omega, \Omega[$, and $\tau\lambda = w_i$. Choose k such that

$$\lim_{N \rightarrow \infty} \frac{(M_N^{(1)})^{2k\alpha}}{n^{2k\alpha-1}} = 0.$$

Then,

i) $E[f_N(\lambda)] - [\phi(\lambda)]^{\frac{p}{\alpha}} = o(1)$

ii) If ϕ satisfies the hypothesis $|\phi(x) - \phi(y)| \leq cste |x - y|^{-\gamma}$, with $\frac{\gamma+1}{2k} < \alpha < 2$, then

$$\left| E[f_N(\lambda)] - [\phi(\lambda)]^{\frac{p}{\alpha}} \right| = \begin{cases} O \left(\frac{1}{(M_N^{(2)})^\gamma} + \frac{(M_N^{(1)})^{2k\alpha}}{n^{2k\alpha-1}} \right) & \text{if } 0 < \gamma \leq 1 \\ O \left(\frac{1}{M_N^{(2)}} + \frac{(M_N^{(1)})^{2k\alpha}}{n^{2k\alpha-1}} \right) & \text{if } 1 < \gamma \leq 2 \end{cases}$$

Proof :

From the definition of the estimator, we have

$$E[f_N(\lambda)] = \int_{-\pi}^{\pi} \frac{W_N^{(2)}(\lambda - u) - \frac{M_N^{(2)}}{M_N^{(1)}} W_N^{(1)}(\lambda - u)}{1 - \frac{M_N^{(2)}}{M_N^{(1)}}} [\psi_N(u)]^{\frac{p}{\alpha}} du$$

$$E[f_N(\lambda)] = \int_{-\pi}^{\lambda - \frac{1}{M_N^{(1)}}} + \int_{\lambda - \frac{1}{M_N^{(1)}}}^{\lambda + \frac{1}{M_N^{(1)}}} + \int_{\lambda + \frac{1}{M_N^{(1)}}}^{\pi} = E_1 + E_2 + E_3$$

$$\lambda - u = v, \quad E_2 = \int_{-\frac{1}{M_N^{(1)}}}^{\frac{1}{M_N^{(1)}}} \frac{M_N^{(2)} [W[M_N^{(2)}v] - W[M_N^{(1)}v]]}{1 - \frac{M_N^{(2)}}{M_N^{(1)}}} [\psi_N(\lambda - v)]^{\frac{p}{\alpha}} dv.$$

Put

for a large N.

Therefore $E_2 = 0$

$$E_1 = \frac{M_N^{(2)}}{1 - \frac{M_N^{(2)}}{M_N^{(1)}}} \int_{-\pi}^{\lambda - \frac{1}{M_N^{(1)}}} W[M_N^{(2)}(\lambda - u)][\psi_N(u)]^{\frac{p}{\alpha}} du -$$

$$\frac{M_N^{(2)}}{1 - \frac{M_N^{(2)}}{M_N^{(1)}}} \int_{-\pi}^{\lambda - \frac{1}{M_N^{(1)}}} W[M_N^{(1)}(\lambda - u)][\psi_N(u)]^{\frac{p}{\alpha}} du$$

Put $M_N^{(2)}(\lambda - u) = v$ in the first integral and put $M_N^{(1)}(\lambda - u) = w$, in the second integral and for a large N , we have $M_N^{(1)}(\lambda + \pi) > 1$ et $M_N^{(2)}(\lambda + \pi) > 1$. As W is null outside of $[-1, 1]$, for large N , the second integral of E_1 is zero. Therefore,

$$E_1 = \frac{1}{1 - \frac{M_N^{(2)}}{M_N^{(1)}}} \int_{\frac{M_N^{(2)}}{M_N^{(1)}}}^1 W(v) \left[\psi_N \left(\lambda - \frac{v}{M_N^{(2)}} \right) \right]^{\frac{p}{\alpha}} dv. \quad (10)$$

$$E_3 = \frac{M_N^{(2)}}{1 - \frac{M_N^{(2)}}{M_N^{(1)}}} \int_{\lambda + \frac{1}{M_N^{(1)}}}^{\pi} W[M_N^{(2)}(\lambda - u)][\psi_N(u)]^{\frac{p}{\alpha}} du - \frac{M_N^{(2)}}{1 - \frac{M_N^{(2)}}{M_N^{(1)}}} \int_{\lambda + \frac{1}{M_N^{(1)}}}^{\pi} W[M_N^{(1)}(\lambda - u)][\psi_N(u)]^{\frac{p}{\alpha}} du$$

Putting $M_N^{(2)}(\lambda - u) = v$ in the first integral and $M_N^{(1)}(\lambda - u) = w$ in the second integral, we obtain

$$E_3 = \frac{1}{1 - \frac{M_N^{(2)}}{M_N^{(1)}}} \int_{\frac{M_N^{(2)}}{M_N^{(1)}}}^{\frac{M_N^{(2)}}{M_N^{(1)}}} W(v) \left[\psi_N \left(\lambda - \frac{v}{M_N^{(2)}} \right) \right]^{\frac{p}{\alpha}} dv - \frac{\frac{M_N^{(2)}}{M_N^{(1)}}}{1 - \frac{M_N^{(2)}}{M_N^{(1)}}} \int_{M_N^{(1)}(\lambda - \pi)}^{-1} W(w) \left[\psi_N \left(\lambda - \frac{w}{M_N^{(1)}} \right) \right]^{\frac{p}{\alpha}} dw.$$

For large N we have, $M_N^{(2)}(\lambda - \pi) < -1$ and $M_N^{(1)}(\lambda - \pi) < -1$.

$$E_3 = \frac{1}{1 - \frac{M_N^{(2)}}{M_N^{(1)}}} \int_{\frac{M_N^{(2)}}{M_N^{(1)}}}^1 W(v) \left[\psi_N \left(\lambda + \frac{v}{M_N^{(2)}} \right) \right]^{\frac{p}{\alpha}} dv. \quad (11)$$

It is easy to show that for a large N

$$\frac{1}{1 - \frac{M_N^{(2)}}{M_N^{(1)}}} \int_{\frac{M_N^{(2)}}{M_N^{(1)}}}^1 W(v) dv = \frac{1}{2}. \quad (12)$$

$$\left| E[f_N(\lambda)] - [\phi(\lambda)]^{\frac{p}{\alpha}} \right| = \left| E_1 + E_3 - [\phi(\lambda)]^{\frac{p}{\alpha}} \right| \leq \left| E_1 - \frac{1}{2} [\phi(\lambda)]^{\frac{p}{\alpha}} \right| + \left| E_3 - \frac{1}{2} [\phi(\lambda)]^{\frac{p}{\alpha}} \right|$$

From (10) and (12), for a large N , we have

$$\left| E_1 - \frac{1}{2} [\phi(\lambda)]^{\frac{p}{\alpha}} \right| \leq \frac{1}{1 - \frac{M_N^{(2)}}{M_N^{(1)}}} \int_{\frac{M_N^{(2)}}{M_N^{(1)}}}^1 W(v) \left| \psi_{N,1} \left(\lambda - \frac{v}{M_N^{(2)}} \right) + \psi_{N,2} \left(\lambda - \frac{v}{M_N^{(2)}} \right) - \phi(\lambda) \right|^{\frac{p}{\alpha}} dv$$

As $\frac{p}{\alpha} < 1$, we obtain

$$\begin{aligned} \left| E_1 - \frac{1}{2} [\phi(\lambda)]^{\frac{p}{\alpha}} \right| &\leq \frac{1}{1 - \frac{M_N^{(2)}}{M_N^{(1)}}} \int_{\frac{M_N^{(2)}}{M_N^{(1)}}}^1 W(v) \left| \psi_{N,1} \left(\lambda - \frac{v}{M_N^{(2)}} \right) - \phi(\lambda) \right|^{\frac{p}{\alpha}} dv \\ &\quad + \frac{1}{1 - \frac{M_N^{(2)}}{M_N^{(1)}}} \int_{\frac{M_N^{(2)}}{M_N^{(1)}}}^1 W(v) \left| \psi_{N,2} \left(\lambda - \frac{v}{M_N^{(2)}} \right) \right|^{\frac{p}{\alpha}} dv. \end{aligned}$$

On the other hand,

$$\int_{\frac{M_N^{(2)}}{M_N^{(1)}}}^1 W(v) \left| \psi_{N,1} \left(\lambda \pm \frac{v}{M_N^{(2)}} \right) - \phi(\lambda) \right|^{\frac{p}{\alpha}} dv \leq \int_0^1 W(v) \left| \psi_{N,1} \left(\lambda \pm \frac{v}{M_N^{(2)}} \right) - \phi(\lambda) \right|^{\frac{p}{\alpha}} dv$$

For all λ belonging to $]-\pi, \pi[$ $\psi_{N,1} \left(\lambda \pm \frac{v}{M_N^{(2)}} \right)$ converges to $\phi(\lambda)$, uniformly in

$v \in [-1, 1]$. Therefore, $\frac{1}{1 - \frac{M_N^{(2)}}{M_N^{(1)}}} \int_{\frac{M_N^{(2)}}{M_N^{(1)}}}^1 W(v) \left| \psi_{N,1} \left(\lambda \pm \frac{v}{M_N^{(2)}} \right) - \phi(\lambda) \right|^{\frac{p}{\alpha}} dv$ converge to zero.

$$\psi_{N,2} \left(\lambda_i \pm \frac{v}{M_N^{(2)}} \right) \leq \sum_{\substack{m=1 \\ m \neq i}}^q \frac{a_m}{B'_{\alpha,N}} \frac{1}{\left| \sin \left[\frac{1}{2} \left(w_i \pm \frac{\tau v}{M_N^{(2)}} - w_m \right) \right] \right|^{2k\alpha}}$$

Since $\tau\lambda = w_i$,

$$+ \frac{a_i}{B'_{\alpha,N}} \frac{1}{\left| \sin \left[\frac{1}{2} \left(\pm \frac{\tau v}{M_N^{(2)}} \right) \right] \right|^{2k\alpha}}.$$

$$\lim_{N \rightarrow +\infty} \sup_{v \in [-1,1]} \frac{1}{\left| \sin \left[\frac{1}{2} \left(w_i \pm \frac{\tau v}{M_N^{(2)}} - w_m \right) \right] \right|^{2k\alpha}} = \frac{1}{\left| \sin \left[\frac{1}{2} (w_i - w_m) \right] \right|^{2k\alpha}}.$$

For all $m \neq i$,

Thus, for large N, we get

$$\sum_{\substack{m=1 \\ m \neq i}}^q \frac{a_m}{B'_{\alpha,N}} \frac{1}{\left| \sin \left[\frac{1}{2} \left(w_i \pm \frac{\tau v}{M_N^{(2)}} - w_m \right) \right] \right|^{2k\alpha}} \leq \left(\varepsilon + \frac{1}{\inf_{m \in \{1,2,\dots,q\} - \{i\}} \left| \sin \left[\frac{1}{2} (w_i - w_m) \right] \right|^{2k\alpha}} \right) \sum_{\substack{m=1 \\ m \neq i}}^q \frac{a_m}{B'_{\alpha,N}}.$$

$$\sum_{\substack{m=1 \\ m \neq i}}^q \frac{a_m}{B'_{\alpha,N}} \frac{1}{\left| \sin \left[\frac{1}{2} \left(w_i \pm \frac{\tau v}{M_N^{(2)}} - w_m \right) \right] \right|^{2k\alpha}} = O \left(\frac{1}{n^{2k\alpha-1}} \right).$$

From the lemma 2.1, we obtain

(13)

$$\sup_{v \in [0,1]} \left| \frac{\tau v}{M_N^{(2)}} \right| = \frac{\tau}{M_N^{(2)}} < \pi$$

On the other hand, for large N, we have . Consequently

$$\frac{a_i}{B'_{\alpha,N}} \frac{1}{\left| \sin \left[\frac{1}{2} \left(\pm \frac{\tau v}{M_N^{(2)}} \right) \right] \right|^{2k\alpha}} \leq \frac{a_i}{B'_{\alpha,N}} \frac{\pi^{2k\alpha}}{\left| \frac{\tau v}{M_N^{(2)}} \right|^{2k\alpha}}.$$

$$\frac{\tau}{M_N^{(1)}} \leq \frac{\tau v}{M_N^{(2)}} \leq \frac{\tau}{M_N^{(2)}}, \text{ we obtain}$$

$$\frac{a_i}{B'_{\alpha,N}} \frac{1}{\left| \sin \left[\frac{1}{2} \left(\pm \frac{\tau v}{M_N^{(2)}} \right) \right] \right|^{2k\alpha}} \leq \frac{a_i}{B'_{\alpha,N}} \frac{\pi^{2k\alpha}}{\left| \frac{\tau}{M_N^{(1)}} \right|^{2k\alpha}}.$$

The lemma 2.2 gives

$$\frac{a_i}{B'_{\alpha,N}} \frac{1}{\left| \sin \left[\frac{1}{2} \left(+ - \frac{\tau v}{M_N^{(2)}} \right) \right] \right|^{2k\alpha}} = O \left(\frac{(M_N^{(1)})^{2k\alpha}}{n^{2k\alpha-1}} \right). \quad (14)$$

Thus, we get

$$\frac{2^{\frac{p}{\alpha}}}{1 - \frac{M_N^{(2)}}{M_N^{(1)}}} \int_{\frac{M_N^{(1)}}{M_N^{(2)}}}^1 W(v) \left| \psi_{N,2} \left(\tau \lambda \pm \frac{\tau v}{M_N^{(2)}} \right) \right|^{\frac{p}{\alpha}} dv = O \left[\frac{1}{n^{\frac{(2k\alpha-1)p}{\alpha}}} + \left(\frac{(M_N^{(1)})^{2k\alpha}}{n^{2k\alpha-1}} \right)^{\frac{p}{\alpha}} \right] \quad (15)$$

Choosing $M_N^{(1)}$ such that $\frac{(M_N^{(1)})^{2k\alpha}}{n^{2k\alpha-1}}$ converges to 0. For example $M_N^{(1)} = n^b$ with $0 < b < 1 - \frac{1}{2k\alpha}$. Thus, $\lim_{N \rightarrow +\infty} E[f_N(\lambda)] - [\phi(\lambda)]^{\frac{p}{\alpha}} = 0$.

Theorem 4.2 Let $-\Omega < \lambda < \Omega$ such that $\phi(\lambda) > 0$. Then, $\text{var}[f_N(\lambda)]$ converges to zero. If $M_N^{(1)} = n^c$ with $\frac{1}{2k^2\alpha^2} < c < \frac{1}{2}$, then $\text{var}[f_N(\lambda)] = O\left(\frac{1}{n^{(1-2c)}}.$

Proof Consider the case where $\tau\lambda \notin \{w_1, w_2, \dots, w_q\}$. It is clear that the variance of f_N can be written as follows:

$$\text{var}[f_N(\lambda)] = \int_{R^2} W_N^{(1)}(\lambda - u) W_N^{(1)}(\lambda_1 - u') \text{cov}[\hat{I}_N(u), \hat{I}_N(u')] du du'.$$

By using the fact that W is zero for $|\lambda| > 1$, for large N , we get

$$\text{var}[f_N(\lambda)] = \int_{-1}^1 \text{cov} \left[\hat{I}_N \left(\lambda - \frac{x_1}{M_N} \right), \hat{I}_N \left(\lambda - \frac{x'_1}{M_N} \right) \right] W(x_1) W(x'_1) dx_1 dx'_1.$$

We define two subsets of the $[-1, 1]^2$ by:

$$\begin{aligned} \bullet L_1 &= \{(x_1, x'_1) \in [-1, 1]^2; \quad |x_1 - x'_1| > \sigma_N\}, \\ \bullet L_2 &= \{(x_1, x'_1) \in [-1, 1]^2; \quad |x_1 - x'_1| \leq \sigma_N\}, \end{aligned}$$

where σ_N is a nonnegative real, converging to 0. We split the integral into an integral over the sub region L_2 and an integral over L_1 : $\text{var}[f_N(\lambda)] = \int_{L_2} + \int_{L_1} \stackrel{\Delta}{=} J_1 + J_2$.

By Cauchy-Schwartz inequality and theorem 3.1, we obtain

$$J_1 \leq C \int_{|x_1 - x_1'| \leq \sigma_N} W(x_1) W(x_1') dx_1 dx_1' \quad \text{where } C \text{ is constant. Thus, we obtain}$$

$$J_1 = O(\sigma_N) \quad (16)$$

It remains to show that J_2 converges to zero. For simplicity, we define

$$\lambda_1 = \lambda - \frac{x_1}{M_N^{(1)}}; \quad \lambda_2 = \lambda - \frac{x_1'}{M_N^{(1)}}, \quad \text{and} \quad C(\lambda) = \text{cov} \left[\hat{I}_N \left(\lambda - \frac{x_1}{M_N^{(1)}} \right), \hat{I}_N \left(\lambda - \frac{x_1'}{M_N^{(1)}} \right) \right].$$

We first show that $C(\lambda)$ converges to zero uniformly in $x_1, x_1' \in [-1, 1]$. Indeed, from lemma 2.3, we have

$$E \hat{I}_N(v) - \hat{I}_N(v) = F_{p,\alpha}^{-1} [C_\alpha]^{-p/\alpha} \int_{-\infty}^{\infty} \frac{\text{Re} \left(e^{i u I_N(v)} \right) - e^{-C_\alpha |u|^\alpha \psi_N(v)}}{|u|^{1+p}} du.$$

Thus, the expression of the covariance becomes

$$\begin{aligned} C(\lambda) &= F_{p,\alpha}^{-2} C_\alpha^{-\frac{2p}{\alpha}} \int_{\mathbb{R}^2} E \left[\prod_{k=1}^2 \cos(u_k I_N(\lambda_k)) \right] \\ &\quad - \exp \left\{ -C_\alpha \sum_{k=1}^2 |u_k|^\alpha \psi_N(\lambda_k) \right\} \frac{du_1 du_2}{|u_1 u_2|^{1+p}} \end{aligned}$$

The following equality : $2 \cos x \cos y = \cos(x+y) + \cos(x-y)$, implies that

$$\begin{aligned} E \left[\prod_{k=1}^2 (\cos u_k I_N(\lambda_k)) \right] &= \frac{1}{2} \exp \left[-C_\alpha \int \left| (\tau)^\frac{1}{\alpha} \sum_{k=1}^2 u_k H_N(\tau \lambda_k - \tau v) \right|^\alpha d\mu(v) \right] \\ &\quad + \frac{1}{2} \exp \left[-C_\alpha \int \left| (\tau)^\frac{1}{\alpha} \sum_{k=1}^2 (-1)^{k+1} u_k H_N(\tau \lambda_k - \tau v) \right|^\alpha d\mu(v) \right]. \end{aligned}$$

By substituting the expression for $C(\lambda)$ and changing the variable u_2 to $(-u_2)$ in the second term, we obtain

$$C(\lambda) = F_{p,\alpha}^{-2} C_\alpha^{-\frac{2p}{\alpha}} \int_{\mathbb{R}^2} \left(e^{-K} - e^{-K'} \right) \frac{du_1 du_2}{|u_1 u_2|^{1+p}}, \quad (17)$$

where $K = C_\alpha \int_{\mathbb{R}} \left| (\tau)^{\frac{1}{\alpha}} \sum_{k=1}^2 u_k H_N(\tau \lambda_k - \tau v) \right|^\alpha d\xi(v)$ and

$$K' = C_\alpha \tau \sum_{k=1}^2 |u_k|^\alpha \int_{\mathbb{R}} |H_N(\tau \lambda_k - \tau v)|^\alpha \xi(v) dv$$

Since $K, K' > 0$, $\left| e^{-K} - e^{-K'} \right| \leq |K - K'| \exp\{|K - K'| - K'\}$, we obtain:

$$|K - K'| \leq 2C_\alpha \tau |u|^\alpha Q_N(\lambda_1; \lambda_2), \quad \text{where}$$

$$Q_N(\lambda_1; \lambda_2) = \int_{-\Omega}^{\Omega} |H_N(\tau \lambda_1 - \tau u)|^{\frac{\alpha}{2}} |H_N(\tau \lambda_2 - \tau u)|^{\frac{\alpha}{2}} d\xi(u)$$

Now, let us show that $Q_N(\lambda_1; \lambda_2)$ converges to zero. Indeed, since ϕ is bounded on $[-\Omega, \Omega]$, we have

$$\begin{aligned} Q_N(\lambda_1; \lambda_2) &\leq \sup(\phi) \int_{-\Omega}^{\Omega} |H_N(\tau \lambda_1 - \tau u) H_N(\tau \lambda_2 - \tau u)|^{\frac{\alpha}{2}} du \\ &\quad + \sum_{i=1}^q c_i |H_N(\tau \lambda_1 - \tau w_i) H_N(\tau \lambda_2 - \tau w_i)|^{\frac{\alpha}{2}} \end{aligned} \quad (18)$$

From the definition of H_N , we can write

$$\int_{-\Omega}^{\Omega} |H_N(\tau \lambda_1 - \tau v) H_N(\tau \lambda_2 - \tau v)|^{\frac{\alpha}{2}} dv = \int_{-\Omega}^{\Omega} \frac{1}{B_{\alpha,N}} \left| \frac{\sin\left[\frac{n}{2}(\tau \lambda_1 - \tau v)\right]}{\sin\left[\frac{1}{2}(\tau \lambda_1 - \tau v)\right]} \right|^{k\alpha} \left| \frac{\sin\left[\frac{n}{2}(\tau \lambda_2 - \tau v)\right]}{\sin\left[\frac{1}{2}(\tau \lambda_2 - \tau v)\right]} \right|^{k\alpha} dv.$$

a) First step:

We show that the denominators of the first and second terms under the last integral do not vanish for the same v , so we suppose that v exists, belonging to $[-\Omega, \Omega]$ and $z, z' \in \mathbb{Z}$ such as: $\tau \lambda_1 - \tau v = 2z\pi$ and $\tau \lambda_2 - \tau v = 2z'\pi$. Since $\lambda_1 \neq \lambda_2$, then z and z' are different.

Therefore, $z - z' = \frac{\tau}{2\pi}(\lambda_1 - \lambda_2)$. Hence, $|z - z'| = \frac{1}{w} |\lambda_1 - \lambda_2|$. As $\lim_{N \rightarrow \infty} |\lambda_1 - \lambda_2| = 0$,

consequently, for a large N we get: $|z - z'| < \frac{1}{2}$.

Thus, we obtain a contradiction with the fact that z and z' are different integers.

b) Second step

We assume there exist q points, $V_1, V_2, \dots, V_q \in [-\Omega, \Omega]$ such as:

for $j = 1, 2, \dots, q$ $\tau\lambda_1 - \tau V_j \in 2\pi Z$ then $\frac{\lambda_1}{w} - \frac{V_j}{w} \in Z$, and we assume that there exist q' points $V'_1, V'_2, \dots, V'_{q'} \in [-\Omega, \Omega]$ such that for $i = 1, 2, \dots, q'$ $\frac{\lambda_2}{w} - \frac{V'_i}{w} \in Z$. Showing that, $-1 < \frac{\lambda - \Omega}{w} < 0$ and $0 < \frac{\lambda + \Omega}{w} < 1$ because $w > 2\Omega$. Hence $\frac{\lambda - \Omega}{w} \notin Z$ and $\frac{\lambda + \Omega}{w} \notin Z$. For a large N , we get that $\left[\frac{\lambda - \Omega}{w} \right]_E < \frac{\lambda_1}{w} - \frac{\Omega}{w} < 1 + \left[\frac{\lambda - \Omega}{w} \right]_E$, where $[x]_E$ is the integer part of x . Hence, $\frac{\lambda_1}{w} - \frac{\Omega}{w} \notin Z$. In the same manner, we show that $\frac{\lambda_1}{w} + \frac{\Omega}{w} \notin Z$. Similarly, it can be shown that $\frac{\lambda_2 + \Omega}{w} \notin Z$. Thus, $|V_j| \neq \Omega$ and $|V'_i| \neq \Omega$.

c) Third step

We classify V_j and V'_i by increasing order: $-\Omega < V_{j_1} < V_{j_2} < \dots < V_{j_{q+q'}} < \Omega$, and we write the integral in the following manner:

$$\int_{-\Omega}^{\Omega} \left| \frac{\sin \left[\frac{n}{2} (\tau\lambda_1 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau\lambda_1 - \tau v) \right]} \right|^{k\alpha} \left| \frac{\sin \left[\frac{n}{2} (\tau\lambda_2 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau\lambda_2 - \tau v) \right]} \right|^{k\alpha} dv = I_1 + \sum_{i=1}^{q+q'} I_{2,i} + \sum_{i=1}^{q+q'-1} I_{3,i} + I_4$$

where

$$\begin{aligned}
I_1 &= \int_{-\Omega_1}^{V_{j_1} - \delta(N)} \left| \frac{\sin \left[\frac{n}{2} (\tau \lambda_1 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau \lambda_1 - \tau v) \right]} \right|^{k\alpha} \left| \frac{\sin \left[\frac{n}{2} (\tau \lambda_2 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau \lambda_2 - \tau v) \right]} \right|^{k\alpha} dv \\
I_{2,i} &= \int_{V_{j_i} - \delta(N)}^{V_{j_i} + \delta(N)} \left| \frac{\sin \left[\frac{n}{2} (\tau \lambda_1 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau \lambda_1 - \tau v) \right]} \right|^{k\alpha} \left| \frac{\sin \left[\frac{n}{2} (\tau \lambda_2 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau \lambda_2 - \tau v) \right]} \right|^{k\alpha} dv \\
I_{3,i} &= \int_{V_{j_i} + \delta(N)}^{V_{j_{i+1}} - \delta(N)} \left| \frac{\sin \left[\frac{n}{2} (\tau \lambda_1 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau \lambda_1 - \tau v) \right]} \right|^{k\alpha} \left| \frac{\sin \left[\frac{n}{2} (\tau \lambda_2 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau \lambda_2 - \tau v) \right]} \right|^{k\alpha} dv \\
I_4 &= \int_{V_{j_{q+q'}} + \delta(N)}^{\Omega} \left| \frac{\sin \left[\frac{n}{2} (\tau \lambda_1 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau \lambda_1 - \tau v) \right]} \right|^{k\alpha} \left| \frac{\sin \left[\frac{n}{2} (\tau \lambda_2 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau \lambda_2 - \tau v) \right]} \right|^{k\alpha} dv
\end{aligned}$$

where $\delta(N)$ is a nonnegative real number converging to zero and satisfying:

$$-\Omega < V_{j_1} - \delta(N) < V_{j_1} + \delta(N) < V_{j_2} - \delta(N) < V_{j_2} + \delta(N) < \dots < V_{j_{q+q'}} - \delta(N) < V_{j_{q+q'}} + \delta(N) < \Omega,$$

$$\delta(N) < \left| \frac{\lambda_1 - \lambda_2}{2} \right|.$$

and

First, we show that the first integral converges to zero. We know that for a large N , we have $\lambda_1 < \Omega$. For simplicity without loss of generality, we assume that for all i , $\tau \lambda_1 - \tau V_{j_i} = 2k_i \pi$ with $k_i \in \mathbb{Z}$. Similarly, we can show the result if we rather assume that $\tau \lambda_2 - \tau V_{j_i} = 2k_i \pi$ with $k_i \in \mathbb{Z}$. Since there is no v between $-\Omega$ and $V_{j_1} - \delta(N)$ on which the denominators are vanishing,

$$I_1 \leq \frac{V_{j_1} - \delta(N) + \Omega}{\inf \left[\left| \sin \frac{\tau \delta(N)}{2} \right|^{k\alpha}, \left| \sin \frac{\tau (\lambda_1 + \Omega)}{2} \right|^{k\alpha} \right]} \frac{1}{\inf \left[\left| \sin \frac{\tau (\lambda_2 - V_{j_1} + \delta(N))}{2} \right|^{k\alpha}, \left| \sin \frac{\tau (\lambda_2 + \Omega)}{2} \right|^{k\alpha} \right]}.$$

By substituting for V_{j_1} in the last inequality, we obtain

$$\left| \sin \frac{\tau (\lambda_2 - V_{j_1} + \delta(N))}{2} \right|^{k\alpha} = \left| \sin \frac{\tau |\lambda_2 - \lambda_1 + \delta(N)|}{2} \right|^{k\alpha}.$$

For a large N , we have

$\frac{\tau |\lambda_2 - \lambda_1 + \delta(N)|}{2} \leq \tau\Omega + \frac{\tau\delta(N)}{2} < \pi - \frac{\tau\delta(N)}{2}$. On the other hand, two cases are possible:

1. $\lambda_2 - \lambda_1 > 0$, then we have $|\lambda_2 - \lambda_1 + \delta(N)| = \lambda_2 - \lambda_1 + \delta(N) > \delta(N)$
2. $\lambda_2 - \lambda_1 < 0$, since $|\lambda_2 - \lambda_1| > 2\delta(N)$, we have

$$|\lambda_2 - \lambda_1 + \delta(N)| = \lambda_1 - \lambda_2 - \delta(N) > \delta(N).$$

Therefore, $\frac{\tau\delta(N)}{2} < \frac{\tau |\lambda_2 - \lambda_1 + \delta(N)|}{2} < \pi - \frac{\tau\delta(N)}{2}$. For a large N , we have $\tau\delta(N) < 2\pi - \tau(\lambda_1 + \Omega)$ and $\tau\delta(N) < 2\pi - \tau(\lambda_2 + \Omega)$. Then,

$$\frac{\tau\delta(N)}{2} < \frac{\tau(\lambda_1 + \Omega)}{2} < \pi - \frac{\tau\delta(N)}{2} \quad \text{and} \quad \frac{\tau\delta(N)}{2} < \frac{\tau(\lambda_2 + \Omega)}{2} < \pi - \frac{\tau\delta(N)}{2}.$$

Consequently,

$$I_1 \leq \frac{V_{j_1} - \delta(N) + \Omega}{\left| \sin \frac{\tau\delta(N)}{2} \right|^{2k\alpha}}.$$

For the integral $I_{2,i}$, we bound the first fraction under integral by $n^{k\alpha}$:

$$I_{2,i} \leq n^{k\alpha} \int_{V_{j_i} - \delta(N)}^{V_{j_i} + \delta(N)} \frac{1}{\left| \sin \left[\frac{1}{2} (\tau\lambda_2 - \tau v) \right] \right|^{k\alpha}} dv.$$

By substituting for V_{j_i} in the last inequality and

putting $v = u - \frac{2k\pi}{\tau}$, we get

$$I_{2,i} \leq n^{k\alpha} \int_{\lambda_1 - \delta(N)}^{\lambda_1 + \delta(N)} \frac{1}{\left| \sin \left[\frac{1}{2} (\tau\lambda_2 - \tau u) \right] \right|^{k\alpha}} du.$$

Since $|\lambda_1 - u| < \delta(N)$, it is easy to note that :

$$|\lambda_2 - u| \geq |\lambda_2 - \lambda_1| - |\lambda_1 - u| \geq |\lambda_2 - \lambda_1| - \delta(N) > \frac{|\lambda_2 - \lambda_1|}{2}$$

Since $\delta(N)$ converges to zero, for a large N we have $\delta(N) < \frac{2}{\tau} \left(\pi - \frac{\tau}{2} |\lambda_2 - \lambda_1| \right)$, therefore

$$0 < \tau \frac{|\lambda_2 - \lambda_1|}{4} < \tau \frac{|\lambda_2 - u|}{2} < \tau \frac{|\lambda_2 - \lambda_1| + \delta(N)}{2} < \pi.$$

Consequently:

$$I_{2,i} \leq \frac{2\delta(N)n^{k\alpha}}{\inf \left[\sin \frac{|\lambda_1 - \lambda_2| \tau}{4}; \sin \frac{|\lambda_1 - \lambda_2| \tau + \delta(N)\tau}{2} \right]},$$

where $|\lambda_2 - \lambda_1| = \left| \frac{x_2 - x_1}{M_N} \right|$. Then, for a large N , we have $\frac{3\tau}{2}|\lambda_2 - \lambda_1| < 2\pi$. Therefore, $\tau \frac{|\lambda_2 - \lambda_1|}{4} < \frac{\tau |\lambda_2 - \lambda_1| + \tau\delta(N)}{2} < \pi - \tau \frac{|\lambda_2 - \lambda_1|}{4}$. Thus, we bound the integral as follows:

$$I_{2,i} \leq \frac{2\delta(N)n^{k\alpha}}{\left| \sin \frac{\tau |\lambda_2 - \lambda_1|}{4} \right|^{k\alpha}}$$

Since there is no v between $V_{j_i} + \delta(N)$ and $V_{j_{i+1}} - \delta(N)$ on which the denominators are vanishing, we get:

$$I_{3,i} \leq \frac{V_{j_{i+1}} - V_{j_i} - 2\delta(N)}{A \times B},$$

where

$$\begin{aligned} A &= \inf \left(\left| \sin \frac{\tau\lambda_1 - \tau V_{j_i} - \tau\delta(N)}{2} \right|^{k\alpha}, \left| \sin \frac{\tau\lambda_1 - \tau V_{j_{i+1}} + \tau\delta(N)}{2} \right|^{k\alpha} \right) \\ B &= \inf \left(\left| \sin \frac{\tau\lambda_2 - \tau V_{j_i} - \tau\delta(N)}{2} \right|^{k\alpha}, \left| \sin \frac{\tau\lambda_2 - \tau V_{j_{i+1}} + \tau\delta(N)}{2} \right|^{k\alpha} \right) \end{aligned}$$

It follows from the hypothesis on $\delta(N)$ that

$$\frac{\tau\delta(N)}{2} < \frac{\tau |\lambda_1 - \lambda_2|}{2} - \frac{\tau |\delta(N)|}{2} < \frac{\tau |\lambda_1 - \lambda_2 - \delta(N)|}{2} < \frac{\tau |\lambda_1 - \lambda_2|}{2} + \frac{\tau |\delta(N)|}{2} < \pi - \frac{\tau\delta(N)}{2}.$$

Thus, by using the definition of V_{j_i} , we obtain

$$\frac{1}{\left| \sin \frac{\tau \lambda_2 - \tau V_{j_i} - \tau \delta(N)}{2} \right|^{k\alpha}} < \frac{1}{\left| \sin \frac{\tau \delta(N)}{2} \right|^{k\alpha}}.$$

We use the same way for bounding the other terms

$$I_{3,i} \leq \frac{V_{j_{i+1}} - V_{j_i} - 2\delta(N)}{\left| \sin \frac{\tau \delta(N)}{2} \right|^{2k\alpha}}.$$

and we get :

Similarly, since there is no v between $V_{j_{q+q'}} + \delta(N)$ and Ω which the denominators are vanishing, we can show that:

$$I_4 \leq \frac{V_{j_{i+1}} - V_{j_i} - 2\delta(N)}{E \times F}, \text{ where}$$

$$E = \inf \left(\left| \sin \frac{\tau \lambda_1 - \tau \Omega}{2} \right|^{k\alpha}, \left| \sin \frac{\tau \lambda_1 - \tau V_{j_{q+q'}} - \tau \delta(N)}{2} \right|^{k\alpha} \right)$$

$$F = \inf \left(\left| \sin \frac{\tau \lambda_2 - \tau \Omega}{2} \right|^{k\alpha}, \left| \sin \frac{\tau \lambda_2 - \tau V_{j_{q+q'}} - \tau \delta(N)}{2} \right|^{k\alpha} \right)$$

Since $\delta(N)$ converges to zero, for a large N , we have $\delta(N) < \frac{2}{\tau} \left(\pi - \frac{\tau}{2} |\lambda_1 - \Omega| \right)$, and

$$I_4 \leq \frac{\Omega - V_{j_{q+q'}} - \delta(N)}{\left| \sin \frac{\tau \delta(N)}{2} \right|^{2k\alpha}}.$$

It follows that: $\delta(N) < \frac{2}{\tau} \left(\pi - \frac{\tau}{2} |\lambda_1 - \lambda_2| \right)$. It follows that:

We recapitulate, from the

previous increases, we obtain

$$\int_{-\Omega}^{\Omega} \left| \frac{\sin \left[\frac{n}{2} (\tau \lambda_1 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau \lambda_1 - \tau v) \right]} \right|^{k\alpha} \left| \frac{\sin \left[\frac{n}{2} (\tau \lambda_2 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau \lambda_2 - \tau v) \right]} \right|^{k\alpha} dv \leq \frac{\Omega + V_{j_1} - \delta(N)}{\left| \sin \frac{\tau \delta(N)}{2} \right|^{2k\alpha}} + \sum_{i=1}^{q+q'} \frac{2n^{k\alpha} \delta(N)}{\left| \sin \frac{\tau |\lambda_2 - \lambda_1|}{4} \right|^{k\alpha}}$$

$$+ \sum_{i=1}^{q+q'-1} \frac{V_{j_{i+1}} - V_{j_i} - \delta(N)}{\left| \sin \frac{\tau \delta(N)}{2} \right|^{2k\alpha}} + \frac{\Omega - V_{j_{q+q'}} - \delta(N)}{\left| \sin \frac{\tau \delta(N)}{2} \right|^{2k\alpha}}$$

After simplification, we have

$$\int_{-\Omega}^{\Omega} \left| \frac{\sin \left[\frac{n}{2} (\tau \lambda_1 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau \lambda_1 - \tau v) \right]} \right|^{k\alpha} \left| \frac{\sin \left[\frac{n}{2} (\tau \lambda_2 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau \lambda_2 - \tau v) \right]} \right|^{k\alpha} dv \leq \frac{2\Omega - 2(q + q' + 1)\delta(N)}{\left| \sin \frac{\tau \delta(N)}{2} \right|^{2k\alpha}} + \frac{2n^{k\alpha} \delta(N)(q + q')}{\left| \sin \frac{\tau |\lambda_2 - \lambda_1|}{4} \right|^{k\alpha}}.$$

By bounding the first term on the right hand side of the last inequality and using the following

inequality $\left| \sin \frac{x}{2} \right| \geq \frac{|x|}{\pi}$, we get

$$\int_{-\Omega}^{\Omega} \left| \frac{\sin \left[\frac{n}{2} (\tau \lambda_1 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau \lambda_1 - \tau v) \right]} \right|^{k\alpha} \left| \frac{\sin \left[\frac{n}{2} (\tau \lambda_2 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau \lambda_2 - \tau v) \right]} \right|^{k\alpha} dv \leq \frac{2\Omega \pi^{2k\alpha}}{(\tau \delta(N))^{2k\alpha}} + \frac{2n^{k\alpha} \delta(N)(q + q')(2\pi)^{k\alpha}}{\left(\frac{\tau |x_2 - x_1|}{M_N} \right)^{k\alpha}}.$$

It follows from the lemma 2.1

$$\begin{aligned} \frac{1}{B'_{\alpha,N}} \int_{-\Omega}^{\Omega} \left| \frac{\sin \left[\frac{n}{2} (\tau \lambda_1 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau \lambda_1 - \tau v) \right]} \right|^{k\alpha} \left| \frac{\sin \left[\frac{n}{2} (\tau \lambda_2 - \tau v) \right]}{\sin \left[\frac{1}{2} (\tau \lambda_2 - \tau v) \right]} \right|^{k\alpha} dv \leq \\ \frac{1}{\pi} \left(\frac{\pi}{2} \right)^{2k\alpha} \left(\frac{2\Omega \pi^{2k\alpha}}{n^{2k\alpha-1} (\tau \delta(N))^{2k\alpha}} + \frac{2\delta(N)(q + q')(2\pi)^{k\alpha}}{n^{k\alpha-1} \left(\frac{\tau \sigma_N}{M_N} \right)^{k\alpha}} \right). \end{aligned} \quad (19)$$

In order to obtain the convergence of the last expression to zero, we choose $\delta(N) = n^{-\beta}$, $\beta > 0$, such as

$$\lim_{n \rightarrow \infty} \frac{n^{2k\alpha\beta}}{n^{2k\alpha-1}} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{1}{n^{\beta+k\alpha-1} \left(\frac{\sigma_N}{M_N} \right)^{k\alpha}} = 0. \quad (20)$$

Thus, from (18) \mathcal{Q}_N converges to zero. On the other hand,

$$C(\lambda) \leq F_{p,\alpha}^{-2} C_{\alpha}^{-\frac{2p}{\alpha}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |K - K'| e^{|K-K'|-K'} \frac{du_1 du_2}{|u_1 u_2|^{1+p}},$$

where $|K - K'| - K' \leq -C_{\alpha} \sum_{k'=1}^2 |u_{k'}|^{\alpha} |\psi_N(\lambda_{k'}) - \tau Q_N(\lambda_1; \lambda_2)|$.

We denote by : $\delta_{(N,k')} = \psi_N(\lambda_{k'}) - \tau Q_N(\lambda_1; \lambda_2)$. It follows from (18) and (20) that $\delta_{(N,k')}$ converges to $\phi(\lambda)$. Hence,

$$C(\lambda) \leq F_{p,\alpha}^{-2} C_\alpha^{-\frac{2p}{\alpha}} 2\tau C_\alpha Q_N(\lambda_1; \lambda_2) 4 \prod_{k'=1}^2 \int_0^\infty \exp[-C_\alpha (u_{k'})^\alpha |\delta_{(N,k')}|] \frac{du_{k'}}{(u_{k'})^{1+p-\frac{\alpha}{2}}}.$$

Putting $u_{k'} (\delta_{(N,k')})^{\frac{1}{\alpha}} = v$, we obtain

$$C(\lambda) \leq F_{p,\alpha}^{-2} C_\alpha^{-\frac{2p}{\alpha}} 2\tau C_\alpha \frac{Q_N(\lambda_1; \lambda_2)}{[\delta_{(N,1)} \delta_{(N,2)}]^{\frac{1-p}{2\alpha}}} \left(\int_{-\infty}^\infty \frac{e^{-C_\alpha |v|^\alpha}}{|v|^{1+p-\frac{\alpha}{2}}} dv \right)^2. \quad (20)$$

Since $\phi(\lambda) > 0$, $C(\lambda)$ converges uniformly in $x_1, x'_1 \in [-1, 1]$ to zero. From (20), we obtain

$$J_2 = O \left(\frac{1}{n^{2k\alpha(1-\beta)-1}} + \frac{1}{n^{\beta+k\alpha-1} \left(\frac{\sigma_N}{M_N} \right)^{k\alpha}} \right). \quad \text{Thus, } \text{var}[f_N(\lambda)] \text{ converges to zero and then, } f_N(\lambda) \text{ is an asymptotically unbiased and consistent estimator.}$$

4. CONCLUSIONS

In this paper, we estimate the spectral density of mixed stable process with continuous time when the process is observed at discrete instants. The aliasing phenomenon is avoided by assuming that the spectral is a compact support. This work can be applied in various fields. For example:

- The study of soil cracking where the observed signal is the resistance of the soil. This resistance can have random jumps that are due to the presence of some stones. Thus, the spectral measurement will be composed of two parts, one continuous and the other discrete corresponding to the resistance jumps encountered during the measurement.

- The growth of fruit on a tree can be considered as a continuous distribution, and when there is a fall of a fruit, the other fruits remaining on the tree absorb more energy and their growth will have a jump in value.

The perspective of this work is to optimize the smoothing parameters to have a better rate of convergence. For this purpose, the cross-validation method will be the most appropriate tool.

This work can also be completed by considering a more general case when we observe the process with random errors. In this case, we will use the deconvolution methods, which have proved their efficiency in the presence of random errors.

It would be interesting to give an estimator of the mode of the spectral density representing the frequency where the spectral density reaches the maximum of energy. For that, we must estimate the derivative of the spectral density function.

REFERENCES

- [1] S. Cambanis, (1983)“Complex symmetric stable variables and processes” In P.K.SEN, ed, *Contributions to Statistics”: Essays in Honour of Norman L. Johnson* North-Holland. New York,(P. K. Sen. ed.), pp. 63-79
- [2] S. Cambanis, and M. Maejima (1989). “Two classes of self-similar stable processes with stationary increments”. *Stochastic Process. Appl.* Vol. 32, pp. 305-329
- [3] M.B. Marcus and K. Shen (1989). “Bounds for the expected number of level crossings of certain harmonizable infinitely divisible processes”. *Stochastic Process. Appl.*, Vol. 76, no. 1 pp 1-32.
- [4] E. Masry, S. Cambanis (1984). “Spectral density estimation for stationary stable processes”. *Stochastic processes and their applications*, Vol. 18, pp. 1-31 North-Holland.
- [5] G. Samorodnitsky and M. Taqqu (1994). “Stable non Gaussian processes ». Chapman and Hall, New York.
- [6] K., Panki and S. Renming (2014). “Stable process with singular drift”. *Stochastic Process. Appl.*, Vol. 124, no. 7, pp. 2479-2516
- [7] C. Zhen-Qing and W. Longmin (2016). “Uniqueness of stable processes with drift.” *Proc. Amer. Math. Soc.*, Vol. 144, pp. 2661-2675
- [8] K. Panki, K. Takumagai and W. Jiang (2017). “Laws of the iterated logarithm for symmetric jump processes”. *Bernoulli*, Vol. 23, n° 4 pp. 2330-2379.
- [9] M. Schilder (1970). “Some Structure Theorems for the Symmetric Stable Laws”. *Ann. Math. Statist.*, Vol. 41, no. 2, pp. 412-421.
- [10] R. Sabre (2012b). “Missing Observations and Evolutionary Spectrum for Random”. *International Journal of Future Generation Communication and Networking*, Vol. 5, n° 4, pp. 55-64.
- [11] E. Sousa (1992). “Performance of a spread spectrum packet radio network link in a Poisson of interferences”. *IEEE Trans. Inform. Theory*, Vol. 38, pp. 1743-1754
- [12] M. Shao and C.L. Nikias (1993). “Signal processing with fractional lower order moments: Stable processes and their applications”, *Proc. IEEE*, Vol.81, pp. 986-1010
- [13] C.L. Nikias and M. Shao (1995). “Signal Processing with Alpha-Stable Distributions and Applications”. Wiley, New York
- [14] S. Kogon and D. Manolakis (1996). “Signal modeling with self-similar alpha- stable processes: The fractional levy motion model”. *IEEE Trans. Signal Processing*, Vol 44, pp. 1006-1010
- [15] N. Azzaoui, L. Clavier, R. Sabre, (2002). “Path delay model based on stable distribution for the 60GHz indoor channel” *IEEE GLOBECOM*, IEEE, pp. 441-467
- [16] J.P. Montillet and Yu. Kegen (2015). “Modeling Geodetic Processes with Levy alpha-Stable Distribution and FARIMA”, *Mathematical Geosciences*. Vol. 47, no. 6, pp. 627-646.
- [17] M. Pereyra and H. Batalia (2012). “Modeling Ultrasound Echoes in Skin Tissues Using Symmetric alpha-Stable Processes”. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. 59, n°. 1, pp. 60-72.
- [18] X. Zhong and A.B. Premkumar (2012). “Particle filtering for acoustic source tracking in impulsive noise with alpha-stable process”. *IEEE Sensors Journal*, Vol. 13, no. 2, pp. 589 - 600.
- [19] Wu. Ligang and W. Zidong (2015). “Filtering and Control for Classes of Two-Dimensional Systems”. *The series Studies in Systems of, Decision and Control*, Vol.18, pp. 1-29.
- [20] N. Demesh (1988). “Application of the polynomial kernels to the estimation of the spectra of discrete stable stationary processes”. (Russian) *Akad.Nauk.ukrain. S.S.R. Inst.Mat. Preprint* 64, pp. 12-36
- [21] F. Brice, F. Pene, and M. Wendler, (2017) “Stable limit theorem for U-statistic processes indexed by a random walk”, *Electron. Commun. Probab.*, Vol. 22, no. 9, pp.12-20.
- [21] R. Sabre (2019). “Alpha Stable Filter and Distance for Multifocus Image Fusion”. *IJSPS*, Vol. 7, no. 2, pp. 66-72.
- [22] JN. Chen, J.C. Coquille, J.P. Douzals, R. Sabre (1997). “Frequency composition of traction and tillage forces on a mole plough”. *Soil and Tillage Research*, Vol. 44, pp. 67-76.
- [23] R. Sabre (1995). “Spectral density estimate for stationary symmetric stable random field”, *Applicaciones Mathematicas*, Vol. 23, n°. 2, pp. 107-133

- [24] R. Sabre (2012a). "Spectral density estimate for alpha-stable p-adic processes". *Revisita Statistica*, Vol. 72, n°. 4, pp. 432-448.
- [25] R. Sabre (2017). "Estimation of additive error in mixed spectra for stable process". *Revisita Statistica*, Vol. LXXVII, n°. 2, pp. 75-90.
- [26] E. Masry, (1978). "Alias-free sampling: An alternative conceptualization and its applications", *IEEE Trans. Information theory*, Vol. 24, pp.317-324.

AUTHOR

Rachid Sabre received the PhD degree in statistics from the University of Rouen, France, in 1993 and Habilitation to direct research (HdR) from the University of Burgundy, Dijon, France, in 2003. He joined Agrosup Dijon, France, in 1995, where he is an Associate Professor. From 1998 through 2010, he served as a member of Institut de Mathématiques de Bourgogne, France. He was a member of the Scientific Council AgroSup Dijon from 2009 to 2013. From 2012 to 2019, he has been a member of Laboratoire Electronique, Informatique, and Image (LE2I), France. Since 2019 he has been a member of Laboratory Biogeosciences UMR CNRS University Burgundy. He is author/co-author of numerous papers in scientific and technical journals and conference proceedings. His research interests lie in areas of statistical process and spectral analysis for signal and image processing.

AN EFFICIENT METHOD FOR A SPECIFIC CASE OF DETECTING IMPULSE NOISE ON SCANNED DOCUMENTS

Petar Prvulović¹, Jelena Vasiljević¹, Dhinakaran Nagamalai²

¹School of Computing, Union University, Belgrade, Serbia

²Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil-626126, India

ABSTRACT

This paper explains a method used to detect the presence of impulse noise in a set of scanned documents as a part of OCR preprocessing. As the document set is supposed to be processed in large scale, the primary concern of the noise detection method was efficiency within existing project constraints. Following the nature of noise, the method seeks to detect the presence of noise in document margins. The method works in two stages. First stage is margin detection, based on color spectre analysis. Second stage is noise recognition in margin samples, based on a pixel contrast score. The resulting implementation proved efficient both in terms of detection accuracy and algorithmic complexity.

KEYWORDS

Impulse noise, Noise detection, Margin detection

1. INTRODUCTION

The problem we faced and developed a method to solve is the detection of specific cases of impulse noise in a set of scanned documents. Solution needed to take in account the boundaries of the project which introduced a method fully adapted to the document set and underlying code base. Document set features a large number of business reports, in the order of millions, scanned as PDF files where each page is a separate JPEG file. Pages were scanned in grayscale CCITTFax format. Data processing program, which the method proposed here is part of, is implemented in Python 3 and uses numpy and OpenCV for other processing steps. These constraints do not affect the essence of the method as the method itself doesn't use any external functions, but they do dictate the coding style [1].

By inspecting the documents manually, several types of noise were noticed [2, 3]. Impulse noise is uniformly distributed over the document. Stains are mostly present in corners of pages as they originate from stapling the paper. Lines, both horizontal and vertical, appear near the edges. Impulse noise presents the largest problem and is in the focus of the presented method. We have taken into account the other two types of noise, so as to make the presented method resistant to their appearance, as it was noticed that it affects noise detection, as explained later in this paper.

When impulse noise is present, the OCR process introduces errors. A variety of despeckling filters is available and some of them proved good in noise removal in this specific case. In a manually processed sample it was noticed that applying a despeckling filter on noisy documents improves

OCR accuracy, but increases error rate if applied on a clean document. Hence it is necessary to accurately check for noise presence, in order to apply a despeckling filter only where needed.

Our aim was to design a method which reduces both the number of passes and the coverage of each pass as a strategy to save on execution time and complexity. Nature of impulse noise in the dataset allows for noise detection in areas without text, where page margins are a good candidate for sampling. On a uniform background it is possible to detect noise by analysing the color spectre of pixels. Both steps, margin detection and noise detection by color spectre, can be implemented in an efficient manner and within given constraints.

1.1. Related Work

There is a variety of noise detection techniques, as noted in [4,5,6]. Most common way of noise detection is to process the document with a noise removal filter and measure the difference of original vs. the filtered version. If there is no significant difference the document is marked as noise-free, because noise removal didn't introduce any changes. Drawback of this approach is that there are several passes over the entire document: to apply noise removal filters, to measure the difference, and more, depending on the nature of the detection process.

Document margin detection, in our specific case, is a special case of a more general text detection problem. As neural networks gained popularity, recent research seem to be focused more on general cases like paper and text detection and recognition in videos and images in a natural scene. There is a number of approaches to this problem, which provide decent results. Some of them can be found in [7, 8]. In [7] authors detect four types of text images: document images, scene images, born-digital images and heterogeneous text images. Our case falls into first type. Methods mentioned include: edge extraction and smoothing, clustering pixels with similar color and other methods aimed towards detecting the text region; detecting discriminative features of characters in text; segmenting the image and detecting the blocks which contain text; etc. As the purpose of these methods is to define a boundary in which the text is contained, they are designed in such a way that they need to pass through the entire image, which is the characteristic we wanted to avoid in our specific case.

1.2. Selected Approach

As previously noted, our aim was to detect page margin area and use it to sample the image and detect if noise is present using the color spectre. Having a very specific case of scanned documents, methods that rely on text boundary detection introduce unneeded processing steps as they need to pass through the entire image. We avoided that, and reduced the page processing to 40% of the page area by using the fact that the pages in our set of scanned documents are scanned straight, without rotation, and that the paper covers the entire image, hence eliminating the need to detect the page text boundary.

The proposed method works in two stages. First stage detects document margins using color spectre. Second stage checks if impulse noise is present within margins, as we treat them as an empty area on a document where noise is easily distinguished. Method is implemented as a main function and a set of helper functions. Main function returns boolean, indicating that the provided bitmap contains noise.

This paper is organised as follows: we first explain prerequisites and define an efficiency boundary which should be met in order to improve execution time of the entire process; then we develop both stages of the method. We conclude the paper with test results over a manually

prepared dataset where documents were separated into two groups: with and without noise. Test results prove that the efficiency of the proposed method is within acceptable boundaries.

2. PREREQUISITES

Manual inspection of the dataset showed that about 3% of documents contain impulse noise. These documents have to be pre-processed using a filter. We aim to save on execution time in the other 97% of documents by introducing a method which is faster than the noise removal process. Let's denote execution time of noise removal process as R and noise detection process as D . In case of using only noise removal process, overall execution time would be $T_1 = 1 * R$. In case of using a detection process, noise removal would be performed only on 3% of samples. Overall execution time in this case is: $T_2 = 1 * D + 0.03 * R$. To achieve $T_2 < T_1$, we need $D < 0.97 * R$. This boundary is used in 6.1. Noise Detection Complexity, to confirm that the proposed method is within acceptable range.

Documents are scanned in portrait orientation, with little to no rotation. Where present, rotation is $< 0.5^\circ$. Page size varies around 3500px in height. Expected displacement of the upper and lower corner of the text area is set to be less than 16px. Each page is scanned as a JPEG picture in RGB format, though they are all in greyscale. Initial unpacking and transformation from PDF file gives us an $n \times m$ matrix of byte values, where each value represents a single pixel. Values range from 0-255, 0 being black and 255 being white. Documents seem to be preprocessed during the scanning, as all of them featured black text, there were no greyish and washed-out samples. Documents contain left aligned or justified text and/or tables, with clearly distinguishable margins. Nature of contents dictates uniform formatting. Typical page in the observed dataset is shown in Figure 1. Few cases were found with contents in header area, which doesn't affect our method as these sections are ignored, so cutting off a part of header or footer, with page number or similar, is acceptable.

Tables are often bordered and OCR showed better results when borders are removed, hence horizontal and vertical line removal is a required step and as such it is not considered as overhead if performed as part of noise detection method. This proved beneficiary in the margin detection stage. Line filtering is performed initially, after the page is loaded into memory as an OpenCV image. We applied a special horizontal and vertical kernel to detect lines and fill them with white pixels [9]. As this is a required pre-processing step for OCR, we do not count it into the noise detection process when we calculate the complexity.



Figure 1. Page example

3. MARGIN DETECTION

The method implements detection of left and right margins. We consider these two sufficient for noise detection in second stage. We expect that the outer side of the margin has significantly more light-colored pixels than the inner side. Margin color spectre is created by counting number of pixels for each of 256 levels of gray in bitmap columns. An example of spectre plot is depicted in Figure 2. and it shows the spectre of the first left quarter of a page depicted in Figure 1. Plot is generated using Python's matplotlib library. Black color indicates 0 values while white color indicates maximum value. Plot is in grayscale representation, which means that spectre values are scaled to [0-255] range. As there is disproportionately more white pixels, white color biases the plot, so to get better plot resolution we put a zero value for white. On Figure 2 we notice a sudden jump around column 350, which corresponds to left margin. x-axis indicates grayscale values (0-black, 255-white). y-axis indicates bitmap column.

As the left and right margins are vertically oriented and taking in account that rotation of text is less than 0.5° we expect that the transition step in color spectre is less than 16px wide, as stated in 2. Prerequisites. In Figure 2. we indeed notice a sharp jump in the darker part of the spectre, on the left side of the plot. Similar follows in the lighter part of the spectre. Gray shades are induced by interpolation and JPEG's lossy nature. Though this confirms the statement it is the consequence and not the reason and we will focus solely on the darker part of the spectre. With this, we reduced the problem of finding a margin to the problem of detecting a step in the spectre plot.

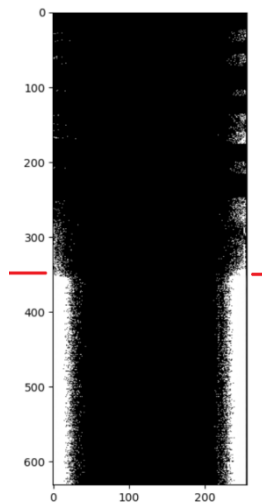


Figure 2. Page color spectre plot



Figure 3. Reduced spectre

We see no difference in black and “almost black” pixels, as they both represent a symbol on a page. Spectre color range is reduced to 8 groups, so the first group represents shades from 0-31, second from 32-63 etc. Figure 3. shows the reduced spectre, with the darker part only. It is noticeable, and expected, that most of the non-zero values fall into the first group. We will extract this group as a vector of values, where each value presents a number of pixels coloured 0-31 in a corresponding bitmap column. Figure 4. shows such vectors of left, right, top and bottom margin, in respective order. For the left margin we see a distinct jump around column 350, where the margin is located. We notice that the right margin follows the same pattern, though we expect more oscillations due to uneven text alignment, as text is not necessarily justified and not all lines end exactly at the margin, but we still notice a jump. Top and bottom margins are hard to distinguish as we cannot be certain if a spike is generated by a line of text or a stain, and instead of a single step, we have multiple small steps as there is empty space between lines. This case is harder to address properly and the method efficiency is within acceptable boundaries with left and right margin only, so we can safely ignore these two cases and focus on left and right margin only. Right margin is symmetrical to the left margin. We can apply the same processing for both margins, as long as we take care of the inverted direction of the right margin.

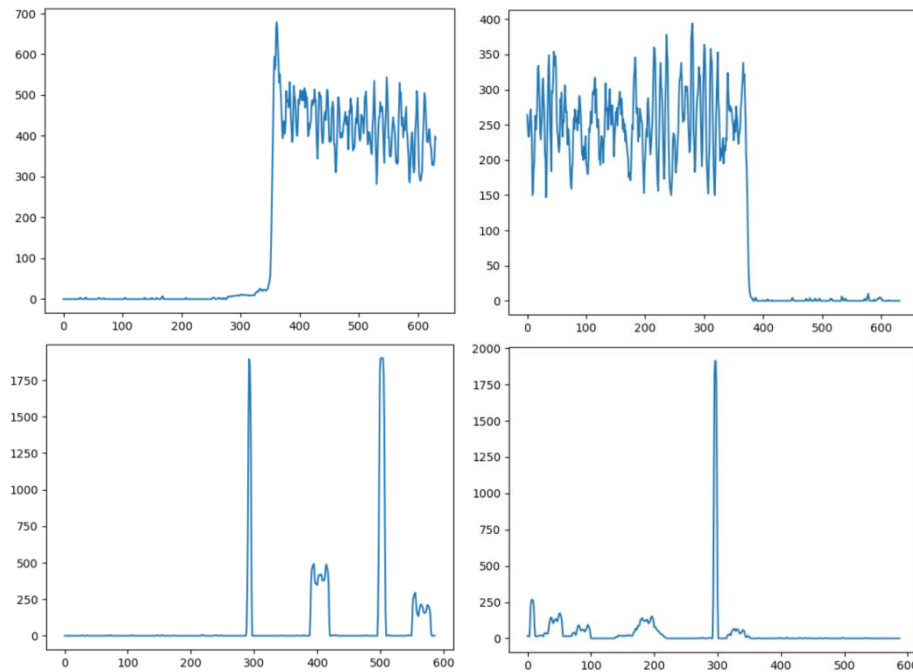


Figure 4. Black color vector line chart of left, right, top and bottom margin

3.1. Step Detection

The problem of finding a margin is now reduced to a problem which is in literature referred to as “step detection”, “step finding” and similar. There are plenty of methods which can be applied, like [10]. As we want to find one single, highest, jump it is possible to use the following steps:

1. Smooth the vector using average filter
2. Calculate a vector of neighbour increments using a 16px slide window
3. Smoothen that vector by cutting off small, below average, values
4. Find the leftmost (rightmost) hill

Average filter takes in account five elements, two on the left, two on the right and the current element. Current element is replaced by the calculated average value. Leftmost and rightmost elements use zero to pad the out-of-bounds vector values.

Neighbour increments are calculated differences between two consecutive vector elements. We expect to have the highest value at the position of step. As the page can be tilted up to 16 pixels, it is possible that the step is spread across these 16 bitmap columns/corresponding vector values. For that reason, we use a slide window to calculate the neighbour increment vector. Instead of using only two consecutive vector elements, we will use previous 8 and next 8, and sum their consecutive differences as a value. This has two useful effects: it will cause further flattening of parts with small oscillations; and it will cause accumulation of consecutive jumps and emphasizing the value in margin position in case of a tilted margin.

Figure 5. shows the original vector in blue, smoothened vector in orange and neighbour increments vector in green. We see negative values which imply drops of vector. As we are looking for the highest increment, we can safely ignore these values and treat them as zero.

Margin position is decided by finding the leftmost hill. We expect it to be the highest, but anticipate that it is possible to have two or more such hills, in case of formatted tabular text, hence the leftmost hill is preferred. It is possible that the neighbour increments vector contains lower hills left of margin position. These need to be removed. Simple and efficient method is to find the average hill height and flatten all the hills below the average. Figure 6. shows the smoothened vector in orange and neighbour increments vector with slide window in green, with all the below-average values trimmed to 0.

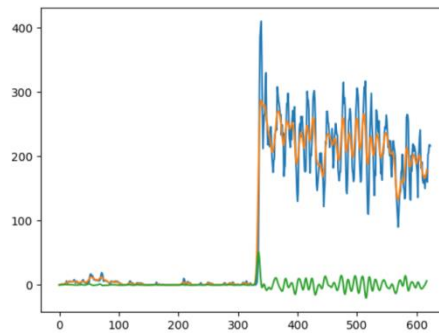


Figure 5. Original vector, smoothened vector and neighbour increments vector

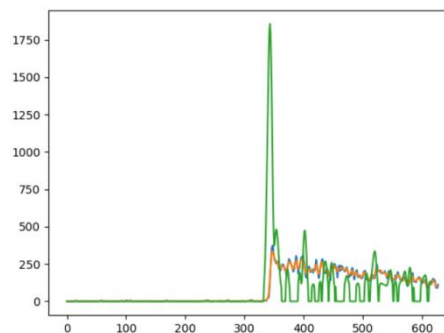


Figure 6. A slide-window neighbour increments vector after trimming

Leftmost hill corresponds to the left margin position. Finding a hill is a simple algorithmic task of iterating through an array and finding the first non-zero value which is greater than its left and right neighbour. Same process follows for right margin. We need to invert the vector and take care to calculate the margin position accordingly, as an offset from right page border.

Accuracy of this method was tested by adapting the code to draw vertical lines on detected margin positions over a set of pages. Manual inspection showed that the method detects margins correctly, with acceptable error rate. In additional calibrating, margins were moved 8 pixels to the outside, to ensure that margin area doesn't contain edges of letters.

4. NOISE RECOGNITION

When present, noise is spread across the entire page with more or less density and granulation. Noise samples vary from a single black pixel to a small group of pixels. Pixel colors vary from black to grayish, but match the previously used first group of quantized colors, with pixel color ranging 0-31. Lighter pixels do not affect OCR efficiency and can be safely ignored. Noise density varies from sample to sample, but on a single sample it is constant.

By applying the previously described method for margin detection we are able to extract two parts of a document which are (mostly) empty. In an empty area noise is easy to detect by simply counting black pixels and setting a threshold. Problem with threshold is that noise density is not the same in all the documents. Setting a high threshold would produce false negatives in documents with low noise density, while setting a low threshold would produce false positives in pages where there is some hanging text or stains. Another problem which we want to address are stains, which should not be counted as noise, as that would distort the process.

By analysing the documents we found the following characteristic of noise: noise is mostly a group of pixels in the form of a dot of varying size, up to 20 pixels. As measuring sizes of each point would be an “expensive” task in terms of execution time, we defined a metric based on contrast and found a threshold value based on statistics of a sample set of documents. A side effect of this method is that it successfully ignores large stains, as these become statistically irrelevant, and thus avoids distortion by this type of noise.

4.1. Measuring Pixel Contrast

Pixel contrast can be expressed and measured in several ways. We have a grayscale image with colors quantized to 8 values which can further be simplified to a monochrome image, as we consider 0-31 as black and the rest as non-black. Contrast of a single pixel is measured by counting how many of 8 surrounding pixels are of the same color. By such measure, a single black pixel on a white background has the highest contrast score of 8. A white pixel on white background has the lowest contrast score of 0, the same as black pixel on a black background – a pixel inside a noisy point or a stain.

Figure 7 shows a zoomed-in part of the margin with noise samples and the same margin after measuring contrast. Contrast values are scaled to match the grayscale spectre, so 0 maps to black and 8 to white. We see that the proposed measure emphasizes edges of noise samples, as this is the area with high contrast.

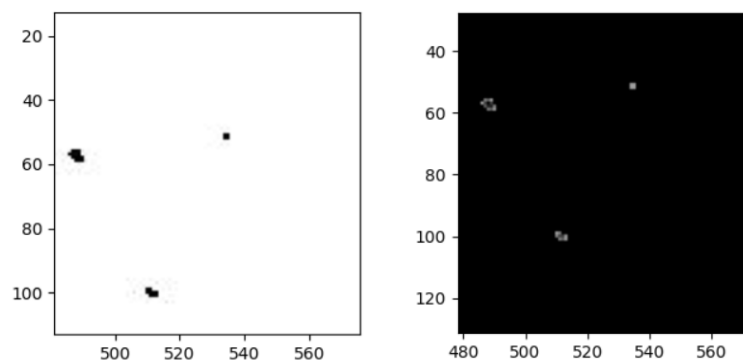


Figure 7. (a) Sample of margin (b) Corresponding pixel contrast values

4.2. Calculating the Noise Granulation Threshold

Margin dimensions vary around 3500x300px, with roughly 1.000.000 pixels in a margin. Noisy documents feature several hundreds of noise samples. Number itself varies and cannot be considered a candidate for a threshold value. Presence of black pixels is an obvious parameter, and we want to determine if black pixels are grouped into small dots, typical to noise, or a large stain which can be ignored as it is not a case of impulse noise.

Size of a stain dictates the ratio of pixels on the edge and pixels inside the stain. Edge pixels have a higher contrast score, while inner pixels have a contrast score of 0. As noise stains are mostly round we expect that the ratio of edge vs. inner pixels be around 0.5. In case of a noise featuring only single black pixels, the ratio is 1, as 100% of pixels are contrasting, edge, pixels. In case of a noise featuring one large stain, the ratio would be low, as the majority of black pixels are inside the stain. We conclude that the smaller the stain the higher the ratio.

To find the ratio threshold we will process a manually separated set of noisy documents. Processing is performed in the following manner: for each document in sample set detect margins; quantize margins to monochrome, by treating 0-31 as black and 32-255 as white; count black pixels; count contrasting pixels; and calculate the ratio as number of contrasting pixels divided by number of black pixels. Results of this process, applied on a sample set, are listed in Table 1. We see that the average ratio is roughly 0.6, which means that the majority of granulation is such that 60% of pixels are on the edge and 40% on the inside of a stain. Ratio of 0.4 is selected as a threshold value of declaring that a margin is filled with the noise in question.

Table 1. Black to contrast pixels ratio in margins of a test set

Margin	Value	Min	Average	Max
Left	Black pixels	1	1328.36	13134
	Contrasting pixels	1	841.52	7862
	Contrasting/black pixels ratio	0.24	0.78	1.00
Right	Black pixels	14	10477.92	66678
	Contrasting pixels	14	6198.08	40126
	Contrasting/black pixels ratio	0.42	0.76	1.00

5. THE NOISE DETECTION METHOD

In 3. we defined the method for detecting left and right margin. In 4. we defined the method to test whether an empty area contains noise. In this specific set of documents margins can be considered as empty, text-free, areas and used to test if document is noisy. Nature of noise is such that the noise, if present, is uniformly spread across the document. Hence, we expect that both margins have similar noise density and granulation. Margins are not the same width, right margin is usually narrower, so margin width ratio needs to be taken in account. We used the ratio of black pixels to all pixels in a sample. Although expected to be approximately the same, experiments showed that this ratio is mostly between 1.1 – 1.7 but varies up to 3. Reason for this could be gradual noise granulation, where granulation decreases from left to right due to the nature of the error source, which is hard to notice and confirm in visual inspection. Experimental results, with testing over the same set where ratio in condition was set to values in range [1,10] showed that the best error-rate is achieved by setting this ratio upper limit to 3, and that for values > 3 it doesn't introduce any changes to method accuracy.

Additionally, we expect a specific type of noise, large stains, to appear in the significant number of documents. This type of noise is to some extent ignored in noise recognition methods, as explained above. To further improve the method we split the margins into quarters and each quarter is processed separately. That way we have four sets of parameters for each margin. Margin sections are sorted and only the middle one in each margin is used for reference. Experiments showed better performance in terms of false positives and false negatives after this step was added, as this step effectively bypasses areas with large stains, if such are present. The obvious prerequisite for noise presence is that both margins contain black pixels. Further analysis of the dataset showed that samples with less than 12 dark pixels in any of the margins are not to be considered noisy, as the noise is sparse enough that it doesn't interfere with OCR.

Following the noise recognition method previously described, the ratio of contrasting pixels to black pixels has to be at least 0.4.

Considering all these conditions, the noise detection method and the following helper functions pseudocode have the following form:

```

function isBlack(bitmap, x,y): return (bitmap[x,y] < 32)
function contrast(bitmap, x, y):
    contrastedNeighbours = 0
    for i in (-1,0,1):
        for j in (-1,0,1):
            if x+i>=0 and x+i<bitmap.width and y+j>=0 and y+j<bitmap.height
               and isBlack(bitmap,i,j) != isBlack(bitmap, x+i, y+j) :
                contrastedNeighbours++
    return contrastedNeighbours
function findMargins(bitmap):
    marginWidth = round(bitmap.width / 5)
    leftMarginSpectre = [0] * marginWidth
    rightMarginSpectre = [0] * marginWidth
    for j=0 to bitmap.height:
        for i=0 to marginWidth:
            if isBlack( bitmap, i, j): leftMarginSpectre[i]++
        for i= (bitmap.width-marginWidth) to bitmap.width:
            if isBlack( bitmap, i, j): rightMarginSpectre[i]++
    rightMarginSpectre = reversed(rightMarginSpectre)
    leftMargin = findStep( leftMarginSpectre, marginWidth ) - 8
    rightMargin = bitmap.width - findStep( rightMarginSpectre, marginWidth ) + 8
    return (leftMargin, rightMargin)
function findStep( spectre, n ):
    spectre[-2] = 0, spectre[-1] = 0, spectre[n] = 0, spectre[n+1] = 0
    for i=0 to n:
        spectre[i]=(spectre[i-2]+spectre[i-1]+spectre[i]+spectre[i+1]+spectre[i+2]) / 5
    increments = [0] * n
    incrementsSum = 0, prev8sum = 0, next8sum = 0
    for i=8 to n-8:
        prev8sum += spectre[i]
        prev8sum -= spectre[i-8]
        next8sum += spectre[i+8]
        increments[i] = next8sum - prev8sum
        incrementsSum += increments[i]
    incrementAverage = incrementsSum / (n-16)
    for i=8 to n-8:
        if increments[i] < incrementAverage:
            increments[i] = 0
    for i=1 to n-1:
        if increments[i-1]<increments[i] and increments[i]>increments[i+1]: return i
    return n
function isNoisy(pageBitmap):
    (marginLeft, marginRight) = findMargins(pageBitmap)
    left, right= [ (0,0), (0,0), (0,0), (0,0) ]
    for j=0 to pageBitmap.height:
        for i=0 to marginLeft:

```

```

    if isBlack(pageBitmap, i,j) :    left[ floor(j/4) ][0]++
    if contrast(pageBitmap, i, j) > 0 : left[ floor(j/4) ][1]++
    for i=marginRight to pageBitmap.width:
        if isBlack(pageBitmap, i,j) :    right[ floor(j/4) ][0]++
        if contrast(pageBitmap, i, j) > 0 : right[ floor(j/4) ][1]++
    sort( left,  lambda a,b: a[0]<b[0] )
    sort( right, lambda a,b: a[0]<b[0] )
    midBlackPixelsLeft, midBlackPixelsRight = left[1][0], right[1][0]
    midContrastingToBlackRatioLeft  = left[1][1] / left[1][0]
    midContrastingToBlackRatioRight = right[1][1] / right[1][0]
    ratioDifference = midContrastingToBlackRatioLeft / midContrastingToBlackRatioRight
    if ratioDifference < 1: ratioDifference = 1/ratioDifference
    if left[1][0] > 12 and right[1][0] > 12
        and midContrastingToBlackRatioLeft > 0.4
        and midContrastingToBlackRatioRight > 0.4
        and ratioDifference < 3:
            return true
    else
        return false

```

5.1. Noise Detection Method Complexity

Pseudocode above initially performs one pass over the marginal area, which is set to be 1/5 of the page on the left and right side. After that pass, a color spectre vector is formed and we have four iterations over that vector and a generated increments vector of the same size. After margins are found, we iterate over both margins to calculate needed pixel counts and generate two vectors with four elements each. Sorting these vectors is considered constant, as four elements can be sorted using a fixed set of nested if conditions.

Let's say we have a bitmap of size $n \times m$. Initial passes take $2 \times n \times m / 5$ iterations. Then we have $2 \times 4 \times m / 5$ iterations over spectre vectors. Finally we have $n \times m_l + n \times m_r$, where m_l is left margin width and m_r is right margin width. In the worst case both can be $m/5$, though we expect less in general case. This gives us $2 \times n \times m / 5 + 2 \times 4 \times m / 5 + 2 \times n \times m / 5 = 4/5 \times n \times m + 8/5 \times m$. Dominating factor here is $4/5 \times n \times m$, where we see that order of the method complexity is 80% of page size. In 2. Prerequisites, we calculated the upper boundary as $D < 0.97 \times R$, where R is the noise removal process. Noise removal process inherently iterates over entire page so order of complexity of R is $n \times m$. As $0.8 \times n \times m < 0.97 \times n \times m$ we conclude that the proposed method provides the required efficiency.

6. EXPERIMENTAL RESULTS AND DISCUSSION

Accuracy of the proposed method was tested on a manually curated set of actual documents. Test set featured two classes of documents: noisy samples and clean samples. We selected 806 samples, of which 301 were noisy and 505 were clean. Both sets were processed and noise detection results were compared to expected results. This enabled us to count false positives and false negatives. Results are presented in Table 2. We see 7% false negatives and 0.8% false positives. We consider these results acceptable, as we are more concerned to not denoise clean documents, as it introduces errors, while denoising noisy documents is considered an advantage, whatever the denoising rate is. To get the real error rate we need to take in account that noisy documents take 3% of overall documents, which means that the effective error rate is 7% of 3%.

Adjusted error rates are also included in Table 2. and we see that overall error rate is below 1%, which we consider a satisfying result.

Table 2. Test results

Set	Samples	Correct		Incorrect		Adjusted error rate	
		Count	%	Count	%	Occurrence	Adjusted %
Noisy	301	280	93%	21	7%	0.03	0.21%
Clean	505	501	99.2%	4	0.8%	0.97	0.77%
Total	806	781	96.9%	25	5.1%		0.98%

Horizontal table borders and lines in header and footer that are outdented and start inside the margin area proved to be a problematic feature in margin detection, notably on the detection of the left margin. Margin detection relies on color spectre and step detection. As text is, by rule, left-aligned, highest step will most often appear at the place where text starts. Hanging parts of borders and lines will, in that case, fall into margin part of the spectre, and will be treated as noise. Such case could skew the noise detection method and produce a false positive result. Test proved this behaviour and the effect to overall error rate was significant. Line removal is a required step in OCR preprocessing. The order of preprocessing steps doesn't affect the efficiency of implementation, so it was decided to apply line removal before the noise detection step. If used in general case, our noise detection method would have to include line removal filter as a first step and add the filter's complexity into noise detection method complexity.

Noise detection method was adjusted so to use left margin only. Initial tests provided good results in terms of low error rate. As this approach processes 50% less page area than initial, future direction of work will be focused on tweaking the noise detection method so to get satisfying error rate and reduce the processing time per page as this is a critical factor in method's design.

7. CONCLUSIONS

This paper introduced a method of noise detection in a specific surrounding. Project constraints, with large scale document processing being most important, affected the method design in such a way that we focused on achieving the efficiency through adapting to the dataset in question. As a result we got an optimised, efficient solution which satisfies project requirements.

Proposed method works in two stages. First stage detects margins by analysing color spectre of the document. We successfully tweaked the process to reduce the number of iterations over a document and constructed the entire process using elementary operations, which further improves code efficiency. Following the nature of the documents, we were able to reduce the color spectre to one-dimensional vector and perform analyses on that vector in a very efficient manner. Second stage detects presence of noise in margin area. By introducing the contrast metric we managed to perform noise detection in the environment with variable noise granulation and density.

Overall complexity of the method is within acceptable boundaries. Success rate was measured on a manually curated set of documents. Results showed a satisfying accuracy.

ACKNOWLEDGEMENTS

We thank Resolution Technology Ltd. (www.companydatashop.com) for providing curated datasets and supporting the open knowledge.

REFERENCES

- [1] Harris, C.R., Millman, K.J., van der Walt, S.J. et al. (2020) "Array programming with NumPy", *Nature* 585, 357–362. doi:10.1038/s41586-020-2649-2
- [2] Verma, Rohit & Ali, J.. (2013). "A comparative study of various types of image noise and efficient noise removal techniques", *International Journal of Advanced Research in Computer Science and Software Engineering*. 3. 617-622.
- [3] Boyat, Ajay Kumar & Brijendra Kumar Joshi (2015) "A review paper: noise models in digital image processing", *Signal & Image Processing: An International Journal (SIPIJ)* Vol.6, No.2
- [4] Saxena, Chandrika, and Deepak Kourav (2014) "Noises and image denoising techniques: a brief survey", *International journal of Emerging Technology and advanced Engineering* Vol.4, No.3
- [5] Sood, Deepika, Anureet Kaur, and Kaushik Adhikary (2013) "A Survey on Despeckling Methods", *International Journal of Emerging Technology and Advanced Engineering* Vol.3, No.7
- [6] Kaur, Jappreet, Jasdeep Kaur, and Manpreet Kaur (2011) "Survey of despeckling techniques for medical ultrasound images", *International Journal of Computer Technology and Applications* Vol.2, No.4
- [7] Karanje, U. B., Dagade, R. (2014) "Survey on text detection, segmentation and recognition from a natural scene images", *International Journal of Computer Applications* Vol.108, No.13
- [8] Agrawal, S.D., Kulliolli, V.C. (2018) "Survey of text detection methods in scene images", *EPRA International Journal of Research and Development* Vol.3, No.11
- [9] "How to remove all the detected lines from the original image using Python?," *Stack Overflow*, 16-Sep-2019. [Online]. Available: <https://stackoverflow.com/questions/57961119/how-to-remove-all-the-detected-lines-from-the-original-image-using-python/57963719#57963719>. [Accessed: 28-Jun-2021].
- [10] Basseville, M., & Benveniste, A. (1983). "Design and comparative study of some sequential jump detection algorithms for digital signals", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(3), 521–535. doi:10.1109/tassp.1983.1164131

AUTHORS

Petar Prvulović is a teaching assistant and doctoral student at Union University, School of Computing and a consultant and software developer working mainly in the Justice and Education sector.

Jelena Vasiljević is a professor at School of Computing, Union University, Belgrade, Serbia.

DCT BASED FUSION OF VARIABLE EXPOSURE IMAGES FOR HDRI

Vivek Ramakrishnan¹ and D. J. Pete²

¹Research Scholar, Department of Electronics Engineering,
Datta Meghe College of Engineering, Sector-3, Airoli,
Navi Mumbai - 400708, India

²Professor and Head, Department of Electronics Engineering,
Datta Meghe College of Engineering, Sector-3, Airoli,
Navi Mumbai - 400708, India

ABSTRACT

Combining images with different exposure settings are of prime importance in the field of computational photography. Both transform domain approach and filtering based approaches are possible for fusing multiple exposure images, to obtain the well-exposed image. We propose a Discrete Cosine Transform (DCT-based) approach for fusing multiple exposure images. The input image stack is processed in the transform domain by an averaging operation and the inverse transform is performed on the averaged image obtained to generate the fusion of multiple exposure image. The experimental observation leads us to the conjecture that the obtained DCT coefficients are indicators of parameters to measure well-exposedness, contrast and saturation as specified in the traditional exposure fusion based approach and the averaging performed indicates equal weights assigned to the DCT coefficients in this non-parametric and non pyramidal approach to fuse the multiple exposure stack.

KEYWORDS

Discrete · Exposure · Cosine · Fusion · Coefficients · Transform · Contrast · Saturation · Weights

1. INTRODUCTION

The classical approach [13] to fuse multiple-exposure image set involves Laplacian pyramid and finding the characteristic parameters like Saturation, Contrast and Well-exposedness and fusing them using the Gaussian weighting scheme. Later approaches like fusing multiple exposure images with like the ones listed in [2], and transform domain approaches [reference] started to evolve. Pixel based representation of images are possible on paper and traditional display devices. For colour image displays we assign three channels R,G and B channel and assign 8 bit per pixel for each channel. So for every colour channel 256 possible values are formed. Figure 1 depicts the variable exposure stack of the Office images. Figure 2 depicts the generated HDR image. Figure 3 shows the tone-mapped LDR image.

Various HDR formats such as the ones stated below are listed in [39]

1. OPENEXR FORMAT (or EXtended Range format or '.exr' format.).

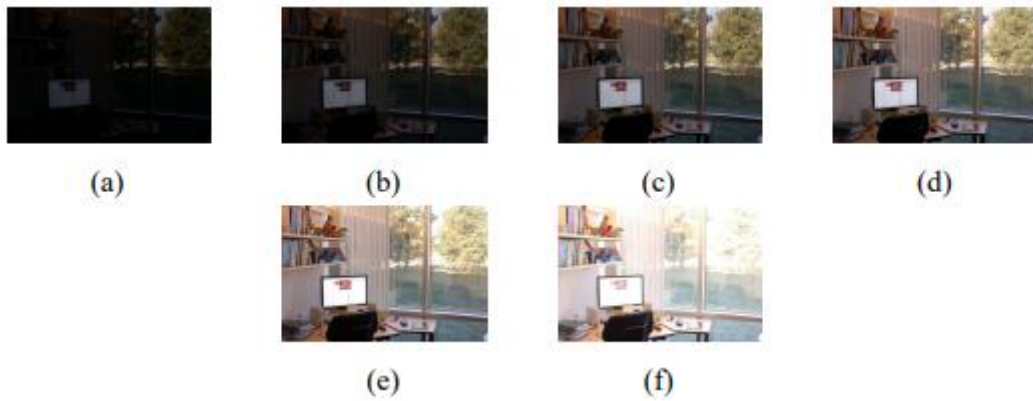


Fig. 1. Set of Six Multiple Exposure Office Images.



Fig. 2. HDR image of Office



Fig. 3. Tone-mapped image of HDR image of Office

2. Other encodings viz. scRGB48, ScRGB-nl, ScYCC-nl
3. TIFF float and the LOGLUV format.
4. HDR format (.hdr, .pic format), originally known as the Radiance picture format and Lossy HDR encodings.

In this work we perform fusion of the transform domain coefficients using a fusion rule. The multiple exposure stack is first transformed into the DCT domain and then an averaging operation is performed of the transform coefficients. This averaged output is then inverse transformed to obtain the fused output. The block variance is allowed to have a statistical distribution, and the DCT coefficients follow a Laplacian nature.[40] this then implies that we exploit the Laplacian nature of the coefficients which follows similar nature of the Laplacian distribution or Laplacian pyramidal decomposition in the spatial domain on which we perform in traditional exposure fusion.

2. CHALLENGES INVOLVED IN HDR IMAGE CAPTURE

Traditional cameras are unable to measure the radiance distribution of the scene and so it becomes unable to capture the full spectral content and dynamic range of the scene. There are inherent limitations in most digital image sensors so as to not make it possible to capture the entire dynamic range of a scene in a single exposure setting. A standard camera records multiple exposures with the right software. Selection of the right pixel to be combined to obtain the fused image is another challenging task.

Inherent assumptions include

1. Capturing device is perfectly linear
2. Each pixel in every exposure needs to be brought into the same irradiance domain
3. Corresponding pixels are averaged across the exposures.
4. Over and under exposed pixels must be excluded

Under practical considerations

1. “Cameras” are not perfectly linear.
2. Registration of various objects in the scene due to camera shake.
3. Moving objects and objects thus the following needs to be considered.
4. Understanding the nature and plotting of the Camera Response CurveFunction (CRF)
5. Involving the computations of the irradiance values.
6. Image registration techniques.
7. Ghost and Lens flare removal.

The further section discusses in detail the classical HDR imaging technique, by Debevec et. al. [2]. Then we describe the classical exposure fusion technique in the spatial domain involving pyramidal decomposition and Gaussian weighting by Mertens et. al. [1] further ahead evolving to the transform domain approaches and to the current work based on a non-pyramidal and uniform weighting approach based on DCT.

3. CRF BASED HDRI

The photochemical and the electronic reciprocity of the imaging system is exploited by the algorithm. The Hurter-Driffield characteristic curve summarizes the response of the film to variations in exposure. The curve gives a plot of the optical density D of the processed film versus the logarithm of its exposure setting. Exposure setting is obtained as the product of the irradiance E and exposure-value or exposure-time. We obtain a quantity Z which is function of the exposure setting X of a particular pixel. Consider a function f which is a conglomeration of characteristic function of the film and the non-linearities introduced by other processing steps. Initially we should recover this function f , once we have f we can recover exposure at each pixel as

$$f^{-1}(Z) \quad (1)$$

The function f is a monotonically increasing function so $f^{-1}(Z)$ is definable. If we know the exposure time $\delta(t)$, the irradiance I is obtained as

$$I = Z/(\delta(t)) \quad (2)$$

The irradiance I is proportional to the radiance L of the scene. So consider a set of exposure times $\delta(t_j)$ for a static scene with constant lighting the film reciprocity equation can be written as

$$Z_{ij} = f(I_i * \delta(t_j)) \quad (3)$$

Assuming monotonicity of f we find that f is an invertible function so we can write equation (previous) as

$$f^{-1}(Z_{ij}) = I_i * \delta(t_j) \quad (4)$$

Taking natural logarithms on both the sides we get

$$\ln(f^{-1}(Z_{ij})) = \ln(I_i) + \ln(\delta(t_j)) \quad (5)$$

Considering $h = f^{-1}(Z)$ we obtain the final equation for the characteristic curve as follows

$$h(Z_{ij}) = \ln(I_i) + \ln(\delta(t_j)) \quad (6)$$

The curve representing h is called as characteristic curve which is generalized and is independent of the input image set. Minimum two photographs with varying exposure settings are required to calculate the characteristic curve. Finally the irradiance is obtained as

$$\ln(I_i) = \frac{(\sum_{n=1}^P w(Z_{ij}) * (h(Z_{ij}) - \ln(\delta(t_j))))}{(\sum_{n=1}^P w(Z_{ij}))} \quad (7)$$

Thus we observe that the CRF based HDRI is computationally intensive and depends on log manipulations.

4. MERTEN'S CLASSICAL EXPOSURE FUSION APPROACH IN HDR

A camera of limited dynamic range is able to capture a Output referred image rather than a Scene referred image which is possible through HDR. This approach works basically in the spatial domain. Multiple exposure image set is obtained and each image gets treated by the estimated inverse camera response function (CRF). Weighted pixel blending is carried out to obtain the exposure fused image [1]. Sometimes a Multi Resolution Analysis (MRA) based blending is carried out which being a pyramidal de- composition approach gives rise to visible seams and blur. There also arises a need to go for higher order pyramidal decomposition.

5. COMPARISON BETWEEN HDR AND LDR ENCODINGS

A scene referred to an output transformation is a one-way irreversible transformation. This transformation is called by a process called as tone mapping. A typical tone mapping operation is a transformation in which the scene referred pixel is transformed into an output referred pixel. In the HDR image encoding scene referred Ness is given more prominence than an output referred Ness. Scene referred HDR encoding can be always converted to an output referred format but vice-versa is not true. So its a one-way transformation.

6. HDR IMAGERY APPLICATIONS

The following are a list of HDR Imagery applications

1. Physically based rendering (Global Illumination)
2. Remote sensing:
3. Digital Photography:
4. Image Editing:
5. Digital Cinema (and Video):
6. Virtual Reality (VR):
7. Movie film stock.

7. RELATED WORK

Debevec et. al. [8] developed a technique in which different exposure images are combined into a single radiance map such that the pixel values are in proportion to the radiance values. Bogoni and Hansen [24] stated an in which the Luma and the Chroma components were decomposed into Laplacian pyramid and Gaussian pyramid respectively. A fusion based on Illumination estimation was suggested by [25], Vonikakis et. al. The problem of motion blur in longer exposure images was addressed by Tico et. al. [26]. In the method developed by Jinho and Okuda [27] they designed weighting functions that not dependant on exposure areas. Shen [28] used Generalized random walks to arrive at an optimal solution. Li and Kang [29] suggested a weighted sum fusion approach. Independent Component Analysis (ICA) was used by Mitianoudis and Stathaki [30] to propose their fusion scheme. Fusion involving weighted averaging of the input image stack was proposed in ([5], [6],[7]). Fusion approaches based on the irradiance values, and not the intensity values of the pixels was developed in ([8], [9]). Transform based approach are discussed in ([38], [4]). Goshtas by developed a block based fusion approach [10]. Compositing without Matte was developed in [12] and the classical exposure fusion approach in [13] are two other important techniques of compositing. Bi lateral filter based method is suggested in [1]. A Bayesian approach for Matte generation is suggested in [11]. Approaches related to Markov random field [14], boundary location information [15], Poisson Solver based [16] and an optimization approach in [17] are other fully developed techniques. Making use of the fore(background) statistics an approach was developed in [18], using motion cues in [19] and through defocus cue in [20]. A generalistic α blending approach is stated in [21].

8. SOME POPULAR HDR IMAGING TECHNIQUES USED FOR COMPARISON IN THIS WORK

We now give a brief account of some state of the art HDR imaging techniques which are popular and can be used for comparison with this technique. Ashikmin et. al.[31]

developed a technique which approach which maps the HDR image input to the luminance values possible to displayed by the display device, Banterlee et. al. [32] presented a luminance zone based approach where every zone is processed by a Tone Mapping Operator (TMO) compatible for each zone. Raman et. al. [1] developed a bilateral filter based compositing. Bruce et. al. [33] developed a fusion approach based on relative entropy. Lischinski et. al. A method based on interactive local adjustment of tonal values was proposed by [34]. Tan et. al. [35] provides a logarithmic tone mapping algorithm. Mertens et. al. [13] present a parametric classical multiple exposure fusion approach. Reinhard et. al. [36] present a traditional tone mapping methodology. In addition we have transform domain approaches [4] and edge preserving Savitsky Golay filtering based approach [37].

9. THE DISCRETE COSINE TRANSFORM (DCT)

The Discrete Cosine Transform used mainly in JPEG compression is a block transform. The basic formulation of the DCT was done by Ahmed et al. [46]. DCT exhibits unitary property and is similar to the Karhunen-Loève Transform (KLT). It has got good energy compaction efficiency and rate distortion function.. In the KLT [47] we observe decorrelated transform coefficients and the signal energy is compacted in the first fewest sub-bands. The $N \times N$ cosine transform matrix $V = v(k, n)$, also known as the Discrete Cosine Transform (DCT) which is defined as follows for $k = 0$ and $0 \leq n \leq (N - 1)$

$$v(k, n) = \frac{1}{\sqrt{N}} \quad (8)$$

and for $0 \leq k \leq (N - 1)$, $0 \leq n \leq (N - 1)$

$$v(k, n) = \frac{\sqrt{2}}{\sqrt{N}} * \cos \frac{\pi(2n + 1)k}{2N} \quad (9)$$

The properties of DCT are

1. DCT is not the real part of the unitary Discrete Fourier Transform (DFT)
2. DCT is real orthogonal
3. DCT is a fast transform.
4. Excellent energy compaction efficiency is seen in the DCT for highly correlated data.
5. The basis vectors of DCT are eigenvectors of the symmetric tridiagonal matrix.
6. DCT is very close to the KLT.

10. REGARDING DISTRIBUTION OF 2-D BLOCK DCT COEFFICIENTS

Pratt (1978) [41] carried out the earlier work on DCT Coefficients, it was conjectured that the DC coefficients are Rayleigh distributed assuming no level shift and the AC coefficients are Gaussian distributed according to the central limit theorem considering each pixel to be statistically independent from another. Tescher (1979) [42] and Mu-rakami, Hatori and Yamamoto (1982) [43] indicated that the AC coefficients are Laplacian distributed, and the DC coefficients are Gaussian distributed. Reininger and Gibson [44] also asserted that DC coefficient follows a Gaussian PDF, and AC coefficients follow a Laplacian PDF. Reininger and Gibson performed the Kolmogorov-Smirnov (KS) test [45] on a 256x256 8-bit PCM grayscale images over multiple modes, and determined the best fit for the different PMFs. Observing the results of the (KS) tests [44] it is concluded that the AC block DCT coefficients follow a Laplacian distribution.

11. MOTIVATION

The drawbacks of fusion in the spatial domain include

1. Brightness changes are vast around the edges.
2. Low pass filtered version is subtracted from the original to get the Laplacian pyramid based decomposition.
3. Pixels with higher frequency that is higher rate of change of illumination are retained.
4. We get reduced number pixels by sub-sampling the edges.

5. Sub-sampling gives disadvantage of approximating the non-edge pixels with the highly varying brightness edge pixels.
6. The weight maps also depict severe variations.
7. There is degradation in the contrast and there are low frequency brightness change in the pixels which are not edge pixels.

Few of the initial steps which are performed to get a low pass filtered version and subtracting it to get the Laplacian decomposition are achieved by the one step DCT transform itself. The finding of the pixel coefficients for compositing the images is simplified due to uniform weights being assigned to each pixel and the averaging operation. We achieve a seamless blending and the results are comparable with standard exposure fusion based approach and NSST based approach.

12. OUR METHOD

The steps involved in our algorithm are as follows

1. Perform the DCT transform of all the images in the image stack.
2. The DCT Coefficients of all these images are averaged out to obtain the resultant final image in the transform domain.
3. The inverse DCT transform of the resultant image in the transform domain is performed to obtain the fused result.

Thus we can note that all the pixels have equal prominence so equal weights are assigned to all the pixels which are captured in the transform domain. The transform coefficients are empowered enough to give prominence in their luma value and chroma values according to the exposure settings.

13. RESULTS AND DISCUSSION

The DCT-based exposure fusion outputs are compared with the techniques described in ([1],[4],[13],[31],[32],[33],[34],[35],[36] and [37]). The PSNR and the SSIM scores are improved. There is a decrease in error i.e. the IMMSE scores. High Q Scores are obtained using the VDP metric [23]. For the following set of images, shown in figures 5, 7, 9, 11 kindly refer to the table 1 shown below for knowing which of the sub-figure was generated using what approach.

Table 1. List of the sub figures indexes, showing results for images generated using various approaches used for comparison in this work.

Sub figure index	Results using method described in
a	S G Filter based [37]
b	Ashikmin et. al. [31]
c	Banterlee et. al. [32]
d	Shanmuga et. al. [1]
e	Bruce et. al. [33]
f	Lischinski et. al. [34]
g	Logarithmic et. al. [35]
h	Mertens et. al. [13]
i	Reinhard et. al. [36]
j	NSST Based [4]
k	Our Method in this paper

The PSNR, SSIM, IMMSE scores and the Q factor score [23] are tabulated in the table 2.



Fig. 4. Variable Exposure House Images.

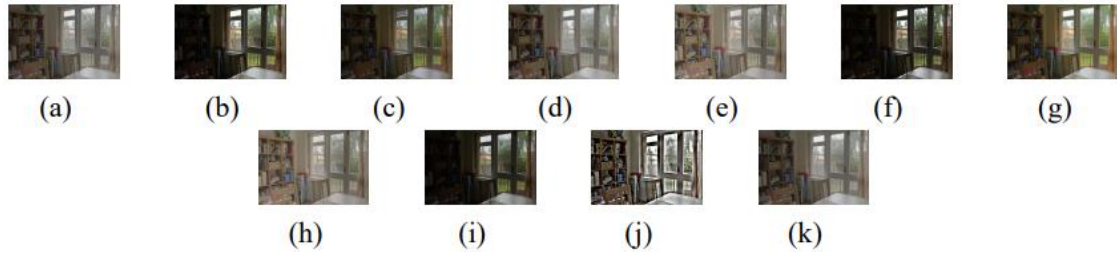


Fig. 5. Results generated for House images with approaches mentioned in table 1.



Fig. 6. Variable exposure Window Images.



Figure 7. Results generated for Windows images with approaches mentioned in table 1.



Figure 8. Variable exposure Door Image set.



Figure 9. Results generated for Door images with approaches mentioned in table 1.



Figure 10. Variable Exposure Garage Image set.

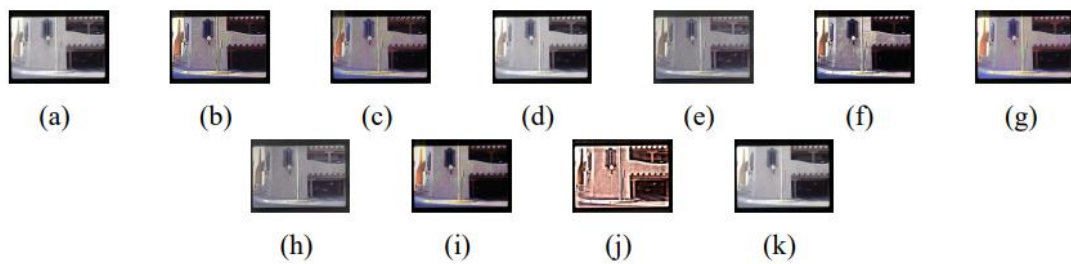


Figure 11. Results generated for Garage images with approaches mentioned in table 1

Table 2. Performance parameters for the DCT-based Exposure Fusion and other state of the art approaches described in ([1],[4],[13],[31],[32],[33],[34],[35],[36]and[37])

Image	Method	PSNR	IMMSE	SSIM	Q
HOUSE IMAGE	Ashikmin[31]	66.1274	0.0300	0.9877	98.0525
	Banterlle[32]	70.1584	0.0179	0.9890	98.1235
	BFbased[1]	67.0624	0.0261	0.9878	98.2014
	Bruce[33]	62.8707	0.0523	0.9839	98.1466
	Lischinski[34]	64.7319	0.0375	0.9865	98.0537
	Logarithmic[35]	68.5648	0.0214	0.9885	98.1503
	Mertens[13]	62.8707	0.0523	0.9839	98.466
	Reinhard[36]	63.5942	0.0366	0.9869	98.1023
	SGFilterBased[37]	68.1417	0.0226	0.9883	98.1874
	NSSTBased[4]	63.6711	0.0452	0.9866	98.0366
	DCTBased(OurMethod)	67.7379	0.0238	0.9881	98.1997
DOOR IMAGE	Ashikmin[31]	61.7326	0.0649	0.9838	98.0960
	Banterlle[32]	61.5589	0.0672	0.9834	98.1026
	BFbased[1]	70.6514	0.0170	0.9892	98.2521
	Bruce[33]	61.1196	0.0733	0.9892	98.1909
	Lischinski[34]	58.2465	0.1326	0.9748	98.0905
	Logarithmic[35]	62.2420	0.0589	0.9842	98.169
	Mertens[13]	61.1196	0.0733	0.9822	98.1909
	Reinhard[36]	58.0458	0.1384	0.9737	98.1865
	SGFilterBased[37]	70.89	0.0167	0.9893	98.2389
	NSSTBased[4]	63.8374	0.0438	0.9864	98.1472
	DCTBased(OurMethod)	71.0176	0.0165	0.9893	98.2925
WINDOW IMAGE	Ashikmin[31]	65.4498	0.0333	0.9875	98.0258
	Banterlle[32]	65.4498	0.0333	0.9875	98.0258
	BFbased[1]	67.2379	0.0255	0.9882	98.0855
	Bruce[33]	61.0632	0.0745	0.9810	98.0566
	Lischinski[34]	65.4698	0.0332	0.9875	98.0036
	Logarithmic[35]	66.7944	0.0271	0.9881	98.0285
	Mertens[13]	61.0368	0.0745	0.9810	98.0566
	Reinhard[36]	66.8427	0.0719	0.9881	98.0362
	SGFilterBased[37]	70.6617	0.0170	0.9892	98.1223
	NSSTBased[4]	66.9991	0.0263	0.9883	98.1374
	DCTBased(OurMethod)	71.7756	0.0154	0.9894	98.1227
GARAGE IMAGE	Ashikmin[31]	66.4497	0.0285	0.9880	98.0795
	Banterlle[32]	65.1389	0.0351	0.9871	98.0327
	BFbased[1]	70.0185	0.0182	0.9890	98.2360
	Bruce[33]	63.7948	0.0442	0.9854	98.1752
	Lischinski[34]	67.6633	0.0240	0.9884	98.0466
	Logarithmic[35]	68.7227	0.0210	0.9888	98.0713
	Mertens[13]	63.7948	0.0442	0.9854	98.1752
	Reinhard[36]	67.2374	0.0255	0.9882	98.1057
	SGFilterBased[37]	70.6275	0.0171	0.9891	98.2424
	NSSTBased[4]	63.6227	0.0455	0.9862	98.1145
	DCTBased(OurMethod)	70.8156	0.0168	0.9891	98.2364

14. CONCLUSIONS

Thus we have developed a method for performing fusion of multiple exposure images using the DCT-based transform domain approach and studied the results. The results favor the use of DCT-based multi-exposure image fusion for the High Dynamic Range Imaging problem.

ACKNOWLEDGEMENTS

Tom Mertens, Frank Van Reeth and Jan Kautz are thanked for the set of multiple exposure images in Fig.(4). The CAVE Computer Vision Laboratory, Columbia University is thanked for the multiple exposure images in the Figures, Fig.(6) , Fig.(8) and Fig.(10).

REFERENCES

- [1] Raman, S. and S. Chaudhuri. "Bilateral Filter Based Compositing for Variable Exposure Photography." Eurographics (2009).
- [2] B. G. Gowri, V. Hariharan, S.Thara, V. Sowmya, S. S. Kumar and K. P. Soman, "2D Image data approximation using Savitzky Golay filter — Smoothing and differencing," 2013 International Mutli-Conference on Automation, Computing, Communication, Control and Com- pressed Sensing (iMac4s), Kottayam, 2013, pp. 365-371, doi: 10.1109/iMac4s.2013.6526438. [3]. T. Porter and T. Duff. Compositing digital images. In ACM Siggraph Computer Graphics, volume 18, pages 253–259 ACM, 1984.
- [4] Ramakrishnan, V., Pete, D.J. Non Subsampled Shearlet Transform Based Fusion of Multiple Exposure Images. SN COMPUTER SCIENCE 1, 326 (2020). <https://doi.org/10.1007/s42979-020-00343-4>.
- [5] Ron Brinkmann. 1999. The art and science of digital compositing. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- [6] Thomas Porter and Tom Duff. 1984. Compositing digital images. SIGGRAPH Comput Graph. 18, 3 (July 1984), 253–259. DOI:<https://doi.org/10.1145/964965.808606>
- [7] J. F. Blinn, "Compositing. 1. Theory," in IEEE Computer Graphics and Applications, vol. 14, no. 5, pp. 83-87, Sept. 1994, doi: 10.1109/38.310740.
- [8] Debevec, and J. Malik. Recovering High Dynamic Range Radiance Maps from Photographs P SIGGRAPH 97 (August 1997).
- [9] MANN S., PICARD R. W.: On being undigital with digital cameras: extending dynamic range by combining exposed pictures. In In Proc. of IS & T 48th annual conference (1995), pp. 422–428.
- [10] GOSHTASBY A.: Fusion of multi-exposure images. Image and Vision Computing 23 (2005), 611–618.
- [11] Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski. A bayesian approach to digital matting. In CVPR, volume 2, pages 264–271, Kauai Marriott, Hawaii, 2001.
- [12] RAMAN S., CHAUDHURI S.: A matte-less, variational approach to automatic scene com- positing. In ICCV (2007).
- [13] MERTENS T., KAUTZ J., REETH F. V.: Exposure fusion. In Pacific Graphics (2007). [14].Y. Guan, W. Chen, X. Liang, Z. Ding, and Q. Peng. Easy matting - a stroke based approach for continuous image matting. Eurographics, 25(3):567–576, 2006.
- [15]. M. Ruzon and C. Tomasi. Alpha estimation in natural images. In CVPR, volume 1, pages 18–25, Hilton Head Island, South Carolina, USA, 2000.
- [16]. J. Sun, J. Jia, C. Tang, and H. Shum. Poisson matting. In SIGGRAPH, pages 315–321, Los Angeles, USA, 2004.
- [17]. J. Wang and M. F. Cohen. An iterative optimization approach for unified image segmentation and matting. In ICCV, pages 936–943, Beijing, China, 2005.
- [18] N. Apostoloff and A. Fitzgibbon. Bayesian video matting using learnt image priors. In CVPR, volume 1, pages 407–414, Washington, DC, USA, 2004.
- [19] Y. Chuang, A. Agarwala, B. Curless, D. H. Salesin, and R. Szeliski. Video matting of com- plex scenes. In SIGGRAPH, pages 243–248, San Antonio, USA, 2002.
- [20] M. McGuire, W. Matusik, H. Pfister, J. F. Hughes, and F. Durand. Defocus video matting. In SIGGRAPH, pages 567–576, Los Angeles, USA, 2005. A. Smith. Alpha and the history of digital compositing. Microsoft Tech Memo 7, 1995.
- [21] 'Savitzky Golay filter for 2D images' - <http://research.microsoft.com/enus/um/people/jckrumm/SavGol/SavGol.htm>, September 2012.
- [22].Mantiuk R., Kim K., Rempel A.,Heidrich W. (2011). HDR-VDP-2: A calibrated visual met- ric for visibility and quality predictions in all luminance conditions ACM Trans. Graph.. 30. 40. 10.1145/1964921.1964935.

- [23] Bogoni, L.; Hansen, M. Pattern-selective color image fusion. *Pattern Recognit.* 2001, 34, 1515–1526.
- [24] Vonikakis, V.; Bouzos, O.; Andreadis, I. Multi Exposure Image Fusion Based on Illumination Estimation. In *Proceedings of the SIPA, Chania, Greece, 22–24 June 2011*; pp. 135–142.
- [25] Tico, M.; Gelfand, N.; Pulli, K. Motion-Blur-Free Exposure Fusion. In *Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010*; pp. 3321–3324.
- [26] Jinno, T.; Okuda, M. Multiple Exposure Fusion for High Dynamic Range Image Acquisition. *IEEE Trans. Image Process.* 2012, 21, 358–365.
- [27] Shen, R.; Cheng, I.; Shi, J.; Basu, A. Generalized Random Walks for Fusion of Multi- Exposure Images. *IEEE Trans. Image Process.* 2011, 20, 3634–3646.
- [28] Li, S.; Kang, X. Fast multi-exposure image fusion with median filter and recursive filter. *IEEE Trans. Consum. Electron.* 2012, 58, 626–632.
- [29] Mitianoudis, N.; Stathaki, T. Pixel-based and Region-based Image Fusion schemes using ICA bases. *Inf. Fusion* 2007, 8, 131–142.
- [30] Michael Ashikhmin. 2002. A tone mapping algorithm for high contrast images. In *Proceedings of the 13th Eurographics workshop on Rendering (EGRW '02)*. Eurographics Association, Goslar, DEU, 145–156.
- [31] Francesco Banterle, Alessandro Artusi, Elena Sikudova, Thomas Bashford-Rogers, Patrick Ledda, Marina Bloj, and Alan Chalmers. 2012. Dynamic range compression by differential zone mapping based on psychophysical experiments. In *Proceedings of the ACM Symposium on Applied Perception (SAP '12)*. Association for Computing Machinery, New York, NY, USA, 39–46. DOI:<https://doi.org/10.1145/2338676.2338685>
- [32] Bruce, N.D. (2014). ExpoBlend: Information preserving exposure blending based on normalized log-domain entropy. *Comput. Graph.*, 39, 12–23.
- [33] Lischinski, D., Farbman, Z., Uyttendaele, M., Szeliski, R. (2006). Interactive local adjustment of tonal values. *ACM Transactions on Graphics (TOG)*, 25(3), 646–653.
- [35] Tan, J., Huang, Y., Wang, K. (2018, July). Logarithmic Tone Mapping Algorithm Based on Block Mapping Fusion. In *2018 International Conference on Audio, Language and Image Processing (ICALIP)* (pp. 168–173). IEEE.
- [36] Reinhard, E., Stark, M., Shirley, P., Ferwerda, J. (2002, July). Photographic tone reproduction for digital images. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques* (pp. 267–276).
- [37] Ramakrishnan, Vivek., Pete, D.. (2021). Savitzky–Golay Filtering-Based Fusion of Multiple Exposure Images for High Dynamic Range Imaging. *SN Computer Science*. 2. 10.1007/s42979-021-00594-9.
- [38] Vivek Ramakrishnan "Exposure Fusion in the Non-Sub-sampled Contourlet Domain " Vol. 9 - No. 2 (Feb 2019), *International Journal of Engineering Research and Applications (IJERA)* ISSN: 2248-9622 , www.ijera.com Last accessed 08 July 2021.
- [39] Erik Reinhard, Greg Ward, Sumanta Pattanaik, and Paul Debevec. 2005. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting* (The Morgan Kaufmann Series in Computer Graphics). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- [40] Lam E. Y. , "Analysis of the DCT coefficient distributions for document coding," in *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 97–100, Feb. 2004, doi: 10.1109/LSP.2003.821789.
- [41] Pratt W. K., *Digital Image Processing*, New York: Wiley-Interscience, 1978, chapter 10.
- [42] Tescher A.G., "Transform image coding," in *Advances in Electronics and Electron Physics. Suppl. 12*. New York: Academic, 1979, pp. 113–115.
- [43] Murakami H, Hatori Y., and Yamamoto H., "Comparison between DPCM and Hadamard transform coding in the composite coding of the NTSC color TV signal," *IEEE Transactions on Communications*, vol COM-30, pp. 469–479, Mar. 1982.
- [44] Reininger R., and Gibson J. : "Distribution of the two-dimensional DCT coefficients for images," *IEEE Transactions on Communications* 31 (6) 1983.
- [45] Weisstein E. W., "Kolmogorov-Smirnov Test." From *Math World—A Wolfram Web Resource*. <http://mathworld.wolfram.com/Kolmogorov-SmirnovTest.html>, last viewed on March 22, 2010.
- [46] Ahmed N., Natarajan T., and Rao K. R., "Discrete Cosine Transform," *IEEE Transactions on Computers*, 90–93, Jan 1974.
- [47] Akansu, Ali N. and Haddad, Richard A., *Multiresolution Signal Decomposition, Transforms, Subbands, Wavelets*, Second Edition, San Diego, CA, Academic Press, 2001.

AUTHORS

Aut Mr. Vivek Ramakrishnan, is a Research scholar in the Electronics engineering department, Datta Meghe College of Engineering, Airoli, Navi-Mumbai, India. He is a researcher in the area of Signal, Image and Video Processing, Computational Photography and Medical Imaging. His current research areas are transform and filtering based approaches for HDRI. He has many research publications to his credit in the said domain.



Dr. D. J. Pete, is Professor and Head, in the Electronics engineering department, Datta Meghe College of Engineering, Airoli, Navi-Mumbai, India. He has teaching experience of over 24 years. His areas of expertise include Communication Engineering, VLSI Reliability and Nano Electronics.



© 2021 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.

THE COMBINATION OF NARRATIVE NEWS AND VR GAMES: COMPARISON OF VARIOUS FORMS OF NEWS GAMES

Xiaohan Feng¹ and Makoto Murakami²

¹Graduate School of Information Sciences and Arts,
Toyo University, Kawagoe, Saitama, Japan

²Dept. of Information Sciences and Arts,
Toyo University, Kawagoe, Saitama, Japan

ABSTRACT

The information explosion makes it easier to ignore information that requires social attention, and news games can make that information stand out. There is also considerable research that shows that people are more likely to remember narrative content. Virtual environments can also increase the amount of information a person can recall. If these elements are blended together, it may help people remember important information. This research aims to provide directional results for researchers interested in combining VR and narrative, enumerating the advantages and limitations of using text or non-text plot prompts in news games. It also provides hints for the use of virtual environments as learning platforms in news games. The research method is to first derive a theoretical derivation, then create a sample of news games, and then compare the experimental data of the sample to prove the theory. The research compares the survey data of a VR game that presents a story in non-text format (Group VR), a game that presents the story in non-text format (Group NVR), a VR game that presents the story in text (Group VRIT), and a game that presents the story in text (Group NVRIT) will be compared and analyzed. This paper describes the experiment. The results of the experiment show that among the four groups, the means that can make subjects remember the most information is a VR news game with a storyline. And there is a positive correlation between subjects' experience and confidence in recognizing memories, and empathy is positively correlated with the correctness of memories. In addition, the effects of "VR," "experience," and "presenting a story from text or video" on the percentage of correct answers differed depending on the type of question.

KEYWORDS

Virtual reality, narratology, news games, interactive, multimedia

1. INTRODUCTION

Telling stories is an indispensable method of human communication. Stories can express emotions, teach methods, and provide experiences. Therefore, the research of narrative theory is essential for any media production.

Virtual reality, in addition to providing viewers with immersive and novel experiences, is even more interactive than other media. News games, on the other hand, span two areas: news reporting and video games, which are fictional experiences base on real-world sources. Although there are already examples of VR news, such as "Project Syria" and "Hunger in LA" [1], [2]. But VR news games still have a lot of room for development in terms of narrative.

As media renewal continues, the amount of information people is faced with is increasing, and along with convenience comes the following problems:

- Social problems that can only be solved by attracting the attention of public opinion are likely to be buried in the vast amount of information, and will not leave a lasting impression.
- The current media environment makes it easy to forget details of information.
- When the same type of information is repeatedly presented, the public's sensitivity to that information decreases and their emotional response gradually weakens.
- It is clearly felt that the online media is more suitable for the masses than the traditional media. This is likely to cause a feeling of dispersed responsibility. This is the so-called bystander effect [3].

According to neurologist Michael Smith, when people watch narrative images that induce empathy, the brain automatically filters out external influencing factors to focus on learning and cognition, indicating that the human mind wants to know the unknown from a narrative perspective. He believes that without a narrative connection, people cannot stay in the brain for long [4]. There is also considerable research showing that memories of events with a strong narrative component are more likely to be remembered [5]. And many studies have revealed that an immersive virtual reality system can better facilitate situational memory performance [6], [7].

Consequently, research reinforce memory through narrative and immersion, and bring players into an independent worldview in the form of VR games to increase the target ability of information transfer and reduce the psychology of distributed responsibility. To address these issues, this research proposes that using virtual reality technology to gamify narrative news is a way to make news more deeply memorable. At the same time, new questions arise, what kind of narrative techniques would be suitable for news games, and could VR affect the outcome?

Therefore, the main purpose of this research is to explore the above questions. Samples will be created from the perspective of narrative theory, and research will examine which sample's story information is most deeply embedded in the subject's mind.

This research aims to provide directional results for researchers interested in combining VR and narrative, enumerating the advantages and limitations of using text or non-text plot prompts in news games. It also provides hints for the use of virtual environments as learning platforms in news games.

2. METHODS

The research method is to first derive a theoretical derivation, then create a sample of news games, and then compare the experimental data of the sample to prove the theory.

Research compared the survey data of a VR game that presents a story in non-text format (Group VR), a game that presents the story in non-text format (Group NVR), a VR game that presents the story in text (Group VRIT), and a game that presents the story in text (Group NVRIT) will be compared and analyzed.

This paper describes the experiment. Specifically, university students will be surveyed, and 30 students in each group will experience the sample. After the experience, they will be asked to fill out a questionnaire and data will be collected. Finally, the collected data will be analyzed and conclusions will be drawn by comparing them with the expected results. All subjects were recruited through convenience sampling and snowball sampling. The original plan was for some

subjects to experience the VR face-to-face with the researchers, but this was all changed to online because of COVID-19 situation. Subjects were asked to download the samples through a link provided by the researcher, experience the samples on their respective devices, and mail answers of a questionnaire. All subjects were aware of and consented to the experiment before it was conducted.

The sample is based on actual events. More than 80% of the text is taken directly from the news interviews. References are to newspaper articles about the incident from 2006 to 2015. These news stories were distributed across diverse media platforms over a long period of time. They were selected as scenarios for the news game because of their integrative nature and the narrative nature required for this experiment.

The synopsis is that the main character was taken to a remote countryside by human traffickers and forced to marry her "husband". The protagonist tried to commit suicide many times without success. She was recruited by the local school as a temporary teacher, and the children and students gave her new hope. A photographer discovered and photographed this, and the incident became known to the public. However, the local government and residents became frustrated with these reports and would not allow the reporter to cover the story, claiming that it exposed human trafficking and the poverty of education, and the school expelled the protagonist. When these facts were uncovered, social discontent rose again, and the local government, under pressure from public opinion, reinstated the protagonist as a teacher.

The sample is divided into four stages, each with one or two enemies and two key items, and the player needs to collect the key items while avoiding enemy pursuit. The enemies are designed to symbolize hostility and the protagonist's fears, such as "husbands," "fear of forced childbirth," and "local government. The key items in each stage are related to the storyline of that stage. In the case of the Group VRIT and Group NVRIT samples, picking up a key item automatically displays text describing the story in the first person of the protagonist, in the case of the Group VR and Group NVR samples, picking up a key item plays a video that is projected full screen, the video is produced in 3ds Max, and the text used in the narration and the text group is the same. Once the items are aligned, the player can move on to the next stage.

The samples in non-VR group are displayed on a traditional screen of PC with a maximum resolution of 1920x1080; VR group use VR headsets and controllers for PC. The subjects in the VR group have full control over navigation and manipulation of objects (key items). Videos for the non-text group were produced using 3ds Max, and the characters are real scale 3D character modeling silhouettes. The same video was used to ensure consistency in the non-text group, given the possible influence of the lens language. The background color of this video is black, and the surroundings are also black when viewed on a VR device, thus creating a fake 360-degree video effect. Since the characters do not come behind the player, the experience is more like watching a play; the difference between the VR group and the NVR group is that the play is viewed on a computer screen and the actors perform the play in front of the subject. In addition to this, the non-text group contains more modality sensory information, such as narration and background music, than the text group.

The questionnaire is divided into four written questionnaires and one recorded questionnaire the written questionnaire consists of basic information, recognition check, correctness check, and empathy check, as shown in Table 1. While the recorded questionnaire consists of subjects telling a story and giving their impressions of the sample. The basic information questionnaire asks the subjects their age, gender, major, gaming experience, experience using VR, and the theme of the sample.






The recognition check and the correctness check use the same 10 questions. The purpose of the recognition check is to find out how much the subjects themselves think they know about the sample story, and their actual cognitive status is not important in this check. In the empathy check, the strength of the empathic emotion for each subject is investigated through a self-report questionnaire.





Table 1. Questionnaire.

Basic information					
Age	Gender	Game experience		major	
Have you ever used VR?		No	Yes		
Have you paid attention to the news of female population sales?		No	Yes		
Have you read related articles on this game?		No	Yes		

Recognition check					
Question	Get it	Maybe Get it	Not sure	Maybe don't get it	Don't get it
Where the protagonist was kidnapped?					
Who bought the protagonist?					
What is the attitude of the protagonist 's husband towards her?					
What are the protagonists' means of suicide?					
Why the protagonists left the village?					
What reminds the protagonist of life's hopes?					
How is the protagonist known to the public?					
Why doesn't the local government want the public to know about the protagonist?					
What is the attitude of the villagers towards this?					
Did you understand the ending?					

Correctness check		
Full score 100	Correct answer +10	score:

Empathy check	
Stage	No feeling  felt a strong emotion
1	
2	
3	
4	

 : No feeling
 : Feeling emotions, but not enough to take action.
 : Feels emotional and takes short-term/single/simple actions.
 : Feels emotional and takes long-term/ continuous/

/complex actions.

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1. Basic Information

In the research of the experiment, 120 Chinese college students participated, of which, 59 were females and 61 males with a mean age of 20.8 years, age range of 18-24 years, and $SD=1.294$. The groups were divided, 19 females and 11 males in the NVR group, age range 19-23 years, mean age 20.733 years, $SD=1.263$; 18 females and 12 males in the NVRIT group, age range 18-24 years, mean age 20.733, $SD=1.263$; 13 females and 17 males in the VR group ranging from 19 to 23 years old, mean age 20.833, $SD=1.240$; and, 9 females and 21 males in the VRIT group ranging from 19 to 24 years old, mean age 20.933, $SD=1.314$.

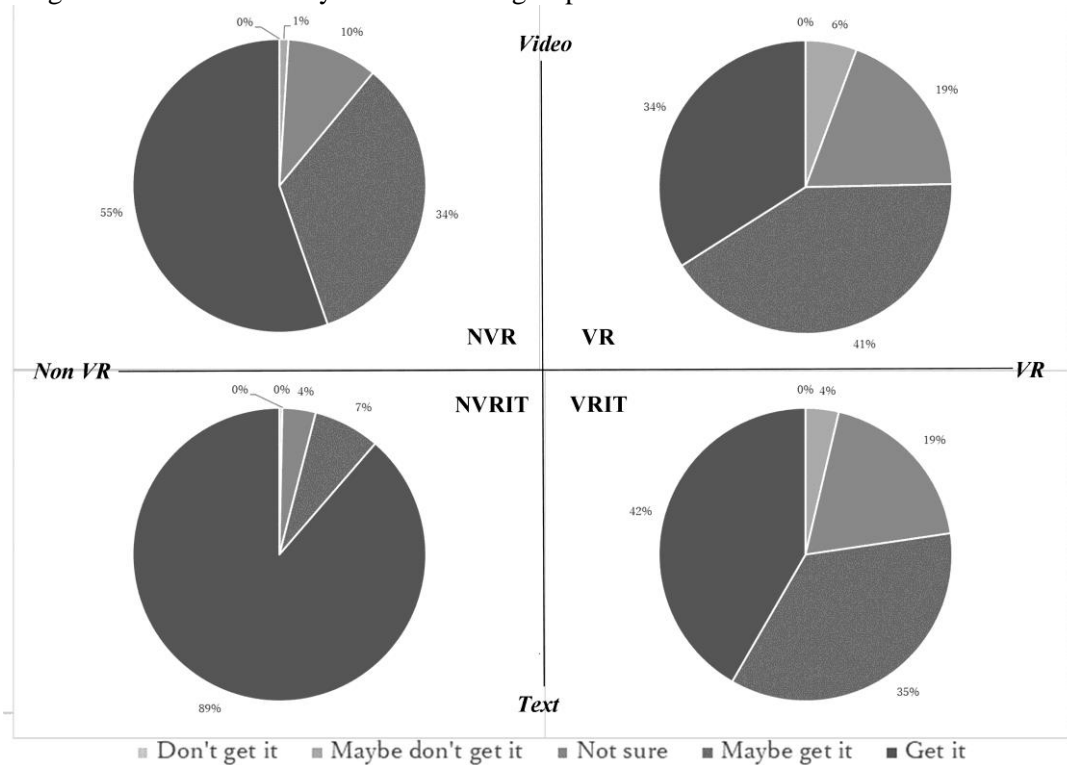
The 120 subjects also included 23 from the Department of Science and Engineering, 18 from the Department of Education, 16 from the Department of Management, 15 from the Department of Arts, 13 from the Department of Architecture, 12 from the Department of Literature, 9 from the Department of Economics, 7 from the Department of Sociology, 3 from the Department of Physical Education, 2 from the Department of History, 1 from the Department of Law, and 1 from the Department of Philosophy. A variety of majors applied. The diversity of subjects' majors is beneficial in getting more diverse perspectives in the recorded survey.

When the subjects self-reported their gaming experience, 87 said they had a lot of gaming experience, 21 said they had normal gaming experience, and 12 said they had little gaming experience. Quite a few of the subjects who considered their gaming experience to be normal and low often watched live game shows, even though they did not play games themselves. On the whole, the subjects have a lot of gaming experience. This also can be seen from the behavior of the subjects. When the sample was given, there was no mention of how to play or operate the game, and as a result, there are no questions asked by the subjects; the subjects sought ways to clear the game from their own game experience. This was one of the reasons why research chose university students as subjects; research deliberately chose targets with a lot of gaming experience and expected more obvious results in terms of recipients filling in the gaps by themselves. In terms of the results, this is the right choice.

3.2. Recognition Check

Fig. 1 shows the percentage and number of times each option was selected in each group. Overall, the NVRIT group was the most confident in their perception, followed by the NVR group, then the VRIT group, and finally the VR group, with a slight difference from the VRIT group. The NVRIT group is more confident in their self-perception than expected. Compared to the NVR group, which is also a non-VR format sample, the number of times the "Get it" option was selected exceeded 100 times. When compared to the VRIT group, which presents the story in the same text, the NVRIT group selected "Get it" more than twice as many times as the same group. The selection patterns of the other three groups are more similar, with most of the selections distributed between "Get it" and "Maybe get it," a certain number of "Not sure," a small number of "Maybe don't get it," and zero selections of "Don't get it." NVRIT group's choices were concentrated on "Get it", with very few "Maybe get it" and "Not sure" choices compared to the other groups. In the NVRIT group, the number of choices for "Maybe don't get

it" is zero, and the number of times "Don't get it" was selected is zero. Among all responses, "Don't get it" was selected only in the NVRIT group.



	Don't get it	Maybe don't get it	Not sure	Maybe Get it	Get it
VRIT	0	11	57	107	125
VR	0	17	57	124	102
NVRIT	1	0	11	22	266
NVR	0	3	30	101	166

Figure 1. The percentage and number of times each option was selected in each group

Of the three groups with similar selection patterns, the NVR group is more confident in their self-perception. The difference in the number of times they selected "Get it" and "Maybe get it" was the largest among the other three groups. The difference in self-perception between the VRIT group and the VR group is not clear. The difference in self-perception between the VRIT group and the VR group is not clear: the two groups chose the same number of "uncertainties," but the VR group was the only group that chose "Maybe get it" more often than "Get it". The VR group also selected "Maybe don't get it" more often than the VRIT group. This indicates that the self-perceived confidence of the VR group is the lowest among all groups.

Through the experimental records of previous research and additional interviews [10], they found that the average number of years of VR purchase for the VR group was two years, with the highest number of people acquiring VR in 2019. The average time spent using VR was 10 hours per month; they had been using VR devices for longer periods of time, but were more dense when they did use them. In many cases, the same was true for the VRIT group; subjects who said they had not recently played any VR games that particularly interested them had consequently

not taken out their VR devices for nearly six months prior to the experiment. Combining the data from the two VR and non-VR samples reconfirmed that subjects' self-reported experience with memory was positively correlated. Studies have shown that the more experience you have, the more confident you are in identifying your memories [11]. This is the reason why the NVRIT group was so confident in their self-identification of memories. Even though video media has become more common, text is still the largest medium through which people access information on a daily basis. Regardless of the ability of the visual media to make viewers remember more information, from the perspective of cumulative experience, people always experience more text than video from birth to adulthood.

The research also took into account that the total amount of text in the VRIT and NVRIT samples was small, divided into eight groups, with only 300-500 words in each group. It remains to be seen whether the same results can be obtained with longer texts. Since this experiment was conducted with college students with gaming experience as subjects, the gaming experience and the experience of reading short texts are one of the reasons for the NVRIT group's confidence in self-awareness.

3.3. Correctness Check

The same question is used for the recognition check and the correctness check. There are five levels of correctness depending on the subject's answer. A high percentage of subjects went beyond the correct answers in certain questions, adding more correct details to their answers. For example, the first question asks where the main character is abducted and sold. The answer is "the train station". Nearly half of the subjects answered not only the train station, but also specific city information. For these answers, the researchers classified them as "more than correct". The answers that matched were considered "correct". Those who gave even slightly wrong answers were placed in the "mostly correct" category and vice versa. Finally, answers that are completely wrong are categorized as "incorrect". Answers that are out of range, but not completely correct, are also classified as "mostly correct".

The score for each level is: "More than correct": 5, "Correct": 4, "Mostly correct": 3, "Somewhat correct": 2, and "Incorrect": 1. The total score is 1500. The group with the highest total score was the VR group with 1248 points. Then came the VRIT group with 1211 points, followed by the NVR group with 1153 points. The last group was the NVRIT group with 1094 points. Fig. 2 shows the total score and the number of times the highest score was obtained for each group. The scores for each question are shown in Table 2, with the highest score for each group shown in bold.

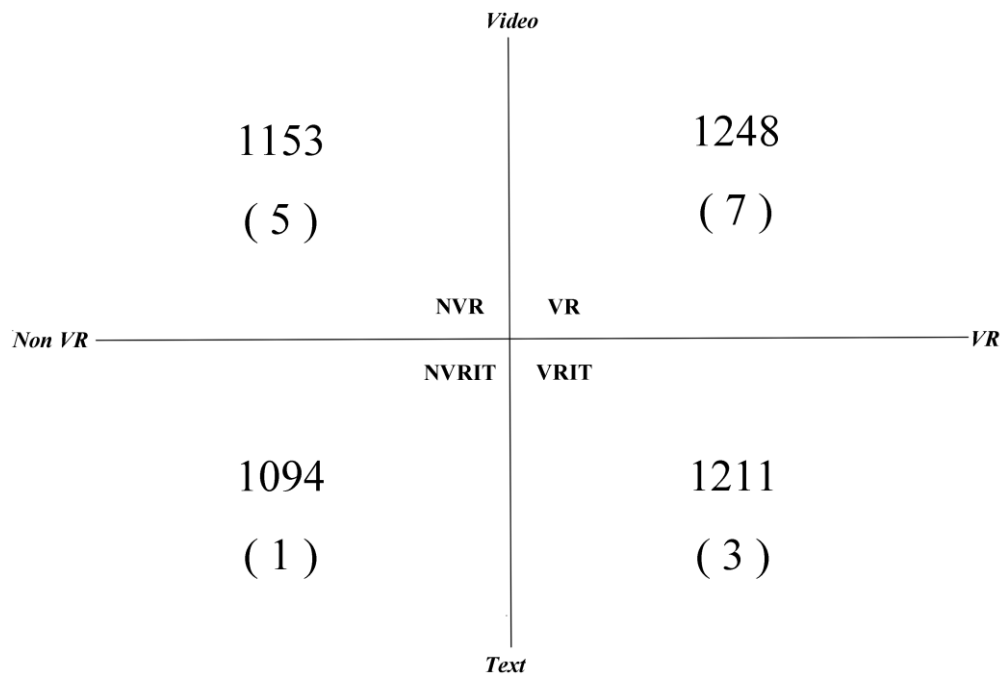


Figure 2. Spam traffic sample

Table 2. Score for each question

	<i>NVRIT</i>	<i>NVR</i>	<i>VRIT</i>	<i>VR</i>
1	124	127	124	138
2	113	120	120	120
3	112	129	136	141
4	120	120	120	120
5	93	102	98	99
6	100	112	119	137
7	105	110	110	94
8	107	135	117	114
9	126	102	131	146
10	94	114	136	139
Total Score	1094	1153	1211	1248

The results also show the following:

For the same data from a non-VR sample, a sample that presents the story through video can make subjects remember the information more accurately than a sample that presents the story through text.

For the same VR sample data, the visual storytelling sample helps subjects remember the information more accurately than the textual storytelling sample.

In the same sample data where the story was presented in text, the VR sample allows subjects to remember the information more accurately.

In the same sample data where the story was presented in video, the VR sample was able to make the participants remember the information more accurately.

The self-reports of the four groups found from the recognition survey are, in order from highest to lowest, "NVRIT> NVR> VRIT> VR". The total result was the exact opposite of the high-low ranking of the correct survey results.

These data validated the prediction that the means of getting subjects to remember the most information among the four groups was the VR news game with a story. When the sample presents the story in text, the VR sample gets subjects to remember the information, and the correct response rate is higher when the story is presented in video than when the sample presents it in text. For those who do not have a VR device, the higher correct response rate of the NVR group indicates that storytelling and gaming also prove that they can reinforce memory.

The NVR group had more questions that received the highest score than the VRIT group. This may mean that, depending on the type of question, "presenting the story in a video" was more effective than in getting the subjects to remember the information correctly.

So, to further analyze the data, the research divided the questions into three types.

- I. Questions with a single answer. This is the simplest type of question, requiring only one answer from the subject. Questions 1, 2 and 4 belong to this type.
- II. Questions that require multiple answers. This question is similar to the Type 1 question, but the experimenter needs to give more than one answer. Questions 3, 5 and 9 belong to this type.
- III. Subjective questions. No clear answer is given in the sample, and the experimenter is required to organize the information by himself/herself and combine it with thinking to come up with an answer. Questions 6, 7, 8 and 10 belong to this type.

The scores are shown in Fig. 3. The first and second types of questions contain three questions, but the third group contains four questions, so the score for the third type of question can only be compared with its internal group.

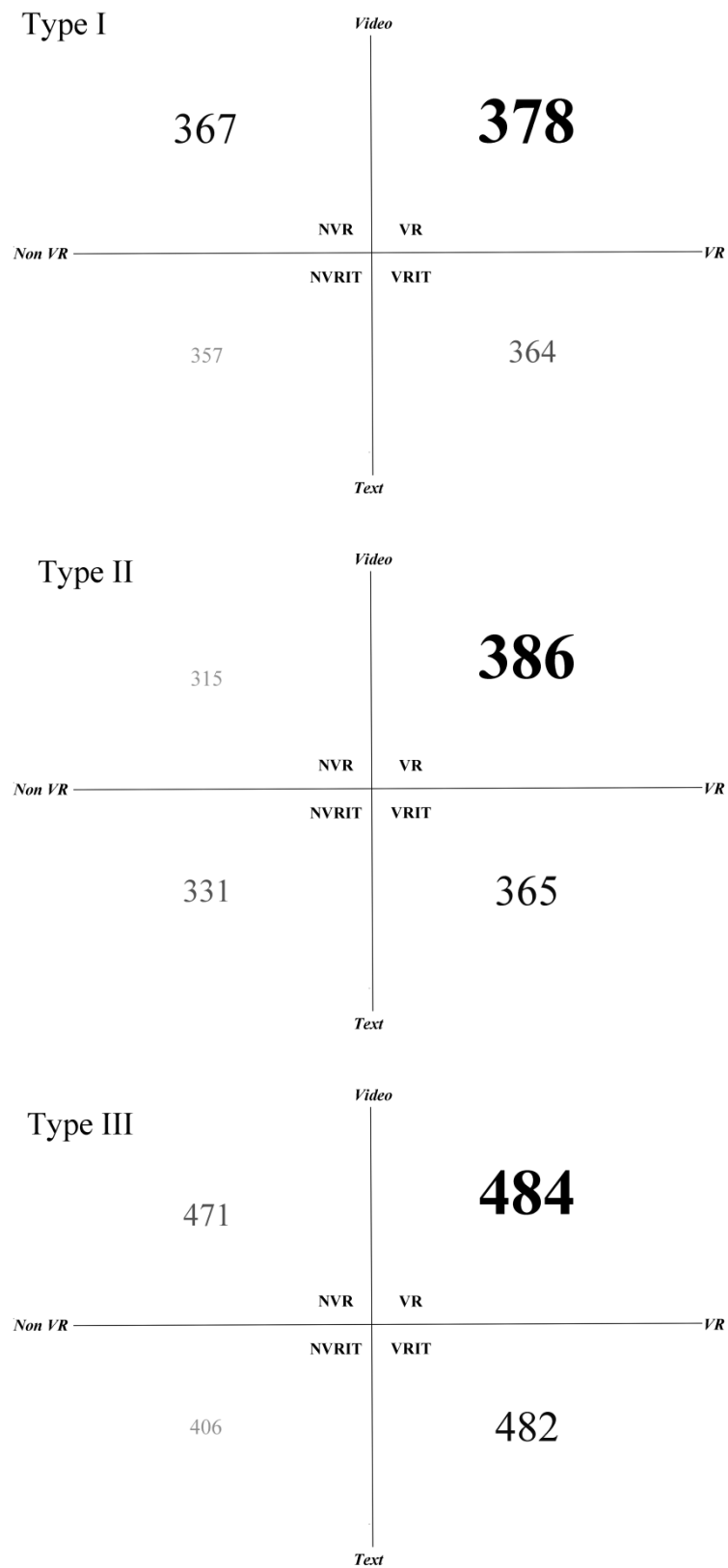


Figure 3. Scores for each type of question

For questions with a single answer, the scoring order and total score results for each group differed, with the VR group scoring the highest, while the NVR group came in second, ahead of the VRIT group. The NVRIT and NVR groups scored higher than for Type 2 questions, while the VRIT group's score was essentially the same. The VRIT group's score is basically the same. This result confirms what has been speculated so far: when there is only one answer, presenting the story in a video is slightly more effective in getting the subject to remember the information correctly. Based on the above results, if someone want to convey a single piece of information accurately but cannot meet the requirements of VR, "present the story in text" is the best choice.

Questions that require multiple answers also differ from the total score ranking. The score ranking for this type is VR>VRIT>NVRIT>NVR for non-VR, presenting the story in text is clearly better at getting the subject to remember the information accurately than presenting the story in video. Based on the above results, "presenting the story with text" is more effective than "presenting the story with video" in a non-VR environment for accurately conveying multiple pieces of information.

Question 9, why does the local government not want the public to know about the protagonist? The VR group had a very high percentage of correct answers, with 23 people answering "more than correct" and the remaining 7 also answering correctly. The VR group had more "more than correct" answers, so the criterion for "more than correct" for this question was to give three reasons for the question answer. The VR group had the highest number of "correct" responses. The VR group averaged 83 characters for this question, while the NVR group averaged only 15 characters. This suggests that the immersive nature of VR made the subjects more emotionally involved, promoted emotional empathy, and strengthened their memory [12].

There is also a high rate of questions and errors that are not fully connected to the story. For example, question 5, three suicide methods to ask the main character. Subjects have difficulty remembering all of them, and in most cases only remember two of the means. This is because these means did not cross over much with the subsequent story, especially when compared to question 4, the number of times the main character committed suicide. The subjects remembered the "suicide" episodes in the story well. The memory of "means" that are not related to the plot is much weaker.

The rankings of the subjective question scores and total scores are consistent: the difference between the VRIT and VR groups is small, and the difference between the NVRIT and NVR groups is the largest among the three types. Thus, in the subjective question, if the sample is in VR format, there is little difference in the amount of information correctly remembered between "presenting the story in text" and "presenting the story in video". However, when the sample is in a non-VR format, there is a large difference in the amount of information correctly remembered between "presenting the story with text" and "presenting the story with video".

In order to get a high score on a subjective question, more test takers need to be able to relate to it. From the data, the NVRIT group clearly has a larger gap in scores than the other three groups, and the difference value is the largest among the three types of questions.

3.4. Empathy Check

As shown in Fig. 4, overall, the four groups tend to have basically high starts and low goals. The four levels of empathy are scored on a scale of 1 to 4, from low to high.

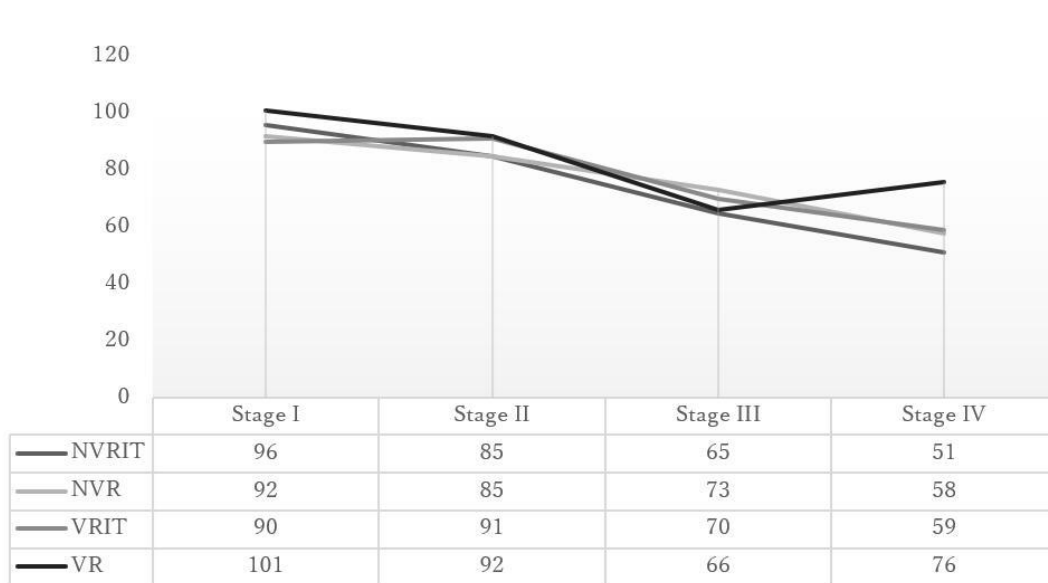


Figure 4. Empathy for each stage

The empathy for Stage 1 was the greatest not only because the first stage was the most novel, one study found that information is more memorable when it appears early in the story [13]. In order from highest to lowest, $VR > NVRIT > NVR > VRIT$.

In Stage 2, the VRIT group has slightly higher empathy, basically the same as the data in Stage 1. In the other groups, empathy is less. However, they are still at the top of the list in the overall data. In order from highest to lowest values, $VR > VRIT > NVR = NVRIT$.

In Stage 3, empathy for all groups drops significantly. In particular, the VR group not only has the largest drop, but also has lower data than the NVR and VRIT groups. From highest to lowest values, $NVR > VRIT > VR > NVRIT$. Stage 3 is the least difficult of all the stages, and also the least fearful. This stage, Jing Wa, has a simple color scheme, with simple scenes and color schemes, mainly white and gray, no background music, relatively slow enemy movement, and an important turning point where the protagonist's attitude toward life becomes more positive. The research speculated that this relationship may have caused a significant drop in empathy in the VR group. The researchers asked the NVR group again. The six subjects who chose Level 4 said that they were alarmed because they thought there was likely to be JUMP SCARE in a quiet environment. These six subjects were not fully immersed in the sample, but were more dependent on the game experience to feel tense. This may reflect the lower level of immersion of the NVR group.

In Stage 4, the empathy of the three groups except the VR group continued to decline, but more slowly than the decline in Stage 3. The empathy of the tension in the VR group increased. The order of values from highest to lowest is $VR > VRIT > NVR > NVRIT$.

3.5. Recording

The recording is divided into two parts: the reproduction part and the impression part. Overall, there was little difference in the results between the four groups, basically focusing on fluency levels, with everyone being able to tell a rough story. A small number of subjects skipped certain episodes in the story, but quickly added more. From the results of this research, it appears that all four groups of subjects did well.

The key words that were mentioned more often in the feedback section are as follows:

" Kidnapping, hope, fear, creepy, background music, sound, children, Stockholm syndrome, angry, uncomfortable, heavy, shock, anxious, helpless, scene, suspicious."

Negative emotional keywords such as "scary" appeared more frequently. In order to increase the accuracy of the research, the research consciously set the game in a horror mood because several studies have shown that horror is helpful for memory [14].

4. CONCLUSIONS AND FUTURE DEVELOPMENT

- In descending order of self-report of the four groups in the recognition check: NVRIT>NVR>VRIT>VR.
- Exactly the opposite in the correctness check: VR>VRIT>NVR>NVRIT.
- The ranking for the empathy check is the same as for the correctness survey: VR>VRIT>NVR>NVRIT.
- The recording is divided into the story retelling part, and all groups can finish telling the story smoothly. The keyword "scary" appears frequently in the impression section. The characteristics of the experimental results for each group can be summarized as follows:
- NVRIT: Among the four groups, the difference in performance between confidence in memory and the information actually remembered is the largest, and the number of "Get it" choices is the largest, which is different from the choice patterns of the other three groups. For questions that require multiple answers, the percentage of correct answers is higher than the NVR group. Empathy can be more successfully elicited by the subjects' experience of playing the game and reading the text. However, as the game progresses, they fall to the bottom of the four groups.
- NVR: The pattern of confidence in memory is similar to the other three groups, except for NVRIT, which ranks first among them. They rank second in the number of times they receive the highest score on the correctness check, and their performance on the single-answer questions. The downward trend in empathy is relatively modest.
- VRIT: Test takers' confidence in their memory is slightly better than in the VR group. The difference in scores between the VRIT group and the VR group on the subjective questions of the correctness check is small for this group. The drop in empathy is the most moderate.
- VR: Of the four groups, the confidence in one's memory is the lowest. In the correctness check, they always rank first, regardless of the type of question. In the empathy check, they are consistently ranked first except for Stage 3. According to the post-experiment interview, the drop to third place is thought to be due to the drop in fear in Stage 3.

Of the four groups, the one that was able to remember the most information was the narrative VR news game. Subjects in this group also tended to be less confident in their own memories. Experiment shows a positive correlation between subjects' experience and their self-reported level of confidence in their memory. Similarly, experience in playing games and reading short texts may be the reason for the lack of confidence in one's memory. Studies have shown that audio enhances memory coding in virtual environments. This is another reason why the VR group always has a higher percentage of correct answers than the other three groups [15]. Non-text samples may yield more multimodal sensory information. However, some studies have shown that reading is also an interactive process [16], and in the case of the group of subjects in the current experiment, they have extensive reading experience. Therefore, the NVRIT group scored higher than the NVR group on the Type 2 question.

Further experiments will be conducted to see if this lack of confidence affects the profile of memory after an extended period of time.

When the sample is presented with a story in text, the VR sample is more likely to remember the information, and the sample presenting the story in video is more likely to answer correctly than the sample presenting the story in text. A number of studies have confirmed that empathy plays a role in the construction of our memories [17], [18]. This, coupled with the fact that it is an experimental result, further proves that the degree of empathy is positively correlated with the correctness of the memory.

Some argue that virtual reality is not ideal for increasing empathy [19] and that the effect of fear on memory cannot be ignored [20]. Taken together with the results of the current experiment, this suggests that immersion in virtual reality does not directly increase empathy, but that immersion can increase the fear of the sample and make them more empathetic in a narrative context. And empathy reinforces the memory that the subject was in the sample.

Studies have shown that people with higher empathy tend to take more helping actions [21]. In other words, the "narrative VR news game" not only helps the recipients to remember the news information better, but also increases their empathy, which makes them more interested in the news.

ACKNOWLEDGEMENTS

The authors would like to thank everyone, just everyone!

REFERENCES

- [1] Ryan Bradley, How Nonny de la Peña, the 'Godmother of VR,' Is Changing the Mediascape, 2018.
- [2] Bryan Bishop, Digital empathy: how 'Hunger in Los Angeles' broke my heart in a virtual world, 2013.
- [3] Darley, J.M.; Latané, B. Bystander intervention in emergencies: diffusion of responsibility. *Journal of Personality and Social Psychology*. 8 (4): 377–383. 1968.
- [4] Michael Smith. "From Theory To Common Practice: Consumer Neuroscience Goes Mainstream". (2016) <https://www.nielsen.com/us/en/insights/article/2016/from-theory-to-common-practice-consumer-neuroscience/>
- [5] Tom Trabasso, Paul van den Broek, Causal thinking and the representation of narrative events, *Journal of Memory and Language*, Volume 24, Issue 5, 1985, Pages 612-630, ISSN 0749-596X.
- [6] Harman, J., Brown, R., & Johnson, D. (2017). Improved memory elicitation in virtual reality: New experimental results and insights. In R. Bernhaupt, G. D. Anirudha, J. Devanuj, K. Balkrishan, J. O'Neill, & M. Winckler (Eds.), *IFIP Conference on Human-Computer Interaction* (pp. 128–146).
- [7] Ruddell, R. A., Volkova, E., Mohler, B., & Bühlhoff, H. H. (2011). The effect of landmark and body-based sensory information on route knowledge. *Memory & Cognition*, 39(4), 686–699.
- [8] J. Clement, Virtual reality (VR) and augmented reality (AR) device ownership and purchase intent among consumers in the United States as of 1st quarter 2017, by gender. (2021).
- [9] J. Clement, Distribution of video gamers in the United States from 2006 to 2020, by gender. Unpublished, (2021).
- [10] Xiaohan Feng, Narrative theory in virtual reality Comparison of VR news game and non-VR news game. (2021).
- [11] Cichoń, E., Gawęda, Ł., Moritz, S. et al. Experience-based knowledge increases confidence in discriminating our memories. *Curr Psychol* 40, 840–852. <https://doi.org/10.1007/s12144-018-0011-8>, 2021
- [12] Buckner, R. L., & Carroll, D. C. Self-projection and the brain. *Trends in Cognitive Sciences*, 11(2), 49–57. <https://doi.org/10.1016/j.tics.2006.11.004>. (2007).

- [13] Abigail C. Doolen & Gabriel A. Radvansky (2021) A novel study: long-lasting event memory, Memory,
- [14] Ginting, Henndy. It is Fear, Not Disgust, That Enhances Memory: Experimental Research on Students in Bandung. *anima*. 31. 77-83. (2016).
- [15] Andreano, J., Liang, K., Kong, L., Hubbard, D., Wiederhold, B. K., & Wiederhold, M. D. (2009). Auditory cues increase the hippocampal response to unimodal virtual reality. *Cyberpsychology & Behavior*, 12(3), 309–313.
- [16] Alan M. Lesgold & Charles A. Perfetti (1978) Interactive processes in reading comprehension, *Discourse Processes*, 1:4, 323-336, DOI: 10.1080/01638537809544443
- [17] Spreng, R. N., & Grady, C. L. Patterns of Brain Activity Supporting Autobiographical Memory, Prospection, and Theory of Mind, and Their Relationship to the Default Mode Network. *Journal of Cognitive Neuroscience*, 22(6), 1112–1123. <https://doi.org/10.1162/jocn.2009.21282>, (2009)
- [18] Spreng, R. N., Mar, A. R., & Kim, A. S. N. The Common Neural Basis of Autobiographical Memory, Prospection, Navigation, Theory of Mind, and the Default Mode: A Quantitative Meta-analysis. *Journal of Cognitive Neuroscience*, 21(3), 489–510. <https://doi.org/10.1162/jocn.2008.21029>, (2009)
- [19] Rueda, Jon & Lara, Francisco, Virtual Reality and Empathy Enhancement: Ethical Aspects. *Frontiers in Robotics and AI*. 7. 10.3389/frobt.2020.506984. (2020)
- [20] Ginting, Henndy. It is Fear, Not Disgust, That Enhances Memory: Experimental Research on Students in Bandung. *anima*. 31. 77-83. (2016).
- [21] Liao, Wan-Ting & Tzeng, Angela. The-Mechanism-Underlying-Empathy-Related-Helping-Behavior-An-Investigation-of-Empathy-Attitude--Action-Model. (2020).

FEATURE FUSION-BASED SIAMESE REGION PROPOSAL NETWORK FOR ULTRASOUND TRACKING

Xinglong Zhu, Ruirui Kang, Yifan Wang, Danni Ai,
Tianyu Fu and Jingfan Fan

Beijing Engineering Research Center of Mixed Reality and
Advanced Display, School of Optics and Photonics,
Beijing Institute of Technology, Beijing 100081, China

ABSTRACT

Object tracking based on ultrasound image navigation can effectively reduce damage to healthy tissues in radiotherapy. In this study, we propose a deep Siamese network based on feature fusion. Whilst adopting MobileNetV2 as the backbone, an unsupervised training strategy is introduced to enrich the volume of the samples. The region proposal network module is designed to predict the location of the target, and a non-maximum suppression-based post-processing algorithm is designed to refine the tracking results. Moreover, the proposed method is evaluated in the Challenge on Liver Ultrasound Tracking dataset and the self-collected dataset, which proves the need for the improvement and the effectiveness of the algorithm.

KEYWORDS

Ultrasound tracking, Siamese network, Respiratory motion estimation, One-shot learning

1. INTRODUCTION

Respiratory motion negatively affects radiotherapy for liver tumors. Doctors typically enlarge the radiation margin to ensure that the tumor receives adequate radiation. However, enlarging the radiation margin can harm surrounding tissues [[1]]. Generally, patients are instructed to hold their breath during radiation. As completion of the radiotherapy in one breath-holding period is impossible, doctors stop the treatment frequently and retarget the tumor with the radiation source at the start of a new round of radiation treatments [[2], [3], [4], [5]]. This approach is time-consuming and difficult. Implantation of invasive markers was also attempted, but invasive surgery causes additional damage to patients [[6], [7], [8]].

In recent years, ultrasound navigation was utilized to predict the location of tumors in real-time, in which the radioactive source is controlled to follow a tumor's movement [[9]]. However, the acoustic reflectivity of liver tumors is similar to that of surrounding tissues [[1]], making it difficult to locate the tumor directly based on ultrasound images. Other anatomical structures were used to predict the location of tumors. Among them, liver vessels have an acoustic reflectivity contrasting that of surrounding tissues; thus, liver vessels are typically chosen as targets for ultrasound tracking [[10], [11]].

Previously, matching or registration algorithms were typically employed to track liver vessels [[12], [13], [14]]. Researchers introduced Siamese networks to ultrasound tracking, as such

networks excel in visual object tracking. Liu et al. (2019) proposed the cascaded SiamFC algorithm and designed a two-stage cascaded Siamese network to improve the tracking accuracy of the network, thereby ranking first in the Challenge on Liver Ultrasound Tracking (CLUST) 2015 competition [[15]].

Recently, a network architecture similar to AlexNet was widely applied as the backbone of network[[15], [16], [17], [18]]. This fact inspires us to apply a highly sophisticated architecture to ultrasound tracking. However, two major obstacles exist in the application of a very deep network in ultrasound tracking. Firstly, the lack of annotated data makes training a general model difficult. Secondly, distractors confuse trackers [[15]]. As it shows in Figure 1, the distractors in the left image are more similar to the target in the right image than that in the left image, because the appearance of the target changes, thereby making tracking difficult.

To overcome the two aforementioned problems, an unsupervised training strategy is introduced to expand the volume of the samples. MobileNetV2 is adopted as the backbone of the SiamRPN-based tracker and the output feature of the backbone is fused for better discrimination. A post-processing algorithm based on non-maximum suppression (NMS) is proposed to eliminate distractors.

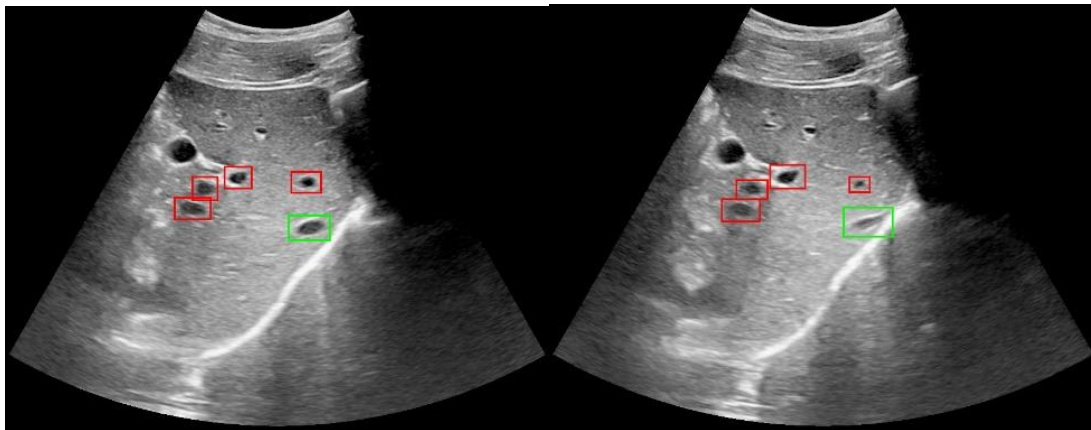


Figure 1. Two frames of an ultrasound sequence, in which the green bounding boxes mark the targets, and the red bounding boxes mark the distractors.

In this work, we propose an original tracking algorithm based on feature fusion to solve the aforementioned problems. The contributions of this work are summarized below.

- 1) A network model based on feature fusion is proposed. Local features and semantic features are integrated to improve the discriminative ability of the algorithm.
- 2) A training strategy combining supervised and unsupervised methods is proposed. Unsupervised training methods can increase the volume of samples, thereby enabling the algorithm to learn substantial general features.
- 3) A post-processing algorithm based on NMS is proposed. Spatial information and temporal features are utilized to eliminate distractors, which can improve the accuracy and robustness of the tracker.

In the next section, the development of visual object tracking and ultrasound tracking is reviewed. The proposed method will be introduced in detail in section 3. The experiments and their results are reported in section 4. In section 5, the advantages and limitations are concluded.

2. RELATED WORKS

Visual target tracking is a classic computer vision problem. Tracking algorithms such as sparse coding [[19]], Kalman filters [[20]], mean shift [[21], [22]] and so on model the target then locate the most similar area in the search image. Algorithms using this tracking method are called generative algorithms. Such algorithms generally do not require training, but their performance depends on the parameters set empirically by the researcher. With the introduction of KCF [[23]], discriminative algorithms attracted researchers' attention. Discriminative algorithms focus on the difference between target and background. Compared with generative algorithms, discriminative algorithms pay attention to negative samples, which leads to better performance. With the increasing numbers of proposed datasets and benchmarks [[24], [25]], deep features based on statistics gradually replaced handcrafted features [[26], [27]]. Deep features are extracted by convolutional neural network (CNN), and weights of network are optimised based on a huge amount of data. Thus, deep features are more robust than handcrafted features. The combination of deep features with discriminative algorithms spawned many remarkable algorithms, such as MDNet [[28]], ECO [[27]], SiamFC [[29]] and SiamRPN++ [[30]], which all achieved SOTA in competitions [[31], [32], [33]].

Ultrasound tracking algorithms combine the characteristics of ultrasound images and are affected by the development of object tracking algorithms for natural images.

A similar process can be seen in ultrasound tracking. Previously, tracking was generally considered as a registration or matching problem in ultrasound sequences [[1]]. Hallack et al. (2015) used LogDemons as a registration framework to solve the problem of target tracking [[12]]. Similarly, Shepard et al. (2017) employed image block matching to track a target [[13]]. Williamson et al. (2017) integrated dense optical flow, template matching, and image intensity information for hybrid tracking [[14]]. The aforementioned algorithms are training-free. However, as most matching and registration tasks are performed offline, such algorithms do not pay attention to real-time performance. Meanwhile, ultrasound tracking also draws on the development of visual object tracking. For example, in 2015, Kondo improved the KCF algorithm for ultrasound tracking [[34]]. Moreover, Shen et al. (2018) and Jeungyoon et al. (2019) adopted a CNN-like architecture to extract features and constructed correlation filters to process the features [[16], [17]]. Gomariz et al. (2018) added prior location information prediction to SiamFC [[18]]. Liu et al. (2019) proposed the cascaded SiamFC algorithm based on SiamFC, which won first place in the CLUST 2015 competition and has yet to be surpassed [[15]].

3. METHOD

The network structure proposed in this study is illustrated in Figure 2. The network uses MobileNetV2 [[35]] as the feature extractor. As the third-, fifth- and seventh-layer features outputted by the network have the same size, they can be stacked easily for feature fusion. Inspired by SiamRPN++ [[30]], a depth wise cross-correlation structure is adopted for the discrimination, and the two branches are designed for precise positioning. The difference between SiamRPN++ and the proposed network structure is that the stacked features are inputted directly into the two branches, which means that convolution layers are utilized to integrate the semantic and local features.

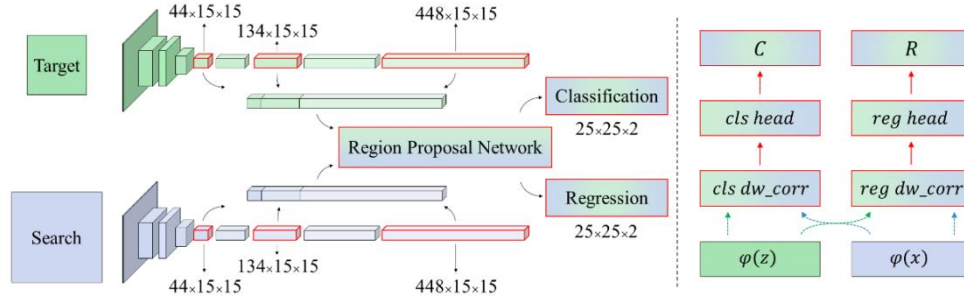


Figure 2. The proposed network architecture; the left side shows the network architecture; the right side shows the classification and regression branches in RPN module.

In addition to the network structure, a post-processing algorithm is proposed based on NMS, which plays a positive role in suppressing distractors.

3.1. Network Architecture

The proposed network structure uses the pretrained MobileNetV2 as the feature extractor. MobileNetV2 employs deep separable convolutions to construct an inverted residual block that maps the high-dimensional image space to the low-dimensional feature space. This design demonstrates a satisfactory balance between performance and computational cost. Meanwhile, the last few inverted residual blocks of MobileNetV2 output tensors with the same scale, which provide convenience for the feature fusion.

Feature fusion of different depths was proven to be effective for tracking. The stacked features are integrated into the RPN. Compared with SiamRPN++, the adjust layers are removed from the RPN module in the proposed structure. Based on experiments, removal of the adjustment layer can prevent overfitting. The depth-wise correlation layer first convolves the stacked features with a 3×3 kernel to 256 channels, integrating the feature output to each layer. After the correlation, fully convolutional layers are built as the head modules. The two-branch head modules predict the position and score for each subregion.

3.2. Mixed Training Strategy

As the network outputs the classification and regression results, the loss function of the network must consider the output of the two branches, and the loss value of the classification branch adopts a cross-entropy form.

$$L_{cls} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)],$$

where L_{cls} represents the loss value of the classification branch of the network, y_i is the result marked at coordinate i and \hat{y}_i is the predicted value of the classification branch.

The loss of the regression branch uses L_1 loss, as follows:

$$L_{reg} = -\frac{1}{N} \sum_{i=1}^N |r_i - \hat{r}_i|,$$

where L_{reg} represents the loss value of the classification branch of the network, y_i is the result marked at coordinate i , and \hat{y}_i is the predicted value of the classification branch.

Finally, the network loss can be written as follows:

$$L = L_{cls} + L_{reg}.$$

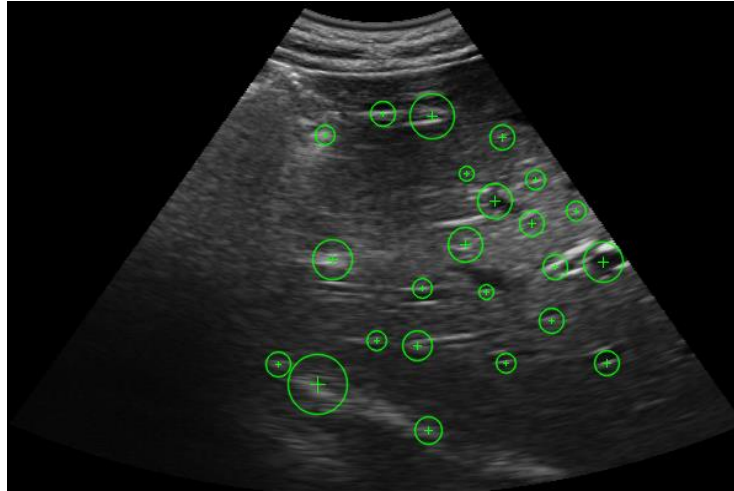


Figure 3. Data generated by the unsupervised strategy; the green crosses mark the key points extracted by SURF, and the green circles indicate the one-quarter size of the key points.

As a typical network using padding, the MobileNetV2 network does not have a shift-invariant characteristic. A large range of random shifts must be set during the training phase to prevent the network from collapsing into the center bias.

To increase the generalization ability of the network, unsupervised training is added to the previous training strategy. SURF algorithm is utilized to extract the key points from the ultrasound image then select the high response key points with large feature size and far from the ultrasound image boundary as the training sample. Figure 3 shows the key points extracted by an unsupervised strategy in an ultrasound image. During the training, the regions around the key points are cropped as target images and search images. The samples generated by the unsupervised algorithm and the samples manually labeled are mixed and added to the data loader. Without the unsupervised strategy, the tracker would propose objects likely to be vessels instead of objects likely to be the target. As the objects labeled in the dataset are nearly all vessels, the addition of the unsupervised strategy will prevent the algorithm from overfitting the labeled objects.

3.3. Tracking Inference Phase

To estimate the location of the target, the region of interest (ROI) is divided into 25×25 subregions, and the RPN outputs a score and a location for each subregion. The score represents the probability of the target appearing in the subregion, and the location indicates where the target is most likely to appear in the subregion.

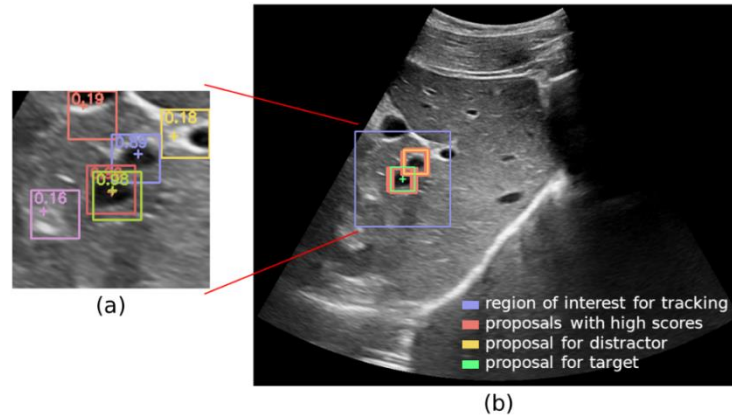


Figure 4. Tracking inference phase; image (a) shows several proposals generated by RPN; image (b) shows the post-processing algorithm dealing with the proposals.

Figure 4 shows the selection of proposals using NMS-based post-processing algorithms. Figure 4(a) shows several proposals generated by RPN. In Figure 4(a), the rectangles represent the subregions, the values annotated in the rectangles indicate the scores of the proposal, the crosses mark the proposed locations. The range and proposal of the same subregion are presented in the same color. Figure 4(b) shows the post-processing algorithm separates different proposals into targets, distractors, and redundant proposals that represent the same object. As there could be several proposals that represents the same objects, the key to improving the accuracy of the tracker is finding the proposal closest to the real target location among them. Inspired by the NMS algorithm, an appropriate algorithm for screening the proposals is designed. Firstly, the low-response proposals are excluded, which are not presented in Figure 4(b). Secondly, the proposals close to the proposal with a maximum response are filtered out, which are marked with red rectangles in Figure 4(b). Finally, the proposal closest to the tracking result of the previous frame is selected as the tracking result of the current frame, which is marked with a green rectangle in Figure 4(b). Those filtered out in the last step are marked in yellow rectangles in Figure 4(b). All rectangles in Figure 4(b) represent the location of the proposals.

NMS-based post-processing strategies can explicitly suppress distractors. When all the proposals have a low response, this strategy ensures that the tracker outputs an appropriate result. As the tracker will extract the ROI based on the result of the previous tracking frame, losing the target would be catastrophic for the subsequent tracking. The proposed method can effectively avoid this problem. Compared with the post-processing strategy based on the Hanning window, the proposed method is more robust.

4. EXPERIMENTS

The trained MobileNetV2 in SiamRPN++ is utilized as the initial weights of the backbone and fine-tune the network with the mixed training strategy mentioned in the previous section. The warm up learning rate is set to make the learning rate decay exponentially from 0.005 to 0.0005 during the training. In addition, the optimizer is SGD.

During the training process, a total of 20 epochs is performed. In the first 10 epochs, only the weights of the RPN are optimized. In the last 10 epochs, the last five inverted residual blocks and the RPN are optimized together. The sizes of the target images and search images are set to 127×127 and 255×255 .

Following the standard of the CLUST 2015 and VOT 2015, the locations predicted by the tracker are compared with the ground truth. Criteria are calculated and plots are drawn below.

The proposed method is built with Python. And the experiments are implemented on a computer with an Intel i7 processor, 32 GB of RAM, and an Nvidia GTX 1080 graphic card.

4.1. Data

Experiments are performed on two datasets, which are, the published CLUST 2D Training Dataset and the self-collected dataset.

The CLUST dataset provides 24 2D ultrasound sequences with public annotations. The ultrasound sequences were acquired from patients during free breathing with various equipment, leading to sequences with different temporal and spatial resolutions. Approximately 10% to 13% of the frames in each sequence are annotated. Moreover, multiple targets may be labeled in a sequence. The dataset is annotated manually with the target location by three observers and verified by an additional observer. The ground truth of the dataset is the mean of the three manual annotations.

The self-collected dataset is acquired using a Philips scanner. The self-collected dataset consists of 10 sequences with temporal resolutions from 27 Hz to 30 Hz and spatial resolutions from $0.16 \text{ mm} \times 0.16 \text{ mm}$ to $0.27 \text{ mm} \times 0.27 \text{ mm}$. All the data are annotated following the method of CLUST dataset. When collecting ultrasound sequences for CLUST 2D Training Dataset, coughing and other emergencies sometimes cause discontinuities in the sequence. In the self-collected dataset, those discontinuous sequences are filtered out for better quality.

4.2. Evaluation Criteria

To evaluate the proposed method, two types of experiments are designed, which are, a cross-validation in the CLUST 2D Training Dataset and an evaluation in the self-collected dataset.

Specifically, a sixfold cross-validation in the CLUST 2D Training Dataset is performed. The dataset is divided into six groups. All the models are fine-tuned into five groups then evaluated in the remaining group. In the evaluation in the self-collected dataset, the model is fine-tuned first in the CLUST 2D Training Dataset.

To compare the various trackers comprehensively, two evaluation methods are designed. Following the criteria of the CLUST dataset, the trackers are evaluated in all the sequences with only one initialization [[1]]. Given annotations l_i and tracking results x_i for target i , tracking error E_i at time t is calculated as

$$E_i(t) = \|l_i - x_i\|,$$

where $\|\cdot\|$ represents the Euclidean distance. The tracking errors are summarised by the mean, standard deviation, and 95th percentile of the Euclidean distance for all the frames.

Inspired by the criteria of VOT 2015 [[31]], an evaluation experiment is designed using a different method. After initialization at the first frame, the tracker is reinitialized when it loses the target. The failures are counted to measure the robustness of the tracker. In addition, the average overlap between the predicted target bounding boxes and annotations is calculated, which is defined as accuracy.

The expected average overlap (EAO) is the average of the expected overlap in an interval $[N_l, N_h]$ of typical sequence lengths. The expected overlap of N_s is calculated by averaging the overlap in all available N_s -length sequences. To get the typical sequence length, the probability density function (PDF) of the sequence lengths is computed via kernel density estimation. Figure 5 presents the estimated PDF of the sequence lengths in the self-collected dataset. As it shows, the typical length is in the interval of $[121, 308]$. The probability of the sequence length being a typical length is 50%.

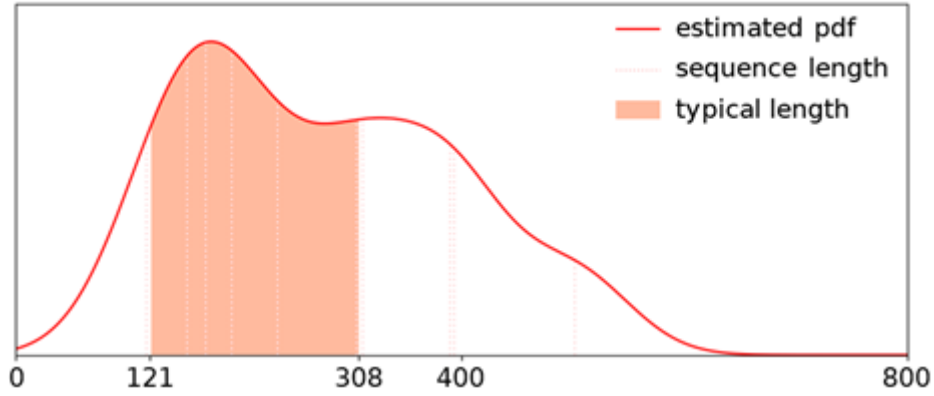


Figure 5. Estimated PDF of sequence lengths in the self-collected dataset; the sequence length in the dataset is marked by dotted lines.

4.3. Cross Validation in CLUST Dataset

A sixfold cross-validation is performed in the CLUST 2D Training Dataset to compare the performance of the proposed model with that of several representative methods. Figure 6 presents the success plots and precision plots of the proposed method and several representative methods, including SiamRPN++, SiamFC, DiMP18, PrDiMP18, and KYS, and shows that the proposed method generates superior results in terms of overlap success.

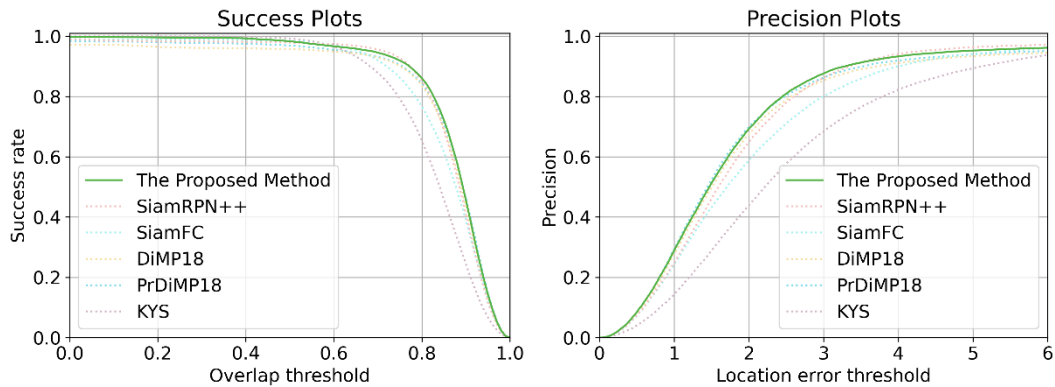


Figure 6. Success plots and precision plots of the proposed method and several representative methods. All the trackers are evaluated in the CLUST dataset via cross-validation.

Table 1 reports the mean, standard deviation, and 95th percentile of the tracking error of each tracker evaluated in the CLUST dataset via cross-validation. Table 2 shows the accuracy, failure, and EAO of each tracker evaluated in the CLUST dataset via cross-validation. The tables reveal that the proposed method obtains a minimal mean error and maximal accuracy.

Table 1. Mean error (Mean), standard deviation (Std) and 95th percentile (TE95th) of each tracker evaluated in CLUST dataset via cross validation

Tracker	Mean (mm)	Std (mm)	TE95th (mm)
The Proposed Method	0.8582	1.7042	1.9410
SiamRPN++	1.3622	4.7684	1.7851
SiamFC	1.4086	3.3518	2.5030
DiMP18	1.3864	3.8717	2.5193
PrDiMP18	1.5818	4.5768	2.7775
KYS	1.0856	0.9283	2.5882

Table 2. Accuracy, failure and EAO of each tracker evaluated in CLUST dataset via cross validation

Tracker	Accuracy	Failure	EAO
The Proposed Method	0.8690	7	0.8322
SiamRPN++	0.8655	8	0.8492
SiamFC	0.8479	14	0.8120
DiMP18	0.8449	2	0.8541
PrDiMP18	0.8564	27	0.7698
KYS	0.8230	2	0.8207

4.4. Evaluation of Trackers in Self-collected Dataset

The trackers are further evaluated by fine-tuning them in the CLUST dataset then evaluating them in the self-collected dataset. Figure 7 presents the success plots and precision plots of the proposed method and several representative methods, including SiamRPN++, SiamFC, DiMP18, PrDiMP18, and KYS, and shows that the proposed method demonstrates the best performance.

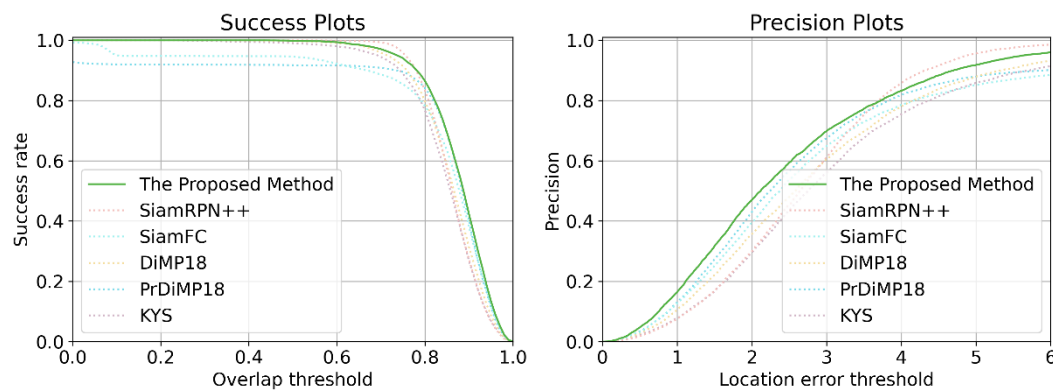


Figure 7. Success plots and precision plots of the proposed method and several representative methods. All the trackers are trained in the CLUST dataset and evaluated in the self-collected dataset.

Table 3 reports the mean, standard deviation, and 95th percentile of the tracking error of each tracker evaluated in the self-collected dataset. Table 4 shows the accuracy, failure, and EAO of each tracker evaluated in the self-collected dataset. Compared with the other methods, the proposed method obtains the best mean error, accuracy, and EAO.

Table 3. Mean error (Mean), standard deviation (Std) and 95th percentile (TE95th) of each tracker trained in the CLUST dataset and evaluated in the self-collected dataset

Tracker	Mean (mm)	Std (mm)	TE95th (mm)
The Proposed Method	0.5745	0.4369	1.4048
SiamRPN++	0.6298	0.3503	1.2469
SiamFC	1.1444	2.3535	8.0855
DiMP18	0.6591	0.4800	1.5631
PrDiMP18	0.8305	0.5736	1.9257
KYS	0.7258	0.5895	1.7611

Table 4. Accuracy, failure and EAO of each tracker trained in the CLUST dataset and evaluated in the self-collected dataset

Tracker	Accuracy	Failure	EAO
The Proposed Method	0.8750	0	0.8782
SiamRPN++	0.8597	0	0.8642
SiamFC	0.8218	2	0.8214
DiMP18	0.8567	0	0.8002
PrDiMP18	0.8357	51	0.3188
KYS	0.8445	0	0.8494

4.5. Visualisation

The tracking results of the proposed method and SiamRPN++ from the cross-validation in the CLUST dataset are visualized for a representative example. The tracking trail of the trackers is plotted and several frames are posted with tracking results in Figure 8. Frames 0570 to 0652 show a distractor approaching the target, leading SiamRPN++ to drift whilst the proposed method tracks steadily. As shown in Figure 8, the proposed method exhibits superior capability in distinguishing the target from the distractors, thereby benefitting from the NMS-based post-processing.

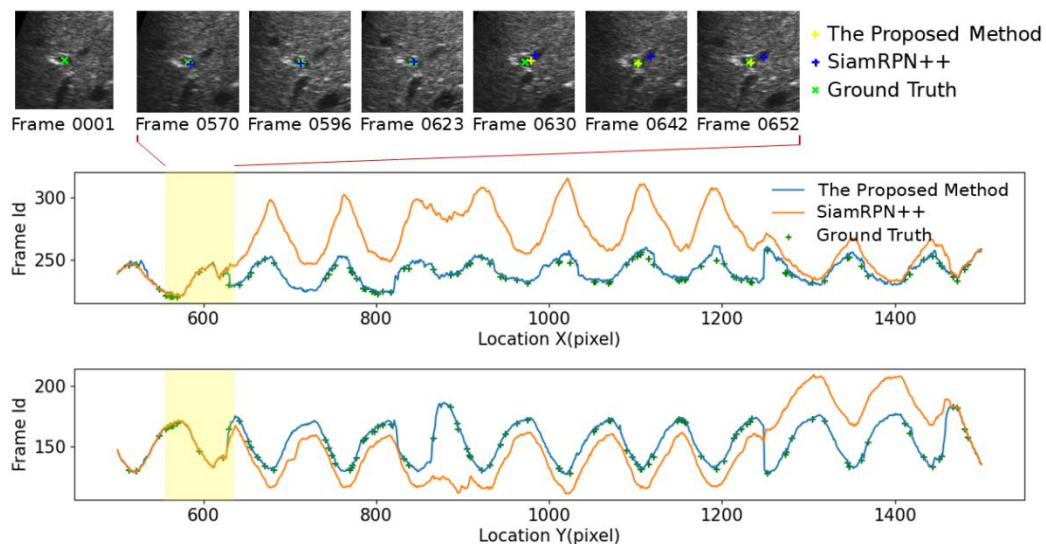


Figure 8. Tracking trail plot and tracking results of several frames.

Table 5 reveals the speed of the proposed method and several representative methods. The proposed method is the second fastest tracker among all the listed trackers and achieves real-time, with 62.12 FPS.

Table 5. Speed of proposed method and several representative methods

Tracker	Speed (FPS)
The Proposed Method	62.12
SiamRPN++	61.14
SiamFC	105.03
DiMP18	31.59
PrDiMP18	19.98
KYS	17.69

4.6. Ablation Study

In this section, the effects of different network factors are validated. Several trackers composed of different factors are evaluated using the same two methods employed in the previous experiments. Table 6 shows several criteria of each tracker evaluated via cross-validation. Table 7 presents several criteria of each tracker evaluated in the self-collected dataset. The tables reveal that the unsupervised strategy, feature fusion, and NMS post-processing contribute to the satisfactory performance of the proposed method.

Table 6: Network factors, accuracy and failure of several trackers evaluated in CLUST dataset via cross validation

Tracker	Unsupervised strategy	Feature fusion	Using NMS	Accuracy	Failure
The Proposed Method	√	√	√	0.8690	7
The Proposed Method (only trk)		√	√	0.8340	43
The Proposed Method (no fuse)	√		√	0.8655	8
The Proposed Method (no nms)	√	√		0.8472	24
SiamRPN++				0.8655	8

Table 7: Network factors, accuracy and failure of several trackers trained in CLUST dataset and evaluated in self-collected dataset

Tracker	Unsupervised strategy	Feature fusion	Using NMS	Accuracy	Failure
The Proposed Method	√	√	√	0.8750	0
The Proposed Method (only trk)		√	√	0.8041	0
The Proposed Method (no fuse)	√		√	0.8597	0
The Proposed Method (no nms)	√	√		0.8746	0
SiamRPN++				0.8597	0

5. CONCLUSIONS

In this paper, the obstacles for utilizing a highly sophisticated architecture in ultrasound tracking are analyzed. Firstly, an unsupervised training strategy is introduced to solve the problem of the lack of labeled data. Secondly, an RPN module is employed to predict the possible location of the target. Thirdly, feature fusion and NMS-based post-processing are proposed to improve the algorithm's robustness to distractors. Finally, an end-to-end network architecture is built with a

unique training strategy. Moreover, a large number of experiments are conducted based on the CLUST and the self-collected dataset, which proves our improvement of the performance of the algorithm.

This work provides a solution to the problem of applying very deep network structures to ultrasound tracking. In addition to the unsupervised training method that solves the lack of samples, other improvements are also proven to refine the accuracy of the algorithm. The proposed method gets accurate tracking results with a speed of 62.12 fps, which is surplus to ensure the real-time performance of the system.

For the lack of utilizing prior knowledge of ultrasound sequences, there is still much room to improve the proposed method. As our algorithm suppresses the distractors with the continuity of the sequence, the problem of distractors is not solved completely. And it also leads to dependence on the continuity of the sequence, which is difficult to guarantee during ultrasound acquisition. In the future, we will combine the characteristics of ultrasound images and the temporal and spatial characteristics of respiratory motion to further improve the accuracy and robustness of the algorithm.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation Program of China (81627803, 81871374)

REFERENCES

- [1] De Luca V, Banerjee J, Hallack A, et al. Evaluation of 2D and 3D ultrasound tracking algorithms and impact on ultrasound-guided liver radiotherapy margins[J]. *Medical physics*, 2018, 45(11): 4986-5003.
- [2] Boda-Heggemann J, Knopf A C, Simeonova-Chergou A, et al. Deep inspiration breath hold—based radiation therapy: a clinical review [J]. *International Journal of Radiation Oncology* Biology* Physics*, 2016, 94(3): 478-492.
- [3] Péguret N, Ozsahin M, Zeverino M, et al. Apnea-like suppression of respiratory motion: First evaluation in radiotherapy[J]. *Radiotherapy and Oncology*, 2016, 118(2): 220-226.
- [4] Parkes M J, Green S, Stevens A M, et al. Safely prolonging single breath-holds to > 5 min in patients with cancer; feasibility and applications for radiotherapy[J]. *The British journal of radiology*, 2016, 89(1063): 20160194.
- [5] Parkes M J. Breath-holding and its breakpoint [J]. *Experimental physiology*, 2006, 91(1): 1-15.
- [6] Takao S, Miyamoto N, Matsuura T, et al. Intrafractional baseline shift or drift of lung tumor motion during gated radiation therapy with a real-time tumor-tracking system [J]. *International Journal of Radiation Oncology* Biology* Physics*, 2016, 94(1): 172-180.
- [7] Hunt M A, Sonnick M, Pham H, et al. Simultaneous MV-kV imaging for intrafractional motion management during volumetric-modulated arc therapy delivery[J]. *Journal of applied clinical medical physics*, 2016, 17(2): 473-486.
- [8] Iwata H, Ishikura S, Murai T, et al. A phase I/II study on stereotactic body radiotherapy with real-time tumor tracking using CyberKnife based on the Monte Carlo algorithm for lung tumors[J]. *International journal of clinical oncology*, 2017, 22(4): 706-714.
- [9] Şen H T, Bell M A L, Zhang Y, et al. System integration and in vivo testing of a robot for ultrasound guidance and monitoring during radiotherapy[J]. *IEEE Transactions on Biomedical Engineering*, 2016, 64(7): 1608-1618.
- [10] De Luca V, Benz T, Kondo S, et al. The 2014 liver ultrasound tracking benchmark[J]. *Physics in Medicine & Biology*, 2015, 60(14): 5571.
- [11] De Luca V, Székely G, Tanner C. Estimation of large-scale organ motion in B-mode ultrasound image sequences: a survey [J]. *Ultrasound in medicine & biology*, 2015, 41(12): 3044-3062.

- [12] Hallack A, Papiez B W, Cifor A, et al. Robust liver ultrasound tracking using dense distinctive image features[J]. MICCAI 2015 Challenge on Liver Ultrasound Tracking, 2015: 28-35.
- [13] Shepard A J, Wang B, Foo T K F, et al. A block matching based approach with multiple simultaneous templates for the real-time 2D ultrasound tracking of liver vessels [J]. Medical physics, 2017, 44(11): 5889-5900.
- [14] Williamson T, Cheung W, Roberts S K, et al. Ultrasound-based liver tracking utilizing a hybrid template/optical flow approach [J]. International journal of computer assisted radiology and surgery, 2018, 13(10): 1605-1615.
- [15] Liu F, Liu D, Tian J, et al. Cascaded one-shot deformable convolutional neural networks: Developing a deep learning model for respiratory motion estimation in ultrasound sequences [J]. Medical Image Analysis, 2020, 65: 101793.
- [16] Shen C, Shi H, Sun T, et al. An Online Learning Approach for Robust Motion Tracking in Liver Ultrasound Sequence[C]//Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Springer, Cham, 2018: 440-451.
- [17] Jeungyoon, L., Euisuk, C., Tai-Kyong, S., Combination of RCNN and KCF for Landmark Tracking in 2D Ultrasound Sequence of Liver. IEEE Engineering in Medicine & Biology Society,
- [18] Gomariz A, Li W, Ozkan E, et al. Siamese networks with location prior for landmark tracking in liver ultrasound sequences[C]//2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, 2019: 1757-1760.
- [19] Zhang S, Yao H, Sun X, et al. Sparse coding based visual tracking: Review and experimental comparison [J]. Pattern Recognition, 2013, 46(7): 1772-1788.
- [20] Xu S, Chang A. Robust object tracking using Kalman filters with dynamic covariance [J]. Cornell University, 2014: 1-5.
- [21] Zeng H, Chen J, Cui X, et al. Quad binary pattern and its application in mean-shift tracking[J]. Neurocomputing, 2016, 217: 3-10.
- [22] Vojir T, Neskova J, Matas J. Robust scale-adaptive mean-shift for tracking[J]. Pattern Recognition Letters, 2014, 49: 250-258.
- [23] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE transactions on pattern analysis and machine intelligence, 2014, 37(3): 583-596.
- [24] Wu Y, Lim J, Yang M H. Online object tracking: A benchmark[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2013: 2411-2418.
- [25] LIRIS F. The Visual Object Tracking VOT2014 challenge results[J].
- [26] Danelljan M, Robinson A, Khan F S, et al. Beyond correlation filters: Learning continuous convolution operators for visual tracking[C]//European conference on computer vision. Springer, Cham, 2016: 472-488.
- [27] Danelljan M, Bhat G, Shahbaz Khan F, et al. Eco: Efficient convolution operators for tracking[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 6638-6646.
- [28] Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 4293-4302.
- [29] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional siamese networks for object tracking[C]//European conference on computer vision. Springer, Cham, 2016: 850-865.
- [30] Li B, Wu W, Wang Q, et al. Siamrpn++: Evolution of siamese visual tracking with very deep networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 4282-4291.
- [31] Kristan M, Matas J, Leonardis A, et al. The visual object tracking vot2015 challenge results[C]//Proceedings of the IEEE international conference on computer vision workshops. 2015: 1-23.
- [32] Kristan M, Leonardis A, Matas J, et al. The visual object tracking vot2017 challenge results[C]//Proceedings of the IEEE international conference on computer vision workshops. 2017: 1949-1972.
- [33] Kristan M, Matas J, Leonardis A, et al. The seventh visual object tracking vot2019 challenge results[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019: 0-0.
- [34] Kondo S. Liver ultrasound tracking using kernelized correlation filter with adaptive window size selection[C]//MICCAI workshop: challenge on liver ultrasound tracking. 2015: 13-19.

- [35] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510-4520.

EXTRACTIVE TEXT SUMMARIZATION USING RECURRENT NEURAL NETWORKS WITH ATTENTION MECHANISM

Shimirwa Aline Valerie and Jian Xu

School of Computer Science and Engineering, Nanjing University of Science
and Technology, Nanjing 210094, China

ABSTRACT

Extractive summarization aims to select the most important sentences or words from a document to generate a summary. Traditional summarization approaches have relied extensively on features manually designed by humans. In this paper, based on the recurrent neural network equipped with the attention mechanism, we propose a data-driven technique. We set up a general framework that consists of a hierarchical sentence encoder and an attention-based sentence extractor. The framework allows us to establish various extractive summarization models and explore them. Comprehensive experiments are conducted on two benchmark datasets, and experimental results show that training extractive models based on Reward Augmented Maximum Likelihood (RAML) can improve the model's generalization capability. And we realize that complicated components of the state-of-the-art extractive models do not attain good performance over simpler ones. We hope that our work can give more hints for future research on extractive text summarization.

KEYWORDS

Extractive summarization, Recurrent neural networks, Attention mechanism, Maximum Likelihood Estimation, Reward Augmented Maximum Likelihood.

1. INTRODUCTION

Automatic text summarization is one of the challenging and interesting tasks of natural language processing, which can help people to obtain important and relevant information from a large number of documents in a short period. It has gained its popularity due to the importance it has in different information access applications such as search engines, information retrieval, recommendation systems, question answering, etc. When it comes to automatic text summarization, there are two approaches extractive text summarization and abstractive text summarization. While in the extractive summarization approaches the most salient sentences or words from the document are selected and concatenated to form a summary, in the abstractive summarization approaches the sentences in the document are paraphrased to make the summary. Even though the abstractive summarization approaches have made steps in recent years, the extractive approaches are still attractive since they can generate coherent and grammatically correct summaries and are computationally efficient [1]. Thus, in this work, we focus on extractive summarization.

A main requirement for the extractive summarization approach is to have a good method to determine the important contents that represent the important information in the document [2].

Several traditional techniques have been used to extract the important sentences to be included in the summary.

These techniques can be categorized into greed-based [3], graph-based [4], hidden Markov models[5], and constraint optimization[6],etc. These traditional extractive techniques use human-crafted features and are complicated. Moreover, they mostly fail to build a good representation of the document. This leads them to fail to generate good summaries.

In recent years, deep learning-based models have been used for extractive text summarization. These models can learn from the input text data directly, and they have attained state-of-the-art results. To create representation of the sentences and documents, neural network-based extractive models are basically constructed using recurrent neural networks [1,7], convolutional neural networks [8,9], the combination of convolutional and recurrent neural networks [10,11] or transformers[12]. While there has been great effort dedicated to designing neural network-based extractive summarization models, there is still a need to explore what makes them work well and how they can be improved. Therefore, in this paper, we present a recurrent neural network-based extractive model that consists of a hierarchical sentence encoder and an attention-based sequence-to-sequence sentence extractor. And we closely explore how the choice of sentence encoder can influence the model's performance.

Since there is little work that has been done on learning approaches for neural extractive summarization, we also examine how different learning approaches can contribute to the performance and generalization of the model. Existing neural-based extractive summarization systems fail to generalize better on the data they have not seen. We introduce the use of the RAML approach to the summarization task with the expectation that it can improve the generalization ability of the model.

The main contributions of this work are: (1) we adopt the RAML optimization approach to the task of extractive summarization; (2) we present two hierarchical neural structures (Avg-Seq_to_Seq and Rnn-Seq_to_Seq) for the extractive summarization task; (3) we perform a multi-domain test, which allows us to better understand how biases in different datasets influence the performance of our models;(4) we analyze the generalization capability of the models on out-of-domain datasets. For example, we train a model on the CNN dataset and test it on the PubMed dataset to see how the model can generalize to other datasets. Additionally, we demonstrate the effect of the position of the sentences on the performance of our models.

The rest of the paper is organized as follows. Section 2 describes the related work. Section 3 demonstrates our model. Section 4 describes our experiments and results. Section 5 demonstrates our discussion. Section 6 concludes our work.

2. RELATED WORK

To identify and select the most important sentences in a document or set of documents to make a summary, researchers have used several methods. These methods can be classified into statistical, graph, machine learning, deep learning-based approaches, etc. In this section, we demonstrate some of these approaches.

2.1. Statistical-Based Summarization Approaches

These approaches mainly use statistical features such as term frequency, sentence position, sentence length, TF-IDF (Term Frequency-Inverse Document Frequency), sentence to centroid

similarity, etc., to score the sentences. Then the sentences with high scores are selected to make the summary. Similarity to centroid sentence was used in [17] to score sentences. In their work TF-IDF is used to get centroid sentence then based on the cosine similarity between each sentence and centroid sentence, each sentence is given a score. Eleven features including document frequency, sentence position, normalized sentence length, proper noun, topic frequency, numerical data, headline frequency, start cluster frequency, and skip Bi-gram topic frequency are used in [18] to score sentences and then sentences with high scores are selected until reaching the length limit of the summary. One of the advantages of statistical-based approaches is that they do not require training data or complex linguistic processing. And one of the limitations of them is that they can generate summaries with redundant information because similar sentences with high scores can be included in the summary.

2.2. Graph-Based Summarization Approaches

Researchers have also used graph-based summarization approaches to perform extractive summarization. In the graph-based method, sentences are represented using nodes of a weighted graph. And the similarities between sentences are represented using edges. Sentence similarity values are obtained based on the overlapping phrases or words between sentences, then the sentences which have high similarity with the other ones are selected to generate the final summary. Two well-known graph-based approaches are Lex Rank and Text Rank. Text Rank was introduced by Mihalcea [4] to extract sentences and keywords from a single document. LexRank was introduced by Erkan [19] to compute the importance of the sentence based on the idea of eigenvector centrality in the sentence representation graph. Graph-based approaches generate summaries with less redundant information and they do not require annotated corpora. One of the disadvantages of these methods is that they do not take into account the importance of the words. They treat the weights of the words equally equal.

2.3. Machine Learning-Based Summarization Approaches

Different machine learning methods have been used to carry out extractive text summarization task. Some of those methods are Support Vector Machine(SVM) [20], Naïve Bayesian [21], Hidden Markov Models [5], etc. A binary classifier is proposed in [21] to score sentences using Bayes' rule. In their work, the probability of each sentence to be included in the summary is obtained by using manually crafted features. In [5] hidden Markov model algorithm identifies the likelihood of each sentence to be select for the summary. SVM is used in [20] for query-based summarization to declare appropriate sentences to put in the summary. The advantage of machine learning-based approaches is that they can explore many features and can represent documents in a better way than statistical and graphical approaches but they also need human crafted features to generate summaries with high accuracy and they need large labeled corpora.

2.4. Deep Learning-Based Summarization Approaches

In recent years, deep learning-based approaches have gained popularity over the above-mentioned traditional approaches because they can directly learn from the data. Neural network-based extractive summarization models have achieved state-of-the-art results. For instance, SummaRuNNer [1] uses bidirectional RNNs at the word level to encode sentences and another bidirectional RNNs at the sentence level to predict which sentences are to be extracted. In their work, the sentence extractor generates document representations and calculates distinct scores for novelty, location and salience of the sentences. In [7], authors propose convolutional neural network (CNN)-based model to encode sentences at the word level and design an extractor to predict which sentence should be included in the summary at the sentence level. Authors [13] propose an end-to-end neural extractive summarization model that learns to score and select

sentences jointly. In their work to obtain sentence representations, they use a hierarchical encoder, and then the output summary is obtained by extracting one sentence at a time. Based on the previous works that used hierarchical architectures [7, 13, 15], we also present a recurrent neural network-based model that consists of hierarchical sentence encoder and sequence-to-sequence based sentence extractor.

Although neural extractive summarization models have achieved great performance, most of the existing works, during training of these models use MLE. MLE approach maximizes the likelihood of the ground truth labels and disregards the structure of the output space by taking all the output which do not match the ground truth labels as equally wrong, irrespective of their structural closeness to the ground truth target. This leads to the inconsistency between training and testing objectives (i.e., during training the model learns to maximize the likelihood of the ground truth labels while during testing the objective is to generate the summaries with a high ROUGE score concerning the reference summary). This inconsistency can cause the overfitting of the ground truth labels and leads to poor generalization capability on test datasets. Some researchers have tried to eliminate this inconsistency by optimizing task reward (ROUGE evaluation metric) directly using Reinforcement learning (RL) approaches. For example, authors [10] proposed an approach that optimizes the ROUGE metric globally and use reinforcement learning objective to rank sentences that can be included in the summary. Authors [11] proposed a consistency model that takes syntactic coherence and cross-sentence semantic patterns. They used the RL objective to train their model. In their work, the output of the model and the reward calculated using the ROUGE package are combined to capture the cross-sentence consistent patterns. The limitation of the reinforcement learning approaches is that they suffer from problems of high variance in the gradients and poor sample effectiveness (sampling from a non-stationary model distribution). In this paper, we adopt a learning approach called RAML to the task of extractive summarization with the expectation that it can improve the performance and generalization of our models. RAML approach was proposed by [16] to include task reward into Maximum-likelihood training. It was successfully applied to machine translation task and speech recognition. It combines the straightforwardness and computational effectiveness of MLE with the advantages of maximizing task reward. Unlike MLE that maximizes the log-likelihood of the ground-truth labels, RAML can sample from the exponentiated payoff distribution which permits the estimation of anticipated maximum likelihood. In this paper, we not only train our models based on the MLE approach but also, train them based on the RAML approach.

3. NEURAL NETWORK-BASED EXTRACTIVE SUMMARIZATION MODEL

In this paper, we treat the task of extractive summarization as a sequence labeling problem or a classification problem. Given a document d consists of n sentences $d = \{s_1, s_2, s_3, \dots, s_n\}$. we aim to generate a summary by predicting the corresponding labels of sequences $y_1, y_2, y_3, \dots, y_n \in \{0, 1\}^n$, where $y_j=1$ denotes that the j^{th} sentence should be included in the summary, otherwise $y_j=0$. Based on the extraction probabilities, sentences are selected until reaching the length limit of the output summary. Since each sentence itself is a sequence of words $s_j = \{w_1, w_2, w_3, \dots, w_L\}$, we set word budget $b \in \mathbb{N}$ to put a constraint on the limit length of the output summary $\sum_j^n y_j \cdot L \leq b$. Generally, our proposed model is suitable for a single document. It consists of the following components, as shown in Figure 1.

3.1. Embedding Layer

The embedding layer is the first layer in our model. It converts positive integers (indices) of the words in the training dataset into dense vector representations of fixed size. These dense vector representations capture the syntactic and semantic potential meaning of the words. Instead of

training our dense vector representation of the words, we initialize the embedding layer with glove pre-trained word embeddings with 200 dimensions.

3.2. Sentence Encoder

The sentence encoder converts the sequence of word embeddings of each sentence into a fixed-length vector. We get sentence representations using two different approaches the first one is by using a Recurrent neural network (Rnn) and another one is simply by averaging word embeddings (Avg). By using the Rnn approach, at each time step, a Bidirectional Recurrent Neural Network (Bi-RNN) runs at the word level of each sentence and then constructs sentence representation s_j . We employ a Bidirectional Gated Recurrent Unit (Bi-GRU) [24] as RNN cells. Bi-directional GRU consists of forwarding and backward GRU. Forward GRU reads the word embeddings in a sentence from left to right to generate a sequence of hidden states $(\vec{h}_1, \vec{h}_2, \vec{h}_3, \dots, \vec{h}_L)$. The backward GRU reads word embeddings in the sentence from right to left to form another sequence of hidden states $(\overleftarrow{h}_1, \overleftarrow{h}_2, \overleftarrow{h}_3, \dots, \overleftarrow{h}_L)$.

$$\vec{h}_j = \overrightarrow{GRU}(w_j, \vec{h}_{j-1}) \quad (1)$$

$$\overleftarrow{h}_j = \overleftarrow{GRU}(w_j, \overleftarrow{h}_{j+1}) \quad (2)$$

Where the initial state of forwarding Bi-GRU is set to zero vector ($\vec{h}_1 = 0$) as well as the initial state of backward Bi-GRU ($\overleftarrow{h}_L = 0$). After reading the words in the sentence, the sentence representation at the word level is constructed by concatenating the hidden states of last forward and backward GRU:

$$s_j = [\vec{h}_L; \overleftarrow{h}_1] \quad (3)$$

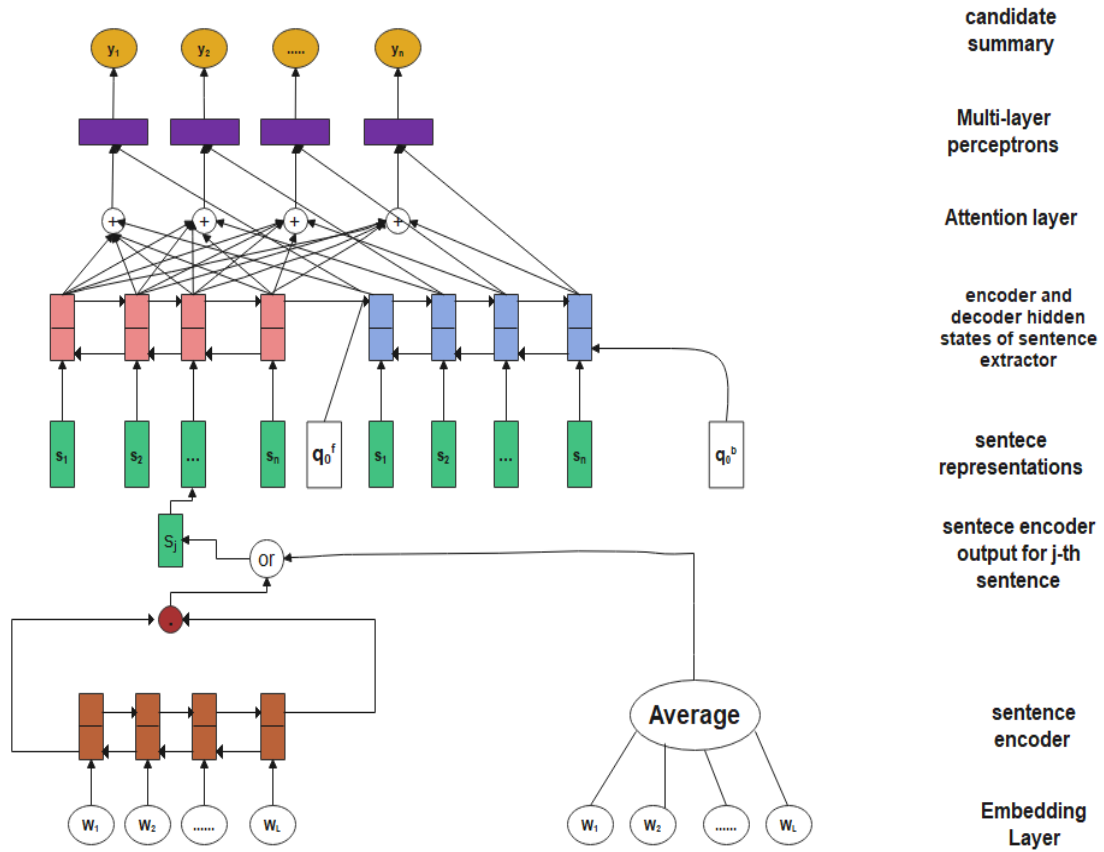


Figure 1. Overview of the two different hierarchical architectures (Rnn- Seq_to_Seq architecture or Avg-Seq_to_Seq architecture) for extractive summarization.

In Figure 1 vertical brown blocks present the sentence encoder's hidden states. \odot presents concatenation. Vertical green blocks indicate the output of the sentence encoder. White block (q_0^f) presents forward learned embeddings (“begin decoding”) and white block (q_0^b) presents backward learned embeddings. Vertical pink boxes present hidden states of the encoder part of the sentence extractor. Vertical blue boxes indicate hidden states of the decoder part of the sentence extractor. \oplus presents attention layer. Horizontal purple blocks indicate multi-layer perceptron. Yellow ovals indicate the candidate summary.

By using the averaging approach, each sentence representation s_j is obtained by averaging its word embeddings:

$$s_j = \frac{1}{L} \sum_{j=1}^L w_j \quad (4)$$

3.3. Sentence Extractor

Sentence extractors take in sentence hidden vectors $s_{1:n}$ and generate the sequence of labels $y_{1:n}$. We use the attention-based sequence-to-sequence sentence extractor (Seq_to_Seq) which consists of an encoder, a decoder, an attention layer, and multi-layer perceptrons.

3.3.1. Encoder Part of Sentence Extractor

The encoder part of the sentence extractor takes sentence representations ($s_1, s_2, s_3, \dots, s_n$) from the sentence encoder as inputs and encodes them using Bi-GRU. Forward and backward hidden states of Bi-GRU are concatenated to produce a sequence of contextualized sentence representations (sentence embedding s'_j).

$$\vec{s}_j = \overrightarrow{GRU_{enc}}(s_j, \vec{s}_{j-1}) \quad (5)$$

$$\overleftarrow{s}_j = \overleftarrow{GRU_{enc}}(s_j, \overleftarrow{s}_{j+1}) \quad (6)$$

We set the initial state of forward Bi-GRU to zero vectors ($\vec{s}_1 = 0$). As well as the initial state of backward Bi-GRU ($\overleftarrow{s}_n = 0$). Sentence representation hidden vectors at sentence level:

$$s'_j = [\vec{s}_j; \overleftarrow{s}_j] \quad (7)$$

3.3.2. Decoder Part of Sentence Extractor

The decoder part of the sentence extractor takes in the sentences from the sentence encoder as inputs and then transforms them into a query vector that sees to the output of the encoder part of the sentence extractor.

$$\vec{q}_j = \overrightarrow{GRU_{dec}}(s_j, \vec{q}_{j-1}) \quad (8)$$

$$\overleftarrow{q}_j = \overleftarrow{GRU_{dec}}(s_j, \overleftarrow{q}_{j+1}) \quad (9)$$

$$q_j = [\vec{q}_j; \overleftarrow{q}_j] \quad (10)$$

The final outputs of the forward and backward encoder are fed to the first decoder steps. q_0^f, q_0^b are learned vectors of the first step of the decoder (i.e., start decoding).

3.3.3. Attention Layer

The attention mechanism is commonly used in abstractive summarization [25] and neural machine translation [26]. It plays a role in enabling models to concentrate on important information of the input while predicting the next output. In the attention layer of our model, given a query vector representation q and a sequence of sentence embeddings $[s'_1, s'_2, s'_3, \dots, s'_n]$, the attention mechanism computes an alignment score between q and each sentence s'_j . The scores are transformed into probabilities by using a SoftMax function. These probabilities are used as weights to sum all sentences and create a contextual embedding for q .

$$\alpha_{j,i} = \frac{\exp(q_j \cdot s_i)}{\sum_{i=1}^n \exp(q_j \cdot s_i)} \quad (11)$$

$$s''_j = \sum_{i=1}^n (\alpha_{j,i} s_i) \quad (12)$$

3.3.4. Multi-Layer Percetrans

Multi-layer perceptrons (MLP) take in a concatenation of the attention-weighted encoder output and decoder output as input to compute the probability of extracting each sentence.

$$a_j = \text{Relu}(U \cdot [s_j''; q_j] + u) \quad (13)$$

$$p(y_j = 1 | s_j) = \sigma(V \cdot a_j + v) \quad (14)$$

where U and V are learned weights, u and v are learned bias.

3.4. Model Training

We first train our model by maximizing the likelihood of the ground truth labels. To achieve this objective, the cross-entropy loss is minimized as follows:

$$\mathcal{L}_{MLE}(\theta) = -\sum_{d=1}^D \sum_j^n \log p(y_j^{(d)} | s_j^{(d)}; \theta) \quad (15)$$

Where D represents the total number of documents in the training dataset. $s^{(d)}$ represents the contextualized sentence vectors, n symbolizes the total number of sentences in the document. $y^{(d)}$ is each document's label vector. θ represents model parameters. When minimizing the above objective, the conditional probability of the output targets is escalated, and at the same time, the conditional probability of alternative wrong outputs is decreased. This can lead to overfitting on target outputs and decreases the generalization capability. To solve this issue, we train our model based on the RAML approach which was proposed in [16]. RAML simply attaches a step of sampling on top of the ordinary maximum likelihood estimation objective. And it can sample from an output distribution called exponentiated payoff distribution which serves as a central to linking between MLE and RL objectives, and is defined as follows:

$$q(y|y'; \tau) = \frac{1}{Z(y'; \tau)} \exp\left\{\frac{r(y, y')}{\tau}\right\} \quad (16)$$

where $Z(y'; \tau) = \sum_{y \in Y} \exp\left\{\frac{r(y, y')}{\tau}\right\}$, hyper-parameter τ that controls the smoothness of the best distribution around correct labels.

The RAML objective is defined as follows:

$$\mathcal{L}_{RAML}(\theta; \tau) = \sum_{d=1}^D \left\{ -\sum_{y \in Y} q(y|y'; \tau) \log p(y|s; \theta) \right\} \quad (17)$$

As stated by [16] RAML approach can be treated as a hybrid between MLE and RL. The connection can be seen by rewriting \mathcal{L}_{MLE} , \mathcal{L}_{RL} , and \mathcal{L}_{RAML} using Kullback-Leibler Divergence:

$$\mathcal{L}_{MLE}(\theta) = \sum_{y \in Y} D_{KL}(\delta(y, y') || p(y|s; \theta)) \quad (18)$$

$$\frac{1}{\tau} \cdot \mathcal{L}_{RL}(\theta; \tau) + \text{constant} = \sum_{y \in Y} D_{KL}(p(y|s; \theta) || q(y|s; \tau)) \quad (19)$$

$$\mathcal{L}_{RAML}(\theta; \tau) + \text{constant} = \sum_{y \in Y} D_{KL}(q(y|s; \tau) || p(y|s; \theta)) \quad (20)$$

Though the core capability of the RAML approach lies in sampling from a static distribution, that distribution is difficult to define and we think that the training process can be destabilized when the sampling is introduced during computing gradients. Therefore, in this paper, instead of sampling, we pre-calculate the reward (ROUGE R1 score) for each possible output summary of each document as $R1(y, y')$. Then after normalizing the scores, the top-scored candidates T (we

use $T=25$ in our experiments) are used to calculate the weighted cross-entropy loss. During optimization, the weighted cross-entropy loss is defined as follows:

$$\mathcal{L}(y, y') = \sum_{i=1}^T -w_i \cdot \sum_{j=1}^n y_j^i \log y_j' \quad (21)$$

where $y' \in \{0,1\}^n$ demonstrates the vector of predicted sentences to be included in the summary, $y^i \in \{0,1\}^n$ denotes candidate(i) labels, and w_i represents weighted vector of the Rouge scores:

$$w_i = \frac{\text{RougeR1}(y', y)}{\sum_j \text{RougeR1}(y', y)} \quad (22)$$

4. EXPERIMENTS

The purpose of our experiments is to answer the following questions: (1) how different architectures of our models influence their performance? (2) how sentence positions affect the performance of the models? (3) how the MLE and RAML influence the generalization capability of the model on out-of-domain datasets? (4) how our models perform compared to the state-of-the-art baselines on CNN and PubMed datasets?

4.1. Datasets

We conduct experiments on two well know datasets from different domains (CNN and PubMed) to evaluate how different biases in each domain can affect the performance of our models. the statistics of the datasets are shown in Table 1.

Table 1. statistics of the datasets used in our experiments (CNN, PubMed): train, valid, and test split. The average number of words in the document and in the summary and the domain they belong to.

Datasets	Number-of-documents			Average Number-of-tokens		Domain
	Train	Validation	Test	Document	Summary	
CNN	90,152	1,220	1,093	761	46	News
PubMed	115,498	6,562	6,602	3,224	203	Scientific paper

CNN is a dataset that was first created by [27] for question answering, then was modified for text summarization task by [28]. This dataset is composed of news articles that are paired with human-generated summaries. For the data preprocessing, the non-anonymized version of the dataset is used in our experiments as in [25].

PubMed is the dataset that was introduced by [29]. the dataset is collected from scientific repositories PubMed.com. The statistics of the dataset are shown in Table 1. In our experiments, we use about 3% of PubMed as validation data and about another 3% for the test; the rest is used for training as in [29].

4.2. Implementation

In our experiments, each document is truncated to 50 sentences and we use padding to keep the lengths of documents. we use pre-trained Glove vectors with 200-dimensions to initialize word embeddings. Weperform mini-batch training with a batch size of 32 documents for 15 training epochs. In the Rnn-based sentence encoder, for each direction, we use a single-layer GRU with 300-dimensional hidden layers, and dropout is applied to GRU with drop probability equals 0.25. In the sentence extractor, for each direction, a single-layered GRU is used with a hidden layer

size of 300. We set the hidden layer size for MLP to 100. The model parameters in the sentence encoder and sentence extractor are initialized using a normal distribution with the Xavier scheme [30]. Our models are optimized using Adam optimizer [31] with an initial learning rate of 0.0001, and momentum parameters $\beta_1 = 0.9, \beta_2 = 0.999$. We use gradient clipping to regularize our models. All our experiments are implemented using Pytorch on the computer that has 256 RAM and NVIDIA GeForce RTX 2080 Ti GPU.

4.3. Evaluation

We use the Rouge metric [32] to evaluate the quality of the summaries. In the reported experimental results, unigram and bigram overlap (R-1, R-2) are reported as a means of evaluating informativeness. And longest common subsequence (R-L) is reported as means of evaluating the fluency.

4.4. Model Comparison

We compare the performance of our models with other well-known extractive models including:

- LSA[33]: Extractive model that uses latent semantic analysis approach to discover important sentences
- SumBasic[34]: A summarization model which can generate summaries for single and multi-document.
- LexRank[35]: Based on the idea of eigenvector centrality in a graph representing sentences, this model computes the importance of the sentences.
- NN-SE[7]: A neural network-based extractive model, which can be used to extract words and sentences.
- Refresh[10]: A neural network-based summarization model trained using Reinforcement learning objective to globally optimize evaluation metric (ROUGE).
- Banditsum[36]: A neural extractive summarization model that treats extractive summarization as a context bandit problem.

4.5. Results and Analysis

4.5.1. Influence of Different Architectures on the Performance of the Model.

To understand how different architectures can influence the performance of the model, we examine the performance of Avg (averaging word embeddings) and the Rnn approach to encode sentences at the word level. As it is shown in Figure 2 the approach of averaging word embeddings of each sentence to obtain its representation performs slightly better than the Rnn based sentence representation on the CNN dataset. Whereas on PubMed dataset, the averaging approach results in high performance than the Rnn. Moreover, the time taken to train our averaging word embedding-based model is less than the time taken to train our Rnn based sentence encoder. This implies that it is not always necessary to build very complex architectures to get good performance.

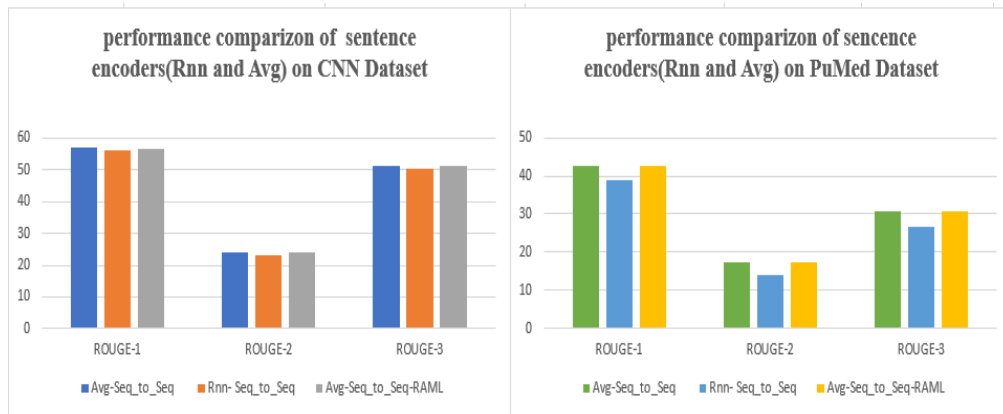


Figure 2. Performance comparison of sentence encoders on CNN and PubMed datasets.

4.5.2. Impact of the Position of the Sentences

When it comes to the extractive summarization for news, the position of the sentence is a very important feature [37]. When deep learning-based summarization models are trained on news datasets, these models mostly select the first sentences in the document and this causes the problem of lead bias. To answer the question (2), we test the performance of our models when the order of the sentences in the document is kept intact and when we shuffle them. As it is shown in Table 2, when sentences are shuffled during training the performance of our model trained based on MLE (Avg-Seq_to_Seq-MLE) drops significantly on CNN and PubMed datasets. This implies that the model has learned the position feature in PubMed/CNN datasets even though the model has no explicit position features. On the other hand, Figure 3 shows that the performance of our model trained using RAML(Avg-Seq_to_Seq-RAML) on shuffle sentences is not significantly dropped. The reason could be that this model is forced to learn from richer distribution of labels. Thus, it is less vulnerable to the lead bias.

Table 2. performance of our models on CNN and PubMed test set using full-length ROUGE F1 scores when using shuffled and in-order sentences during model training.

Models	Sentence shuffling	CNN			PubMed		
		R-1	R-2	R-3	R-1	R-2	R-3
Avg-Seq_to_Seq-MLE	shuffled	52.84	20.14	45.16	39.71	14.53	29.82
	normal	56.84	24.04	51.15	42.71	17.33	30.82
Avg-Seq_to_Seq-RAML	shuffled	56.69	23.76	51.06	42.55	17.26	30.77
	normal	56.71	23.97	51.09	42.69	17.29	30.79

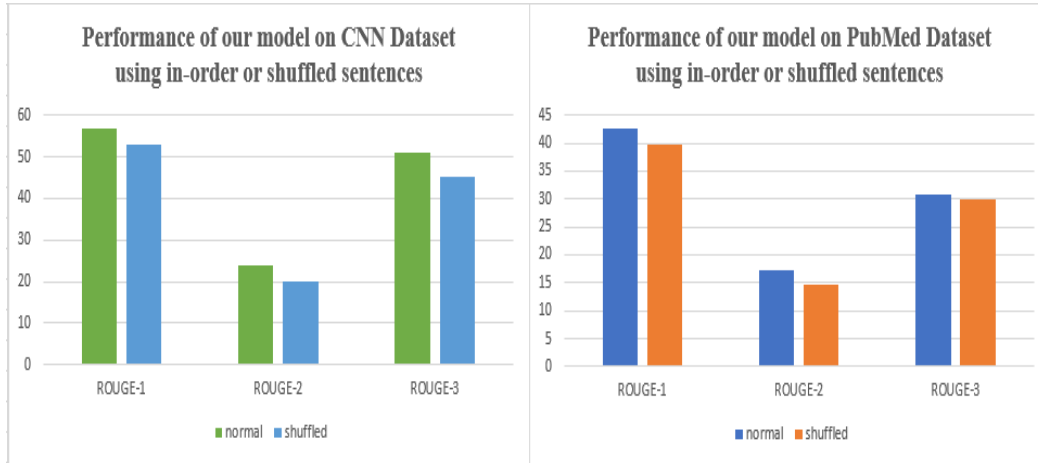


Figure 3. performance of our model trained using MLE approach on CNN and PubMed test set. where shuffled or in-order sentences are used during model training.

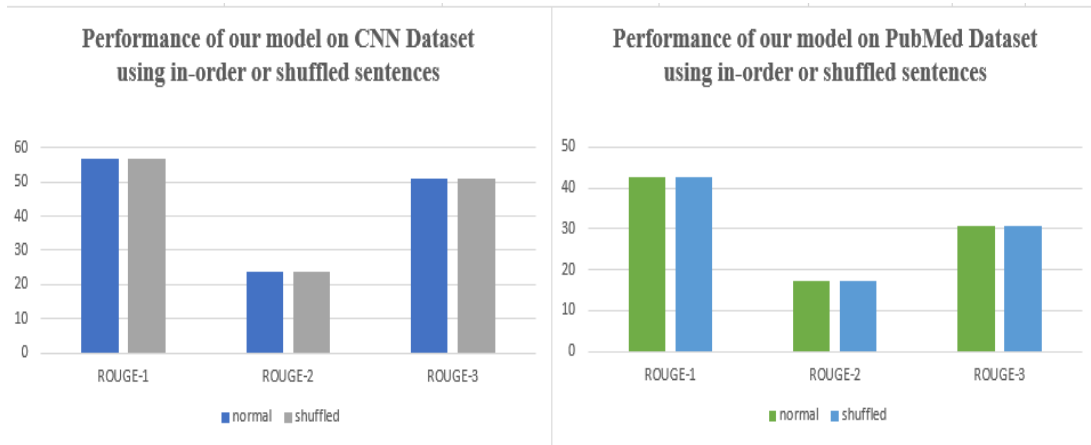


Figure 4. performance of our model trained using RAML approach on CNN and PubMed test set. where shuffled or in-order sentences are used during model training.

4.5.3. Generalization Capability on Out of Domain Dataset

Domain transfer is when a model is trained on one dataset but needs to have a better performance on the other datasets from different domains. Most of the time we want to train the model on a particular domain and be able to reuse it in another domain without retraining it. Let's say, for example we train our model to summarize a dataset of news articles. we do not want to retrain the model if we want to summarize research papers, personal stories, blogs, etc. To answer question (3), we first choose the Avg-Seq_to_Seq model and then train it on the CNN dataset using the MLE approach then transfer this model to the PubMed dataset. As it is shown in Table 3, our Avg-Seq_to_Seq model trained on the CNN dataset achieves 56.84% according to the R-1(ROUGE-1) measure on the test set of the CNN dataset. And the model achieves 35.17% on PubMed test data. The performance of the model drops almost 21.67%. Training models based on the MLE approach tend to cause poor generalization because these models mostly tend to overfit particular features in the training set that might not be in other datasets. We then train our model based on the RAML approach with the expectation that the model will perform better than its counterpart trained based on vanilla Maximum likelihood estimation. RAML approach can incorporate reward task into MLE training and this can improve the generalization ability of the

model. To our surprise, the model trained on the CNN dataset using RAML has slightly better performance on the PubMed dataset compared to the model trained on the CNN dataset using MLE and transferred to PubMed.

Table 3. experiment results for domain transfer, where we transfer a model trained on CNN dataset to PubMed dataset. Meaning we train a model using CNN dataset and test it using the PubMed dataset.

Models (CNN to PubMed)	R-1	R-2	R-3
Avg-Seq_to_Seq-MLE	35.17	11.59	25.26
Avg-Seq_to_Seq-RAML	35.61	11.73	25.41

4.5.4. Performance Comparison of the Summarization Models

To answer the fourth question (4), we inspect the ROUGE scores of the summaries generated by our models (Avg-Seq_to_Seq, Rnn- Seq_to_Seq trained using MLE or RAML) and the baselines that have achieved state-of-the-art results on CNN and PubMed datasets. As it is shown in Table 4, our models have attained significant results in terms of R-1,2, L on CNN dataset compared to Refresh [10], Banditsum [36], and NN-SE [7]. These results approve the efficacy of our models. Our sentence encoder (Avg or Rnn) followed by the attention-based sequence-to-sequence sentence extractor helps to get better representations of the documents and generate good summaries. We also examine our models on the PubMed dataset and compare our results with other summarization models. As it is also shown in Table 4, NN-SE [7] has slightly outperformed our models. This might be caused by the fact that during training, the NN-SE model takes into account previous predictions to inform future predictions while our models do not.

Table 4. the performance comparison of our models with other different extractive summarization modelson the CNN and PubMed test set using full-length ROUGE F-1 scores.

Models	CNN			PubMed		
	R-1	R-2	R-L	R-1	R-2	R-L
SumBasic ⁺ [34]	–	–	–	37.15	11.36	33.43
LSA ⁺ [33]	–	–	–	33.89	9.93	29.70
LexRank ⁺ [35]	–	–	–	39.19	13.89	34.59
Refresh [*] [10]	30.40	11.70	26.90	–	–	–
Banditsum [*] [36]	30.70	11.60	27.40	–	–	–
NN-SE ^{*~} [7]	28.40	10.00	25.00	43.89	18.78	30.17
Avg- Seq_to_Seq (ours)	56.84	24.04	51.16	42.71	17.33	30.82
Rnn- Seq_to_Seq (ours)	56.18	23.27	50.35	38.95	14.04	26.70
Avg-Seq_to_Seq-RAML (ours)	56.71	23.97	51.09	42.69	17.29	30.79

In Table 4, the result with * are obtained from [36], results with + are gotten from [29], – illustrates that the correlated result is not reported. and results with *~ are reported from [36] for CNN and from [38] for PubMed respectively. The top section of the table represents traditional approaches; the second and the third sections represent other deep learning-based extractive models and our models respectively.

4.5.5. Run Time Comparison of Our Models based on Sentence Encoder

To train Avg-Seq_to_Seq on CNN and PubMed took 8 and 10 hours respectively on a single GPU. Training Rnn-Seq_to_Seq on CNN dataset took 12 hours and 15 hours on PubMed (i.e., training Rnn-Seq_to_Seq took about 1.5 times as much time as training Avg-Seq_to_Seq). Moreover, our models trained using RAML took much time compared to when we train them based on MLE.

5. DISCUSSION

On the CNN dataset, our models generate summaries with sentences from the top of the document rather than from other parts. This means that they are severally hindered by lead bias. We think the lead bias problem is caused by the fact that in news documents most important information is mostly at the beginning of the document, and the details come after. Our RNN-based sequence-to-sequence extractor eagerly learns the position features and heavily relied on them. Shuffling sentences in documents reduces the lead bias however, the overall performance of the models drops; without position, our models are not capable to identify important sentences in news domain. Additionally, there is a drop in the performance of our models on the PubMed dataset. The reason of this, is that the PubMed dataset contains long documents. Our models were not able to learn better representation for these long documents. Graph-based neural network approach and incorporating semantic units such as latent topics, entities and queries can improve extractive summarizer on long documents. We leave this for our future work. Results shown in Table 3 show that our model trained with RAML has potential because it seems to perform better on out-of-domain datasets compared to the model trained with MLE. However, more work is needed to fine-tune hyper-parameter that controls the smoothness of the best distribution around correct labels.

6. CONCLUSIONS

In this paper, we develop a recurrent neural network-based extractive text summarization model and investigate two kinds of hierarchical network structures, to see the effect of different model architectures on the performance of the model. Our experimental results on two datasets from different domains show that our model attains results that are comparable to other deep learning-based state-of-the-art extractive models as well as the state-of-the-art models that use manually engineered features. By comparing the two different approaches of sentence encoders, the performance of Avg-Seq_to_Seq architecture is slightly better than that of Rnn-Seq_to_Seq architecture. We adopt the RAML approach to the task of extractive summarization expecting that it can improve the performance and generalization of our models. Though the RAML approach does not improve the performance of our models over the MLE approach, it poses a potential behavior of improving the generalization ability of the models on out-of-domain datasets. In the future, we plan to investigate more on how different learning criteria used to train neural summarization models influence the generalization and performance of the model. Also, we want to join the extractive approach and abstractive approach to build a model which can generate abstractive summaries. Moreover, we plan to explore more on how different deep learning architectures such as CNNs, RNNs, and transformers influence the performance of other different NLP tasks.

ACKNOWLEDGMENTS

We would like to thank two anonymous reviewers for their helpful comments on various aspects of this work. The work was supported in part by the National Natural Science Foundation of China under grant number 61872186 and Science and Technology on Information System Engineering Laboratory (No.05201901).

REFERENCES

- [1] R. Nallapati, F. Zhai, and B. Zhou, "SummaRuNNer: A recurrent neural network based sequence model for extractive summarization of documents," 31st AAAI Conf. Artif. Intell. AAAI 2017, pp. 3075-3081, 2017.

- [2] M. Isonuma, T. Fujino, J. Mori, Y. Matsuo, and I. Sakata, "Extractive summarization using multi-task learning with document classification," EMNLP 2017 - Conf. Empir. Methods Nat. Lang. Process. Proc., pp. 2101-2110, 2017, doi: 10.18653/v1/d17-1223.
- [3] J. Carbonell and J. Goldstein, "Use of MMR, diversity-based reranking for reordering documents and producing summaries," SIGIR Forum (ACM Spec. Interes. Gr. Inf. Retrieval), pp. 335-336, 1998, doi: 10.1145/290941.291025.
- [4] D. R. Radev and G. Erkan, "LexRank : Graph-based Centrality as Salience in Text Summarization," J. Artif. Intell. Res., vol. 22, no. 1, pp. 457-479, 2004, [Online]. Available: <https://arxiv.org/abs/1109.2128>.
- [5] J. M. Conroy and D. P. O'leary, "Text summarization via hidden Markov models," SIGIR Forum (ACM Spec. Interes. Gr. Inf. Retrieval), no. August 2014, pp. 406-407, 2001, doi: 10.1145/383952.384042.
- [6] R. McDonald, "A study of global inference algorithms in multi-document summarization," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 4425 LNCS, pp. 557-564, 2007, doi: 10.1007/978-3-540-71496-5_51.
- [7] J. Cheng and M. Lapata, "Neural summarization by extracting sentences and words," 54th Annu. Meet. Assoc. Comput. Linguist. ACL 2016 - Long Pap., vol. 1, pp. 484-494, 2016, doi: 10.18653/v1/p16-1046.
- [8] W. Yin and Y. Pei, "Optimizing sentence modeling and selection for document summarization," IJCAI International Joint Conference on Artificial Intelligence, vol. 2015-Janua. pp. 1383-1389, 2015.
- [9] Z. Cao, F. Wei, S. Li, W. Li, M. Zhou, and H. Wang, "Learning summary prior representation for extractive summarization," ACL-IJCNLP 2015 - 53rd Annu. Meet. Assoc. Comput. Linguist. 7th Int. Jt. Conf. Nat. Lang. Process. Asian Fed. Nat. Lang. Process. Proc. Conf., vol. 2, pp. 829-833, 2015, doi: 10.3115/v1/p15-2136.
- [10] S. Narayan, S. B. Cohen, and M. Lapata, "Ranking sentences for extractive summarization with reinforcement learning," arXiv, pp. 1747-1759, 2018.
- [11] Y. Wu and B. Hu, "Learning to extract coherent summary via deep reinforcement learning," 32nd AAAI Conf. Artif. Intell. AAAI 2018, pp. 5602-5609, 2018.
- [12] M. Zhong, P. Liu, D. Wang, X. Qiu, and X. Huang, "Searching for effective neural extractive summarization: What works and what's next," arXiv, pp. 1049-1058, 2019.
- [13] Q. Zhou, N. Yang, F. Wei, S. Huang, M. Zhou, and T. Zhao, "Neural document summarization by jointly learning to score and select sentences," arXiv, pp. 654-663, 2018.
- [14] F. Mohsen, J. Wang, and K. Al-Sabahi, "A hierarchical self-attentive neural extractive summarizer via reinforcement learning (HSASRL)," Appl. Intell., vol. 50, no. 9, pp. 2633-2646, 2020, doi: 10.1007/s10489-020-01669-5.
- [15] K. Al-Sabahi, Z. Zuping, and M. Nadher, "A hierarchical structured self-attentive model for extractive document summarization (HSSAS)," arXiv, pp. 1-8, 2018.
- [16] D. S. Mohammad Norouzi, Samy Bengio, Zhifeng Chen, Navdeep Jaitly, Mike Schuster, Yonghui Wu, "Reward Augmented Maximum Likelihood for Neural Structured Prediction," arXiv, no. ML, 2017.
- [17] D. R. Radev, H. Jing, M. Styś, and D. Tam, "Centroid-based summarization of multiple documents," Inf. Process. Manag., vol. 40, no. 6, pp. 919-938, 2004, doi: 10.1016/j.ipm.2003.10.006.
- [18] M. Afsharizadeh, H. Ebrahimpour-Komleh, and A. Bagheri, "Query-oriented text summarization using sentence extraction technique," 2018 4th Int. Conf. Web Res. ICWR 2018, pp. 128-132, 2018, doi: 10.1109/ICWR.2018.8387248.
- [19] P. T. Rada Mihalcea, "TextRank: Bringing Order into Texts," Proc. 2004 Conf. Empir. methods Nat. Lang. Process., 2004, doi: 10.1016/0305-0491(73)90144-2.
- [20] M. Fuentes, E. Alfonseca, and H. Rodríguez, "Support vector machines for query-focused summarization trained and evaluated on pyramid data," no. June, p. 57, 2007, doi: 10.3115/1557769.1557788.
- [21] J. Kupiec, J. Pedersen, and F. Chen, "Trainable document summarizer," SIGIR Forum (ACM Spec. Interes. Gr. Inf. Retrieval), pp. 68-73, 1995, doi: 10.1145/215206.215333.
- [22] M. Ranzato, S. Chopra, M. Auli, and W. Zaremba, "Sequence level training with recurrent neural networks," 4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc., pp. 1-16, 2016.
- [23] K. Yao, L. Zhang, T. Luo, and Y. Wu, "Deep reinforcement learning for extractive document summarization," Neurocomputing, vol. 284, pp. 52-62, 2018, doi: 10.1016/j.neucom.2018.01.020.

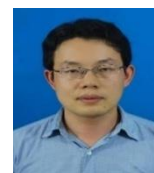
- [24] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," EMNLP 2014 - 2014 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf., pp. 1724-1734, 2014, doi: 10.3115/v1/d14-1179.
- [25] A. See, P. J. Liu, and C. D. Manning, "Get To The Point: Summarization with Pointer-Generator Networks," arXiv, 2017.
- [26] D. Bahdanau, K. H. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1-15, 2015.
- [27] K. M. Hermann et al., "Teaching machines to read and comprehend," Adv. Neural Inf. Process. Syst., vol. 2015-Janua, pp. 1693-1701, 2015.
- [28] R. Nallapati and B. Xiang, "Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond Cicero dos Santos," pp. 280-290, 2016.
- [29] A. Cohan et al., "A discourse-aware attention model for abstractive summarization of long documents," NAACL HLT 2018 - 2018 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf., vol. 2, pp. 615-621, 2018, doi: 10.18653/v1/n18-2097.
- [30] Y. B. Xavier Glorot, "Understanding the difficulty of training deep feedforward neural networks Xavier," AISTATS, 2010, doi: 10.1109/LGRS.2016.2565705.
- [31] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1-15, 2015.
- [32] C.-Y. Lin and E. Hovy, "Automatic evaluation of summaries using N-gram co-occurrence statistics," no. June, pp. 71-78, 2003, doi: 10.3115/1073445.1073465.
- [33] J. Steinberger and K. Ježek, "Using Latent Semantic Analysis in text summarization and summary evaluation," Proc. ISIM, pp. 93-100, 2004.
- [34] L. Vanderwende, H. Suzuki, C. Brockett, and A. Nenkova, "Beyond SumBasic: Task-focused summarization with sentence simplification and lexical expansion," Information Processing and Management, vol. 43, no. 6, pp. 1606-1618, 2007, doi: 10.1016/j.ipm.2007.01.023.
- [35] G. Erkan and D. R. Radev, "LexRank: Graph-based lexical centrality as salience in text summarization," J. Artif. Intell. Res., vol. 22, pp. 457-479, 2004, doi: 10.1613/jair.1523.
- [36] Y. Dong, Y. Shen, E. Crawford, H. van Hoof, and J. C. K. Cheung, "Banditsum: Extractive summarization as a contextual bandit," Proc. 2018 Conf. Empir. Methods Nat. Lang. Process. EMNLP 2018, pp. 3739-3748, 2020, doi: 10.18653/v1/d18-1409.
- [37] K. Hong and A. Nenkova, "Improving the estimation of word importance for news multi-document summarization," 14th Conference of the European Chapter of the Association for Computational Linguistics 2014, EACL 2014, pp. 712-721, 2014, doi: 10.3115/v1/e14-1075.
- [38] W. Xiao and G. Carenini, "Extractive summarization of long documents by combining global and local context," EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf., pp. 3011-3021, 2020, doi: 10.18653/v1/d19-1298.

AUTHORS

Shimirwa Aline Valerie received her Bachelor's degree in Software Engineering from Nanjing University of Science and Technology in 2019. Currently, she is working toward a Master's degree in the Department of Computer Science at Nanjing University of Science and Technology. Her research interests include natural language processing, machine learning, data mining, and deep learning.



Jian Xu received a Ph.D. in Computer Science in 2007 from Nanjing University of Science and Technology, Nanjing, China. Now he holds the position of a professor at Nanjing University of Science and Technology. His research interests are event mining, log mining, and their applications to system management. He has published about 30 papers in journals and refereed conference proceedings in those areas.



THE DIFFERENCE OF MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

Yew Kee Wong

School of Information Engineering, HuangHuai University, Henan, China

ABSTRACT

In the information era, enormous amounts of data have become available on hand to decision makers. Big data refers to datasets that are not only big, but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques. Due to the rapid growth of such data, solutions need to be studied and provided in order to handle and extract value and knowledge from these datasets. Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. Such minimal human intervention can be provided using big data analytics, which is the application of advanced analytics techniques on big data. This paper aims to analyse some of the different machine learning algorithms and methods which can be applied to big data analysis, as well as the opportunities provided by the application of big data analytics in various decision making domains.

KEYWORDS

Artificial Intelligence, Machine Learning, Big Data Analysis.

1. INTRODUCTION

Resurging interest in machine learning is due to the same factors that have made data mining and Bayesian analysis more popular than ever. Things like growing volumes and varieties of available data, computational processing that is cheaper and more powerful, and affordable data storage. All of these things mean it's possible to quickly and automatically produce models that can analyse bigger, more complex data and deliver faster, more accurate results – even on a very large scale. And by building precise models, an organization has a better chance of identifying profitable opportunities – or avoiding unknown risks [1].

Because of new computing technologies, machine learning today is not like machine learning of the past. It was born from pattern recognition and the theory that computers can learn without being programmed to perform specific tasks; researchers interested in artificial intelligence wanted to see if computers could learn from data. The iterative aspect of machine learning is important because as models are exposed to new data, they are able to independently adapt. They learn from previous computations to produce reliable, repeatable decisions and results. It's a science that's not new – but one that has gained fresh momentum. While many machine learning algorithms have been around for a long time, the ability to automatically apply complex mathematical calculations to big data - over and over, faster and faster – is a recent development [2]. This paper will look at some of the different machine learning algorithms and methods which can be applied to big data analysis, as well as the opportunities provided by the application of big data analytics in various decision making domains.

2. HOW MACHINE LEARNING WORKS

Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy [3].

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, uncovering key insights within data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics [4]. As big data continues to expand and grow, the market demand for data scientists will increase, requiring them to assist in the identification of the most relevant business questions and subsequently the data to answer them.

2.1. Machine Learning Algorithms

Machine learning algorithms can be categorized into three main parts:

1. A Decision Process:

In general, machine learning algorithms are used to make a prediction or classification. Based on some input data, which can be labelled or unlabelled, your algorithm will produce an estimate about a pattern in the data.

2. An Error Function:

An error function serves to evaluate the prediction of the model. If there are known examples, an error function can make a comparison to assess the accuracy of the model.

3. A Model Optimization Process:

If the model can fit better to the data points in the training set, then weights are adjusted to reduce the discrepancy between the known example and the model estimate. The algorithm will repeat this evaluate and optimize process, updating weights autonomously until a threshold of accuracy has been met.

2.2. Types of Machine Learning Methods

Machine learning classifiers fall into three primary categories [5]:

Supervised machine learning

Supervised learning also known as supervised machine learning, is defined by its use of labelled datasets to train algorithms that to classify data or predict outcomes accurately. As input data is fed into the model, it adjusts its weights until the model has been fitted appropriately. This occurs as part of the cross validation process to ensure that the model avoids over fitting or under-fitting. Supervised learning helps organizations solve for a variety of real-world problems at scale, such as classifying spam in a separate folder from your inbox. Some methods used in supervised learning include neural networks, naïve bayes, linear regression, logistic regression, random forest, support vector machine (SVM), and more.

Unsupervised machine learning

Unsupervised learning, also known as unsupervised machine learning, uses machine learning algorithms to analyse and cluster unlabelled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention. Its ability to discover similarities and differences in information make it the ideal solution for exploratory data analysis, cross-selling strategies, customer segmentation, image and pattern recognition. It's also used to reduce the number of features in a model through the process of dimensionality reduction; principal component analysis (PCA) and singular value decomposition (SVD) are two common approaches for this. Other algorithms used in unsupervised learning include neural networks, k-means clustering, probabilistic clustering methods, and more [6].

Semi-supervised learning

Semi-supervised learning offers a happy medium between supervised and unsupervised learning. During training, it uses a smaller labelled data set to guide classification and feature extraction from a larger, unlabelled data set. Semi-supervised learning can solve the problem of having not enough labelled data (or not being able to afford to label enough data) to train a supervised learning algorithm.

2.3. Practical Use of Machine Learning

Here are just a few examples of machine learning you might encounter every day [7]:

Speech Recognition: It is also known as automatic speech recognition (ASR), computer speech recognition, or speech-to-text, and it is a capability which uses natural language processing (NLP) to process human speech into a written format. Many mobile devices incorporate speech recognition into their systems to conduct voice search—e.g. Siri—or provide more accessibility around texting.

Customer Service: Online chatbots are replacing human agents along the customer journey. They answer frequently asked questions (FAQs) around topics, like shipping, or provide personalized advice, cross-selling products or suggesting sizes for users, changing the way we think about customer engagement across websites and social media platforms. Examples include messaging bots on e-commerce sites with virtual agents, messaging apps, such as Slack and Facebook Messenger, and tasks usually done by virtual assistants and voice assistants.

Computer Vision: This AI technology enables computers and systems to derive meaningful information from digital images, videos and other visual inputs, and based on those inputs, it can take action. This ability to provide recommendations distinguishes it from image recognition tasks. Powered by convolutional neural networks, computer vision has applications within photo tagging in social media, radiology imaging in healthcare, and self-driving cars within the automotive industry.

Recommendation Engines: Using past consumption behaviour data, AI algorithms can help to discover data trends that can be used to develop more effective cross-selling strategies. This is used to make relevant add-on recommendations to customers during the checkout process for online retailers.

Automated stock trading: Designed to optimize stock portfolios, AI-driven high-frequency trading platforms make thousands or even millions of trades per day without human intervention.

3. WHAT IS DEEP LEARNING

Deep learning is one of the foundations of artificial intelligence (AI), and the current interest in deep learning is due in part to the buzz surrounding AI. Deep learning techniques have improved the ability to classify, recognize, detect and describe – in one word, understand [8]. For example, deep learning is used to classify images, recognize speech, detect objects and describe content.

Several developments are now advancing deep learning:

- Algorithmic improvements have boosted the performance of deep learning methods.
- New machine learning approaches have improved accuracy of models.
- New classes of neural networks have been developed that fit well for applications like text translation and image classification.
- We have a lot more data available to build neural networks with many deep layers, including streaming data from the Internet of Things, textual data from social media, physicians notes and investigative transcripts.
- Computational advances of distributed cloud computing and graphics processing units have put incredible computing power at our disposal. This level of computing power is necessary to train deep algorithms.

At the same time, human-to-machine interfaces have evolved greatly as well. The mouse and the keyboard are being replaced with gesture, swipe, touch and natural language, ushering in a renewed interest in AI and deep learning [9].

This paper will look at some of the different deep learning algorithms and methods which can be applied to AI analysis, as well as the opportunities provided by the application in various decision making domains.

3.1. How Deep Learning Works

Deep learning changes how you think about representing the problems that you're solving with analytics. It moves from telling the computer how to solve a problem to training the computer to solve the problem itself.

A traditional approach to analytics is to use the data at hand to engineer features to derive new variables, then select an analytic model and finally estimate the parameters (or the unknowns) of that model. These techniques can yield predictive systems that do not generalize well because completeness and correctness depend on the quality of the model and its features [10]. For example, if you develop a fraud model with feature engineering, you start with a set of variables, and you most likely derive a model from those variables using data transformations. You may end up with 30,000 variables that your model depends on, then you have to shape the model, figure out which variables are meaningful, which ones are not, and so on. Adding more data requires you to do it all over again.

The new approach with deep learning is to replace the formulation and specification of the model with hierarchical characterizations (or layers) that learn to recognize latent features of the data from the regularities in the layers. The paradigm shift with deep learning is a move from feature engineering to feature representation. The promise of deep learning is that it can lead to predictive systems that generalize well, adapt well, continuously improve as new data arrives, and are more dynamic than predictive systems built on hard business rules. You no longer fit a model. Instead, you train the task.

Deep learning is making a big impact across industries. In life sciences, deep learning can be used for advanced image analysis, research, drug discovery, prediction of health problems and disease symptoms, and the acceleration of insights from genomic sequencing. In transportation, it can help autonomous vehicles adapt to changing conditions [11]. It is also used to protect critical infrastructure and speed response.

Most deep learning methods use neural networks architectures, which is why deep learning models are often referred to as deep neural networks. The term “deep” usually refers to the number of hidden layers in the neural network. Traditional neural networks only contain 2-3 hidden layers, while deep networks can have as many as 150. Deep learning models are trained by using large sets of labelled data and neural network architectures that learn features directly from the data without the need for manual feature extraction.

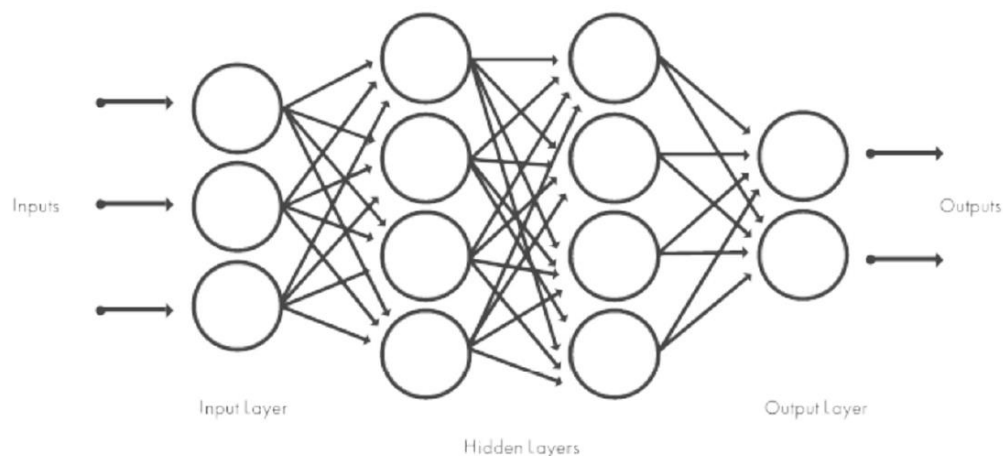


Figure 1: Neural networks, which are organized in layers consisting of a set of interconnected nodes. Networks can have tens or hundreds of hidden layers.

3.2. How Deep Learning Being Used

To the outside eye, deep learning may appear to be in a research phase as computer science researchers and data scientists continue to test its capabilities. However, deep learning has many practical applications that businesses are using today, and many more that will be used as research continues [12]. Popular uses today include:

Speech Recognition

Both the business and academic worlds have embraced deep learning for speech recognition. Xbox, Skype, Google Now and Apple's Siri, to name a few, are already employing deep learning technologies in their systems to recognize human speech and voice patterns.

Natural Language Processing

Neural networks, a central component of deep learning, have been used to process and analyse written text for many years. A specialization of text mining, this technique can be used to discover patterns in customer complaints, physician notes or news reports, to name a few.

Image Recognition

One practical application of image recognition is automatic image captioning and scene description. This could be crucial in law enforcement investigations for identifying criminal activity in thousands of photos submitted by bystanders in a crowded area where a crime has occurred. Self-driving cars will also benefit from image recognition through the use of 360-degree camera technology.

Recommendation Systems

Amazon and Netflix have popularized the notion of a recommendation system with a good chance of knowing what you might be interested in next, based on past behaviour. Deep learning can be used to enhance recommendations in complex environments such as music interests or clothing preferences across multiple platforms.

Recent advances in deep learning have improved to the point where deep learning outperforms humans in some tasks like classifying objects in images [13]. While deep learning was first theorized in the 1980s, there are two main reasons it has only recently become useful:

1. Deep learning requires large amounts of labelled data. For example, driverless car development requires millions of images and thousands of hours of video.
2. Deep learning requires substantial computing power. High-performance GPUs have a parallel architecture that is efficient for deep learning. When combined with clusters or cloud computing, this enables development teams to reduce training time for a deep learning network from weeks to hours or less.

When choosing between machine learning and deep learning, consider whether you have a high-performance GPU and lots of labelled data. If you don't have either of those things, it may make more sense to use machine learning instead of deep learning. Deep learning is generally more complex, so you'll need at least a few thousand images to get reliable results. Having a high-performance GPU means the model will take less time to analyse all those images [14].

3.3. Deep Learning Opportunities and Applications

A lot of computational power is needed to solve deep learning problems because of the iterativenature of deep learning algorithms, their complexity as the number of layers increase, and the large volumes of data needed to train the networks.

The dynamic nature of deep learning methods – their ability to continuously improve and adapt to changes in the underlying information pattern – presents a great opportunity to introduce more dynamic behaviour into analytics [15]. Greater personalization of customer analytics is one possibility. Another great opportunity is to improve accuracy and performance in applications where neural networks have been used for a long time. Through better algorithms and more computing power, we can add greater depth.

While the current market focus of deep learning techniques is in applications of cognitive computing, there is also great potential in more traditional analytics applications, for example, time series analysis. Another opportunity is to simply be more efficient and streamlined in existing analytical operations. Recently, some study showed that with deep neural networks in speech-to-text transcription problems [16]. Compared to the standard techniques, the word-error-rate decreased by more than 10 percent when deep neural networks were applied. They also

eliminated about 10 steps of data preprocessing, feature engineering and modelling. The impressive performance gains and the time savings when compared to feature engineering signify a paradigm shift.

Here are some examples of deep learning applications are used in different industries:

Automated Driving:

Automotive researchers are using deep learning to automatically detect objects such as stop signs and traffic lights. In addition, deep learning is used to detect pedestrians, which helps decrease accidents.

Aerospace and Defence:

Deep learning is used to identify objects from satellites that locate areas of interest, and identify safe or unsafe zones for troops.

Medical Research:

Cancer researchers are using deep learning to automatically detect cancer cells. Teams at UCLA built an advanced microscope that yields a high-dimensional data set used to train a deep learning application to accurately identify cancer cells [17].

Industrial Automation:

Deep learning is helping to improve worker safety around heavy machinery by automatically detecting when people or objects are within an unsafe distance of machines.

Electronics:

Deep learning is being used in automated hearing and speech translation. For example, home assistance devices that respond to your voice and know your preferences are powered by deep learning applications.

3.4. How to Create and Train Deep Learning Models

The three most common ways people use deep learning to perform object classification are:

Training from Scratch

To train a deep network from scratch, you gather a very large labelled data set and design a network architecture that will learn the features and model. This is good for new applications, or applications that will have a large number of output categories. This is a less common approach because with the large amount of data and rate of learning, these networks typically take days or weeks to train [18].

Transfer Learning

Most deep learning applications use the transfer learning approach, a process that involves fine-tuning a pre-trained model. User can start with an existing network, such as AlexNet or GoogLeNet, and feed in new data containing previously unknown classes [19]. After making some tweaks to the network, user can now perform a new task, such as categorizing only dogs or

cats instead of 10,000 different objects. This also has the advantage of needing much less data (processing thousands of images, rather than millions), so computation time drops to minutes or hours.

Feature Extraction

A slightly less common, more specialized approach to deep learning is to use the network as a feature extractor. Since all the layers are tasked with learning certain features from images, user can pull these features out of the network at any time during the training process [20]. These features can then be used as input to a machine learning model such as support vector machines (SVM).

4. CONCLUSIONS

So this study was concerned by understanding the interrelation between machine learning and big data analysis, what frameworks and systems that worked, and how machine learning can impact the big data analytic process whether by introducing new innovations that foster advanced machine learning process and escalating power consumption, security issues and replacing human in workplaces. The advanced big data analytics and machine learning algorithms with various applications show promising results in artificial intelligence development and further evaluation and research using machine learning are in progress.

REFERENCES

- [1] Shi, Z., (2019). Cognitive Machine Learning. *International Journal of Intelligence Science*, 9, pp. 111-121.
- [2] Lake, B.M., Salakhutdinov, R. and Tenenbaum, J.B., (2015). Human-Level Concept Learning through Probabilistic Program Induction. *Science*, 350, pp. 1332-1338.
- [3] Silver, D., Huang, A., Maddison, C.J., et al., (2016). Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature*, 529, pp. 484-489.
- [4] Fukushima, K., (1980). Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*, 36, pp. 193-202.
- [5] Lecun, Y., Bottou, L., Orr, G.B., et al., (1998). Efficient Backprop. *Neural Networks Tricks of the Trade*, 1524, pp. 9-50.
- [6] McClelland, J.L., et al., (1995). Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and Failures of Connectionist Models of Learning and Memory. *Psychological Review*, 102, pp. 419-457.
- [7] Kumaran, D., Hassabis, D. and McClelland, J.L., (2016). What Learning Systems Do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences*, 20, pp. 512-534.
- [8] S. Del. Rio, V. Lopez, J. M. Bentez and F. Herrera, (2014). On the use of mapreduce for imbalanced big data using random forest, *Information Sciences*, 285, pp. 112-137.
- [9] MH. Kuo, T. Sahama, A. W. Kushniruk, E. M. Borycki and D. K. Grunwell, (2014). Health big data analytics: current perspectives, challenges and potential solutions, *International Journal of Big Data Intelligence*, 1, pp. 114-126.
- [10] R. Nambiar, A. Sethi, R. Bhardwaj and R. Vargheese, (2013). A look at challenges and opportunities of big data analytics in healthcare, *IEEE International Conference on Big Data*, pp. 17-22.
- [11] Z. Huang, (1997). A fast clustering algorithm to cluster very large categorical data sets in data mining, *SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*.
- [12] M. D. Assuno, R. N. Calheiros, S. Bianchi, M. a. S. Netto and R. Buyya, (2015). Big data computing and clouds: Trends and future directions, *Journal of Parallel and Distributed Computing*, 79, pp. 3-15.
- [13] I. A. T. Hashem, I. Yaqoob, N. Badrul Anuar, S. Mokhtar, A. Gani and S. Ullah Khan, (2014). The rise of big data on cloud computing: Review and open research issues, *Information Systems*, 47, pp.

- 98-115.
- [14] L. Wang and J. Shen, (2013). Bioinspired cost-effective access to big data, International Symposium for Next Generation Infrastructure, pp. 1-7.
 - [15] Mouha, R., (2021). Internet of Things (IoT). Journal of Data Analysis and Information Processing, 9, pp. 77-101.
 - [16] Manyika, et al., (2015). The Internet of Things: Mapping the Value Beyond the Hype. Mckinsey Global Institute, San Francisco.
 - [17] Bradbury, D., (2015). How Can Privacy Survive in the Era of the Internet of Things? The Guardian, April 7, Sec. Technology.
 - [18] Gaona-Garcia, P., Montenegro-Marin, C.E., Prieto, J.D., Nieto, Y.V., (2017). Analysis of Security Mechanisms Based on Clusters IoT Environments. International Journal of Interactive Multimedia and Artificial Intelligence, 4, pp. 55-60.
 - [19] Alavi, A.H., Jiao, P., Buttlar, W.G. and Lajnef, N., (2018). Internet of Things-Enabled Smart Cities: State-of-the-Art and Future Trends. Measurement, 129, pp. 589-606.
 - [20] Zanella, A., Bui, N., Castellani, A., Vangelista, L. and Zorgi, M., (2014). Internet of Things for Smart Cities. IEEE Internet of Things Journal, 1, pp. 22-32.

AUTHOR

Prof. Yew Kee Wong (Eric) is a Professor of Artificial Intelligence (AI) & Advanced Learning Technology at the HuangHuai University in Henan, China. He obtained his BSc (Hons) undergraduate degree in Computing Systems and a Ph.D. in AI from The Nottingham Trent University in Nottingham, U.K. He was the Senior Programme Director at The University of Hong Kong (HKU) from 2001 to 2016. Prior to joining the education sector, he has worked in international technology companies, Hewlett-Packard (HP) and Unisys as an AI consultant. His research interests include AI, online learning, big data analytics, machine learning, Internet of Things (IOT) and blockchain technology.



UNDERSTANDING THE FEATURES OF INTERNET OF THINGS (IoT) AND BIG DATA ANALYSIS

Yew Kee Wong

School of Information Engineering,
Huang Huai University, Henan, China

ABSTRACT

In the information era, enormous amounts of data have become available on hand to decision makers. Big data refers to datasets that are not only big, but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques. Due to the rapid growth of such data, solutions need to be studied and provided in order to handle and extract value and knowledge from these datasets. The Internet of Things, or "IoT" for short, is about extending the power of the internet beyond computers and smartphones to a whole range of other things, processes and environments. IoT is at the epicentre of the Digital Transformation Revolution that is changing the shape of business, enterprise and people's lives. This transformation influences everything from how we manage and operate our homes to automating processes across nearly all industries. This paper aims to analyse the relationships of AI, big data and IoT, as well as the opportunities provided by the applications in various operational domains.

KEYWORDS

Artificial Intelligence, Big Data, IoT, Digital Transformation Revolution, Machine Learning.

1. INTRODUCTION

When something is connected to the internet, which means that it can send information or receive information, or both. This ability to send and/or receive information makes things smart, and smarter is better. To be smart, a thing doesn't need to have super storage or a supercomputer inside of it. All a thing has to do is *connect* to super storage or to a supercomputer. Being connected is awesome. Connecting things to the internet yields many amazing benefits. We've all seen these benefits with our smartphones, laptops, and tablets, but this is true for everything else too. And yes, I do mean everything. The Internet of Things (IoT) means taking all the things in the world and connecting them to the internet.

With the development and increase of apps and social media and people and businesses moving online using IoT, there's been a huge increase in data. If we look at only social media platforms, they interest and attract over a million users daily, scaling up data more than ever before. The next question is how exactly is this huge amount of data handled and how is it processed and stored. This is where AI, big data and IoT all 3 components come into play.

2. BENEFITS OF ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) is a branch of computer science. AI technologies aim to reproduce or surpass abilities in computational systems that are generally deemed intelligent if performed by a human [1]. These abilities include:

- learning
- reasoning
- pattern-recognition
- problem-solving
- visual perception
- language-understanding

2.1. Different Types of AI

There are two main types of AI [2]:

Applied AI: is more common and includes systems designed to intelligently carry out a single task, eg move a driverless vehicle, or trade stocks and shares. This category is also known as 'weak' or 'narrow' AI.

Generalised AI: is less common and includes systems or devices that can theoretically handle any task, as they carry enough intelligence to find solutions to unfamiliar problems. Generalised AI is also known as 'strong' AI. Examples of true strong AI don't currently exist, as these technologies are still in very early stages of development.

Many modern AI applications are enabled through a sub-field of AI known as 'machine learning'. What is machine learning? The roots of machine learning (ML) are in statistics. ML uses algorithms and statistical models to perform a specific task without using explicit instructions, instead relying on patterns and inference [3]. For example, ML applications can:

- read a text and decide if the author is making a complaint or a purchase order
- listen to a piece of music and find other tunes to match the mood
- recognise images and classify them according to the elements they contain
- translate large volumes of text in real time
- accurately recognise faces, speech and objects

2.2. How are AI and Machine Learning used in Business

Over the years, AI research has enabled many technological advances [4], including:

- virtual agents and chatbots
- suggestive web searches
- targeted advertising
- pattern recognition
- predictive analytics
- voice and speech recognition
- face recognition
- machine translation
- autonomous driving
- automatic scheduling

Many of these are now commonplace and provide solutions to a great number of business challenges and complex, real-world problems.

2.3. How are Businesses using AI

AI is steadily passing into everyday business use. From workflow management to trend predictions, AI has many different uses in business [5]. It also provides new business opportunities.

Application of AI in business:

- **Improve customer services** – e.g. use virtual assistant programs to provide real-time support to users (for example, with billing and other tasks).
- **Automate workloads** – e.g. collect and analyse data from smart sensors, or use machine learning (ML) algorithms to categorise work, automatically route service requests, etc.
- **Optimise logistics** – e.g. use AI-powered image recognition tools to monitor and optimise your infrastructure, plan transport routes, etc.
- **Increase manufacturing output and efficiency** – e.g. automate production line by integrating industrial robots into your workflow and teaching them to perform labour-intensive or mundane tasks [6].
- **Prevent outages** – e.g. use anomaly detection techniques to identify patterns that are likely to disrupt your business, such as an IT outage. Specific AI software may also help you to detect and deter security intrusions.
- **Predict performance** – e.g. use AI applications to determine when you might reach performance goals, such as response time to help desk calls.
- **Predict behaviour** – e.g. use ML algorithms to analyse patterns of online behaviour to, for example, serve tailored product offers, detect credit card fraud or target appropriate adverts.
- **Manage and analyse your data** – e.g. AI can help you interpret and mine your data more efficiently than ever before and provide meaningful insight into your assets, your brand, staff or customers [7].
- **Improve your marketing and advertising** – e.g. effectively track user behaviour and automate many routine marketing tasks.

3. WHAT IS BIG DATA AND WHAT ARE ITS BENEFITS

Big data analytics has revolutionized the field of IT, enhancing and adding added advantage to organizations. It involves the use of analytics, new age tech like machine learning, mining, statistics and more. Big data can help organizations and teams to perform multiple operations on a single platform, store Tbs of data, pre-process it, analyse all the data, irrespective of the size and type, and visualize it too [8].

The sources of big data:

Black Box Data

This is the data generated by air planes, including jets and helicopters. Black box data includes flight crew voices, microphone recordings, and aircraft performance information.

Social Media Data

This is data developed by such social media sites as Twitter, Facebook, Instagram, Pinterest, and Google+.

Stock Exchange Data

This is data from stock exchanges about the share selling and buying decisions made by customers.

Power Grid Data

This is data from power grids. It holds information on particular nodes, such as usage information.

Transport Data

This includes possible capacity, vehicle model, availability, and distance covered by a vehicle.

Search Engine Data

This is one of the most significant sources of big data. Search engines have vast databases where they get their data.

The speed at which data is streamed, nowadays, is unprecedented, making it difficult to deal with it in a timely fashion. Smart metering, sensors, and RFID tags make it necessary to deal with data torrents in almost real-time. Most organizations are finding it difficult to react to data quickly. Not many years ago, having too much data was simply a storage issue [9]. However, with increased storage capacities and reduced storage costs are now focusing on how relevant data can create value.

There is a greater variety of data today than there was a few years ago. Data is broadly classified as structured data (relational data), semi-structured data (data in the form of XML sheets), and unstructured data (media logs and data in the form of PDF, Word, and Text files). Many companies have to grapple with governing, managing, and merging the different data varieties [10].

3.1. Advantages of Big Data

1. Today's consumer is very demanding. All customer wants to be treated as an individual and to be thanked after buying a product. With big data, supplier will get actionable data that they can use to engage with their customers one-on-one in real-time [11]. One way bigdata allows supplier to do this is that they will be able to check a complaining customer's profile in real-time and get info on the product(s) the customer is complaining about. Supplier will then be able to perform reputation management.
2. Big data allows supplier to re-develop the products/services they are selling. Information on what others think about their products, such as through unstructured social networking site text helps supplier in product development.
3. Big data allows supplier to test different variations of CAD (computer-aided design) images to determine how minor changes affect their process or product. This makes big data

invaluable in the manufacturing process.

4. Predictive analysis will keep supplier ahead of their competitors. Big data can facilitate this by, as an example, scanning and analysing social media feeds and newspaper reports. Big data also helps supplier do health-tests on their customers, suppliers, and other stakeholders to help supplier reduce risks such as default.
5. Big data is helpful in keeping data safe. Big data tools help supplier map the datalandscape of their company, which helps in the analysis of internal threats. As an example, supplier will know if their sensitive information has protection or not. A more specific example is that supplier will be able to flag the emailing or storage of 16 digit numbers (which could, potentially, be credit card numbers) [12].
6. Big data allows supplier to diversify their revenue streams. Analysing big data can give supplier trend-data that could help the supplier come up with a completely new revenue stream.
7. The supplier website needs to be dynamic if it is to compete favourably in the crowded online space. Analysis of big data helps supplier personalize the look/content and feel of their site to suit every visitor based on, for example, nationality and sex. An example of this is Amazon's IBCF (item-based collaborative filtering) that drives its "People you may know" and "Frequently bought together" features [13].
8. If the supplier is running a factory, big data is important because the supplier will not have to replace pieces of technology based on the number of months or years they have been in use. This is costly and impractical since different parts wear at different rates. Big data allows supplier to spot failing devices and will predict when the supplier should replace them.
9. Big data is important in the healthcare industry, which is one of the last few industries still stuck with a generalized, conventional approach. Big data allows a cancer patient to get medication that is developed based on his/her genes.

3.2. Challenges of Big Data

1. One of the issues with big data is the exponential growth of raw data. The data centres and databases store huge amounts of data, which is still rapidly growing. With the exponential growth of data, organizations often find it difficult to rightly store this data [14].
2. The next challenge is choosing the right big data tool. There are various big data tools, however choosing the wrong one can result in wasted effort, time and money too.
3. Next challenge of big data is securing it. Often organizations are too busy understanding and analysing the data, that they leave the data security for a later stage, and unprotected data ultimately becomes the breeding ground for the hackers.

4. WHAT IS INTERNET OF THINGS (IoT)

In the Internet of Things (IoT), all the things can be put into three categories [15]:

1. Sensors that collect information and then send it.

2. Computers that receive information and then act on it.
3. Things that do both.

And all three of these have enormous benefits that feed on each other:

Collecting and Sending Information

Sensors can measure temperature, motion, moisture, air quality, light, and almost anything else you can think of. Sensors, when paired with an internet connection, allow us to collect information from the environment which, in turn, helps make better decisions [16]. On a farm, automatically getting information about soil moisture can tell farmers exactly when crops need to be watered. Instead of watering too much or too little (either of which can lead to bad outcomes), the farmer can ensure that crops get exactly the right amount of water. Just as our senses allow us to collect information, sensors allow machines to make sense of their environments [17].

Receiving and Acting on Information

We're all very familiar with machines acting on input information. A printer receives a document and then prints it. A garage door receives a wireless signal and the door opens. It's commonplace to remotely command a machine to act. The real power of IoT arises when things can both collect information and act on it [18].

Doing Both

Let's go back to farming. The sensors collect information about the soil moisture. Now, the farmer could activate the irrigation system, or turn it off as appropriate. Instead, the irrigation system can automatically act as needed, based on how much moisture is detected. If the irrigation system receives information about the weather from its internet connection, it can also know when it's going to rain and decide not to water the crops when they'll be watered by the rain any ways [19].

4.1. How Does IoT Impact You

The new rule for the future is going to be, "Anything that can be connected, will be connected." But why on earth would a person want so many connected devices talking to each other? There are many examples for what this might look like or what the potential value might be. Say for example you are on your way to a meeting; your car could have access to your calendar and already know the best route to take. If the traffic is heavy your car might send a text to the other party notifying them that you will be late. What if your alarm clock wakes up you at 6 a.m. and then notifies your coffee maker to start brewing coffee for you? What if your office equipment knew when it was running low on supplies and automatically re-ordered more? What if the wearable device you used in the workplace could tell you when and where you were most active and productive and shared that information with other devices that you used while working? On a broader scale, the IoT can be applied to things like transportation networks: "smart cities" which can help the society to reduce waste and improve efficiency for things such as energy use; this helping the government to understand and improve how everyone work and live [20].

The reality is that the IoT allows for virtually endless opportunities and connections to take place, many of which we can't even think of or fully understand the impact of today. It's not hard to see how and why the IoT is such a hot topic today; it certainly opens the door to a lot of opportunities but also to many challenges. Security is a big issue that is often times brought up [21]. With billions of devices being connected together, what can people do to make sure that their

information stays secure? Will someone be able to hack into your toaster and thereby get access to your entire network? The IoT also opens up companies all over the world to more security threats. Then we have the issue of privacy and data sharing [22]. This is a hot-button topic even today, so one can only imagine how the conversation and concerns will escalate when we are talking about many billions of devices being connected. Another issue that many companies specifically are going to be faced with is around the massive amounts of data that all of these devices are going to produce. Companies need to figure out a way to store, track, analyse and make sense of the vast amounts of data that will be generated [23].

5. CONCLUSIONS

So this study was concerned by understanding the interrelation between AI, big data and IoT, what frameworks and systems that worked, and how AI can impact the big data analytic process whether by introducing new innovations that foster advanced IoT development process and escalating power consumption, security issues and replacing human in workplaces [24]. The advanced big data analytics and algorithms with various applications show promising results in artificial intelligence development and further evaluation and research using IoT are in progress.

REFERENCES

- [1] Kovach, D., (2017). The Computational Theory of Intelligence: Feedback. *International Journal of Modern Nonlinear Theory and Application*, 6, pp. 70-73.
- [2] J. Wang, (2013). "On the Limit of Machine Intelligence, " *International Journal of Intelligence Science*, Vol. 3 No. 4, pp. 170-175.
- [3] A. Turing, (1950). "Computing Machinery and Intelligence," *Mind*, Vol. 59, pp. 433-466.
- [4] M. Minsky, (1986). "The Society of Mind," Touchstone, Simon & Schuster, New York, p. 19.
- [5] A. Newell and H. Simon, (1958). "Heuristic Problem-Solving: The Next Advance in Operation Research," *Operations Research*, Vol. 6, No. 6.
- [6] A. Clark, (2001). "Mindware: In Introduction to the Philosophy of Cognitive Science," Oxford University Press, New York.
- [7] S. Russell and P. Norvig, (2010). "Artificial Intelligence—A Modern Approach," 3rd Edition, Prentice Hall, New Jersey.
- [8] S. Del Rio, V. Lopez, J. M. Bentez and F. Herrera, (2014). On the use of mapreduce for imbalanced big data using random forest, *Information Sciences*, 285, pp. 112-137.
- [9] MH. Kuo, T. Sahama, A. W. Kushniruk, E. M. Borycki and D. K. Grunwell, (2014). Health big data analytics: current perspectives, challenges and potential solutions, *International Journal of Big Data Intelligence*, 1, pp. 114-126.
- [10] R. Nambiar, A. Sethi, R. Bhardwaj and R. Vargheese, (2013). A look at challenges and opportunities of big data analytics in healthcare, *IEEE International Conference on Big Data*, pp.17-22.
- [11] Z. Huang, (1997). A fast clustering algorithm to cluster very large categorical data sets in data mining, *SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*.
- [12] M. D. Assuno, R. N. Calheiros, S. Bianchi, M. a. S. Netto and R. Buyya, (2015). Big data computing and clouds: Trends and future directions, *Journal of Parallel and Distributed Computing*, 79, pp. 3-15.
- [13] I. A. T. Hashem, I. Yaqoob, N. Badrul Anuar, S. Mokhtar, A. Gani and S. Ullah Khan, (2014). The rise of big data on cloud computing: Review and open research issues, *Information Systems*, 47, pp. 98-115.
- [14] L. Wang and J. Shen, (2013). Bioinspired cost-effective access to big data, *International Symposium for Next Generation Infrastructure*, pp. 1-7.
- [15] Mouha, R., (2021). Internet of Things (IoT). *Journal of Data Analysis and Information Processing*, 9, pp. 77-101.
- [16] Manyika, et al., (2015). The Internet of Things: Mapping the Value Beyond the Hype. McKinsey Global Institute, San Francisco.
- [17] Bradbury, D., (2015). How Can Privacy Survive in the Era of the Internet of Things? *The Guardian*, April 7, Sec. Technology.
- [18] Gaona-Garcia, P., Montenegro-Marin, C.E., Prieto, J.D., Nieto, Y.V., (2017). Analysis of Security

- Mechanisms Based on Clusters IoT Environments. *International Journal of Interactive Multimedia and Artificial Intelligence*, 4, pp. 55-60.
- [19] Alavi, A.H., Jiao, P., Buttler, W.G. and Lajnef, N., (2018). Internet of Things-Enabled Smart Cities: State-of-the-Art and Future Trends. *Measurement*, 129, pp. 589-606.
 - [20] Zanella, A., Bui, N., Castellani, A., Vangelista, L. and Zorzi, M., (2014). Internet of Things for Smart Cities. *IEEE Internet of Things Journal*, 1, pp. 22-32.
 - [21] Khajenasiri, I., Estebasari, A., Verhelst, M. and Gielen, G., (2017). A Review on Internet of Things for Intelligent Energy Control in Buildings for Smart City Applications. *Energy Procedia*, 111, pp. 770-779.
 - [22] Liu, T., Yuan, R. and Chang, H., (2012). Research on the Internet of Things in the Automotive Industry. *ICMeCG 2012*, Beijing, pp. 2-13.
 - [23] Yan, Z., Zhang, P. and Vasilakos, A.V., (2014). A Survey on Trust Management for Internet of Things. *Journal of Network and Computer Applications*, 42, pp. 120-134.
 - [24] Palattella, M.R., Dohler, M., Grieco, A., Rizzo, G., Torsner, J., Engel, T. and Ladid, L., (2016). Internet of Things in the 5G Era: Enablers, Architecture and Business Models. *IEEE Journal on Selected Areas in Communications*, 34, pp. 510-527.

AUTHOR

Prof. Yew Kee Wong (Eric) is a Professor of Artificial Intelligence (AI) & Advanced Learning Technology at the HuangHuai University in Henan, China. He obtained his BSc (Hons) undergraduate degree in Computing Systems and a Ph.D. in AI from The Nottingham Trent University in Nottingham, U.K. He was the Senior Programme Director at The University of Hong Kong (HKU) from 2001 to 2016. Prior to joining the education sector, he has worked in international technology companies, Hewlett-Packard (HP) and Unisys as an AI consultant. His research interests include AI, online learning, big data analytics, machine learning, Internet of Things (IOT) and blockchain technology.



AUTHOR INDEX

<i>Afef Saihi</i>	81
<i>Areej Alhothali</i>	139
<i>Asmaa Hakami</i>	139
<i>Batool Madani</i>	103
<i>Cem Ata Baykara</i>	119
<i>D. J. Pete</i>	189
<i>Danni Ai</i>	219
<i>David Noever</i>	63
<i>Dhinaharan Nagamalai</i>	174
<i>Esam Alzahrani</i>	01
<i>Hosna Ghandeharioun</i>	21
<i>Hussam Alshraideh</i>	81, 103
<i>İlgin Şafak</i>	119
<i>Imran N. Junejo</i>	09
<i>Jelena Vasiljević</i>	174
<i>Jian Xu</i>	233
<i>Jingfan Fan</i>	219
<i>Kübra Kalkan</i>	119
<i>Leon Jololian</i>	01
<i>Mahila Almutairi</i>	139
<i>Makoto Murakami</i>	203
<i>Min Guk I. Chi</i>	71
<i>Petar Prvulović</i>	174
<i>Philipp Bolte</i>	43
<i>Prudhvi Parne</i>	37
<i>Rachid Sabre</i>	149
<i>Raneem Alqarni</i>	139
<i>Rolf Morgenstern</i>	43
<i>Ruirui Kang</i>	219
<i>Samantha E. Miller Noever</i>	63
<i>Shimirwa Aline Valerie</i>	233
<i>Tianyu Fu</i>	219
<i>Ulf Witkowski</i>	43
<i>Vittalis Ayu</i>	93
<i>Vivek Ramakrishnan</i>	189
<i>Xiaohan Feng</i>	203
<i>Xinglong Zhu</i>	219
<i>Yew Kee Wong</i>	249, 259
<i>Yifan Wang</i>	219