

Computer Science & Information Technology

178

Artificial Intelligence and Applications

David C. Wyld,
Dhinaharan Nagamalai (Eds)

Computer Science & Information Technology

- 8th International Conference on Artificial Intelligence and Applications (AI 2022)
- 2nd International conference on AI, Machine Learning in Communications and Networks (AIMLNET 2022)
- 8th International Conference on Bioinformatics & Biosciences (BIOS 2022)
- 2nd International Conference on Big Data, IOT & NLP (BINLP 2022)
- 8th International Conference on Computer Science and Information Technology (CSTY 2022)
- 8th International Conference of Managing Value and Supply Chains (MaVaS 2022)
- 8th International Conference on Signal and Image Processing (SIGI 2022)

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: 978-1-925953-78-7

DOI: 10.5121/csit.2022.121801 - 10.5121/csit.2022.121825

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

8th International Conference on Artificial Intelligence and Applications (AI 2022), October 29~30, 2022, Vienna, Austria, 2nd International conference on AI, Machine Learning in Communications and Networks (AIMLNET 2022), 8th International Conference on Bioinformatics & Biosciences (BIOS 2022), 2nd International Conference on Big Data, IOT & NLP (BINLP 2022), 8th International Conference on Computer Science and Information Technology (CSTY 2022), 8th International Conference of Managing Value and Supply Chains (MaVaS 2022), 8th International Conference on Signal and Image Processing (SIGI 2022) was collocated with 8th International Conference on Artificial Intelligence and Applications (AI 2022). 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 AI 2022, AIMLNET 2022, BIOS 2022, BINLP 2022, CSTY 2022, MaVaS 2022 and SIGI 2022. 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, AI 2022, AIMLNET 2022, BIOS 2022, BINLP 2022, CSTY 2022, MaVaS 2022 and SIGI 2022 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 AI 2022, AIMLNET 2022, BIOS 2022, BINLP 2022, CSTY 2022, MaVaS 2022 and SIGI 2022

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,
Abdelmounaim Aggour,
Abdessamad Belangour,
Abdulhamit Subasi,
Abhishek Roy,
Adnan Salman,
Ahm Shamsuzzoha,
Ahmed Farouk AbdelGawad,
Ahmed Y. Nada,
Ajayeb Salama Abu-Daibes,
Ajit Kumar Singh,
Alba Córdoba-Cabús,
Ali Abdrhman Mohammed Ukasha,
Alireza Valipour Baboli,
Alla Anohina-Naumeca,
Alvaro Lopez-Martin,
Amal Azeroual,
Amari Farouk,
Amel Ourici,
Amine Khaldi,
Anand Nayyar,
Anirut Satiman,
Anouar Abtoy,
Apurv Asthana,
Archit Yajnik,
Aridj Mohamed,
Ashkan Ebadi,
Ashwin Viswanathan Kannan,
Atul Garg,
Aurel Vlaicu,
Avinash Waikar,
Ayad Ghany Ismaeel,
B. K. Tripathy,
Batuhan Ayhan,
Bejoy Abraham,
Bhagyashree SR,
Bhavtosh Rath,
Bilal Abu-Salih,
Bilal Alatas,
Bin Cao,
Binod Kumar Pattanayak,
Brahim Lejdel,
Carlos Guardado da Silva,
Chang-Wook Han,

Ain Shams University, Egypt
Hassan II University, Morocco
University Hassan II Casablanca, Morocco
University of Turku, Finland
Adamas University, India
An-najah National University, Palestine
University of Vaasa, Finland
Zagazig University, Egypt
Al - Quds University, Palestine
Emirates Collage of Technology, UAE
Patna Women's College, India
University of Malaga, Spain
Sebha University, Libya
Technical and Vocational University, Iran
Riga Technical University, Latvia
University of Malaga, Spain
Mohammed V University, Morocco
University of Sfax, Tunisia
University Badji Mokhtar Annaba, Algeria
University in Ouargla, Algeria
Duy Tan University, Viet Nam
Silpakorn University, Thailand
Abdelmalek Essaadi University, Morocco
University of Michigan, USA
Sikkim Manipal Institute of technology, India
Hassiba Benbouali University, Algeria
National Research Council Canada, Canada
Oklahoma State University, USA
Chitkara University, India
University of Arad, Romania
Southeastern Louisiana University, USA
Al-Kitab University College, Iraq
VIT, India
Marmara University, Turkey
College of Engineering Perumon, India
ATMECE, India
Target Corporation, USA
The University of Jordan, Jordan
Firat University, Turkey
Hebei University of Technology, China
Siksha O Anusandhan University, India
University of El-Oued, Algeria
Universidade de Lisboa, Portugal
Dong-Eui University, South Korea

Charles Kalinzi,	Kyambogo University, Uganda
Cheng Siong Chin,	Newcastle University, Singapore
Chew Fong Peng,	University of Malaya, Malaysia
Christian Mancas,	Ovidius University, Romania
Chung-Shun Feng,	Chaoyang University of Technology, Taiwan
Claudio Cuevas,	Universidade Federal de Pernambuco, Brazil
Daniela Lopez De Luise,	Universidad de Palermo, Argentina
Dario Ferreira,	University of Beira Interior, Portugal
Dariusz Jacek Jakobczak,	Koszalin University of Technology, Poland
Deepak Kumar,	University of Delhi, India
Denivaldo Lopes,	Federal University of Maranhao, Brazil
Dibya Mukhopadhyay,	Principal Data Scientist at Zuora, Inc., USA
Dmitry A. Zaitsev,	Odessa State Environmental University, Spain
Dmitry Korzun,	Petrozavodsk State University, Russia
Ebenezer Bonyah,	Kumasi polytechnic, Ghana
El Manaa Barhoumi,	Dhofar University, Oman
Ela Goyal,	SIES College of Management Studies, India
Elboukhari Mohamed,	University Mohamed First, Morocco
Elena Avksentieva,	ITMO University, Russia
Elzbieta Macioszek,	Silesian University of Technology, Poland
Emad Awada,	Applied Science University, Jordan
F. M. Javed Mehedi Shamrat,	Daffodil International University, Bangladesh
Faisal AbuRub,	University of Petra, Jordan
Faouzi Kamoun,	Zayed University, UAE
Faouzia Benabbou,	University Hassan II of Casablanca, Morocco
Farhad Soleimanian Gharehchopogh,	Hacettepe University, Turkey
Fariba Azizzadeh,	Islamic Azad University, Iran
Felix J. Garcia Clemente,	University of Murcia, Spain
Francesco Zirilli,	(retired) Sapienza Universita Roma, Italy
Gelayol Safavi,	Semnan University, Iran
Ghani Albaali,	Princess Sumaya University for Technology, Jordan
Ghasem Mirjalily,	Yazd University, Iran
Gitesh K. Raikundalia,	Victoria University, Australia
Giulio Concas,	DIEE Univeristy of Cagliari, Italy
Gniewko Niedbala,	Poznan University of Life Sciences, Poland
Grigorios N. Beligiannis,	University of Patras - Agrinio Campus, Greece
Grzegorz Sierpiński,	Silesian University of Technology, Poland
Gülden Köktürk,	Dokuz Eylül University, Turkey
Guoqian Chen,	Peking University, China
Gyanendra Verma,	National Institute of technology Raipur, India
Gyu Myoung Lee,	Liverpool John Moores University, UK
Habil. Gabor Kiss,	J. Selye University, Slovakia
Habil. Ioan-Gheorghe Rotaru,	University of Arad, Romania
Hala Abukhalaf,	Palestine Polytechnic University, Palestine
Hamid Ali Abed AL-Asadi,	Basra University, Iraq
Hamid Khemissa,	USTHB University Algiers, Algeria
Hamidreza Rokhsati,	Sapienza university of Rome
Hanan Alghamdi,	King Abdulaziz University, Saudi Arabia
Heba Mohammad,	Higher College of Technology, UAE
Hedayat Omidvar,	Research & Technology Dept., Iran
Hon Hai,	Research Institute, Taiwan
Ilham Huseyinov,	Istanbul Aydin University, Turkey

Ivana Burgetová,	Brno University of Technology, Czech Republic
Iyad Alazzam,	Yarmouk University, Jordan
Janaki Raman Palaniappan,	Brunswick Corporation, USA
Janusz Kacprzyk,	Systems Research Institute, Poland
Jawad K. Ali,	University of Technology, Iraq
Jesuk Ko,	Universidad Mayor de San Andres, Bolivia
Jia Ying Ou,	York University, Canada
Jonah Lissner,	Technion - israel institute of technology, Israel
José Manuel Fonseca,	NOVA University of Lisbon, Portugal
Jun Zhang,	South China University of Technology, China
K. S. Shivaprakasha,	NMAM Institute of Technology, India
Kamel Hussein Rahouma,	Nahda University, Egypt
Karim El Moutaouakil,	FPT/USMBA, Morocco
Karim Mansour,	Salah Boubenider University, Algeria
Kazuyuki Matsumoto,	Tokushima University, Japan
Kire Jakimoski,	FON University, Republic of Macedonia
Klenilmar Lopes Dias,	Federal Institute of Amapa, Brazil
Koh You Beng,	University of Malaya, Malaysia
Krzysztof Piech,	Lazarski University, Poland
Kurdi Zakaria,	University of Lynchburg, USA
Lei Zhang,	Qinghai Normal University, China
Lihong Zheng,	Charles Sturt University, Australia
Ljiljana Trajkovic,	Simon Fraser University, Canada
Loc Nguyen,	Loc Nguyen's Academic Network, Vietnam
Lucian Lupu-Dima,	University of Petrosani, Romania
Luisa Maria Arvide Cambra,	University of Almeria, Spain
M V Ramana Murthy,	Osmania University, India
M. Mujiya Ulkhaq,	Diponegoro University, Indonesia
M.Madheswaran,	Muthayammal Engineering College, India
MA. Jabbar,	Vardhaman College of Engg, India
Manal Mostafa,	Al- azher University, Egypt
Manish Kumar Mishra,	University of the People, USA
Marco Javier Suarez Baron,	UPTC, Colombia
Marcos Roberto Tovani Palone,	University of São Paulo, Brazil
Mario Versaci,	Mediterranea University of Reggio Calabria, Italy
Masoomah Mirrashid,	Semnan University, Iran
Md. Sadique Shaikh,	AIMSR, India
Michail Kalogiannakis,	University of Crete, Greece
Min Zhang,	Universidade NOVA de Lisboa, Portugal
Miral M. Desai,	Department of EC Engineering, India
Mohamed Anis Bach Tobji,	University of Manouba, Tunisia
Mohamed Hamlich,	ENSAM, Morocco
Mohamed Khalil,	MSA University, Egypt
Mohamed-Khireddine,	Echahid Hamma Lakhdar d'El-Oued, Algeria
Mohammed Al-Sarem,	Taibah University, KSA
Mu-Chun Su,	National Central University, Taiwan
Mueen Uddin,	Universiti Brunei Darussalam, Brunei
Muhammad Aslam Javed,	University of Central Punjab, Pakistan
Mu-Song Chen,	Da-Yeh University Dacun, Taiwan
N. Jeyanthi,	VIT University Vellore, India
N. Yamuna Devi,	Coimbatore Institute of Technology, India
Nadia Abd-alsabour,	Cairo University, Egypt

Naren.J,	Jain (Deemed-to-be University), India
Ndia G. John,	Murang'a University of Technology, Kenya
Neel Rajpurohit,	Parul University, Baroda, India
Ngoc Hong Tran,	Vietnamese-German University, Vietnam
Noura Taleb,	Badji Mokhtar University, Algeria
Noureddine Aboutabit,	Sultan moulay sliman university, Morocco
Nur Eiliyah Wong,	Senior Lecturer/ Researcher, Malaysia
Oleksii K. Tyshchenko,	University of Ostrava, Czech Republic
Olha Kanishcheva,	National Technical University "KhPI", Ukraine
Ong Pauline,	Universiti Tun Hussein Onn Malaysia, Malaysia
Otilia Manta,	Romanian Academy, Italy
P Venkata Suresh,	Director of Socis Ignou, India
P.V.Siva Kumar,	VNR VJIET, India
Panagiotis Fotaris,	University of Brighton, UK
Panagiotis Varzkas,	University of Thessaly, Lamia, Greece
Pasupuleti Venkata Siva Kumar,	Vnr Vjiet, India
Patrick Fiati,	Patrick Fiati Engineering, Ghana
Paulo Jorge dos Mártires Batista,	University of Évora, Portugal
Pavel Loskot,	Swansea University, UK
Pokkuluri Kiran Sree,	Sri Vishnu Engineering College for Women, India
Pr. Smain Femmam,	UHA University, France
Preeti Mehta,	Senior scientific officer, India
Preeti Solanki,	Apex Professional University, India
Quang Hung Do,	University of Transport Technology, Vietnam
Rabah Mohammed Amin,	Algerian Space Agency, Algeria
Radu Godina,	Universidade NOVA de Lisboa, Portugal
Rahul B. Chauhan,	Parul University, India
Ramadan Elaïess,	University of Benghazi, Libya
Ramgopal Kashyap,	Amity University Chhattisgarh, India
Ridda Laouar,	LAMIS Laboratory, Algeria
Rodolfo Delmonte,	Ca Foscari University, Italy
Rogério Rossi,	University of São Paulo, Brazil
Saad Aljanabi,	Alhikma College University, Iraq
Safawi Abdul Rahman,	Universiti Teknologi Mara, Malaysia
Sahar Saoud,	Ibn Zohr University, Morocco
Sai kumar Tara,	CMR Technical Campus, India
Saif aldeen Saad Obayes Alkadhim,	Shiite Endowment Diwan, Iraq
Samarendra Nath Sur,	Sikkim Manipal Institute of Technology, India
Samir Kumar Bandyopadhyay,	University of Calcutta, India
Samrat Kumar Dey,	Bangladesh Open University, Bangladesh
Sanjeevikumar Padmanaban,	Aarhus University, India
Sasikumar P,	Vellore Institute of Technology, India
Sathyendra Bhat J,	St Joseph Engineering College, India
Satish Gajawada,	IIT Roorkee Alumnus, India
Sebastian Floerecke,	University of Passau, Germany
Shah Khalid Khan,	RMIT University, Australia
Shahid Ali,	AGI Education Ltd, New Zealand
Shahram Babaie,	Islamic Azad University, Iran
Shahzad Ahmed,	Comsats University Islamabad, Pakistan
Shashikant Patil,	ViMEET Khalapur Raigad, India
Shing-Tai Pan,	National University, Taiwan
Siarry Patrick,	Universite Paris-Est Creteil, France

Siddhartha Bhattacharyya,
Sidi Mohammed Meriah,
Simone Nasser Matos,
Suhad Faisal Behadili,
Sunanda Dixit,
Svetla Koeva,
T.T. Dhivyaprabha,
Takfarinas Saber,
Taleb Zouggar Souad,
Tanik Saikh,
Tariq Tawfeeq Yousif Alabdullah,
Tasher Ali Sheikh,
Tatyana A. Komleva,
Taufiq Hidayat,
Thidawan Klaysri,
Tolga Ensari,
Toufik Bakir,
Unger,
V. Dinesh Reddy,
V. Ilango,
Valerianus Hashiyana,
Venkata Duvvuri,
Vilem Novak,
Wanyang Dai,
Xiaochun Cheng,
Yousef Farhaoi,
Youssef Ounejjar,
Yuan-Kai Wang,
Yu-Chen Hu,
Yugen Yi,
Yuping Yan,
Zati Hakim Azizul Hasan,
Zoran Bojkovic,

Christ University, India
University of Tlemcen, Algeria
Universidade Tecnológica Federal do Paraná, Brazil
University of Baghdad, Iraq
BMS Institute of Technology and Management, India
Institute for Bulgarian Language, Bulgaria
Avinashilingam Institute, India
National University of Ireland, Ireland
Oran 2 University, Algeria
Indian Institute of Technology, India
University of Basrah, Iraq
Residential Girls' Polytechnic, India
Odessa State Academy, Ukraine
International Islamic University, Malaysia
RMUTP, Thailand
Arkansas Tech University, USA
University of Burgundy, France
FernUni in Hagen, Germany
SRM University, India
CMR Institute of Technology, India
University of Namibia, Namibia
Northeastern University, USA
University of Ostrava, Czech Republic
Nanjing University, China
Middlesex University, UK
Moulay Ismail University, Morocco
Moulay Ismail University, Morocco
Fu Jen Catholic University, Taiwan
Providence University, Taiwan
Jiangxi Normal University, China
ELTE, Hungary
Universiti Malaya, Malaysia
University of Belgrade, Serbia

Technically Sponsored by

Computer Science & Information Technology Community (CSITC)



Artificial Intelligence Community (AIC)



Soft Computing Community (SCC)



Digital Signal & Image Processing Community (DSIPC)



8th International Conference on Artificial Intelligence and Applications (AI 2022)

Improving Explanations of Image Classification with Ensembles of Learners...01-16

Aadil Ahamed, Kamran Alipour, Sateesh Kumar, Severine Soltani and Michael Pazzani

The NP-Completeness of Quay Crane Scheduling Problem17-27

Ali Skaf, Samir Dawaliby, and Arezki Aberkane

A Context-Aware and Adaptive System to Automate the Control of the AC Windshield using AI and Internet of Things.....29-35

Joshua Tian and Yu Sun

Robust Discriminative Non-Negative Matrix Factorization with Maximum Correntropy Criterion.....37-49

Hang Cheng, Shixiong Wang and Naiyang Guan

A Unity Microscope Simulation to Help Students Get More Access to Lab Equipment Online during Covid-19 Pandemic.....51-61

Kaiwen Chen and Yu Sun

Trustworthy Artificial Intelligence for Blockchain-based Cryptocurrency.....63-68

Tiffany Zhan

An Intelligent Video Editing Automate Framework using AI and Computer Vision69-79

Mei Yi Yang and Yu Sun

A Single Level Detection Model for Traffic Sign Detection using Channel Shuffle Residual Structure.....81-93

Yuanzhi Luo and Jie Hao

Hyper-Parameter Tuning in Deep Neural Network Learning.....95-101

Tiffany Zhan

2nd International conference on AI, Machine Learning in Communications and Networks (AIMLNET 2022)

Anomaly Detection based on Alarms Data.....103-114

Michel Kamel, Anis Hoayek and Mireille Batton-Hubert

8th International Conference on Bioinformatics & Biosciences (BIOS 2022)

Machine Learning based to Predict B-Cell Epitope Region Utilizing Protein Features.....115-122
Fatema Nafa and Ryan Kanoff

Tensor-based Multi-Modality Feature Selection and Regression for Alzheimer's Disease Diagnosis123-134
Jun Yu, Zhaoming Kong, Liang Zhan, Li Shen and Lifang He

Machine Learning GUI based For Detecting Alzheimer's.....135-143
Fatema Nafa, Evelyn RodriguezArgueta, Annie Dequit and Changqing Chen

LabBuddy: A Game-based Interactive and Immersive Educational Platform for Physics Lab Learning using Artificial Intelligence and 3D Game Engine.....145-154
Yuxing Ji, Mingze Gao and Yu Sun

An Approach using Machine Learning Model for Breast Cancer Prediction..155-161
Fatema Nafa, Enoc Gonzalez and Gurpreet Kaur

Converting Real Human Avatar to Cartoon Avatar using CycleGAN163-175
Wenxin Tian

2nd International Conference on Big Data, IOT & NLP (BINLP 2022)

An Interactive and Scenario-based Simulation Gaming System for Business Education using Game Engine and Machine Learning.....177-193
Yingzhi Ma and Yu Sun

Word Predictability is Based on Context - and/or Frequency.....195-211
Rodolfo Delmonte and Nicolò Busetto

An Intelligent and Social-Oriented Sentiment Analytical Model for Stock Market Prediction using Machine Learning and Big Data Analysis.....213-223
Muqing Bai and Yu Sun

The Problem Solver: A Mobile Platform to Mediate Teenager Family Relationship using Dart and Machine Learning.....285-295
Ziheng Guan and Ang Li

8th International Conference on Computer Science and Information Technology (CSTY 2022)

Towards Devising a Fund Management System using Blockchain.....225-238

*Nibula Bente Rashid, Joyeeta Saha, Raonak Islam Prova, Nowshin Tasfia,
Md. Nazrul Huda Shanto and Jannatun Noor*

F-low: A Promising Countermeasure Against DDoS Attacks based on Split Sketch and PCA.....239-250

Fei Wang, Zhenxing Li and Xiaofeng Wang

8th International Conference of Managing Value and Supply Chains (MaVaS 2022)

An Automatic Sheet Music Generating Algorithm based on Machine Learning and Artificial Intelligence.....251-258

Ruize Yu and Yu Sun

Low-Carbon Innovation Decision Considering Quality Differences and Government Subsidies under the Three-Party Trading Platform259-278

Xu Wang, Longzhen Zhou, Zusheng Zhang, Yingbo Wu and Longxiao Li

8th International Conference on Signal and Image Processing (SIGI 2022)

Image Segmentation in Shape Synthesis, Shape Optimization, And Reverse Engineering279-284

Milan Ćurković, Andrijana Ćurković, Damir Vučina and Domagoj Samardžić

IMPROVING EXPLANATIONS OF IMAGE CLASSIFICATION WITH ENSEMBLES OF LEARNERS

Aadil Ahamed¹, Kamran Alipour¹, Sateesh Kumar¹,
Severine Soltani² and Michael Pazzani³

¹Department of Computer Science and Engineering,
University of California, San Diego, La Jolla, CA, USA

²Department of Bioengineering,
University of California, San Diego, La Jolla, CA, USA

³Information Sciences Institute, Marina Del Rey, CA, USA

ABSTRACT

In explainable AI (XAI) for deep learning, saliency maps, heatmaps, or attention maps are commonly used to identify important regions for the classification of images of explanations. Recent research has shown that many common XAI methods do not accurately identify the regions that human experts consider important. We propose averaging explanations from ensembles of learners to increase the accuracy of explanations. Our technique is general and can be used with multiple deep learning architectures and multiple XAI algorithms. We show that this method decreases the difference between regions of interest of XAI algorithms and those identified by human experts. Furthermore, we show that human experts prefer the explanations produced by ensembles to those of individual networks.

KEYWORDS

Neural Networks, Machine Learning, Explainable AI, Image Classification, Computer Vision.

1. INTRODUCTION

A variety of eXplainable Artificial Intelligence (XAI) methods have emerged for explaining image classification [15, 17] to developers or end-users [16, 6]. These approaches typically locate and highlight regions of the image that are important to the classification decision. Recently, several papers have called into question the ability of existing XAI methods to accurately identify regions that are meaningful to human experts such as radiologists, dermatologists, neurologists, oncologists, ophthalmologists, or even bird watchers [30, 8, 33, 22]. For example, [2] discusses the substantial differences between the regions on an x-ray that radiologists find important and those found by XAI algorithms.

We explore the use of ensemble learning [9] of neural networks to increase the accuracy of identifying regions of interest for any XAI algorithm by combining explanations from multiple neural networks. To illustrate, Figure 1(a-d) shows the heatmap of four neural networks trained on the same image data starting with different initial random weights. The task was to determine the wing pattern (e.g., striped, solid, spotted, wingbar). Figure 1(e) shows the heatmap produced by averaging the heatmaps of 11 networks. Of course, a disadvantage of our approach is that it requires more computation to create an ensemble than a single network. This linear increase in

computation can be mitigated by coarse-grained parallel training of N networks.

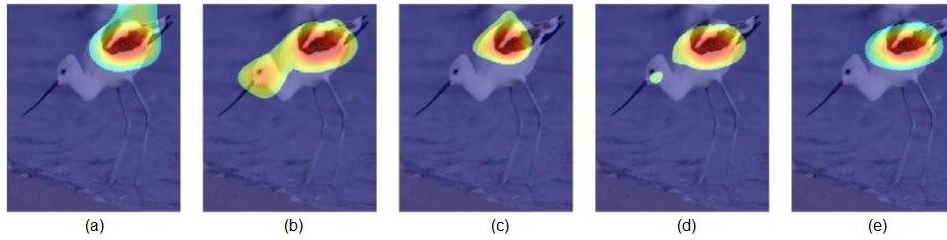


Figure 1. Saliency maps from an ensemble of classifiers. (a-d) are individual networks trained to identify the wing pattern. (e) is an average of 11 networks.

There are multiple purposes for XAI. One is to inform developers and perhaps validators how the deep learning system is working. Figure 2a shows an example heatmap produced by XAI algorithms to illustrate how the underlying deep learner operates. A second use which becomes important as applications of deep learning are deployed is to describe to end users the location of important diagnostic features. For example, Figure 2b shows an image from a dermatology journal [21] with three regions identified and labeled with “*milky pink structureless areas centrally (*), white streaks (^) and atypical pigment network (arrows).*” In this case, it is important for a deep learner and XAI system to not only get the correct diagnosis but also to correctly identify where the white streaks are.

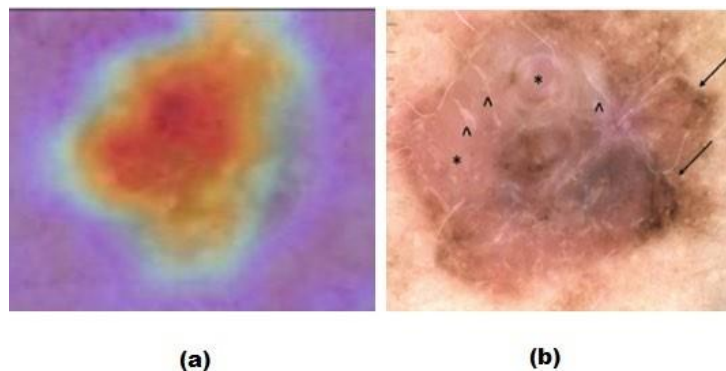


Figure 2. (a) Heatmap for explaining melanoma classification. (b) Image from journal explaining melanoma diagnosis.

In the remainder of this paper, we first describe the methods we use to generate an ensemble of networks. Second, we discuss our evaluation methods which compare the regions of interest of an XAI algorithm to the regions of interest identified by people. Third, we describe the databases used in evaluation. Fourth, we describe the results using ensembles starting with different random weights on several problems. Fifth, we generalize our results by using two additional approaches to generating an ensemble of networks. Furthermore, we show the impact of varying the number of networks in the ensemble. Finally, we present results from an experiment with experienced bird watchers that show that they prefer the explanations from ensembles to those of individual networks.

2. PRIOR WORK

There are two main approaches in XAI to identify regions of interest in deep learning for image classification:

- a. Model agnostic methods, such as LIME [23], manipulate inputs (e.g., pixels, regions or more commonly superpixels) and measure how changes in input affect output. If an input perturbation has no effect, it is not relevant to findings. If a change has a major impact (e.g., changing the classification from pneumonia to normal), then the region is important to the classification. Shapley Additive Explanations (SHAP) [20] uses a game-theoretic measure to assign each feature or region an importance value for a particular prediction.
- b. Other methods examine the activations or weights of the deep network to find regions of importance. Grad-CAM [26], Integrated Gradients [31], Saliency [28], GradientShap [19], and Layerwise Relevance Propagation LRP [25] are examples of such methods.

Ensemble learning has long been used to reduce the error of machine learning methods, including neural networks [11]. This error reduction is due to reduction in variance in the learned models [10]. Ensemble learning reduces errors most when the errors of the individual models are not highly correlated [1, 18]. Recent work [32] has shown that XAI methods for deep learners trained under slightly different circumstances produce explanations that are not highly correlated. The variability occurs due to the initial random parameter selection or the random order of training examples. This suggests that we can reduce the error of an explanation by combining explanations from an ensemble of networks.

Reiger [24] proposed combining different XAI methods such as Grad-CAM and LRP since each method has their own strengths and weakness and found this increased the stability of the XAI output. This does not use an ensemble of learners and cannot increase the accuracy of the classifications.

In learning from tabular data, ensembles have been shown to improve accuracy at the expense of interpretability. In contrast, our goal with image data is to both improve classification accuracy and the explanation.

3. METHODS

3.1. Generating Ensemble Explanations

For this paper, we use two common image classifiers: VGG16 [29] and ResNet [14]. We consider three methods of generating a diverse ensemble of classifiers.

1. **Different Random Weights.** We start with N identical base networks and then initialize each of the N classification heads with different random weights and present the same training data to each network. The idea here is that based on the initial conditions, the network will find a slightly different solution [3]. We evaluate whether on average the ensemble produces a better explanation than the members of the ensemble.
2. **Leave Out One Bucket.** We divide the training data into N buckets and train the N identical architectures on $N-1$ buckets [13]. We evaluate whether the ensemble produces a better explanation than a single network trained on all the data.
3. **Bootstrap Aggregation.** We use bagging [4] which creates N training sets by sampling with replacement from the original training data. This leaves out some of the original training data and places additional weight on other examples by replicating them. We evaluate whether the ensemble produces a better explanation than network trained on all the data. The motivation of the latter two methods is that slightly different training data will

result in slightly different solutions that may be averaged. Even if each of the individual networks is less accurate than training on all the data, the consensus on the ensemble often exceeds the classification accuracy of a single network on all data. We anticipate this will hold with the accuracy of the explanation as well.

Once the models are trained, we generate an ensemble explanation by averaging the relevance score for each pixel (i, j) in the input image as shown in Eq. 1.

$$score_{ens}(i, j) = \frac{1}{n} \sum_{k=1}^n score_k(i, j) \quad (1)$$

3.2. Metrics

We consider three measures of explanation accuracy: Intersection over Union (IoU), correlation, and the center of mass distance. In all cases, the ground truth region is collected from human annotators. It is worth stressing that this region information is used only in evaluation, not in training.

1. **Intersection over Union** For IoU, we binarize the generated explanation by normalizing it between 0 and 1 (or $[-1, 1]$ for some XAI algorithms) and setting a threshold to find regions of interest. The IoU score is obtained by computing the intersection between the explanation and ground truth mask and then dividing it by their union. We use a default threshold of 0.3 in this work. Figure 3 illustrates how we compute the IoU score for a single image. The IoU metric allows us to compare how well an XAI algorithm identifies region of interest found by human annotators. A higher value shows more agreement between the algorithm and the annotator.
2. **Correlation** To quantify the similarities between explanation maps and ground truth masks, we consider the two as jointly distributed random variables and use the Pearson correlation between them. This metric is obtained by down sampling the masks to a lower resolution (e.g., 14×14) to reduce noise errors. Then we flatten the 2D masks into a 1D vectors and compute the correlation as:

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (2)$$

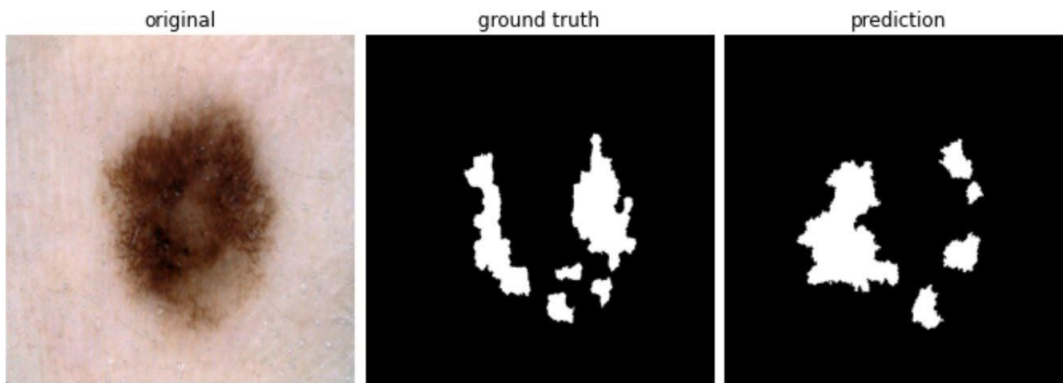


Figure 3. Left: Example of an image used to evaluate ensembles of explanations Middle: Ground-truth masks for an image in the ISIC-2018 melanoma dataset. Right: Explanation generated by averaging explanations generated by an ensemble of models.

- 3. Center of Mass Distance** We compute the center of mass of a saliency or heatmap identified by the XAI algorithm. The center of mass R is computed as a function of each weight w at point r .

$$R = \frac{1}{W} \sum_{i=1}^n w_i r_i \quad \text{where} \quad W = \sum_{i=1}^n w_i \quad (3)$$

Figure 3 illustrates the center of mass identified by ten networks for the wingbar of a bird. The black arrows indicate the center of mass of the heatmap of individual networks, and the red arrow indicates the center of mass of the ensemble explanation.

We compute the Euclidean distance from the center of mass to the key point identified by human annotator. The distance between the center of mass and a key point is useful for two reasons. First, it is quicker to collect key points vs. regions, i.e., pointing to a bird's bill vs. tracing it. Second, many explanations in medical journals or bird watching guides use arrows instead of heatmaps to identify features, and the center of mass can be used as the endpoint of the arrowhead.

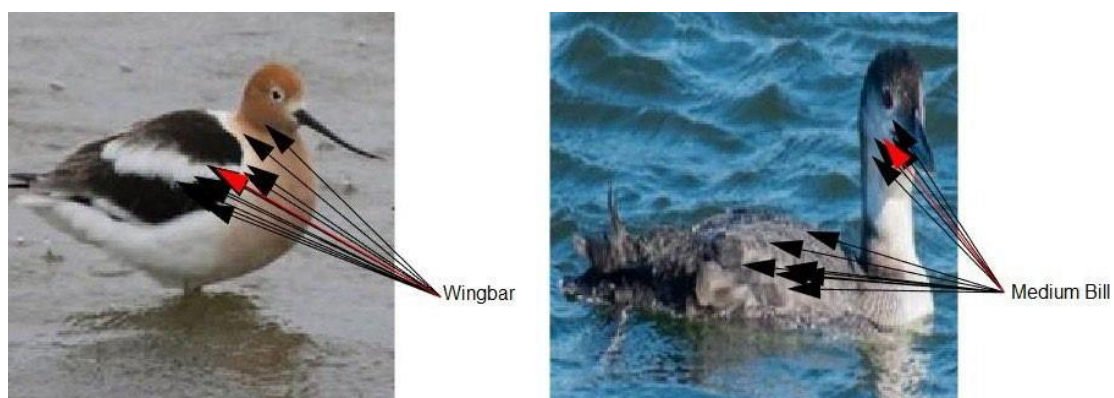


Figure 4. Use of key points that show center of mass (CoM) for heatmaps. The red arrowhead points to the CoM of the average heatmap while black arrows point to the CoM by each model in the ensemble.

3.3. XAI Methods

We evaluated the averaging approach on multiple XAI methods to show the generality of the approach. The methods that we explored are surveyed in [5] and include, GradCAM and Guided GradCAM [26], LIME [23], Input gradient [27], Gradient SHAP [19], Integrated Gradients [31], and Saliency [28].

4. DATASETS

We train and evaluate our averaging algorithm on two datasets. Note that a dataset for evaluation must include ground truth labels and ground truth regions or key points. The regions or key points are used in evaluation, but not in training.

HiRes Birds HiRes Birds is a new dataset we introduce of 14,380 images of birds divided into 66 species. Additionally, we have collected data on various attributes of each bird, such as its bill length, wing pattern and location of the bill. We continue to collect additional feature and location information for this dataset. In this paper, we use this dataset to learn to identify attributes of the bird, such as whether it has a striped wing and evaluate whether the XAI algorithm focuses on the wing when making this classification.

ISIC-2018 ISIC-2018[7] is a melanoma detection dataset consisting of skin lesions obtained from a variety of anatomic sites and patients spread out across multiple different institutions. It consists of 2594 images along with 5 segmentation masks per image to identify the location of attributes of the region such as streaks and milia-like cysts. We only use the segmentation masks in evaluation. We train networks to recognize the presence or absence of features such as “milia-like cysts” and evaluate whether XAI algorithms find the region identified by the segmentation masks.

5. RESULTS

In this section, we first present data that shows averaging explanations from an ensemble of learners improves the explanation for a variety of XAI algorithms using initial random weights to create the ensemble. Next, we show results that vary the conditions under which the ensemble is learned to go deeper into the conditions under which the approach is effective. To assess the performance of our proposed method, we define two quantities: Individual Average (Ind-Avg) and Ensemble Average (Ens-Avg). Individual Average is the average metric on the evaluation set of each of the individual networks in the ensemble in the case that the ensemble is trained with random weights, or the individual network trained on all the training data in the case of bagging and leave-out-one-fold. Ensemble Average calculates an average heatmap and we report the target metric between the average heatmap and the ground truth.

5.1. HiRes Birds Results

We consider two different classification tasks with the HiRes Birds dataset. In both cases, we create ensembles starting with different random weights.

First, we train on a multi-class problem of identifying the bill length of the birds. For this task, there are 3 classes: Large, Medium and Small. We train on 4763 examples from the dataset using 530 as a validation set and 2322 as an evaluation set. We used Hive Data (a crowd-sourcing platform for data labelling) to collect both the bill length data and the bill location data. Our dataset includes a key point for each bill.

Table 1 shows the distance between the center of mass of the region found by 5 XAI algorithms using VGG16 as the deep learning classifier. Table 2 shows the data using ResNet as the deep learning classifier. In these tables, statistically significant results using a paired t-test are shown in bold with p-value < 0.01 indicated by ** and p-value < 0.0001 indicated by ****. The results show that for five commonly used XAI algorithms and two commonly used deep learning architectures, our ensemble method results in better identification of the center of mass that can be compared to a key point.

In our second use of HiRes Birds, we train on a multi-class problem of distinguishing the wing pattern of birds. For this task, there are 6 classes. We train on 12221 examples from the dataset using 2156 as a validation set. We have collected wing patterns for our data, but not the wing location. Therefore, we evaluate on the location of the wing using the bird data from the PartImageNet which contains wing locations but not patterns. The evaluation is based on the ground truth locations of wings in 594 bird images from the PartImageNet dataset.

Table 1. Ensembling improves explanation center of mass distance on the beak size identification task with VGG16 [29]. Ind-Avg refers to average performance of individual models evaluated separately. Ens-Avg refers to performance of the ensemble model. Lower is better. Best results are in bold

Table 1. Ensembling improves explanation center of mass distance on the beak size identification task with VGG16 [29].

Method	Ind-Avg	Ens-Avg
GradCAM [26]	0.395	0.328****
Input gradient [27]	0.331	0.320****
Gradient SHAP [19]	0.346	0.334****
Integrated Gradients [31]	0.346	0.334****
Saliency [28]	0.328	0.322**

Table 2. Ensembling improves explanation center of mass distance on the beak size identification task with ResNet18.

Method	Ind-Avg	Ens-Avg
GradCAM [26]	0.236	0.216****
Input gradient [27]	0.315	0.300****
Gradient SHAP [19]	0.320	0.309****
Integrated Gradients [31]	0.330	0.313****
Saliency [28]	0.313	0.309**

Table 3 shows the correlation between the importance of pixels identified by XAI algorithms and the wing region in PartImageNet using VGG16. As before, averaging over an ensemble improves the XAI algorithms we tested. In addition, we computed the center of the PartImageNet wing regions and compared to the center of mass of the regions found by various XAI algorithms. The results shown in Table 4 indicate improvement for this metric as well.

Table 3. Ensembling improves explanation correlation on the wing pattern identification task with ResNet18.

Method	Ind-Avg	Ens-Avg
GradCAM [26]	0.207	0.256****
Input gradient [27]	0.263	0.324****
Gradient SHAP [19]	0.263	0.315****
Integrated Gradients [31]	0.265	0.312****
Saliency [28]	0.218	0.337****

Table 4. Ensembling improves explanation center of mass distance on the wing pattern identification task with ResNet18.

Method	Ind-Avg.	Ens-Avg.
GradCAM [26]	0.157	0.142****
Input gradient [27]	0.153	0.145****
Gradient SHAP [19]	0.152	0.146****
Integrated Gradients [31]	0.152	0.146****
Saliency [28]	0.158	0.147****

5.2. ISIC-2018 Results

We also tested our method on the ISIC2018 dataset for melanoma lesion detection. For this task, each training image is paired with 5 dermatologist annotated masks that identify important attributes for melanoma detection. We train a five different multi-label binary classifier starting with random weights to output a binary variable for each of the 5 attributes of an image. We generate the ground-truth binary labels for an image by checking whether the associated ground-truth mask is non-zero or not. Once trained, we run different XAI algorithms to generate regions for each attribute that we compare to the segmentation mask in the evaluating data. We evaluate the quality of the generated explanations of individual models against our proposed method which averages the explanation across the ensemble. Results are reported in Table 5. We observe that for GradCAM and LIME we see a noticeable and statistically significant increase in explanation quality with ensemble explanations. Integrated-gradients and Saliency show small but nonetheless statistically significant improvements. We believe this is because Integrated-gradients and Saliency generate pixel level explanations whereas GradCAM and LIME generate region-based explanations leading to better IoU scores when comparing regions to segmentation masks.

Table 5. IoU between the ground truth and explanations generated from XAI methods on ISIC2018 dataset for the task of lesion identification.

Method	Ind-Avg.	Ens-Avg.
GradCAM [26]	0.17	0.19****
Integrated Gradients [27]	0.04	0.05****
Saliency [28]	0.03	0.04****
LIME [23]	0.11	0.13****

Our trained classifiers achieve a mean Area under the ROC curve (AUC) score of 0.82 on the validation set. Ensembling the models leads to an improved AUC score of 0.86, showing that in addition to increasing explanation accuracy, classification accuracy is also increased. The increase in accuracy is statistically significant with a p-value of 0.03.

5.3. Separating the Effect on Intersection and Union

Here, we investigate the effect of averaging explanations on the intersection and union metrics independently using the same training procedure described in section 5.2. A greater intersection means that the XAI algorithm finds more of the area of interest identified by an annotator. A smaller union indicates that fewer regions are found outside the relevant region. Both intersection and union are measured in pixels.

Table 6 shows that the average explanation increases the intersection and decreases the union for LIME, integrated gradients, and saliency. This indicates that averaging finds more relevant regions and fewer irrelevant regions. For GradCAM, both the intersection and union increase but the increase in the intersection is more significant, leading to an overall improvement in the IoU score.

Table 6. Intersection (I) and Union (U) metrics for melanoma dataset. Ind-Avg is the average of the metric of individual models. Ens-Avg is the metric for our proposed technique. The unit of measurement for both intersection and union is pixels.

Method	Ind-Avg-I	Ens-Avg-I	Ind-Avg-U	Ens-Avg-U
GradCAM [26]	1344.82	1668.59	7968.59	8602.46
LIME [23]	904.44	999.31	8262.70	7897.76
Integrated Gradients [31]	95.20	118.46	3470.67	3463.11
Saliency [28]	150.01	174.93	3574.14	3556.70

5.4. Other Approaches to Creating an Ensemble

We also tried training an ensemble on the HiRes Birds beaks dataset using the Leave-Out-One method of creating an ensemble. In this setting, we divide the training data into $K = 10$ folds and then train each model in an ensemble of size 10 with a unique combination of 9 folds. In a similar experiment, we trained another ensemble of size 10 using a bagging algorithm where the training set for each model is generated by sampling from the original training set with replacement. The results for both experiments are reported in Table 7. We observe that both K-fold and bagging lead to improved explanation accuracy when compared to explanations generated by individual models trained on the entire training data.

Table 7. Center of mass distance between explanations and ground-truth annotations for an ensemble trained using the K-fold and Bagging algorithms. Ind-Avg is the CoM distance of a single model trained on the entire training data. Lower is better.

Method	Ind-Avg	K-fold	Bagging
Guided GradCAM [26]	0.169	0.153**	0.157
GradCAM [26]	0.297	0.273**	0.277
Integrated Gradients [31]	0.278	0.217**	0.219
Saliency [28]	0.280	0.222**	0.225
Input gradient [27]	0.293	0.241**	0.243
Gradient SHAP [19]	0.278	0.218**	0.219

The results show that other ways of creating diverse models also works with our ensemble averaging method.

5.5. Varying the Size of the Ensemble

Next, we investigated the effect of ensemble size on the quality of the averaged explanation using the melanoma dataset. Fig. 5 plots the IoU score of the averaged explanation and the average IoU score of individual models against the size of the ensemble using the random weights method. We observe that IoU scores tend to increase with ensemble size and plateau after a certain threshold.

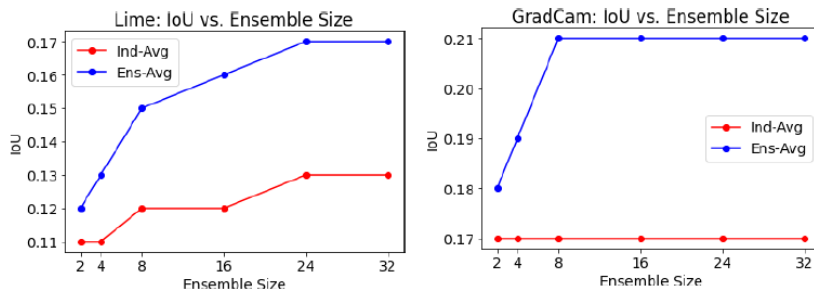


Figure 5. IoU between ground truth and generated explanations on ISIC-2018 dataset as a function of ensemble size. The blue curve shows the IoU score from our proposed technique which averages the explanations across the ensemble. The red curve shows the average of the individual IoU scores in the ensemble.

6. HUMAN EVALUATION

In this section, we show that people can notice the difference in the quality of explanations produced by ensembles when compared to individual models. In our computational experiments, the ensemble results in greater identification of regions of interest and less identification of irrelevant regions. The stimuli for the experiment consisted of images annotated by LIME or averaging of an ensemble of eleven LIME classifiers starting with different random weights. The stimuli were generated using the version of imageLIME in MATLAB. Figure 6 shows examples of the stimuli used in the experiment.

In Figure 6, the top bird is a common goldeneye. Its distinguishing characteristics (called field marks by birders) include a gold-colored eye, and it is distinguished from the similar Barrow’s goldeneye by having a round vs. kidney-shaped patch on the cheek and striped vs. checkered wings. The averaged annotation on the right picks up both distinguishing characteristics, showing

that averaging finds more relevant features. The middle bird is the Barrow’s Goldeneye and again the ensemble focuses on both the wing pattern and the cheek patch. The lower bird is a western grebe. It is distinguished from the similar Clark’s grebe by having a yellow vs. orange bill and having the black on the head extend below the eye. The average annotation on the right picks up both and furthermore does not emphasize a patch on the back that is not relevant to the classifications.

6.1. Participant Recruitment

This study was approved by UCSD’s IRB. Participants for this study were expert bird watchers who were recruited from mailing lists that report rare bird sightings in Southern California. We recruited 28 participants for a LIME vs. averaged LIME study: one participant self-excluded due to a lack of familiarity with the bird species included in the studies and did not complete the study, and one was excluded due to being under the age of 18, which may point to less real-world bird watching experience. In total, 26 participants were included in the analyses. Participants in the study had a median of 15 years of bird-watching experience.

6.2. Study Design

Prior to beginning the study, participants were shown an example of LIME used on an image of a dog to identify the features most important to classifying its breed. This example was intended to familiarize participants with how to interpret the colors on a LIME-generated heatmap. Each study contained 24 unique bird images, each of which was shown once to each subject with LIME-generated annotations and another time with averaged LIME-generated annotations. This results in 48 trials evenly split between the base LIME method and the averaged LIME method.

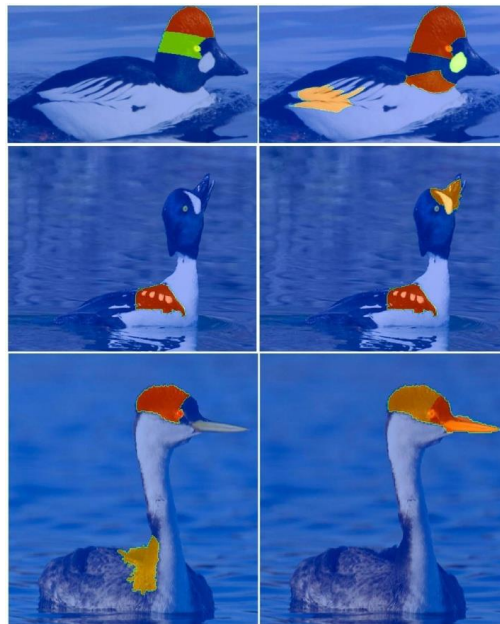


Figure 6. The figures on the left were annotated by LIME on a single VGG16 network. The annotations of the figures on the right are the average of 11 VGG16 networks.

In addition, these 24 unique bird images consisted of 10 different bird species, and these ten bird species were selected to include five pairs of similar-looking birds:

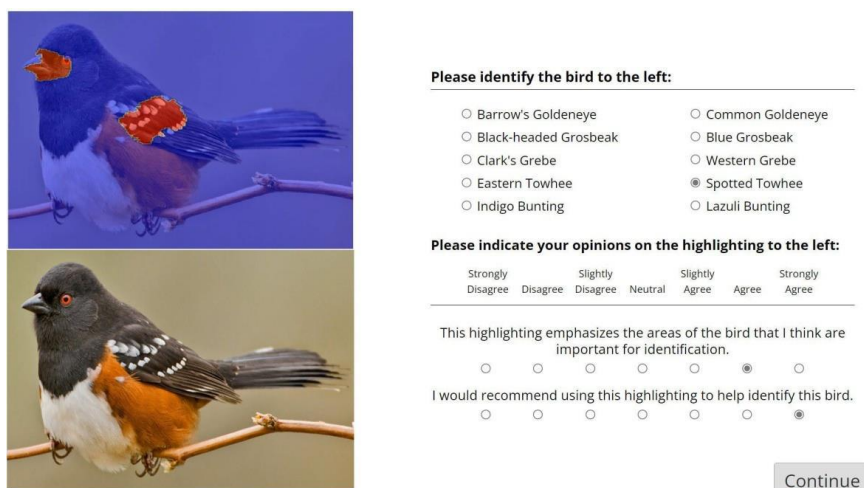
- Black-headed grosbeak and blue grosbeak.
- Clark’s grebe and Western grebe.
- Eastern towhee and spotted towhee.
- Indigo bunting and lazuli bunting.
- The Barrow’s goldeneye and common goldeneye.

6.3. Results of Study

Each trial displayed heatmap-annotated bird images alongside unannotated images, a bird species classification task, and questions about annotation preferences. A screen capture of the study interface is shown in Figure 7. Participants were asked to classify the bird species in the image by selecting 1 of 10 radio buttons corresponding to the ten unique bird species. In each study, participants were asked to provide their opinions on the novel highlighting method by answering two questions: “This highlighting emphasizes the areas of the bird that I think are important for identification” (Question 1), and “I would recommend using this highlighting to identify this bird” (Question 2). Participants indicated their responses using a 7-point Likert scale ranging from “Strongly Agree” (a value of 7) to “Strongly Disagree” (a value of 1). The midpoint (a value of 4) indicates a “Neutral” sentiment.

We compared the median preference ratings for LIME trials to averaged LIME trials. We only include trials in which the subject correctly identified the bird species in calculating median preference ratings. The median classification accuracy across all participants was greater than 90%. Thus, participants were generally very accurate in their bird species classification, resulting in excluding a few trials per participant.

Bird images annotated by averaged LIME were preferred over annotations from LIME by a significant margin for both questions (p -value < 0.001 for both Question 1 and Question 2). All reported p -values are Bonferroni-corrected for 3 pairwise Mann–Whitney U tests. Broadly, subjects exhibit a significant preference for averaged LIME over standard LIME. Question 1 was given a median rating of 4.0 (“Neutral”) for LIME images. This increased to 5.5 (“Slightly Agree”/“Agree”) for averaged LIME images. Question 2 was given a median rating of 3.0 (“Slightly Disagree”) for LIME images. Averaging over an ensemble increased this to (“Slightly Agree”).



Please identify the bird to the left:

Barrow's Goldeneye Common Goldeneye
 Black-headed Grosbeak Blue Grosbeak
 Clark's Grebe Western Grebe
 Eastern Towhee Spotted Towhee
 Indigo Bunting Lazuli Bunting

Please indicate your opinions on the highlighting to the left:

Strongly Disagree	Slightly Disagree	Neutral	Slightly Agree	Strongly Agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

This highlighting emphasizes the areas of the bird that I think are important for identification.

I would recommend using this highlighting to help identify this bird.

[Continue](#)

Figure 7. An example screen capture from the study.

An alternative method of analyzing the responses keeps all trials, including those in which the subject incorrectly identified the bird's species. This may be reasonable since the birder may know for example that the coloring around the eye is important for distinguishing grebes, but not recall whether the Western or Clark's grebe has black below the eye. Including all trials, even those with incorrect classifications do not significantly differ in the distribution of preference ratings for any question for any highlighting method; averaged LIME remains significantly preferred over regular LIME for both questions at least with p -value < 0.001 .

Subjects in general thought the ensembles led to improved highlighting areas that are important for identification and would recommend using that over the regions identified by a single model.

7. FUTURE WORK

We have shown that a simple average of networks increases the accuracy of explanations. However, in ensemble learning for classification accuracy, more complex methods such as boosting [12] have been developed. Boosting generates models that focus on correcting errors of other ensemble members and differentially weights the contributions of different ensemble members. This leads us to consider alternative combination algorithms.

One challenge is that boosting requires a ground truth classification, while we do not use ground truth explanations in training. A possible approach is to adopt methods in which weights are a function of correlations between classifications [1] to the situation where there are correlations across explanations.

Our user-study used LIME as the XAI algorithm. Our future work will explore other XAI algorithms. Our ultimate goal is to develop and evaluate with end users a new method that produces explanations similar to the expert explanation in Figure 2b with labeled regions. To achieve this, we need to accurately identify the regions, and the ensemble technique in this paper is an important step toward this goal.

8. CONCLUSION

XAI algorithms were developed to increase trust in deep learning algorithms for tasks such as image classification. However, XAI algorithms themselves need to be trustworthy. It has been shown that differences in training and initial conditions can produce different explanations and that the explanations of current XAI systems fail to identify all regions of importance used by human experts.

In this paper, inspired by the success of ensembles to increase classification accuracy, we proposed using ensembles to improve the explanation accuracy of saliency-based XAI algorithms. We show, through empirical results, that ensembles can improve the accuracy of explanations when measured using metrics such as IoU, correlation, and center of mass distance. Furthermore, we showed that explanations produced by ensembles are preferred by people over explanations produced by a single network. By looking for areas of consensus across multiple networks, ensembles reduce the irrelevant areas and increase the relevant areas in explanation.

ACKNOWLEDGEMENTS

This work was supported with funding from the DARPA Explainable AI Program under a contract from NRL and from NSF grant 2026809. We would like to thank Dave Gunning for stimulating XAI research. Discussions with David Aha, Albert Hsiao, and Justin Huynh were helpful in developing the ideas in this paper.

REFERENCES

- [1] K. M. Ali and M. J. Pazzani, "Error reduction through learning multiple descriptions," *Machine learning*, vol. 24, no. 3, pp. 173–202, 1996.
- [2] N. Arun, N. Gaw, P. Singh, K. Chang, M. Aggarwal, B. Chen, K. Hoebel, S. Gupta, J. Patel, M. Gidwani et al., "Assessing the trustworthiness of saliency maps for localizing abnormalities in medical imaging," *Radiology: Artificial Intelligence*, vol. 3, no. 6, p. e200267, 2021.
- [3] P. P. Brahma, D. Wu, and Y. She, "Why deep learning works: A manifold disentanglement perspective," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 10, pp. 1997–2008, 2016.
- [4] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [5] S. Chakraborty et al., "Interpretability of deep learning models: A survey of results," in *2017 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, Internet of people and smart city innovation (smartworld/SCALCOM/UIC/ATC/CBDcom/IOP/SCI)*. IEEE, 2017, pp. 1–6.
- [6] S. Chakraborty, R. Tomsett, R. Raghavendra, D. Harborne, M. Alzantot, F. Cerutti, M. Srivastava, A. Preece, S. Julier, R. M. Rao et al., "Interpretability of deep learning models: A survey of results," in *2017 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, Internet of people and smart city innovation (smartworld/SCALCOM/UIC/ATC/CBDcom/IOP/SCI)*. IEEE, 2017, pp. 1–6.
- [7] N. C. F. Codella et al., "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC)," *CoRR*, vol. abs/1902.03368, 2019. [Online]. Available: <http://arxiv.org/abs/1902.03368>
- [8] L. A. de Souza Jr, R. Mendel, S. Strasser, A. Ebigbo, A. Probst, H. Messmann, J. P. Papa, and C. Palm, "Convolutional neural networks for the evaluation of cancer in barrett's esophagus: Explainable ai to lighten up the black-box," *Computers in Biology and Medicine*, vol. 135, p. 104578, 2021.
- [9] T. G. Dietterich, "Ensemble methods in machine learning," in *International workshop on multiple classifier systems*. Springer, 2000, pp. 1–15.
- [10] P. Domingos, "A unified bias-variance decomposition," in *Proceedings of 17th International Conference on Machine Learning*, 2000, pp. 231–238.
- [11] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," *Frontiers of Computer Science*, vol. 14, no. 2, pp. 241–258, 2020.
- [12] Y. Freund, R. E. Schapire et al., "Experiments with a new boosting algorithm," in *icml*, vol. 96. Citeseer, 1996, pp. 148–156.
- [13] M. Gams, "New measurements highlight the importance of redundant knowledge," in *Proceedings of the 4th European Working Session on Learning (EWSL89)*, 1989, pp. 71–80.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.
- [16] S. Lapuschkin, A. Binder, G. Montavon, K.-R. Müller, and W. Samek, "The lrp toolbox for artificial neural networks," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 3938–3942, 2016.
- [17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [18] Y. Liu and X. Yao, "Ensemble learning via negative correlation," *Neural networks*, vol. 12, no. 10, pp. 1399–1404, 1999.
- [19] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.
- [20] S. M. Lundberg and S. Lee, "A unified approach to interpreting model predictions," *CoRR*, vol. abs/1705.07874, 2017. [Online]. Available: <http://arxiv.org/abs/1705.07874>
- [21] Mar, V.J., Soyer, H., Button-Sloan, A., Fishburn, P., Gyorki, D.E., Hardy, M., Henderson, M. and Thompson, J.F., 2020. Diagnosis and management of cutaneous melanoma. *Australian journal of general practice*, 49(11), pp.733-739.
- [22] M. Pazzani, R. K. Severine Soltani, S. Qian, and A. Hsiao, "Expert-informed, user-centric explanations for machine learning," in *Proceedings of the AAAI Conference on Artificial Intelligence-2022*. IOS Press, 2022.
- [23] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you?" explaining the predictions of

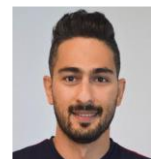
- any classifier,” in Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144.
- [24] L. Rieger and L. K. Hansen, “Aggregating explainability methods for neural networks stabilizes explanations,” arXiv preprint arXiv:1903.00519, 2019.
- [25] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K.-R. Müller, “Evaluating the visualization of what a deep neural network has learned,” IEEE transactions on neural networks and learning systems, vol. 28, no. 11, pp. 2660–2673, 2016.
- [26] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in Proceedings of the IEEE international conference on computer vision, 2017, pp. 618–626.
- [27] A. Shrikumar, P. Greenside, A. Shcherbina, and A. Kundaje, “Not just a black box: Learning important features through propagating activation differences,” arXiv preprint arXiv:1605.01713, 2016.
- [28] K. Simonyan, A. Vedaldi, and A. Zisserman, “Deep inside convolutional networks: Visualising image classification models and saliency maps,” arXiv preprint arXiv:1312.6034, 2013.
- [29] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.
- [30] A. Singh, S. Sengupta, A. R. Mohammed, I. Faruq, V. Jayakumar, J. Zelek, V. Lakshminarayanan et al., “What is the optimal attribution method for explainable ophthalmic disease classification?” in International Workshop on Ophthalmic Medical Image Analysis. Springer, 2020, pp. 21–31.
- [31] M. Sundararajan, A. Taly, and Q. Yan, “Axiomatic attribution for deep networks,” in International conference on machine learning. PMLR, 2017, pp. 3319–3328.
- [32] M. Watson, B. A. S. Hasan, and N. Al Moubayed, “Agree to disagree: When deep learning models with identical architectures produce distinct explanations,” in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 875–884.
- [33] X. L. Weina Jin and G. Hamarneh, “Evaluating explainable ai on a multi-modal medical imaging task: Can existing algorithms fulfill clinical requirements?” in Proceedings of the AAAI Conference on Artificial Intelligence-2022. IOS Press, 2022.

AUTHORS

Aadil Ahamed received his M.S. in Computer Science, Specialization in Artificial Intelligence from University of California, San Diego and his B.S. in Computer Engineering from University of California, Irvine.



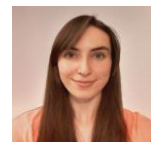
Kamran Alipour received his PhD in Computer Science from University of California, San Diego in 2022 and M.S. in Aerospace Engineering from Sharif University of Technology and a B.S. in Aerospace from K. N. Toosi University of Technology



Sateesh Kumar is a Master's student at the Computer Science and Engineering department of UC San Diego. He obtained his bachelor's in Computer Science from National University of Computer and Emerging Sciences, Pakistan.



Severine Soltani completed her B.S. in Cognitive Science with a focus on machine learning and neural computation at UC San Diego in 2020. She is currently a Bioinformatics and Systems Biology Ph.D. student at UC San Diego, examining the effects of environmental perturbations on physiological data via wearable devices.



Michael Pazzani received his Ph.D. in Computer Science from University of California, Los Angeles. He is director of the Artificial Intelligence Research for Health Center at the Information Sciences Institute in Marina Del Rey, CA, USA



© 2022 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.

The NP-completeness of quay crane scheduling problem

Ali Skaf, Samir Dawaliby, and Arezki Aberkane

Caplog R&D, Paris, France

Abstract. This paper discusses the computational complexity of the quay crane scheduling problem (QCSP) in a maritime port. To prove that a problem is NP-complete, there should be no polynomial time algorithm for the exact solution, and only heuristic approaches are used to obtain near-optimal solutions but in reasonable time complexity. To address this, first we formulate the QCSP as a mixed integer linear programming to solve it to optimal, and next we theoretically prove that the examined problem is NP-complete.

Keywords: Quay crane, Container, Scheduling, Optimization, MILP, NP-complete.

1 Introduction

In any maritime port, several machines exist to unload/load containers, and quay cranes are allowed to unload containers from the vessels or load containers into the vessels. Each vessel is divided into several bays, and each bay contains several containers. (Fig. 1)

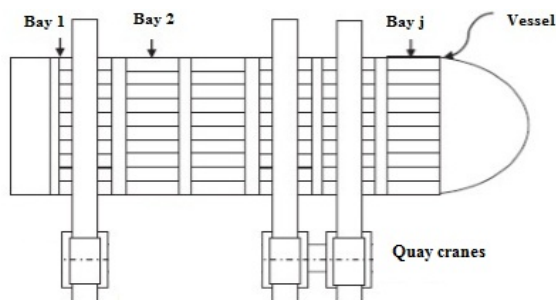


Fig. 1. Vessel-Quay cranes-Bays-Containers

In this paper, we consider the quay crane scheduling problem (QCSP) in the port of Tripoli-Lebanon. First, we present a novel mathematical Mixed Integer Linear Programming (MILP) model, and next we prove that the QCSP is NP-complete. A problem is NP-complete (Fig. 2) when it can be solved by a restraint class of

search algorithms and can be used to simulate any other problem with a similar algorithm. More precisely, each entry in the problem must be associated with a set of solutions of polynomial length, in which the validity can be tested quickly (in polynomial time for example) as the output for any input is "yes" if the solution set is not empty and "no" if it is empty. The complexity class of problems of this form is called NP, which stands for "non-deterministic polynomial time".

The theory of NP-completeness is designed by [1] HR Lewis (1983).

A problem is NP-complete if:

- it is in NP
- it is NP-hard

For several problems that have only NP solutions, there exist polynomial time algorithms that produce good approximations of the optimal solutions.

For a given NP problem, if there is a correct answer in polynomial time, then there is a correct answer in polynomial time for all the full NP problems.

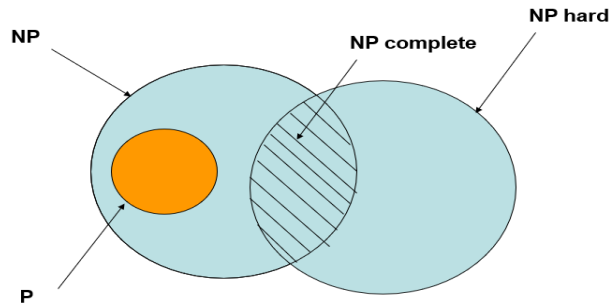


Fig. 2. NP-complete

Formally, we define the notion of polynomial reduction as follows: let A and B be two problems, with the aim of using the solution of A to solve B. (Fig. 3)

- A reduction of A to B is a polynomial time algorithm transforming any instance of B into an instance of A.
- Thus, if we have an algorithm to solve A, we also know how to solve B.
- A is therefore at least as difficult to solve as B.
- B is at most as difficult to solve as A.

We also write $B \leq A$.

The remainder of this paper is organized as follows: In section 2, we present a

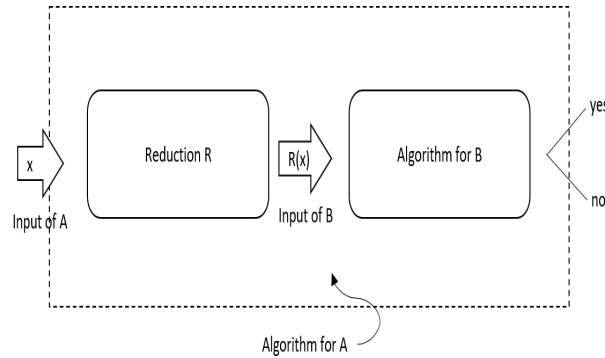


Fig. 3. Polynomial

literature review for the NP-complete problems, then in section 3, we describe in detail the new proposed mathematical model formulation for the QCSP problem. Next, we provide the NP-Completeness proof and discussion in section 4. Finally, some concluding remarks are made in section 5.

2 Literature review

[2] Plesn et al. (1979) proved that the Hamiltonian cycle problem is NP-complete even in the case of planar digraphs with out-degrees at most 2. Clearly, the bound two on degrees is the best possible. A Hamiltonian cycle in a graph or digraph is a cycle containing all the points. Thus any such cycle has p points as well as p lines (arcs) if the graph (digraph) has p points. No elegant characterization of the graphs or digraphs which possess Hamiltonian cycles exists, although the problem is at least one hundred years old. As the problem is a special case of the famous traveling salesman problem, it is interesting also from the computational viewpoint.

[3] Tovey (1984) proved that 3-SAT is NP-complete when restricted to instances where each variable appears in at most four clauses. When no variable appears in more than three clauses, 3-SAT is trivial and SAT is NP-complete. When no variable appears in more than two clauses, SAT may be solved in linear time.

[4] Ouyang et al. (1997) represent further evidence for the ability of DNA computing to solve NP-complete search problems. A pool of DNA molecules corresponding to the total ensemble of six-vertex cliques was built, followed by a series of selection processes. The algorithm is highly parallel and has satisfactory fidelity. This work represents further evidence for the ability of DNA computing to solve NP-complete search problems.

[5] Kaye (2000) demonstrates that a puzzle based on the Minesweeper game is NP-complete makes this important computer science topic accessible to high school students. The resource described here is a set of slides showing the detailed solution

of two introductory puzzles, following by the step-by-step simulation of digital circuit elements required for proving NP-completeness.

[6] Liu et al. (2002) show further evidence for the ability of DNA computing to solve NP-Complete problems. The graph-theoretic parameter that has probably received the most attention over the years is the chromatic number. As is well-known, the coloring problem is an NP-Complete problem. In their study, it has been solved by means of molecular biology techniques. The algorithm is highly parallel and has satisfactory fidelity. This work shows further evidence for the ability of DNA computing to solve NP-Complete problems.

[7] Kellerer et al. (2004) present an introduction to NP-Completeness of Knapsack Problems. The reader may have noticed that for all the considered variants of the knapsack problem, no polynomial time algorithm have been presented which solves the problem to optimality. Indeed all the algorithms described are based on some kind of search and prune methods, which in the worst case may take exponential time. It would be a satisfying result if we somehow could prove it is not possible to find an algorithm which runs in polynomial time, somehow having evidence that the presented methods are "as good as we can do". However, no proof has been found showing that the considered variants of the knapsack problem cannot be solved to optimality in polynomial time.

[8] Bresar et al. (2011) prove that the problem of determining is NP-complete for every path ≥ 2 . For path =2 this equals to the vertex cover problem (shortly VCP) which is known to be NP-complete.

The importance of the TSP presented by [9] Hoffman et al. (2013) is that it is representative of a larger class of problems known as combinatorial optimization problems. The TSP problem belongs in the class of such problems known as NP-complete. Specifically, if one can find an efficient (i.e., polynomial-time) algorithm for the traveling salesman problem, then efficient algorithms could be found for all other problems in the NP-complete class.

In our previous papers, [11] Skaf et al. (2018) suggested two methods to solve the quay crane scheduling problem at port of Tripoli-Lebanon, to minimize the total completion time for all containers from the vessel to the storage location. Later [12] Skaf et al. (2019) proposed for the same problem a genetic algorithm to obtain near-optimal solutions in an acceptable CPU time. After that, [13] Skaf et al. (2021) detailed the model with results comparison with literature and benchmark.

In this study, we present a mixed-integer linear programming model to solve the quay crane scheduling problem and a proof that this problem is NP-complete.

3 Mathematical formulation

In this section, we present the mathematical formulation of the QCSP with the different variables, parameters and the mathematical model.

3.1 Assumptions

We consider in this study a single vessel to transport containers. It is divided into several bays which in turn contains several containers. We also consider multi-quay cranes to unload/load container from/to the vessel with interference constraints between them (each bay is handled by at most one quay crane at a time). Each quay crane can unload/load one container at a time and the most important constraint here is that a quay crane should complete its unloading/loading in a current bay before passing to another one. Finally, we ignore the travel time between bays, and we do not consider any idle quay crane in this study.

3.2 Data

The notation that was used in the proposed formulation of the problem is shown below:

- Q : set of quay cranes
- $|Q|$: number of quay cranes
- i : index of quay cranes i ($\forall i \in Q$)
- B : set of bays
- $|B|$: number of bays
- j, j' : index of bays j and j' ($\forall j, j' \in B$)
- C_j : number of containers in bay j .
- Tc : time required for a quay crane to unload a container and place it in the storage location. It is not the same for all containers in the different bays
- M : large integer number

3.3 Decision variables

The decision variables that was used in the proposed formulation of the problem is shown below:

- $x_{(j,i)} = \begin{cases} 1 & \text{if quay crane } i \text{ unload container from bay } j \\ 0 & \text{otherwise} \end{cases}$
- $z_{(j,j')} = \begin{cases} 1 & \text{if the containers unloading in bay } j \text{ finishes before the starting} \\ & \text{of the containers unloading in bay } j' \\ 0 & \text{otherwise} \end{cases}$
- t_j : completion time of bay j
- C_{max} : makespan

3.4 Mixed integer linear programming

The mixed integer linear formulation (MILP) was founded by [10] Leonid Kantorovich (1939). It is a technique for optimizing an objective function, subject to linear inequality and equality constraints.

The MILP we propose is as follows:

Objective

$$\text{Minimize } C_{max} \quad (1)$$

Equation (1) is the objective function : minimize the makespan (completion time of all bays).

Subject to

$$\sum_{i=1}^{|Q|} x_{(j,i)} = 1 \quad \forall j \in B \quad (2)$$

Constraint (2) : each bay must be handled only by one quay crane.

$$t_j \geq (Tc \cdot C_j) \quad \forall j \in B \quad (3)$$

Constraint (3) : the completion time is bigger than the working time in each bay (working time = number of containers \times time needed to unload a container by a quay crane and store it into the storage location).

$$t_j - t_{j'} + (Tc \cdot C_{j'}) + z_{(j,j')} \cdot M > 0 \quad \forall j, j' \in B \quad (4)$$

Constraint (4) : if $z_{(j,j')} = 1$, then bay j finishes before bay j' starts.

$$t_j - t_{j'} + (Tc \cdot C_{j'}) - (1 - z_{(j,j')}) \cdot M \leq 0 \quad \forall j, j' \in B \quad (5)$$

Constraint (5) : if $z_{(j,j')} = 0$, bay j finishes after bay j' starts.

$$\sum_{i=1}^{|Q|} i \cdot x_{(j,i)} + 1 \leq \sum_{i'=1}^{|Q|} i' \cdot x_{(j',i')} + (z_{(j,j')} + z_{(j',j)}) \cdot M$$

$$\forall j, j' \in B, j < j' \quad (6)$$

Constraint (6) avoid the interference between the quay cranes. If bays j and j' are performed simultaneously, this means that $(z_{(j,j')} + z_{(j',j)}) = 0$.
If quay crane i and i' work in bay j and j' respectively, then $i+1 \leq i'$.

$$C_{max} = \max_j t_j \quad (7)$$

Constraint (7) defines the C_{max} value.

$$x_{(j,i)} = \{0, 1\} \quad \forall j \in B, \forall i \in Q \quad (8)$$

$$z_{(j,j')} = \{0, 1\} \quad \forall j, j' \in B, j < j' \quad (9)$$

Constraints (8) and (9) define the property of the decision variables.

4 NP-completeness

In this section we discuss the complexity of the quay crane scheduling problem.

Before getting in details, we need some parameters in this discussion. Let's suppose that X is a set of positive integer, the working time $(Tc \cdot C_j) \in X$ ($1 \leq j \leq |B|$) and Y is a given number which $Y \in X$.

The objective is to find all the schedules of the $|Q|$ quay cranes which unload containers from $|B|$ bays without interfering with each other and whose total completion time is less or equal to Y .

In the following, we present 4 phases to proof that the quay crane scheduling problem is NP-complete:

Phase 1: Indicate that the QCSP is in NP.

The feasibility of a given quay crane schedule for the QCSP problem can be analyzed in polynomial time, taking into account that the schedule must satisfy the non-interference constraints and thus can be realized in $O(|B|^2)$ time.

After that, we check that the completion time (makespan) $\leq Y$ which can be done in $O(|B|)$ time. Therefore, the QCSP is in NP.

Phase 2: Choosing an NP-complete problem.

The partition problem is a very known NP-complete problem ([1] HR Lewis (1983)).

The steps of this problem are defined as follows:

There are $|B|$ elements in a fixed set $L = \{l_1, l_2, \dots, l_{|B|}\}$. For each element $l_j \in L, l_j \in X$ and the sum of all elements $\sum_{l_j \in L} l_j = Z$.

→ Can the set L be divided into two separate sub-sets L_1 and L_2 such that $\sum_{l_j \in L_1} l_j = \sum_{l_j \in L_2} l_j = Z/2$?

Problem explanation with a numerical example: We suppose that the set $L = \{80, 117, 57, 100, 58, 110, 174\}$ and the sum of all elements $\sum_{l_j \in L} l_j = 696$. The answer to the previous question is Yes because the set L can be divided into two separate sub-sets $L_1 = \{80, 110, 58, 100\}$ and $L_2 = \{57, 117, 174\}$ such that $\sum_{l_j \in L_1} l_j = \sum_{l_j \in L_2} l_j = Z/2 = 348$. (See Fig. 4.)

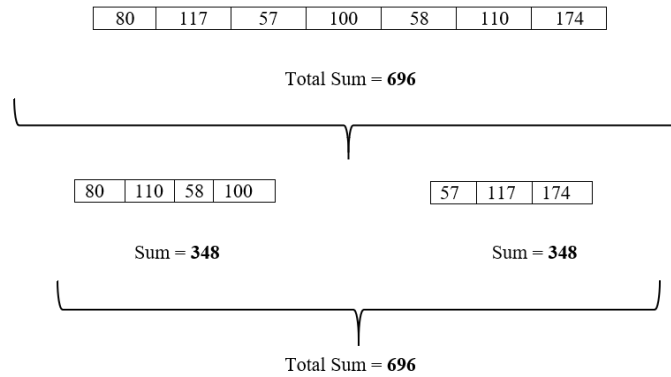


Fig. 4. Two subsets

The partition is obtained using the function below (developed in JAVA):

```
static int SplitPoint(int array [], int n){
    int lSum = 0;
    for (int i = 0; i < n; i++) {
        lSum += array[i];
        int rSum = 0;
        for (int j = i+1 ; j < n ; j++ )
            rSum += array[j];
        if (lSum == rSum)
            return i+1;
    }
    return -1;
}
```

The above function performs the following features:

- Traversing the array elements
- Adding the current element to a variable
- Finding the sum of the rest of the array elements
- Splitting the point index
- Returning -1 if it is not possible to split the array into two parts

Phase 3: Making a conversion from the partition problem to the quay crane scheduling problem.

The partition problem is switched to the QCSP as follows: An instance of the QCSP corresponding to an instance of the partition problem, has Q quay cranes and $|B| + |Q|$ bays. The given positive number X is set as Z , then the following equations (10, 11 and 12) show the processing time of each bay, in another meaning the processing time of bay 1 and bay $|B| + 2$ is set as $Z/2$, the processing time of bay 2 to bay $|B| + 1$ is set as l_1 to $l_{|B|}$, respectively, and the processing time of bay $|B| + 3$ to bay $|B| + |Q|$ is set as Z .

$$Tc \cdot C_1 = Tc \cdot C_{|B|+2} = Z/2 \quad (10)$$

$$Tc \cdot C_{j+1} = l_j \quad 1 \leq j \leq |B| \quad (11)$$

$$Tc \cdot C_j = Z \quad \forall |B| + 3 \leq j \leq |B| + |Q| \quad (12)$$

Table 1 shows this conversion. It indicates $|Q|$ quay cranes, $|B| + |Q|$ bays and the processing time of each bay.

After that, we must prove that the set L can be divided into two separate sub-sets L_1 and L_2 such that $\sum_{l_j \in L_1} l_j = \sum_{l_j \in L_2} l_j = Z/2$ **if** all the $|B| + |Q|$ bays can be completed by $|Q|$ quay cranes in Z time without interference between all quay cranes.

Firstly, we suppose that the set L can be divided into two separate sub-sets L_1 and L_2 such that $\sum_{l_j \in L_1} l_j = \sum_{l_j \in L_2} l_j = Z/2$.

Then, the scheduling of the $|Q|$ quay cranes without interference is as follows:

- Quay Crane 1 unloads containers from all the bays $j + 1$, where $l_j \in L_1$ and then bay 1.
- Quay Crane 2 unloads containers from bay $|B| + 2$, and then all the bays $j + 1$, where $l_j \in L_2$.
- Quay Cranes 3 unloads containers from bay $|B| + 3$.
- ...
- Quay Crane $|Q|$ unloads containers from bay $|B| + |Q|$.

Clearly, in this schedule the latest completion time between all bays is Z without any interference. Therefore, if the set L can be divided into two separate sub-sets L_1 and L_2 such that $\sum_{l_j \in L_1} l_j = \sum_{l_j \in L_2} l_j = Z/2$, all the $|B| + |Q|$ bays can be completely handled by $|Q|$ quay cranes in Z time without interference between all quay cranes.

Table 1. Phase 3 : Transformation

Quay crane	Bay Number	Bay Processing time
1	1	$D/2$
	2	l_1
	3	l_2

	$ B $	$l_{ B -1}$
	$ B + 1$	$l_{ B }$
2	$ B + 2$	$D/2$
3	$ B + 3$	D
4	$ B + 4$	D
...
$ Q $	$ B + Q $	D

In addition, suppose all the $|B| + |Q|$ bays can be completely handled by $|Q|$ quay cranes in Z time without any interference between them, then all the Q quay cranes are fully used as the sum of the processing time of all the bays is $|Q| \cdot |B|$.

Then, the completion time (makespan) of each quay crane must be equal to Z . Moreover, there is no interference in the previous quay crane schedule. Accordingly, the sum of the processing time of all bays, excepting bay 1 handled by Quay Crane 1 should be $Z/2$ and the sum of the processing time of all bays, excepting bay $|B| + 2$ handled by Quay Crane 2 should be $Z/2$, which means that the set L can be

divided into two separate sub-sets L_1 and L_2 such that $\sum_{l_j \in L_1} l_j = \sum_{l_j \in L_2} l_j = Z/2$. Subsequently, if all the $|B| + |Q|$ bays can be completely handled by $|Q|$ quay cranes in Z time without interference between them, the set L can be divided into two separate sub-sets L_1 and L_2 such that $\sum_{l_j \in L_1} l_j = \sum_{l_j \in L_2} l_j = Z/2$.

Phase 4: Proof that the above conversion is polynomial. The above conversion can be done in $O(|B| + |Q|)$ time. Therefore, PARTITION PROBLEM \propto QCSP, and finally the theory is proved.

5 Conclusion

In this paper, we have investigated the problem of quay crane scheduling in a maritime port. The objective is how to orchestrate the planning of quay cranes that unload containers from bays without interfering with each other so that the total completion time is minimized. To this end, first we proposed a novel mixed integer linear programming model to solve the studied problem to optimality, and then we provided a theoretical analysis to prove that QCSP is NP-complete.

References

1. H. R. Lewis, "Michael r. πgarey and david s. johnson. computers and intractability. a guide to the theory of np-completeness. wh freeman and company, san francisco 1979, x+ 338 pp.," *The Journal of Symbolic Logic*, vol. 48, no. 2, pp. 498–500, 1983.
2. J. Plesn *et al.*, "The np-completeness of the hamiltonian cycle problem in planar diagraphs with degree bound two," *Information Processing Letters*, vol. 8, no. 4, pp. 199–201, 1979.
3. C. A. Tovey, "A simplified np-complete satisfiability problem," *Discrete applied mathematics*, vol. 8, no. 1, pp. 85–89, 1984.
4. Q. Ouyang, P. D. Kaplan, S. Liu, and A. Libchaber, "Dna solution of the maximal clique problem," *Science*, vol. 278, no. 5337, pp. 446–449, 1997.
5. R. Kaye, "Minesweeper is np-complete," *The Mathematical Intelligencer*, vol. 22, no. 2, pp. 9–15, 2000.
6. Y. Liu, J. Xu, L. Pan, and S. Wang, "Dna solution of a graph coloring problem," *Journal of chemical information and computer sciences*, vol. 42, no. 3, pp. 524–528, 2002.
7. H. Kellerer, U. Pferschy, and D. Pisinger, "Introduction to np-completeness of knapsack problems," pp. 483–493, 2004.
8. B. Brešar, F. Kardoš, J. Katrenič, and G. Semanišin, "Minimum k-path vertex cover," *Discrete Applied Mathematics*, vol. 159, no. 12, pp. 1189–1195, 2011.
9. K. L. Hoffman, M. Padberg, G. Rinaldi, *et al.*, "Traveling salesman problem," *Encyclopedia of operations research and management science*, vol. 1, pp. 1573–1578, 2013.
10. L. V. Kantorovich, "Mathematical methods of organizing and planning production," *Management science*, vol. 6, no. 4, pp. 366–422, 1960.
11. A. Skaf, S. Lamrous, Z. Hammoudan, and M.-A. Manier, "Exact method for single vessel and multiple quay cranes to solve scheduling problem at port of tripoli-lebanon," pp. 457–461, 2018.
12. A. Skaf, S. Lamrous, Z. Hammoudan, and M.-A. Manier, "Genetic algorithm to optimize unloading of large containers vessel in port of tripoli-lebanon," pp. 569–574, 2019.
13. A. Skaf, S. Lamrous, Z. Hammoudan, and M.-A. Manier, "Solving methods for the quay crane scheduling problem at port of tripoli-lebanon," *RAIRO-Operations Research*, vol. 55, no. 1, pp. 115–133, 2021.

A CONTEXT-AWARE AND ADAPTIVE SYSTEM TO AUTOMATE THE CONTROL OF THE AC WINDSHIELD USING AI AND INTERNET OF THINGS

Joshua Tian¹ and Yu Sun²

¹Arnold O. Beckman High School, 3588 Bryan Ave, Irvine, CA 92602

²California State Polytechnic University, Pomona, CA, 91768, Irvine, CA 92620

ABSTRACT

In recent years, we have seen a huge increase in air conditioning usage [4]. However, much of this energy put into air conditioning is being wasted, which contributes to a far less environmentally friendly world and is inconvenient for many [5][6]. This paper develops a smart vent and a mobile app to regulate temperatures in different rooms of a home to create an efficient solution to save energy. This conservation of energy allows both the environment to be preserved as well as the financial burden of families in need to be alleviated. Controlled studies of the system provide evidence of the system's automated ability to be energy efficient.

KEYWORDS

AI, AC, Raspberry Pi.

1. INTRODUCTION

Air Conditioners are a commonality throughout the world right now, with 87 percent of US homes equipped with some sort of cooling. However, with such a large amount of air conditioning every year comes a ridiculously high 100 million tons of carbon dioxide being released every year [10].

To address the issue of wasted energy, while there are many methods to harness more energy efficient forms of energy such as solar energy, these methods not only are extremely expensive, but also fairly inefficient, which makes using these alternative sources of energy impractical [7]. Although there are other practices to save energy without becoming more energy-efficient, these practices are purely based on the user's choice of using less energy and are therefore often deemed inconvenient. Without accountability, this inconvenience leads to many of these energy-conserving practices to be discontinued after a short period of time.

In this paper, we aim to create an energy efficient smart vent capable of being placed into an everyday home. This smart vent is controlled by a mobile app which allows the user to view the current temperature of the house as well as setting the target temperature of the room that the vent is in. By setting the app into heating or cooling mode, the smart vent will open or close to let the temperature of the house reach the user's target temperature. In addition, the user can manually open or close the vent by simply clicking a button on the app. Compared to alternative methods of conserving energy, the smart vent is a far more sophisticated solution, as it not only impacts

the environment by conserving energy, but also lets the user adjust to different situations by using a mobile app [8]. Therefore, we believe that the smart vent is one of the best ways to help cut down wasted energy.

In multiple application scenarios, we demonstrate how the smart vent saves energy. Our goal is to compare the smart vent with non-smart vent energy levels. After finding the BTU for our AC system and how efficient it is, we pick a temperature that is common during the week, and then make note of this temperature as the starting temperature [9]. Next, we set the thermostat 5 degrees below the starting temperature. First, we use the smart vent with it being closed and then measure the time it takes for the thermostat to reach the target temperature. Then we use the same approach with the non-smart vent which is always opened. Finally, we have the target temperature in the room at a midway point so that the vent opens or closes in the middle. Then using the amount of time that the AC was running, along with the previous information, we calculate the amount of energy used. Through the final experiment, we can also see how after the smart vent closes halfway through the thermostat target temperature, the room cools or heats far quicker because it no longer needs to alter the temperature of that room. This is especially important, as it allows a house's many different rooms to stay at different temperatures.

The rest of the paper is organized as follows: Section 2 gives the details on the challenges that we met during the experiment and designing the sample; Section 3 focuses on the details of our solutions corresponding to the challenges that we mentioned in Section 2; Section 4 presents the relevant details about the experiment we did, following by presenting the related work in Section 5. Finally, Section 6 gives the conclusion remarks, as well as pointing out the future work of this project.

2. CHALLENGES

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

2.1. Data reading and input

One of the most fundamental parts of this project is being able to read the temperature in the room and then making decisions off of it. Initially, in order to effectively read the temperature repeatedly over a long period of time, we decided to create a complex circuit system. This system is composed of a raspberry pi connected to a breadboard [11]. The breadboard would have all of the pins for the motor, and more importantly the pins for the digital temperature sensor. This temperature sensor would be able to constantly take in the temperature and then immediately have it be sent to the Raspberry Pi which would ultimately use that information for a multitude of things. However, this solution was far too complex and took up too much space to the point where we couldn't fit it in the enclosed area in which we wanted to test it in. This becomes a problem because it ends up extremely difficult to test and apply to real world settings. To combat this, we have tried to use a shorter breadboard.

2.2. Data communication and storage

In order for the motor to be powered and the app to function, data communication and storage is essential [12]. In the cloud database Firebase, we store information such as whether the thermostat is heating or cooling, whether the vent is open or closed, the current and target temperature, and more. The Raspberry Pi also plays a huge role in this, as it both receives information from Firebase and sends information into Firebase. When the digital thermometer gets an updated temperature, the Raspberry Pi immediately transfers that data to Firebase and our

backend application program likewise immediately takes that information to update the app. In addition, along with the temperature given by the thermometer, the Raspberry Pi also checks the many different conditions that we use to ensure the vent is opened and closed properly.

2.3. Updating the information on the app live

As an application that is centered around providing and changing information, a key part of that is to be able to update the information on the app live. We accomplished this by creating database reference variables which corresponded to a specific cell in the Firebase database. This allowed us to store all of our information in one place and be able to constantly update each variable by just editing the reference variable and uploading it to the database. Likewise, we were able to take information from the database and immediately reflect those changes on the app. In addition, using different kinds of functions.

3. SOLUTION

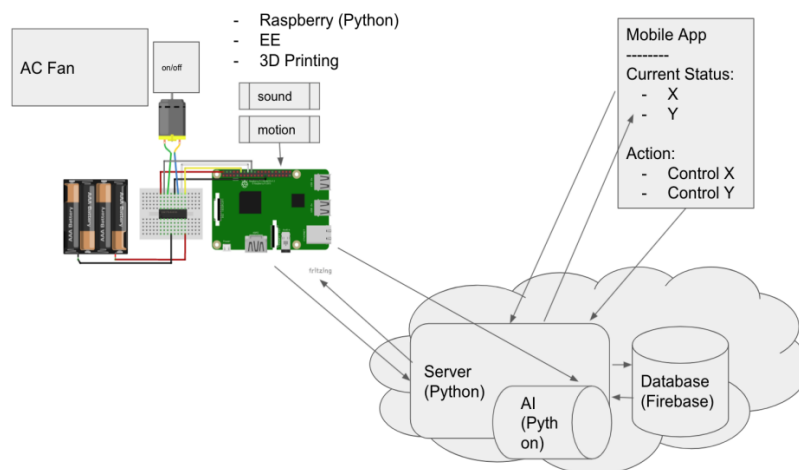


Figure 1. Overview of the solution

This solution is a smart vent system that allows the user to regulate the air flow in their home through altering different settings in a mobile app. The system has many components to ensure its proper function: The hardware and motorized parts, the raspberry pi and its code, the frontend application, and the backend cloud database. The application written in Dart in Flutter, can run on both iOS and Android devices, which allows for a diverse user base [13]. To start, we have modified a standard AC vent by connecting a 5V servo motor using linkage rods. The servo motor is attached to three of the pins on a breadboard which is circuited properly to supply the sufficient amount of power to the motor. In addition, a digital thermometer pin is also connected to the breadboard, allowing us to read the temperature of the surroundings. This breadboard is then naturally connected to the Raspberry Pi, which uses the built-in Python IDE to open or close the vent after reading the temperature. Subsequently, the Raspberry Pi will then upload the temperature reading to the Firebase database. The database stores many important pieces of information, such as the ID number of each vent, whether the thermostat system is cooling or heating, the current temperature, target temperature, and more. This information is not only used when determining whether the vent should be open or not, but is also the basis for the mobile app. The mobile app displays the current temperature and allows users to both manually control the vent or set it on auto mode. When the vent is in auto mode, the user will set a target

temperature and the vent will open or close depending on whether the thermostat is heating or cooling, and then maintain the target temperature until a new one is set. This app is developed in the Android studio IDE using Google’s open-source UI development kit Flutter. Flutter uses the object-oriented programming language, dart.

The tangible system of parts in this project is mainly composed of a 5V servo motor and a Raspberry Pi. The servo motor is fixed to the vent using a 3d printed motor mold that has M3 screw holes. The motor arm is connected to an M2 linkage rod which extends to screw into the arm of the vent. When the motor arm turns, it pulls or pushes the linkage rod which in turn pulls or pushes the vent arm to open or close the vent. The Raspberry Pi is attached to a breadboard which has the motor pins(GPIO18, 5V, and GND) connected. In addition, the digital thermometer used to intake the temperature readings is also connected to the breadboard.

When first opening the app, and logging in to an account, you are brought to a page that lists all the devices that you have created. At the bottom there is also a button that allows you to create a new device by naming it and giving it a unique 6 digit device ID. Next, by selecting one of your devices you can see the current temperature of the room, the different settings, and the target temperature which you can increase or decrease using buttons on the screen.

```

Future<List<Device>> getDevices() async{
  String uid = FirebaseAuth.instance.currentUser!.uid;
  print(uid);
  DataSnapshot ownerDeviceSnapshot = await FirebaseDatabase.instance.ref('userDevices/$uid').get();
  print(ownerDeviceSnapshot);
  List<Device> deviceList = [];
  if(ownerDeviceSnapshot.value != null){
    Map<dynamic,dynamic> deviceMap = ownerDeviceSnapshot.value as Map;

    for(String id in deviceMap.keys){

      DataSnapshot deviceInfoSnapshot = await FirebaseDatabase.instance.ref('devices/$id').get();

      if(deviceInfoSnapshot != null){
        Map<dynamic,dynamic> deviceInfo = deviceInfoSnapshot.value as Map;

        String deviceID = id;

        String deviceLabel = deviceInfo['label'];

        bool deviceOpen = deviceInfo['isOpen'];

        Device d = Device(deviceID, deviceLabel, deviceOpen);

        deviceList.add(d);

      }

    }
  }
  return deviceList;
}

```

Figure 2. Screenshot of App code

In this figure, we get all of the devices under a user by creating a snapshot and setting it as a reference to the user id in Firebase. We loop through all of the items in the deviceMap and for each item we store its device id, device label, and whether it is open or not. Next, using this information, we transfer those properties to a temporary device “d” and then add it to the device list which we ultimately end up returning. This device list is what you see on the homepage.

```

def readTemp():
    lines = readRawTemp()
    while lines[0].strip()[-3:] != 'YES':
        time.sleep(0.5)
        lines = readRawTemp()
    position = lines[1].find('t=')
    if position != -1:
        tempString = lines[1][position+2:]
        tempC = float(tempString)/1000
        tempF = tempC*9.0/5.0 +32.0
        return tempF
    return None

def controlVents(isHeating, targetTemp, currentTemp):
    if isHeating:
        if targetTemp>currentTemp:
            return ventOpen()
        else:
            return ventClose()
    else:
        if targetTemp<currentTemp:
            return ventClose()
        else:
            return ventOpen()

while True:
    temp = readTemp()
    if temp != None:
        temp = round(temp,1)
        print(temp)
        requests=requests.get(firebasepath)
        firebasejson= request.json()
        firebasejson['currentTemperature'] = temp

        if firebasejson['isManual']:
            if firebasejson['isOpen']:
                ventOpen()
            else:
                ventClose()
        else:
            targetTemp = firebasejson['targetTemperature']
            isHeating = firebasejson['isHeating']
            firebasejson['isOpen'] = controlVents(isHeating, targetTemp, temp)

        requests.put(firebasepath, json = firebasejson)

    else:
        print('temp is none')

    time.sleep(1.0)

```

Figure 3. Screenshot of RPi Code

This figure shows the main parts of the Raspberry Pi code which is constantly running to check the temperature and adjust the motor position as needed. As shown in the while loop, we call the readTemp function at the beginning and using this temperature reading, we first update it into the database so that it can be used for the app. Next we check if the vent is being manually controlled, and if it is, then we won't change anything. However, if the vent is on auto mode, then we take the target temperature and the thermostat setting from the database and then call controlVents, which uses some boolean comparisons to determine whether the vent should open or close.



Figure 4. Diagram of different parts of the database

These are the properties of one of my devices in Firebase. As shown, all of the necessary pieces of information that are shown on the app and used to open or close the vent are stored here. These

variables are constantly updated both by the user in the app and by the Raspberry Pi when it gets a new temperature reading.

4. EXPERIMENT

4.1. Experiment 1

To evaluate the effectiveness of our solution, we compared the energy cost for cooling down the home using a normal vent and our smart vent. We created a situation in which the room that the smart vent was in didn't need to be cooled down, and as such, the smart vent closed by itself. In order to accurately determine the effectiveness of our solution, we ran this experiment multiple times in different rooms.

The results of the experiment show that when using a standard AC vent, the house cooled by 3 degrees in a span of roughly an hour. However, with the smart vent in this situation being closed, it took between roughly 54-58 minutes to cool the house by 3 degrees depending on which room the smart vent was in.

The experiment results shows that by using the smart vent, under the scenario where the specific room reaches its desired temperature and the smart vent closes, it lets the rest of the house heat or cool roughly 5% quicker. This decrease in time to change the temperature in the house means that the AC will be running for a shorter period of time which leads to lower energy cost.

5. RELATED WORK

K.J. Chua presented a solution that decreases the base energy output from a standard chiller plant [1]. This solution aims to save energy by decreasing the amount of energy that was needed to cool down a home normally, rather than trying to decrease energy waste. This is a unique approach that can be extremely impactful, but by ignoring the wasted energy, the potential of this solution is limited.

Adnan A.Kinsara, Moustafa M.Elsayed, and Omar M.Al-Rabghi created a solution that used Carbon Chloride as a liquid desiccant [2]. This liquid desiccant can hold water vapor which allows it to be able to overcome the latent part of the air conditioning load. This solution makes use of the characteristics of a typical air conditioning system and creates a chemical solution to make the system energy efficient.

P. Hengjinda and Dr. Joy Chen proposed a solution using an arduino microcontroller that monitors both the room and temperature outside a window [3]. Their intelligent controller continuously monitors the surrounding temperature and opens the window when the surrounding temperature equals the room temperature. This solution is very effective because it makes use of the natural environment to help speed up the air conditioning process.

6. CONCLUSIONS

In this project, we proposed an energy efficient air conditioning system to address the immense amount of wasted energy that air conditioning systems create. This system consists of a 5V servo motor attached to a standard AC vent's arm which is then attached to a Raspberry Pi. There is also a digital thermometer pin attached to the Raspberry Pi which supplies information to a backend database [14]. In addition, we have a mobile app where users can put certain settings into place. Using the temperature readings, as well as the target temperature and other factors

determined by the user, the motor will open or close the vent. Experiments show that we save roughly 5% more energy by being able to close one vent in a room that doesn't need to be heated or cooled. By closing off a certain area of your home, it takes less time to cool the rest of the house which saves a lot of energy.

Currently the vent is a little bit inconvenient as there are many parts to the system which causes it to be extremely difficult to assemble. In addition, because it is battery reliant, it doesn't run for a very long time. There is also the issue of optimization with the thermometer readings, which could change depending on the location of the thermometer.

In the future, in order to find a more secure and optimal spot to place the Raspberry Pi and its thermometer, we will experiment with 3D printing as well as other methods to create a more stable system [15]. We will also try soldering the pins straight to the Raspberry Pi which will make it much more compact. In addition, I plan to add an A.I. based video recognition system that can recognize entry into the room and will open or close the vent based on user decision.

REFERENCES

- [1] Chua, Kian Jon, et al. "Achieving better energy-efficient air conditioning—a review of technologies and strategies." *Applied Energy* 104 (2013): 87-104.
- [2] Kinsara, Adnan A., Moustafa M. Elsayed, and Omar M. Al-Rabghi. "Proposed energy-efficient air-conditioning system using liquid desiccant." *Applied Thermal Engineering* 16.10 (1996): 791-806.
- [3] Hengjinda, P., Dr Chen, and Joy Iong Zong. "An Intelligent Feedback Controller Design for Energy Efficient Air Conditioning System." *Journal of Electronics and Informatics* 2.3 (2020): 168-174.
- [4] Ren, Xiaoxin, Da Yan, and Chuang Wang. "Air-conditioning usage conditional probability model for residential buildings." *Building and Environment* 81 (2014): 172-182.
- [5] Clark, James H., and Duncan J. Macquarrie. "Environmentally friendly catalytic methods." *Chemical Society Reviews* 25.5 (1996): 303-310.
- [6] Traore, Moussa K., and Gisela Buschle-Diller. "Environmentally friendly scouring processes." *Textile Chemist & Colorist & American Dyestuff Reporter* 32.12 (2000).
- [7] Cuéllar, Amanda D., and Michael E. Webber. "Wasted food, wasted energy: the embedded energy in food waste in the United States." *Environmental science & technology* 44.16 (2010): 6464-6469.
- [8] Neumerkel, René, and Stephan Groß. "A sophisticated solution for revealing attacks on wireless LAN." *International Conference on Trust, Privacy and Security in Digital Business*. Springer, Berlin, Heidelberg, 2006.
- [9] Kakigano, H., M. Nomura, and T. Ise. "Loss evaluation of DC distribution for residential houses compared with AC system." *The 2010 International Power Electronics Conference-ECCE ASIA-IEEE*, 2010.
- [10] Buchanan, Andrew H., and Brian G. Honey. "Energy and carbon dioxide implications of building construction." *Energy and Buildings* 20.3 (1994): 205-217.
- [11] Zhao, Cheah Wai, Jayanand Jegatheesan, and Son Chee Loon. "Exploring iot application using raspberry pi." *International Journal of Computer Networks and Applications* 2.1 (2015): 27-34.
- [12] Prabhu, Narahari Umanath. *Stochastic Storage Processes: queues, insurance risk, and dams, and data communication*. No. 15. Springer Science & Business Media, 1998.
- [13] Tracy, Kim W. "Mobile application development experiences on Apple's iOS and Android OS." *Ieee Potentials* 31.4 (2012): 30-34.
- [14] Hadley, I. C. D., and R. D. Gould. "Inexpensive digital thermometer for measurements on semiconductors." *International Journal of Electronics Theoretical and Experimental* 70.6 (1991): 1155-1162.
- [15] Shahrubudin, Nurhalida, Te Chuan Lee, and Rhaizan Ramlan. "An overview on 3D printing technology: Technological, materials, and applications." *Procedia Manufacturing* 35 (2019): 1286-1296.

ROBUST DISCRIMINATIVE NON-NEGATIVE MATRIX FACTORIZATION WITH MAXIMUM CORRENTROPY CRITERION

Hang Cheng, Shixiong Wang and Naiyang Guan

National Innovation Institute of Defense Technology,
Academy of Military Sciences, Beijing, China

ABSTRACT

Non-negative matrix factorization (NMF) is an effective dimension reduction tool widely used in pattern recognition and computer vision. However, conventional NMF models are neither robust enough, as their objective functions are sensitive to outliers, nor discriminative enough, as they completely ignore the discriminative information in data. In this paper, we proposed a robust discriminative NMF model (RDNMF) for learning an effective discriminative subspace from noisy dataset. In particular, RDNMF approximates observations by their reconstructions in the subspace via maximum correntropy criterion to prohibit outliers from influencing the subspace. To incorporate the discriminative information, RDNMF builds adjacent graphs by using maximum correntropy criterion based robust representation, and regularizes the model by margin maximization criterion. We developed a multiplicative update rule to optimize RDNMF and theoretically proved its convergence. Experimental results on popular datasets verify the effectiveness of RDNMF comparing with conventional NMF models, discriminative NMF models, and robust NMF models.

KEYWORDS

Dimension reduction, non-negative matrix factorization, maximum correntropy criterion, supervised learning, margin maximization.

1. INTRODUCTION

Dimension reduction plays an important role in pattern recognition, computer vision and information retrieval. It projects samples from high-dimensional space onto a low-dimensional space, and thus reveals the intrinsic structure of a dataset to boost the subsequent processing. Recently, non-negative matrix factorization (NMF, [1]) has been proven to be a powerful dimension reduction method which approximates a non-negative data matrix by the product of two lower dimensional non-negative matrices. Since NMF learns a natural parts-based representation, it has been widely used in many tasks such as data mining [2], pattern recognition [3,4], and computer vision [5].

Since traditional NMF methods cannot take advantage of the labels of a dataset, they usually perform unsatisfactorily in classification tasks. To overcome this deficiency, Zaferiou et al. [6] proposed discriminant NMF (DNMF) to incorporate the Fisher's criterion into NMF. However, DNMF intrinsically assumes that samples obey Gaussian distribution, and this assumption is sometimes improper because NMF itself does not assume samples are Gaussian distributed. To overcome this problem, Guan et al. [7] proposed manifold regularized discriminative NMF (MD-NMF) to retain discriminative information for subsequent classification by marginal

maximization. Neither DNMF nor MD-NMF can perform well on some seriously noisy datasets because their Frobenius norm based [8] or Kullback-Leibler (KL) divergence based [9] loss functions are sensitive to outliers.

In this paper, we propose a correntropy supervised NMF (CSNMF) to overcome this deficiency. In particular, inspired by [12], CSNMF measures the loss of NMF by the well-known correntropy induced metric (CIM, [10]) instead of Frobenius-norm and KL-divergence. CIM is controlled by a kernel size and approximates L_0 -norm when the loss is relatively large, and thus it is robust to noise of large magnitudes or outliers. Assuming even noisy samples have correct labels, to utilize the labels of the dataset, CSNMF narrows the distance between any samples of the same class in the lower dimensional space. Since this discriminative information is noise-free and the utilized CIM-based loss function filters out any noise of large magnitude in the dataset, CSNMF can boost subsequent classification performance on the noisy datasets. In addition, we developed a multiplicative update rule to optimize CSNMF and theoretically proved its convergence. The experimental results on several popular face image datasets confirm the effectiveness of our CSNMF comparing with the supervised NMF variants and the robustified NMF variants.

The rest of this paper is organized as follows. Section 2 briefly reviews NMF and its variants. Section 3 presents the proposed CSNMF, the multiplicative update rule and its convergence. Section 4 shows the experimental results on popular face image datasets. We conclude this paper in Section 5.

2. RELATED WORKS

2.1. NMF

Given any non-negative matrix, i.e., $X \in \mathbb{R}_+^{m \times n}$, non-negative matrix factorization (NMF, [1]) aims at finding two lower dimensional non-negative matrices, i.e., $U \in \mathbb{R}_+^{m \times r}$ and $V \in \mathbb{R}_+^{r \times n}$, by minimizing the distance between X and UV . Conventional NMF methods measure the distance by using either Kullback-Leibler (KL) divergence [9] or squared Frobenius-norm [8], and thus they are not robust enough because their underlying distributions cannot model outliers. In addition, traditional NMF methods are not supervised because they completely ignore labels of a dataset.

2.2. NMF's Variants

From the seminal work of Lee and Seung [1] until now on, many NMF variants have been developed to deal with various practical tasks. Guan *et al.* [11] proposed a non-negative patch alignment framework (NPAF) to unify the popular NMF-related dimension reduction methods. The objective function of NPAF is

$$\min_{U \geq 0, V \geq 0} \frac{\gamma}{2} \text{tr}(VLV^T) + D(X, UV), \quad (1)$$

where $D(\cdot, \cdot)$ measures the loss of such factorization, γ is a positive tradeoff parameter, and L is an alignment matrix that encodes the statistical information of the datasets.

Based on NPAF, one can easily develop novel NMF-related dimension reduction method. For example, Guan *et al.* [11] developed a manifold regularized discriminative non-negative matrix factorization (MD-NMF) method to preserve the local geometric structure and incorporate the discriminative information of the dataset. The objective function of MD-NMF is

$$\min_{U \geq 0, V \geq 0} \frac{\gamma}{2} \text{tr}(VLV^T) + \sum_{ij} \left(X_{ij} \log \frac{X_{ij}}{(UV)_{ij}} - X_{ij} + (UV)_{ij} \right). \quad (2)$$

However, like NMF, MD-NMF is not robust enough because its loss function is sensitive to outliers.

To enhance the robustness of NMF, Du *et al.* [12] proposed a correntropy induced metric (CIM) based NMF (CIMNMF), which introduce CIM to measure the loss of the factorization. The objective function of CIMNMF is

$$\min_{U \geq 0, V \geq 0} \text{CIM}^2(X, UV), \quad (3)$$

where $\text{CIM}(X, UV) = \left(k_\sigma(0) - \frac{1}{mn} \sum_{ij} k_\sigma(X_{ij} - (UV)_{ij}) \right)^{1/2}$ and $k_\sigma(z) = e^{-z^2/2\sigma^2}$ is a Gaussian kernel function and σ is the kernel size.

Li *et al.* [13] proposed a graph regularized nonnegative matrix factorization method by maximizing the correntropy criterion (MCCGR) to incorporate the local geometric structure into CIMNMF, i.e.,

$$\max_{U \geq 0, V \geq 0} \sum_{i=1}^m k_\sigma \left(\sqrt{\sum_{j=1}^n \left(X_{ij} - \sum_{k=1}^r U_{ik} V_{kj} \right)^2} \right) - \frac{1}{2} \text{tr}(VLV^T), \quad (4)$$

where L is the graph Laplacian of the constructed adjacent graph.

Huang *et al.* [14] proposed robust manifold NMF (RMNMF) to preserve the local geometric structure in their previously developed robust NMF (RNMF) with $L_{2,1}$ -norm, i.e.,

$$\min_{U \geq 0, V \geq 0, V^T V = I} \|X - UV\|_{2,1} + \gamma \text{tr}(VLV^T), \quad (5)$$

where $\|Y\|_{2,1} = \sum_j \|Y_{:,j}\|_2$ signifies the $L_{2,1}$ -norm. Both MCCGR and RMNMF can be easily unified under NPAF.

Although both MCCGR and RMNMF show promises by the authors, they still have flaws. Since outliers might violate the intrinsic geometric structure of the clean dataset without outliers, the alignment matrix constructed on the noisy observations might be inaccurate from the view point of NPAF. Therefore, there are still some space to develop a novel robust NMF variant by simultaneously considering the robustness of loss function and alignment matrix.

3. CORRENTROPY SUPERVISED NMF

In this section, we first described a novel correntropy supervised NMF (CSNMF) based on NPAF to overcome the deficiencies of NMF and its variants. Then we developed a multiplicative update rule (MUR) to solve CSNMF. At last, we theoretically proved that the objective function of CSNMF is non-increasing under MUR.

3.1. Correntropy Induced Metric

In information-theoretic learning (ITL), one often uses correntropy to process noise [10]. Correntropy is defined as a generalized similarity between two variables

$$C_\sigma(x, y) = E(k_\sigma(x - y)), \quad (6)$$

where k_σ is the kernel function, and both x and y represent random variables. Given n samples, the estimator of correntropy is

$$\hat{C}_\sigma(x, y) = \frac{1}{n} \sum_{i=1}^n k_\sigma(x_i, y_i). \quad (7)$$

Based on ITL, Liu *et al.* [10] proposed the correntropy induced metric (CIM)

$$CIM(x, y) = \left(k_\sigma(0) - \frac{1}{n} \sum_{i=1}^n k_\sigma(e_i) \right)^{1/2}, \quad (8)$$

where $e_i = x_i - y_i$ represents the reconstruction error. The CIM value of large error in (8) is upper bounded by 1 regardless the scale of error. Therefore, CIM is less influenced by outliers. Due to its robustness, CIM has been widely used in many signal processing [10] and face recognition [15-16] tasks.

3.2. The CSNMF Model

Given n samples arranged in a non-negative matrix, i.e., $X \in \mathbb{R}_+^{m \times n}$, correntropy supervised NMF (CSNMF) decomposes it into the product of two lower-rank matrices, i.e., $X \approx UV$, by minimizing the CIM between X and UV , i.e.,

$$\min_{U \geq 0, V \geq 0} CIM^2(X, UV), \quad (9)$$

where $U \in \mathbb{R}_+^{m \times r}$ signifies the bases, and $V \in \mathbb{R}_+^{r \times n}$ signifies the coefficients. Based on ITL [10][12], CSNMF succeeds to filter out outliers.

For supervised learning, we assume that all samples including noisy samples are correctly labeled. Such assumption makes sense in some situations. Taking face recognition system for example, training images might be corrupted by illuminations but the subjects' name are known and correct. CSNMF expect to dig the discriminative information from labels of the dataset by incorporating the labels of a dataset, i.e., it narrows the distance between samples of the same class, i.e.,

$$\min_{V \geq 0} \sum_{i \neq j} \|v_i - v_j\|_2^2 S_{ij}, \quad (10)$$

where $\|\cdot\|_2$ signifies the L_2 -norm, and S_{ij} reflects the similarity between x_i and x_j , i.e.,

$$S_{ij} = \begin{cases} 1, & l(x_i) = l(x_j), i \neq j \\ 0, & \text{otherwise} \end{cases}, \quad (11)$$

where $l(\cdot)$ means the label of a sample. Although samples x might be corrupted, considering the robustness of CIM, it is reasonable to trust that the coefficients v is much less sensitive to outliers than x . Therefore, we can measure the distance between coefficients of two samples by the L_2 -norm in (10) to benefit from its nice mathematic property.

By simple algebra, we can rewrite the objective function in (10) as $\sum_{i \neq j} \|v_i - v_j\|_2^2 S_{ij} = \text{tr}(VLV^T)$, where $L = D - S$ and D is a diagonal matrix with $D_{ii} = \sum_{ij} S_{ij}$. By combing (9) and (10), we obtain the objective function of CSNMF as follows:

$$\min_{U \geq 0, V \geq 0} \text{CIM}^2(X, UV) + \frac{\gamma}{2} \text{tr}(VLV^T), \quad (12)$$

where γ is a positive trade-off parameter. Obviously, CSNMF can be unified by NPAF and L is considered an alignment matrix. CSNMF is jointly non-convex, and thus it is impossible to find its global optimum in polynomial time. In the following section, we developed a multiplicative update rule (MUR) to find its local minimum.

3.3. MUR for Optimizing CSNMF

For solving the constrained optimization problem (12), by using the Lagrangian multiplier method [8], we can obtain the Lagrangian function as follows:

$$L = \text{CIM}^2(X, UV) + \frac{\gamma}{2} \text{tr}(VLV^T) + \text{tr}(\alpha U) + \text{tr}(\beta V), \quad (13)$$

where α and β are the Lagrangian multipliers for the constraints $U \geq 0$ and $V \geq 0$, respectively.

To solve (12), we firstly calculate the first-order derivatives of L with respect to U and V , i.e.,

$$\begin{aligned} \frac{\partial L}{\partial v_{ab}} &= -\frac{1}{mn\sigma^2} \sum_k u_{ka} \left(X_{kb} - \sum_r u_{kr} v_{rb} \right) e^{\frac{\left(X_{kb} - \sum_r u_{kr} v_{rb} \right)^2}{2\sigma^2}} + \gamma \sum_r v_{ar} L_{rb} + \beta_{ab} \\ &= -\frac{1}{mn\sigma^2} \left[U^T \left((X - UV) \otimes W \right) \right]_{ab} + \gamma (VL)_{ab} + \beta_{ab}. \end{aligned} \quad (14)$$

And

$$\begin{aligned} \frac{\partial L}{\partial u_{ij}} &= -\frac{1}{mn\sigma^2} \sum_k v_{jk} \left(X_{ik} - \sum_r u_{ir} v_{rk} \right) e^{\frac{\left(X_{ik} - \sum_r u_{ir} v_{rk} \right)^2}{2\sigma^2}} + \alpha_{ij} \\ &= -\frac{1}{mn\sigma^2} \left[W \otimes (X - UV) V^T \right]_{ij} + \alpha_{ij}, \end{aligned} \quad (15)$$

The above two equations can be further written as

$$\frac{\partial L}{\partial v_{ab}} = -\frac{1}{mn\sigma^2} \sum_k u_{ka} X_{kb} e^{\frac{\left(X_{kb} - \sum_r u_{kr} v_{rb} \right)^2}{2\sigma^2}} + \frac{1}{mn\sigma^2} \sum_k u_{ka} e^{\frac{\left(X_{kb} - \sum_r u_{kr} v_{rb} \right)^2}{2\sigma^2}} \sum_r u_{kr} v_{rb} + \gamma (VL^+)_{ab} - \gamma (VL^-)_{ab} + \beta_{ab}, \quad (16)$$

and

$$\frac{\partial L}{\partial u_{ij}} = -\frac{1}{mn\sigma^2} \sum_k v_{jk} X_{ik} e^{\frac{(X_{ik} - \sum_r u_{ir} v_{rk})^2}{2\sigma^2}} + \frac{1}{mn\sigma^2} \sum_k v_{jk} e^{\frac{(X_{ik} - \sum_r u_{ir} v_{rk})^2}{2\sigma^2}} \sum_r u_{ir} v_{rk} + \alpha_{ij}, \quad (17)$$

where $L = L^+ - L^- = D - S$, and $D \geq 0$ and $S \geq 0$ construct the positive and negative components of L , respectively.

By using the K.K.T. conditions [17], any stationary point of (12) satisfies the following conditions

$$\begin{cases} \frac{\partial L}{\partial u_{ij}} = 0, \frac{\partial L}{\partial v_{ab}} = 0 \\ \alpha_{ij} u_{ij} = 0, \beta_{ab} v_{ab} = 0. \\ u_{ij} \geq 0, v_{ab} \geq 0 \\ \alpha_{ij} \geq 0, \beta_{ab} \geq 0 \end{cases} \quad (18)$$

By combing the first two conditions in (18), we can obtain the following equations

$$\left(-\frac{1}{mn\sigma^2} \sum_k v_{jk} X_{ik} e^{\frac{(X_{ik} - \sum_r u_{ir} v_{rk})^2}{2\sigma^2}} + \frac{1}{mn\sigma^2} \sum_k v_{jk} e^{\frac{(X_{ik} - \sum_r u_{ir} v_{rk})^2}{2\sigma^2}} \sum_r u_{ir} v_{rk}\right) u_{ij} = 0, \quad (19)$$

$$\left(-\frac{1}{mn\sigma^2} \sum_k u_{ka} X_{kb} e^{\frac{(X_{kb} - \sum_r u_{kr} v_{rb})^2}{2\sigma^2}} + \frac{1}{mn\sigma^2} \sum_k u_{ka} e^{\frac{(X_{kb} - \sum_r u_{kr} v_{rb})^2}{2\sigma^2}} \sum_r u_{kr} v_{rb} + \gamma(VL^+)_{ab} - \gamma(VL^-)_{ab}\right) v_{ab} = 0. \quad (20)$$

From (19-20), we can obtain the following multiplicative update rules:

$$V^{t+1} \leftarrow V^t \otimes \frac{\gamma V^t S + (U^{tT} (W^t \otimes X)) / mn\sigma^2}{\gamma V^t D + (U^{tT} (W^t \otimes (U^t V^t))) / mn\sigma^2}, \quad (21)$$

where $w_{ij}^t = e^{-\frac{(X_{ij} - (U^t V^t)_{ij})^2}{2\sigma^2}}$, and \otimes signifies element-wise multiplication. Then

$$U^{t+1} \leftarrow U^t \otimes \frac{(W^{t+1} \otimes X) V^{t+1T}}{(W^{t+1} \otimes U^t V^{t+1}) V^{t+1T}}, \quad (22)$$

where $w_{ij}^{t+1} = e^{-\frac{(X_{ij} - (U^{t+1} V^{t+1})_{ij})^2}{2\sigma^2}}$. According to [12], the kernel size can be adaptively updated by

$$\sigma^2 = \frac{1}{2mn} \sum_{i=1}^n \sum_{j=1}^m (X_{ij} - (U^t V^t)_{ij})^2. \quad (23)$$

We summarized the total procedure of MUR in **Algorithm 1**. The stopping condition is written

as follow: $\frac{F(U^{t+1}, V^{t+1}) - F(U^t, V^t)}{F(U^{t+1}, V^{t+1}) - F(U^0, V^0)} < tol$, where $F(U, V)$ is the objective value in (12). It accepts the

non-negative samples and outputs its factorization results. Although the MUR is derived from the Lagrangian multiplier method, we can theoretically analyze its convergence by using the auxiliary function in the following section.

Algorithm 1: The MUR for Optimizing CSNMF**Input :** $X \in \mathbb{R}_+^{m \times n}$, $r \in \mathbb{N} \cap \min\{m, n\}$.**Output:** $U \in \mathbb{R}_+^{m \times r}$, $V \in \mathbb{R}_+^{r \times n}$.**1. Calculate the alignment matrix L according to (12).****2. Initialize U^0 , V^0 , and calculate $(W_{ij}^0)_U = \exp\left(-\frac{(X_{ij} - (U^0 V^0)_{ij})^2}{2\sigma^2}\right)$.****3. Repeat**

$$V^{t+1} \leftarrow V^t \otimes \frac{\gamma V^t S + (U^{tT} (W^t \otimes X)) / mn\sigma^2}{\gamma V^t D + (U^{tT} (W^t \otimes (U^t V^t))) / mn\sigma^2},$$

$$W_{ij}^{t+1} = \exp\left(-\frac{(X_{ij} - (U^t V^{t+1})_{ij})^2}{2\sigma^2}\right).$$

$$U^{t+1} = U^t \otimes \frac{(W^{t+1} \otimes X)(V^{t+1})^T}{(W^{t+1} \otimes (U^t V^{t+1}))(V^{t+1})^T}.$$

$$W_{ij}^{t+1} = \exp\left(-\frac{(X_{ij} - (U^{t+1} V^{t+1})_{ij})^2}{2\sigma^2}\right).$$

$$\sigma^2 = \frac{1}{2mn} \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - (U^{t+1} V^{t+1})_{ij})^2.$$

 $t \leftarrow t+1$.**4. Until {The stopping condition is satisfied.}****5. $U = U^t$, $V = V^t$.**

The computational complexity of **Algorithm 1** is dominated by two parts: the updating statements and the construction of the alignment matrix. The complexity of the first part is $\#iter \times O(mnr + n^2r)$, where the $\#iter$ is the iteration of the algorithm and $O(mnr + n^2r)$ is the time cost of each iteration. The complexity of the second part is $O(n^2)$. Therefore, the total time complexity of **Algorithm 1** is $\#iter \times O(mnr + n^2r) + o(n^2)$.

3.4. Theoretical Analysis

Here we use the auxiliary function technique to prove the convergence of **Algorithm 1**. Let

$$F(U, V) = CIM^2(X, UV) + \frac{\gamma}{2} tr(VLV^T)$$
 denote the objective function in (12).

Lamma 1. If there exists an function G for $F(x)$ which satisfies $G(x, x') \geq F(x)$ and $G(x, x) = F(x)$, then we call it auxiliary function, and F is non-increasing under the following update rule:

$$x^{t+1} = \arg \min_x G(x, x'). \quad (24)$$

Let $J(V) = F_U(U, V)$ denotes the function with respect to V when U is fixed, and $J(U) = F_V(U, V)$ denotes the function with respect to U with V fixed.

Lamma 2. Given U^t , the following function

$$\begin{aligned}
G(u, u'_{ij}) &= J_{u_{ij}}(u'_{ij}) + J'_{u'_{ij}}(u'_{ij})(u - u'_{ij}) + \frac{1}{2mn\sigma^2} \frac{\sum_k v_{jk} w_{ik} \sum_r u_{ir} v_{rk}}{u'_{ij}} (u - u'_{ij})^2 \\
&= J_{u_{ij}}(u'_{ij}) + J'_{u'_{ij}}(u'_{ij})(u - u'_{ij}) + \frac{1}{2mn\sigma^2} \frac{\{[W \otimes (UV)]V^T\}_{ij}}{u'_{ij}} (u - u'_{ij})^2,
\end{aligned} \tag{25}$$

where $J'_{u'_{ij}}$ is the first order derivative with respect to U , is an auxiliary function of $J_{u_{ij}}$.

Lamma 3. Given V^t , the following function

$$\begin{aligned}
G(v, v'_{ab}) &= J_{v_{ab}}(v'_{ab}) + J'_{v'_{ab}}(v'_{ab})(v - v'_{ab}) + \frac{1}{2} \frac{\sum_k u_{ka} w_{kb} \sum_r u_{kr} v_{rb} / mn\sigma^2 + \gamma \sum_k v_{ak} L_{kb}}{v'_{ab}} (v - v'_{ab})^2 \\
&= J_{v_{ab}}(v'_{ab}) + J'_{v'_{ab}}(v'_{ab})(v - v'_{ab}) + \frac{1}{2} \frac{\{U^T [(X - UV) \otimes W]\}_{ab} / mn\sigma^2 + \gamma (VL)_{ab}}{v'_{ab}} (v - v'_{ab})^2,
\end{aligned} \tag{26}$$

where $J'_{v'_{ab}}$ is the first order derivative with respect to V , is an auxiliary function of $J_{v_{ab}}$.

The proofs of both **Lemma 2** and **Lemma 3** are deduced in **Appendix A** and **Appendix B** for the smoothness of logic.

According to **Lamma 1** and **Lemma 2**, $F_U(\arg \min_U G(U, U'), V) \leq F_U(U, V)$. Let $\frac{\partial G(U, U')}{\partial u_{ij}} = 0$, we have the multiplicative update rule (22). According to **Lamma 1** and **Lemma 3**, $F_V(U, \arg \min_V G(V, V')) \leq F_V(U, V)$. Let $\frac{\partial G(V, V')}{\partial v_{ab}} = 0$, we have the multiplicative update rule (21).

4. EXPERIMENTS

In this section, we evaluate the effectiveness of the proposed CSNMF on several popular face datasets by comparing with CIMNMF [12] and MCCGR [13]. We also compared CSNMF with a supervised NMF (SNMF) to demonstrate the robustness of CSNMF. SNMF incorporates the identical discriminative information like CSNMF, i.e.,

$$\min_{U \geq 0, V \geq 0} \frac{1}{2} \|X - UV\|_F^2 + \frac{\gamma}{2} \text{tr}(VLV^T), \tag{27}$$

where $\|\cdot\|_F$ signifies the Frobenius norm, and L is defined as the same alignment matrix as (12).

4.1. Experimental Setting

Our experiments are conducted on the Yale [18], UMIST [19] and ORL [20] datasets. Each experiment was test on the dataset which was divided to training set and test set. To sufficiently compare the performance, we use different sizes of training sets to learn the lower dimensional space. And in order to obtain better performance, it is important to choose a proper tradeoff parameter γ in (12), here we set it to be 10^{-5} in all experiments. In the classification stage, we applied the nearest neighbor (NN) rule as a classifier to determine the labels of test samples.

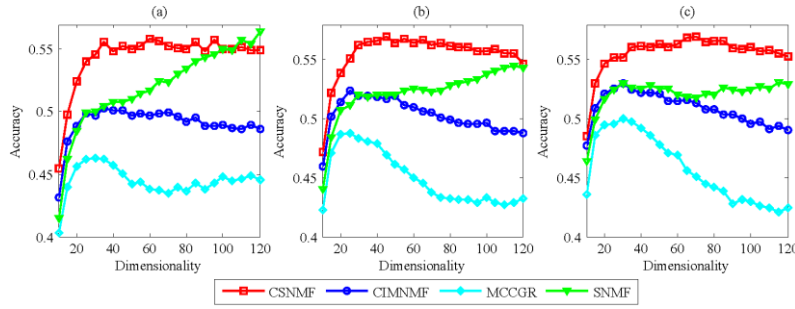


Figure 1. Face recognition accuracy on the Yale dataset. We randomly selected (a) 4, (b) 5, and (c) 6 images from each subject to learn the lower dimensional space and evaluate on the remaining images. The average results of 10 runs are reported.

4.2. Face Recognition

Yale dataset: The Yale [18] dataset contains 165 frontal view images from 15 subjects. Each one was taken 11 photos with varying facial expressions. And each image was normalized to 32 x 32 pixel array then reformulated to a vector form. Figure 1 shows the average accuracy of CSNMF, MCCGR, SNMF and CIMNMF. Their dimensions ranged from 10 to 120, and we random selected 4, 5 and 6 images for each individual for training. Table 1 records the highest average classification accuracy and their corresponding dimension. The experimental results show that our CSNMF is significantly superior to other algorithms in most cases. The comparison between CSNMF and SNMF shows that the utilized CIM successfully filters out outliers. The comparison between CSNMF and CIMNMF shows that the incorporated labels enhances the discriminative ability of the learned subspace. The comparison between CSNMF and MCCGR reflects that the adjacent graph constructed purely on labels are more robust than that constructed on features. In summary, CSNMF can fully takes advantage of the labels of dataset meanwhile take off the influence of the noises from both samples and constructed adjacent graphs.

Table 1. The highest average face recognition accuracies on the Yale dataset.

Algorithm	6	7	9
CSNMF	0.5705(35)	0.5994(30)	0.6453(40)
CIMNMF	0.5229(30)	0.5694(30)	0.6060(35)
MCCGR	0.4719(30)	0.5400(15)	0.5820(20)
SNMF	0.5852(115)	0.5661(120)	0.6033(65)

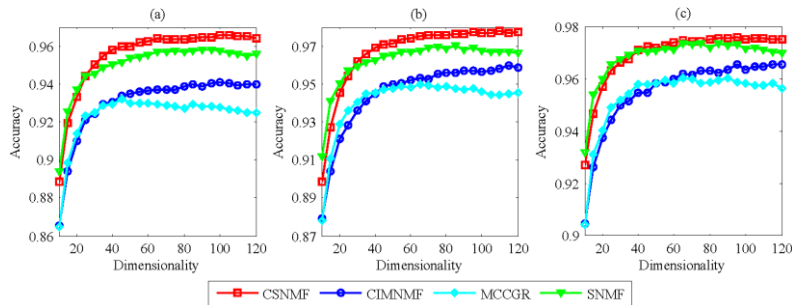


Figure 2. Face recognition accuracy on the UMIST dataset. We randomly selected (a) 6, (b) 7, and (c) 9 images from each subject to learn the lower dimensional space and evaluate on the remaining images. The average results of 10 runs are reported.

UMIST dataset: The UMIST [19] dataset contains 575 images from 20 subjects. Each one holds 41 to 82 images which varying in poses from profile to frontal views, and each image was normalized to 40 x 40 pixel array then reformulated to a vector form. Fig.2 shows the average accuracy of CSNMF, SNMF, MCCGR and CIM-NMF. Their dimensions range from 10 to 120, and we random selected 6, 7 and 9 images for each subject to comprise the training set. Table 3 records the highest average face recognition accuracy and their corresponding dimension. The experimental results show that our CSNMF is superior to other two algorithms mostly because it simultaneously takes advantages of the robustness of CIM and the discriminative information of labels of the dataset.

Table 2. The highest average face recognition accuracies on the UMIST dataset.

Algorithm	6	7	9
CSNMF	0.9839(85)	0.9972(75)	1.0000(75)
CIMNMF	0.9750(115)	0.9869(95)	0.9942(110)
MCCGR	0.9697(45)	0.9800(105)	0.9812(45)
SNMF	0.9769(75)	0.9931(110)	0.9963(60)

ORL dataset: The Cambridge ORL dataset consists of 400 images from 40 subjects, and each subject hold 10 images with varying lighting, facial expressions, and facial details. Each image was normalized to 32 x 32 pixel array then reformulated to a vector form. Fig.3 shows the average accuracy of CSNMF, SNMF, MCCGR and CIMNMF. Their dimensions range from 10 to 120, and we random selected 4, 5 and 6 images for each subject in the learning procedure. Table 4 records the highest average face recognition accuracy and their corresponding dimension. The experimental results show that our CSNMF is superior to SNMF, MCCGR and CIMNMF, and it confirms that CSNMF simultaneously takes their advantages without introducing their disadvantages.

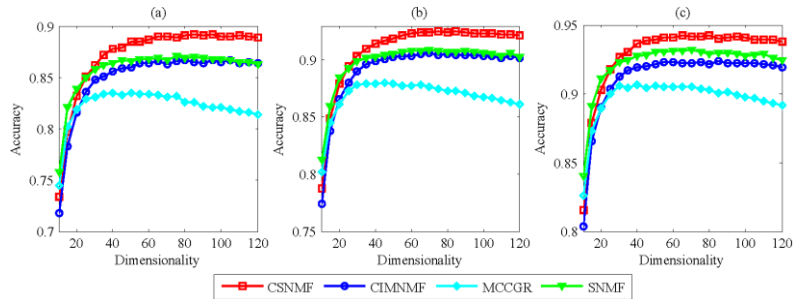


Figure 3. Face recognition accuracy on the ORL dataset. We randomly selected (a) 4, (b) 5, and (c) 6 images from each subject to learn the lower dimensional space and evaluate on the remaining images. The average results of 10 runs are reported.

Table 3. The highest average face recognition accuracy on the ORL dataset.

Algorithm	6	7	9
CSNMF	0.8900(110)	0.9552(110)	0.9544(90)
CIMNMF	0.8640(105)	0.9320(100)	0.9422(90)
MCCGR	0.8363(45)	0.9137(70)	0.9284(45)
SNMF	0.8869(80)	0.9400(75)	0.9478(55)

5. CONCLUSIONS

In this paper, we proposed a correntropy supervised non-negative matrix factorization (CSNMF) method for learning the discriminative lower dimensional space from noisy datasets. Since CSNMF can take advantages of the robustness of maximum correntropy criterion and discriminant power of the labels of dataset under the non-negative patch alignment framework, it outperforms both robust variant and supervised variant of NMF. We developed a multiplicative update rule to solve CSNMF and proved that it can monotonically decrease the objective function of CSNMF.

REFERENCES

- [1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization." *Nature*, vol. 401, no. 6755, pp. 781-791, 1999.
- [2] C. Liu, H. Yang, J. Fan, L. He, and Y. Wang, "Distributed nonnegative matrix factorization for web-scale dyadic data analysis on mapreduce," *Proceedings of the 19th international conference on World wide web* 10, 2010.
- [3] S. Li, X. W. H. X. W. Hou, H. J. Z. H. J. Zhang, and Q. S. C. Q. S. Cheng, "Learning spatially localized, parts-based representation," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.*, vol. 1, 2001.
- [4] A. Pascual-Montano, J. M. Carazo, K. Kochi, D. Lehmann, and R. D. Pascual-Marqui, "Nonsmooth nonnegative matrix factorization (nsNMF)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 403-415, 2006.
- [5] H. Liu, Z. Wu, X. Li, D. Cai, and T. S. Hunag, "Constrained nonnegative matrix factorization for image representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, vol. 34, pp. 1299-1311, 2012.
- [6] S. Zafeiriou, "Discriminant nonnegative tensor factorization algorithms," *IEEE Transactions on Neural Networks*, vol. 20, pp. 217-235, 2009.
- [7] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 2030-2048, 2011.
- [8] D. D. Lee and H. S. Seung, "Algorithms for nonnegative matrix factorization," *Advances in neural information processing systems*, vol. 13, pp. 556-562, 2001.
- [9] Z. Yang, H. Zhang, Z. Yuan, and E. Oja, "Kullback-Leibler divergence for nonnegative matrix factorization," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6791 LNCS, 2011, pp. 250-257.
- [10] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286-5298, 2007.
- [11] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Non-negative patch alignment framework," *IEEE Transactions on Neural Networks*, vol. 22, no. 8, pp. 1218-1230, Aug 2011.
- [12] L. Du, X. Li, and Y. D. Shen, "Robust nonnegative matrix factorization via half-quadratic minimization," in *Proceedings of IEEE International Conference on Data Mining, ICDM, 2012*, pp. 201-210.
- [13] L. Le, Y. J, Z. K, and et al, "Graph Regularized Non-negative Matrix Factorization By Maximizing Correntropy." *Ar Xiv preprint arXiv*, vol. 1405, no. 2246, 2014.
- [14] H. Jin, N. Feiping, H. Heng, and C. Ding, "Robust Manifold Nonnegative Matrix Factorization," *ACM Transactions Knowledge Discovery Data*, vol. 8, no.3, 2014.
- [15] X. Yuan and B. Hu, "Robust Feature Extraction via Information Theoretic Learning," in *Proceedings of the 26th International Conference on Machine Learning, 2009*, pp. 1193-1200.
- [16] R. He, W. S. Zheng, and B. G. Hu, "Maximum correntropy criterion for robust face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1561-1576, 2011.
- [17] D. P Bertsekas, *Nonlinear Programming*, 2nd ed. MA : Athena Scientific: Belmont, 1999.

- [18] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, 1997.
- [19] D. B. Graham and N. M. Allinson, "Characterizing virtual eigensignatures for general purpose face recognition," Proceedings of Face Recognition: Theory Application., vol. 163, pp. 446-456, 1998.
- [20] F.S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," Proceedings of 1994 IEEE Workshop on Applications of Computer Vision, vol. 13, pp. 138-142, Dec 1994.

APPENDICES

Appendix A: Proof of Lemma 2

It is obvious that $G(u, u) = J_{u_j}(u)$. With the Taylor series expansion of $J_{u_j}(u)$, we can obtain

$$J_{u_j}(u) = J_{u_j}(u'_j) + J'_{u_j}(u'_j)(u - u'_j) + \frac{1}{2} J''_{u_j}(u - u'_j)^2 \quad (28)$$

And

$$\begin{aligned} J''_{u_j} &= \frac{\partial^2 \mathbf{L}}{\partial u_{ij}^2} \\ &= \frac{1}{mn\sigma^2} \sum_k \left\{ e^{\left[\frac{X_{ik} - \sum_r u_{ir} v_{rk}}{2\sigma^2} \right]^2} \left[1 - \frac{1}{\sigma^2} \left(X_{ik} - \sum_r u_{ir} v_{rk} \right)^2 \right] \right\} v_{jk}^2 \\ &= \frac{1}{mn\sigma^2} \left\{ \left[W \otimes \left(I_W - \frac{1}{\sigma^2} (X - UV) \otimes (X - UV) \right) \right] (V \otimes V) \right\}_{ij} \end{aligned}$$

Then we can see that,

$$\begin{aligned} [W \otimes (UV) V^T]_{ij} &= \sum_k w_{ik} v_{jk} \sum_r u_{ir} v_{rk} \geq \sum_k w_{ik} v_{jk} v_{jk} u_{ij} = [W (V \otimes V)]_{ij} u_{ij} \\ &\geq \left\{ \left[W \otimes \left(I_W - \frac{1}{\sigma^2} (X - UV) \otimes (X - UV) \right) \right] (V \otimes V) \right\}_{ij} u_{ij} \end{aligned} \quad (27)$$

So we have $G(u, u'_j) \geq J_{u_j}(u)$. This completes the proof. \square

Appendix B: Proof of Lemma 3

It is obvious that $G(v, v) = J_{v_{ab}}(v)$. With the Taylor series expansion of $J_{v_{ab}}(v)$, we have

$$J_{v_{ab}}(v) = J_{v_{ab}}(v'_{ab}) + J'_{v_{ab}}(v'_{ab})(v - v'_{ab}) + \frac{1}{2} J''_{v_{ab}}(v - v'_{ab})^2, \quad (29)$$

And

$$\begin{aligned} J''_{v_{ab}} &= \frac{\partial^2 \mathbf{L}}{\partial v_{ab}^2} \\ &= \frac{1}{mn\sigma^2} \sum_k u_{ka}^2 e^{\left[\frac{X_{ik} - \sum_r u_{ir} v_{rk}}{2\sigma^2} \right]^2} \left[1 - \frac{1}{\sigma^2} \left(X_{kb} - \sum_r u_{kr} v_{rb} \right)^2 \right] + \gamma L_{bb} \\ &= \frac{1}{mn\sigma^2} \left\{ (U \otimes U)^T \left[\left(I_W - \frac{1}{\sigma^2} (X - UV) \otimes (X - UV) \right) \otimes W \right] \right\}_{ab} + \gamma L_{bb} \end{aligned}$$

Then we can see that

$$\begin{aligned}
& \left\{ U^T [(X - UV) \otimes W] \right\}_{ab} / mn\sigma^2 + \gamma (VL)_{ab} = \sum_k u_{ka} w_{kb} \sum_r u_{kr} v_{rb} / mn\sigma^2 + \gamma \sum_k v_{ak} L_{kb} \\
& \geq \sum_k u_{ka} w_{kb} u_{ka} v_{ab} / mn\sigma^2 + \gamma v_{ab} L_{bb} = \left[(U \otimes U)^T W \right]_{ab} v_{ab} / mn\sigma^2 + \gamma L_{bb} v_{ab} \\
& \geq \left\{ \left[I_w - \frac{1}{\sigma^2} (X - UV) \otimes (X - UV) \right] \otimes W \right\}_{ab} v_{ab} / mn\sigma^2 + \gamma L_{bb} v_{ab}.
\end{aligned} \tag{30}$$

So we have $G(v, v'_{ab}) \geq J_{v_{ab}}(v)$. This completes the proof. \square

AUTHORS

Hang Cheng: He is currently studying for a master's degree at the Research Centre for Artificial Intelligence, National Innovation Institute of Defense Technology. His research interests include computer vision, few-shot object detection and machine learning.



Shixiong Wang: received the B.S. degree in 2013 from Tsinghua University, Beijing, China, and the Ph.D. degree in 2019 from National University of Defense Technology, Changsha, China. He is now an assistant professor with National Innovation Institute of Defense Technology, Beijing, China. His research fields include computer science and machine learning.



Naiyang Guan: received the B.S., M.S., and Ph.D. degrees from the National University of Defense Technology, Changsha, China, in 2004, 2006, and 2011, respectively. He is currently an Associate Professor with the Defense Innovation Institute (DII), Academy of Military Sciences, Beijing, China. He has authored or coauthored over 60 research articles on top-tier journals, including the IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), IEEE Transactions on Neural Networks and Learning Systems (T-NNLS), IEEE Transactions on Image Processing (T-IP), and IEEE Transactions on Signal Processing (T-SP), and top-tier conferences, including the IEEE International Conference on Data Mining (ICDM), International Joint Conference on Artificial Intelligence (IJCAI), European Conference on Computer Vision (ECCV), and International Joint Conference on Neural Networks (IJCNN). His research interests include machine learning, computer vision, and data mining.



A UNITY MICROSCOPE SIMULATION TO HELP STUDENTS GET MORE ACCESS TO LAB EQUIPMENT ONLINE DURING COVID-19 PANDEMIC

Kaiwen Chen¹ and Yu Sun²

¹Santa Margarita Catholic High School, 22062 Antonio Pkwy,
Rancho Santa Margarita, CA 92688

²California State Polytechnic University, Pomona,
CA, 91768, Irvine, CA 92620

ABSTRACT

Something that still remains an issue to this day is how students and other individuals can become educated in matters that are generally taught in person and are difficult to translate to an online environment in particular [5]. In particular, teaching how to operate lab equipment without having hands-on experience is incredibly difficult. With the COVID-19 pandemic, the need for sufficient online learning materials and tools has become much greater in recent years [6]. To resolve this issue, a simulation was made in Unity that aims to educate its users on how to work with a microscope [7]. Sliders are provided in the simulation to control the X-axis, Y-axis, Z-axis, and focus. The simulation was tested for its effectiveness by gathering fifteen participants to download and test the simulation, then asking each participant to fill out a survey. In the survey, the participants graded the educational value and convenience of using the application on a scale from one to ten, and they were encouraged to leave any other feedback in a free-response section of the survey [8]. Results indicated that the general public would find this simulation practical in daily life, as participants generally rated the simulation as both educational and convenient to use.

KEYWORDS

Simulation, Microscope, Unity.

1. INTRODUCTION

Microscopes have been an important invention that has been widely used throughout history. The concept started with two lenses placed on opposite sides of the tube, and Galileo Galilei perfected the invention in 1609. When using microscopes, we can observe the world much closer and through a perspective that would not have been possible with the naked human eye. In particular, microscopes have paved the way for numerous advancements in science and medicine. With the help of microscopes, humanity has been able to discover smaller forms of life such as bacteria, viruses, fungi, and other microorganisms, and doctors have been able to identify the source of illnesses and even develop cures for illnesses [1]. With its impact on our society, the ability to use them can be crucial to those who wish to pursue careers in relevant fields.

The topic of how to use microscopes is significant because many students use them, whether they are taking major courses in college or doing a lab in a middle school science class. Whenever

microscopes had to be used, they were commonly done so within a classroom or laboratory setting, in which students and researchers would be able to handle the microscopes physically. However, after the COVID-19 pandemic hit, using physical microscopes became much more difficult [9]. And social distancing guidelines made it so that in-person instruction can be quite limited. With the rapid development of technology in recent years, a new popular path can be opened for teaching individuals how to operate a microscope.

There are currently many guides on how to operate a microscope. They exist in mostly text or video form, and many of them are easily accessible through the internet. Simulations that teach users the intricacies of using a microscope are also present, which can give visual learners an easier time understanding what to do. The way that is believed to be best for students' understanding is to have a hands-on demonstration of how to use the microscope and have the students follow along with a physical microscope. Depending on what the situation calls for, different methods of teaching may have their advantages and be preferable. Unfortunately, these methods also come with fairly significant drawbacks. With text-based guides on how to use microscopes, the biggest limitation is being able to visualize the instructions and apply them in a lab environment. Video-based guides can solve this issue, but what both guides lack is the involvement of the learners. This concept of active learning is applied in simulations [10]. Unfortunately, many simulations suffer from being either too simple or too complex. Simulations can simplify the action of using a microscope far too much when they only require a click of a button. Others may require the user to make so many complicated motions with the microscope and show so many numbers on the screen that the details detract from the main points. Furthermore, the graphics of some older simulations may be jarring and ward away potential users that have been accustomed to better and cleaner graphics. The method of getting hands-on experience with a physical microscope may no longer be feasible in recent times, when concerns over public health and distance may restrict people to only using online resources. With advancements in technology, a new method may be able to address the downsides of previously used methods.

A Unity game was created that was meant to educate its players on how to properly use a microscope in a lab environment [11]. The primary features of this game are sliders that the player is able to control using the mouse. There are four sliders in the simulation: the first one controls the X-axis, the second one controls the Y-axis, the third one controls the Z-axis, and the final one controls the focus. The X-axis determines how far to the left or to the right the image is shifted, and the Z-axis determines how far above or below the image is shifted. The Y-axis determines how closely the image is zoomed in. Finally, the focus slider will only allow the image to show up clearly on the microscope if the focus value is within a certain interval. Otherwise, the image will appear blurry.

Unlike guides on how to operate a microscope, the method of using a simulation is much more interactive and may allow for deeper understanding. Looking at text or a video of how to use a microscope may not let the concepts stick in people's minds as well as playing an active role in their own learning. While this simulation may be similar to other simulations, it also does its best to strike the perfect balance between being easy to use and understand and being educational.

The method we used to prove the effectiveness of the microscope simulation is by creating a survey on Google Forms. The Google Forms would have two questions: "How educational was this simulation?" and "How convenient was this simulation to use?". The first question asks the participants how much understanding and knowledge they can derive from the simulation, and the second question asks the participants how well the knowledge can be absorbed by the individuals that use them in a timely manner. Both of these qualities would be vital to the success of the simulation, as high scores in both would indicate that the general public would deem the

simulation worthy of use, whether it be in a classroom setting or at an individual's own leisure. At the bottom of the survey, an optional free-response section would be available for the participants to send any custom feedback that they may have about the simulation.

Fifteen participants will be gathered, and each of them would be asked to download the ZIP file that includes the Unity simulation. They would be instructed to unzip the file, then run the simulation inside the unzipped folder. From there, the participants would spend at least one minute testing each of the sliders inside of the simulation and understanding how the microscope works as a whole. After they are finished using the simulation, they would be provided with a link to the Google Forms survey. Whatever responses the participants gave would be recorded in a linked Google Sheets table, including the feedback in the optional free-response section at the bottom of the survey.

The remainder of the paper will be divided into five sections, which will be labeled with the numbers 2 through 6. Section 2 describes the hurdles that had to be overcome when coming up with ideas for the simulation and experimenting on the simulation. Section 3 offers an explanation for how the simulation as a whole, as well as specific sections of the simulation, was implemented. Section 4 brings up how effective the simulation is using a survey that tests for important qualities that are desirable for users, and Section 5 introduces related works and how they compare to this work. The final section, Section 6, provides a conclusion that includes the summary of the simulation, some current limitations to the application, and what can be done in the future to address these limitations.

2. CHALLENGES

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

2.1. How education regarding science labs was limited by the COVID-19 pandemic

The problem to be addressed was how education regarding science labs was limited by the COVID-19 pandemic, leaving many classes only available online. Unfortunately, bringing science labs to an online environment is a more difficult transition than some other classes, as lab equipment that students would normally be able to get a hands-on experience with would no longer be accessible. When searching for the most common lab equipment that students would use, the microscope was consistently regarded as a useful and popular tool in modern science education. A comprehensive and well-written guide seemed like a good idea at first, but such a solution would not be able to emulate the active learning that students would be able to attain by directly operating a microscope. The next idea that was eventually implemented as the final solution was to create a simulation that could educate users on how to utilize a microscope.

2.2. How exactly to make the simulation

After figuring out what to make to address the issue of students having a lack of access to lab equipment, the issue of how exactly to make the simulation arose. There are several different kinds of simulations of microscopes online, with some that can even simulate multiple types of microscopes. To stand out from the others, the original plan was to make something that could provide the history of microscopes and other various information regarding microscopes that other simulations might not have. However, after envisioning the final product, such a feature might be overwhelming and might even be seen as irrelevant for users that are new to operating microscopes and are only using the simulation for the sole purpose of understanding how to use a microscope. With this in mind, the implementation of the simulation was made to be as easy to

understand and clear as possible to use, which would allow the target demographic of students or other individuals who have never used microscopes before to absorb the information in a much more efficient manner.

2.3. Experiments for the microscope simulation

The final challenge was coming up with experiments for the microscope simulation. Since the microscope is simple, it is difficult to test the simulation for how accurate it is. The only features to test within the simulation are the sliders that control the X-axis, the Y-axis, the Z-axis, and the focus. Therefore, a different form of experimentation was needed. After some brainstorming, the path of experimentation that seemed best was surveying participants and asking them to try the simulation. The survey would be split into two main sections, which are how educational it is and how convenient it is to use. Participants would be given the option to rate each section on a scale from one to ten. These sections would be crucial for our results, as having positive scores in both sections would mean that the simulation would likely be practical as a tool to be used by the general public.

3. SOLUTION

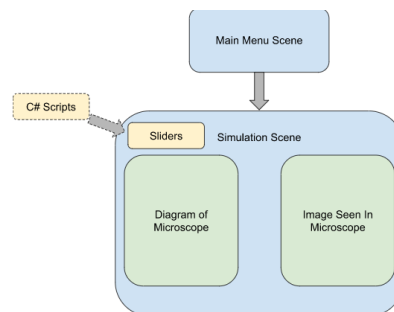


Figure 1. Overview of the solution

The Unity microscope simulation has two main scenes. The first scene is the main menu screen, in which the user is met with the title “Microscope” and a button labeled “Start” right below the title. To proceed with the next screen, the user will click the start button, which will transfer the simulation to the second scene. In this scene, an image of a microscope can be seen on the left side of the screen, and the image that can be seen through the microscope is present on the right side of the screen. On the top left of the screen, four sliders are available for the user to interact with. The first slider is the X-axis, and adjusting the X-axis value causes the stage of the microscope to move left and right in the diagram of the microscope on the left side of the screen. The image that can be seen through the diagram of the microscope on the right side of the screen is shifted horizontally. The second slider is the Y-axis, in which adjusting the Y-axis value moves the microscope stage up and down. The image seen from the microscope zooms in when the stage is moved up and zooms out when the stage is moved down. The Z-axis slider moves the stage forward and backward, and the image in the microscope shifts vertically to account for this change. Finally, the focus slider makes it so the image in the microscope shows up clearly when at a certain interval of values, but shows up as blurry for all other values. The functionality of the simulation is implemented using scripts written in the C# programming language [15].

In order to switch from the main menu scene to the simulation scene, a button is placed in the main menu scene. When the user presses the button, the button is able to transfer to the second

scene because the button was assigned a C# script. In this script, the SceneManager called the load scene function with the desired scene to be switched to in the argument of the function. As C# scripts provide much of the functionality in many Unity games and applications, this is a common way to perform the task of switching scenes.

The C# scripts play more of a crucial role in the simulation scene. The four sliders that are present control the X-axis, Y-axis, Z-axis and focus. All of these sliders have different effects on the image that can be seen through the microscope, and the first three sliders control the diagram of the microscope. The sliders in the simulation were created using the built-in slider element that can be selected from the user interface section of Unity. Just like with the button that transfers scenes, the sliders also have a script attached to them. The script controlling the X-axis slider controls both the stage of the microscope diagram and the image seen through the microscope as two separate tasks. The slider for the X-axis is directly linked to the position of the stage, which changes the X-axis of the stage object depending on the value of the X-axis slider. The X-axis of the image through the microscope is also shifted, which made the implementation of this slider surprisingly convenient. The same procedure is applied to the Z-axis. When it comes to the Y-axis, the script causes the stage to move up as the slider is moved from left to right. To match the stage moving up, the image seen through the microscope becomes zoomed in the farther the value is moved right along the slider. For the script controlling the focus slider, the diagram of the microscope does not change at all no matter what value the focus is set to. Rather, only the image seen through the microscope is affected. When the user first enters the simulation scene, the image appears blurry, as the focus slider is set all the way to the left. However, the only way to make the image appear clear in the simulation is to move the slider somewhere near the middle. This is performed by using a blur effect. When images are displayed in Unity, the quality of the image will be completely clear and unblurred by default. However, the script is coded so that the image's blur value is set to zero as long as the focus value resides within a certain interval, then the blur value becomes greater the farther the focus value is from that interval.

```
// Update is called once per frame
void Update()
{
    float x = startPosition.x;
    float y = startPosition.y;
    float z = startPosition.z;

    if (xSlider != null)
    {
        x = (startPosition.x - xlimit / 2) + xSlider.value * xlimit;
    }
    if (ySlider != null)
    {
        y = startPosition.y + ySlider.value * heightlimit;
    }
    if (zSlider != null)
    {
        z = (startPosition.z - zlimit / 2) + zSlider.value * zlimit;
    }

    gameObject.transform.position = new Vector3(x, y, z);
}
```

Figure 2. Screenshot of code 1

```

public void moveforwardBack(float value)
{
    value -= 0.5f;
    if (backgroundImage)
    {
        Vector3 pos = backgroundImage.localPosition;
        pos.y = value * backgroundSpeed;
        backgroundImage.localPosition = pos;
    }
}

public void movelDown(float value)
{
    value = 1 - value;
    if (backgroundImage)
    {
        float scaleFactor = 1 - (value * backgroundScalingFactor);
        backgroundImage.localScale = Vector3.one * scaleFactor;
    }
}

```

Figure 3. Capture Device Screenshot

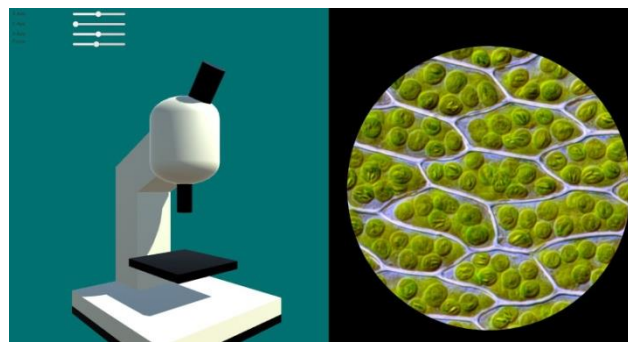


Figure 4. Image through microscope 1

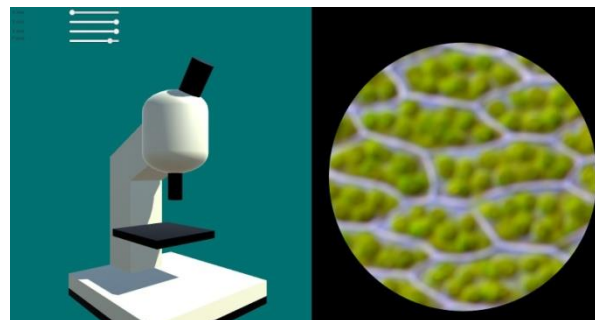


Figure 5. Image through microscope 2

4. EXPERIMENT

4.1. Experiment 1

The experiment that was used to test the effectiveness of the microscope simulation was a survey that was created in Google Forms. First, the participants would download a ZIP file, unzip the file, and run the Unity simulation from inside the folder. The participants would then spend at least one minute fiddling with the sliders and examining how changing certain values affects the position and image quality of the image and the stage of the microscope. After the participants are done, they would answer the first question in the Google Forms survey about how educational they felt the simulation was, which was recorded on a scale from one to ten. Since there are

fifteen participants taking the survey, there is a large enough sample size to account for variability.

Participant Number	Educational Value Rating of Simulation
1	10
2	9
3	9
4	5
5	8
6	7
7	8
8	9
9	8
10	10
11	9
12	9
13	7
14	8
15	7
Average	8.4

Figure 6. Table of participant number and education value rate

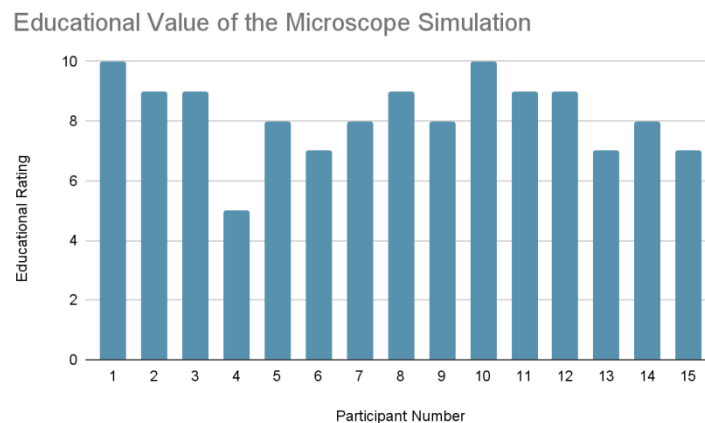


Figure 7. Graph of educational value of the microscope simulation

According to the results of the participants' perceived educational value of the simulation, it appears that the educational value ratings were fairly high. With the exception of one of the ratings at a 5, the rest of the ratings were 7 or above. Such results could signify that the participants were pleased with how much knowledge they received when using the simulation. Some feedback that participants sent stated that the implementation of the sliders in the simulation was well-made, as the sliders captured the feeling of having to use "guesswork" that people would experience when turning the knobs on an actual microscope and using trial and error to eventually get the best view from the microscope. However, two pieces of feedback also indicated that the coarse adjustment knob and the fine adjustment knob that are usually present on microscopes could have been emphasized more, and separate sliders to represent each knob could have been more informative.

4.2. Experiment 2

The second question in the same Google Forms survey asked how convenient using the simulation was. To keep the style consistent throughout the survey, a scale from one to ten was also used to record the results. At the end of the survey, an optional free-response section was provided to the participants. This would allow the participants more freedom to express their thoughts about the simulation and share more valuable insight that would not be possible with the previous two questions. The measurement for the convenience of the application was also not altered significantly by variability, as the sample size of fifteen would still be enough to mitigate its effects.

Participant Number	Convenience Rating of Simulation
1	10
2	8
3	9
4	6
5	9
6	7
7	9
8	10
9	8
10	10
11	8
12	9
13	8
14	8
15	9
Average	8.53

Figure 8. Table of participant number and convenience rating

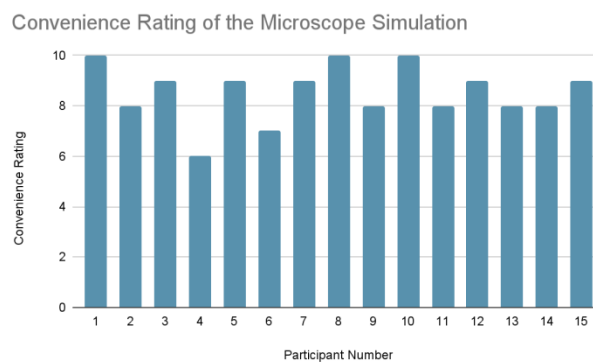


Figure 9. Graph of convenience rating of the microscope simulation

The results regarding the convenience of the microscope simulation indicate that the simulation is effective at helping its users absorb knowledge more easily. Three participants rated the convenience a perfect score of 10, and all of the ratings were 6 or above. Furthermore, an average result of 8.53 means that participants were generally very satisfied with how user-friendly the simulation is. The vast majority of participants who left feedback reported that the simulation was incredibly simple and easy to use, which allowed for a more convenient and enjoyable time learning with the simulation. The user interface was also generally praised by the participants for

its clean and minimal aesthetic. On the other hand, one participant also noted that the concept of the focus only reaching maximum image sharpness at a certain interval could be confusing to new users, and there should be at least some indicator or note that informs the users of this.

In the first experiment, the vast majority of the participants believed the implementation of the microscope simulation made sense and provided accurate information regarding how a microscope would work. As the participants are a subset of the general public, it is believed that the general public would likely find this to be educational for both students and other curious individuals.

With the results of the second experiment, it appears that the general public would also find the simulation to be very intuitive and convenient to use. With both qualities of the simulation being rated very highly, the simulation seems to be a valuable and helpful tool overall that can improve the learning process when it comes to using a microscope. These results were expected, as the simulation was specifically designed to optimize the learning of users who have never had experience operating a microscope before. Therefore, it could be concluded that this simulation has been a success so far.

5. RELATED WORK

One article reports on the use of an Arduino as a piece of lab equipment that can take on multiple roles within a lab environment, which is possible by loading scripts onto the board's memory. The Arduino is significant because it is more affordable and portable than dedicated hardware [2]. Both the Arduino article and this work mention the usage of lab equipment and the effectiveness of implementations. While the Arduino article focuses on the physical implementation of lab equipment, this work emphasizes the implementation of a microscope through software that can act as an online learning tool for the general public.

Another work states that the use of algorithms and computations can be integrated into a microscope in order to create a more advanced microscope with improved accuracy. Recent improvements in simulation speed can allow the viewing of important biochemical processes such as protein folding and membrane transport in milliseconds [3]. The related work and this work both share the topic of simulations involving microscopes. In particular, the related work offers an in-depth view of how computational microscopes can improve the field of science through various examples. On the other hand, this work involves the creation of a simulation of a traditional optical microscope and how effective it is at educating the general public.

A third related work covers the history of the optical microscope in cell biology and medicine. Only with the large magnification of microscopes, along with proper contrast and resolution, have we been able to see cells and their components. Recently, light microscopy has advanced to the point that live cell imaging is possible [4]. This work is similar to the related work because the topics of both works are related to optical microscopes. However, while the related work emphasizes how the optical microscope has been improved and utilized in making discoveries for cell biology, this work places its focus on the creation and experimentation of an optical microscope simulation.

6. CONCLUSIONS

In light of the COVID-19 pandemic and the difficulty of learning information from online science labs that would be best done in person, a simulation was created to educate the general public on how to operate a microscope [14]. The simulation features four sliders, which each

affects the image shown by the microscope in a different way. The X-axis and the Z-axis control how far the image in the microscope is shifted in a certain direction, the Y-axis controls how zoomed in the image in the microscope is, and the focus controls how clear the image appears in the microscope. The simulation was designed to teach students and other members of the general public the core concepts of how to use a microscope in a simple and digestible form.

The experiment that was performed was a survey that gauged how useful the simulation would be if it was released to the general public. Fifteen participants were asked to download the Unity simulation, use the simulation for at least one minute, then submit their responses on a Google Forms survey as to how educational the simulation was and how convenient the simulation was to use on a scale from one to ten. The responses were then recorded into a Google Sheets table that was linked to the Google Forms survey [13]. According to the results, the simulation was effective at providing some educational value to its users as well as offering a clean and intuitive user interface for the users to work with. The feedback indicated that the ease of use was what the participants generally appreciated the most about the simulation.

A current limitation of the simulation is that the simulation only educates the user on how to use a microscope. However, in a lab environment, microscopes are not the only instruments that are used to conduct research or perform experiments; many others are used as well, such as volumetric flasks and spring balances. Although the simulation serves its current purpose fairly well, the scope of the simulation is limited by its sole topic of how to use a microscope. Ideally, the simulation could be expanded to become a more general-purpose tool that could offer its users more knowledge about various sets of lab equipment.

To address this limitation, some work regarding the simulation that is planned for the near future is adding some of the most common lab instruments [12]. In the main menu of the screen, rather than just having a button to start the simulation, there would be multiple buttons that correspond to each instrument. Furthermore, the option to go back to the main menu would be added as well, in case the users decide that they have tried a certain instrument enough and want to try another one.

REFERENCES

- [1] History of the microscope - who invented the microscope? MicroscopeMaster. (n.d.). Retrieved September 17, 2022, from <https://www.microscopemaster.com/history-of-the-microscope.html>
- [2] D'Ausilio, A. (2011). Arduino: A low-cost multipurpose lab equipment. *Behavior Research Methods*, 44(2), 305–313. <https://doi.org/10.3758/s13428-011-0163-z>
- [3] Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H., & Shaw, D. E. (2012). Biomolecular simulation: A computational microscope for molecular biology. *Annual Review of Biophysics*, 41(1), 429–452. <https://doi.org/10.1146/annurev-biophys-042910-155245>
- [4] Masters, B. R. (2008). History of the optical microscope in Cell Biology and medicine. *Encyclopedia of Life Sciences*. <https://doi.org/10.1002/9780470015902.a0003082>
- [5] Sun, Anna, and Xiufang Chen. "Online education and its effective practice: A research review." *Journal of Information Technology Education* 15 (2016).
- [6] Ciotti, Marco, et al. "The COVID-19 pandemic." *Critical reviews in clinical laboratory sciences* 57.6 (2020): 365-388.
- [7] Helander, M. G., E. J. Grossmith, and P. Prabhu. "Planning and implementation of microscope work." *Applied ergonomics* 22.1 (1991): 36-42.
- [8] Raia, Anthony P. "A study of the educational value of management games." *The Journal of Business* 39.3 (1966): 339-352.
- [9] Pokhrel, Sumitra, and Roshan Chhetri. "A literature review on impact of COVID-19 pandemic on teaching and learning." *Higher Education for the Future* 8.1 (2021): 133-141.
- [10] Banks, Jerry. "Introduction to simulation." *Proceedings of the 31st conference on Winter simulation: Simulation---a bridge to the future-Volume 1*. 1999.

- [11] Craighead, Jeff, Jennifer Burke, and Robin Murphy. "Using the unity game engine to develop sarge: a case study." Proceedings of the 2008 Simulation Workshop at the International Conference on Intelligent Robots and Systems (IROS 2008). Vol. 4552. 2008.
- [12] Siegel, M. "What to do when the teaching lab instruments are too sophisticated for the students?." 2002 IEEE International Symposium on Virtual and Intelligent Measurement Systems (IEEE Cat. No. 02EX545). IEEE, 2002.
- [13] Vasantha Raju, N., and N. S. Harinarayana. "Online survey tools: A case study of Google Forms." National Conference on Scientific, Computational & Information Research Trends in Engineering, GSSS-IETW, Mysore. 2016.
- [14] Chatterjee, Suparna. "A primer for transitioning to online science labs: "Identifying potential types of guidance for supporting student inquiry when using virtual and remote labs in science"." Educational Technology Research and Development 69.1 (2021): 249-253.
- [15] Al-Bastami, Bashar G., and Samy S. Abu Naser. "Design and Development of an Intelligent Tutoring System for C# Language." European academic research 4.10 (2017).

TRUSTWORTHY ARTIFICIAL INTELLIGENCE FOR BLOCKCHAIN-BASED CRYPTOCURRENCY

Tiffany Zhan

USAOT, Las Vegas, Nevada, USA

ABSTRACT

Blockchain-based cryptocurrency has attracted the immersive attention of individuals and businesses. With distributed ledger technology (DLT) consisting of growing list of record blocks and securely linked together using cryptography, each block contains a cryptographic hash of the previous block, a timestamp, and transaction data. The timestamp proves that the transaction data existed when the block was created. Since each block contains information about the block previous to it, they effectively form a chain, with each additional block linking to the ones before it. Consequently, blockchain transactions are irreversible in that, once they are recorded, the data in any given block cannot be altered retroactively without altering all subsequent blocks. The blockchain-based technologies have been emerging with a fleet speed. In this paper, the trustworthy Artificial Intelligence will be explored for blockchain-based cryptocurrency where the prohibitive price leap creates a challenge for financial analysis and prediction.

KEYWORDS

Trustworthy Artificial Intelligence, Blockchain, Cryptocurrency, Financial Prediction.

1. TRUSTWORTHY ARTIFICIAL INTELLIGENCE

The flotilla development of Artificial Intelligence (AI) technology has enabled numerous applications in the world, including AI in astronomy, AI in healthcare, AI in gaming, AI in data security, AI in social media, AI in travel and transportation, AI in automotive industry, etc. [1, 2] However, many AI systems are vulnerable to indiscernible attacks which degrade people's trust in AI systems [3, 4]. One inspiring question is what does it mean to be trustworthy?

In April 2019, the European Commission's High-Level Expert Group on Artificial Intelligence (AI HLEG) published Ethics Guidelines for Trustworthy AI, stating that human beings will be able to confidently and fully reap the benefits of AI only if they have trust in it [5, 6]. There are three aspects to enforce the trustworthy AI: 1) robust and reliable technology to avoid unintentional damage due to lack of technological mastery; and 2) behave ethically and morally; 3) human in the loop. Four principles including human autonomy, prevention of harm, fairness, and explicability were set.

One definition is that to be trustworthy, an AI system should operate competently, behave ethically and morally, and interact appropriately with humans. Is this sufficient? Based on the definition provided by the Merriam-Webster dictionary, it says that firm belief in the character, strength, or truth of someone or something. Can we have a firm belief for an AI system?

Holton [7] stated that in order to trust one need not believe. He used an example of a shopkeeper who decides to trust his employee, although the latter has been convicted of petty theft. Holton argued that the shopkeeper can decide to trust the man without believing that he will not steal. He may trust him because he wants to give him moral support, a new chance to earn trust. This sort of trust has been called “therapeutic trust”. However, contrary to Holton, I think that a firm belief must be implemented to be trustworthy.

Carsel [8] proposed a social-cognitive theory of trust which states that, in social contexts, trust is a fundamental element of relationships. Without trust, people may not be able to pursue valued interdependent goals [9] and meet relationship satisfaction [10]. To understand trust formally, a coherent theoretical framework is needed [11,12]. The question that typically guides psychological research and theory in trust is “Does Human A trust AI B?” An important implication is that current paradigms overlook the possibility that Human A might trust AI B differently across various contexts. For example, imagine AI B achieved a great job for Human A in Task X, Human A may trust AI in the context of Task X. However, AI performs poorly in Task Y due to data adversarial attacks. Consequently, Human A may not trust AI B in Task Y. Such possibilities raise the challenges to scientific community. One of important focus in trust community is to predict whether or not Person (Human) A will trust Person (AI) B, such as generalized anxiety [13], attachment style [14], and group membership [15]. However, such an analysis does not consider the social contexts and can only examine average levels of trust between people (human and AI).

By being able to identify when, why, and how human come to trust AI in context, practitioners need to implement policies that facilitate trust between human and AI on how to regain trust that was lost in their relationship or identify levels of (dis)trust in specific contexts that facilitate maladaptive behaviour within those contexts. An alternative question that draws attention to the potential variability in trust between human and AI across contexts is “When does Human A trust AI B?”

When discussing the trust between humans, researchers link the concept of risk [16, 17], ranging from situating trust as simply a subset of risk to locating risk as an antecedent to trust, or the current risks to the individual do indeed affect the individual’s trust in others in a trust game. In the person to person trust, the risks to the trustor are not constant across contexts, even if they are similar to past interactions between the trustor and trustee, we should expect interpersonal trust to calibrate to the specific demands of the interaction. How about the trust situation between human and AI?

Instead of examining whether and to what degree an individual trusts others, the focus should be on the potentially varying levels of trust across the various contours within human and AI’s relationships. In other words, the motivating question becomes “Human A trusts AI B for what?” A careful reading of “trust” indicates that we often say “trust” when people share a goal. I think that we should propose a novel theoretical orientation to trust between human and AI and formulate a contextual theory of trust using a new theoretical lens. Following this idea, I will apply the trustworthy AI in the blockchain-based cryptocurrency systems.

2. BLOCKCHAIN-BASED CRYPTOCURRENCY

A blockchain is a distributed ledger technology (DLT) that consists of growing list of records, called blocks, that are securely linked together using cryptography [18, 19, 20]. Each block contains a cryptographic hash of the previous block, a timestamp, and transaction data (generally represented as a Merkle tree, where data nodes are represented by leafs). The timestamp proves that the transaction data existed when the block was created. Since each block contains

information about the block before it, they effectively form a chain (compare linked list data structure), with each additional block linking to the ones before it. Consequently, blockchain transactions are irreversible in that, once they are recorded, the data in any given block cannot be altered retroactively without altering all subsequent blocks.

Cryptographer David Chaum first proposed a blockchain-like protocol in his 1982 dissertation [21] "Computer Systems Established, Maintained, and Trusted by Mutually Suspicious Groups." Further work on a cryptographically secured chain of blocks was described in 1991 by Stuart Haber and W. Scott Stornetta [22, 23]. A blockchain was created by a person (or group of people) using the name (or pseudonym) Satoshi Nakamoto in 2008 [24] to serve as the public distributed ledger for bitcoin cryptocurrency transactions, based on previous work by Stuart Haber, W. Scott Stornetta, and Dave Bayer [22, 23] The identity of Satoshi Nakamoto remains unknown to date. The implementation of the blockchain within bitcoin made it the first digital currency to solve the double-spending problem without the need of a trusted authority or central server. The bitcoin design has inspired other applications and blockchains that are readable by the public and are widely used by cryptocurrencies. The blockchain may be considered a type of payment rail [25, 26, 27].

The Bitcoin protocol is the consensus mechanism that allows users to send and receive a digital like currency called Bitcoin. Because the transfer of Bitcoin requires intensive use of cryptography, Bitcoin is referred to as a cryptocurrency. Just in this year, Bitcoin price rose to near \$65k, then fell \$29k the next. The current price as of today is \$19,000. Additionally, the volatility of Bitcoin and other cryptocurrency is highly compared to traditional stock and indexes. Normal stock prediction is a non-trivial task, but to add extreme volatility and parameters that are internal only to Blockchain, a question of whether an AI algorithm such as a Deep Neural Network (DNN) can learn the behaviour of Bitcoin is a scorching topic in Cryptocurrency. Previous studies have shown a deep neural network is no better than traditional statistical methods [28, 29, 30].

3. CRYPTOCURRENCY PREDICTION

Stock trend prediction is challenging because there are many factors can influence the price [33]. The factors may be internal or external (or both) events to the given company. The events may not visible before it occurs. The problem of predicting Bitcoin prices is even more challenging in that Bitcoin prices may not adhere to outside business influence and government, but only on limit of coins [34, 35]. With increasing business using cryptocurrency, the United States Exchange Commission (USEC) has enforced rules to regulate cryptocurrency. To predict cryptocurrency prices, linear regression, support vector machine, logistic regression, and time series analysis have been used [36, 37, 38]. Among them, linear regression provides decent results. In this study, I would like to use deep neural network learning.

4. DATA DESCRIPTION

The data are obtained via blockchain.com (<https://www.blockchain.com/charts>) for blockchain data, bitcoinnity.org (<https://data.bitcoinnity.org/markets/price>) for bitcoin data, and Yahoo Finance (www.yahoofinance.com) for indexes data. The datasets are collected in CSV format. The data range from September 2011 to September 2022.

5. ALGORITHMIC AND EXPERIMENTAL IMPLEMENTATION

In these experiments, I used the Scikit Learn Library and trained different neural network models on the datasets, including 1-hidden, 2 hidden, and 3-hidden layer deep neural network (DNN) models. I used 80% of the dataset was used for training and 20% used for testing. The DNN models are trained through a 10-fold cross-validation and 100 epochs. As discussed in class 10-fold cross validation and a high epoch were important to implement to ensure the consistency and validity of the results. I then calculated the RMSE (Root-mean-square deviation) and MAPE (Mean absolute percentage error)

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{N}}$$

RMSE = root-mean-square deviation
i = variable *i*
N = number of non-missing data points
x_i = actual observations time series
ŷ_i = estimated time series

$$M = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

M = mean absolute percentage error
n = number of times the summation iteration happens
A_t = actual value
F_t = forecast value

6. CONCLUSION

Trustworthy AI becomes increasingly critical for the current world applications. The scientific community should explore further about the fundamental concepts on trustworthy AI in various social contexts. This paper explores trustworthy AI for blockchain-based cryptocurrency where the prohibitive price leap creates a challenge financial analysis and prediction. Blockchain is an emerging technology which promises security and true decentralization with cryptocurrency being the first widely used application. In this paper, I used deep neural networks trained with blockchain and macroeconomic variables which provides stronger predicting power than linear regression. In the future, I would like to see how a DNN model would predict Bitcoin price with its current trend. Another idea I plan to explore is to model other cryptocurrencies such as Ethereum or Dogecoin. As each blockchain has unique parameters that could affect the coin prices differently. Such studies would contribute sanguinely to the blockchain and cryptocurrency resources.

REFERENCES

- [1] Bezboruah, T., & Bora, A. (2020). Artificial intelligence: the technology, challenges and applications. *Trans Mach Learn Artif Intell*, 8(5), 44-51.
- [2] Kaushik, M. (2022). Artificial Intelligence (Ai). In *Intelligent System Algorithms and Applications in Science and Technology* (pp. 119-133). Apple Academic Press.
- [3] Dixon, R. B. L. (2022). Artificial Intelligence Governance: A Comparative Analysis of China, the European Union, and the United States.
- [4] Petersen, E., Potdevin, Y., Mohammadi, E., Zidowitz, S., Breyer, S., Nowotka, D., ... & Herzog, C. (2022). Responsible and Regulatory Conform Machine Learning for Medicine: A Survey of Challenges and Solutions. *IEEE Access*.
- [5] The European Commission's Artificial Intelligence Act, Available online at https://hai.stanford.edu/sites/default/files/2021-06/HAI_Issue-Brief_The-European-Commissions-Artificial-Intelligence-Act.pdf
- [6] EU Commission Ethics guidelines for trustworthy AI. Available online at <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

- [7] Holton, R. (1994). Deciding to trust, coming to believe. *Australasian Journal of Philosophy* 72, 63–76.
- [8] Carsel, T. (2020) *In Context We Trust: A Social-Cognitive Theory of Trust*, PhD Dissertation in Psychology, University of Illinois at Chicago, 2020.
- [9] Rusbult, C. E., & Van Lange, P. A. (2003). Interdependence, interaction, and relationships. *Annual Review of Psychology*, 54(1), 351-375.
- [10] Sanderson, C. A., & Cantor, N. (1997). Creating satisfaction in steady dating relationships: The role of personal goals and situational affordances. *Journal of Personality and Social Psychology*, 73(6), 1424-1433.
- [11] Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, 84(1), 123-136. Mayer, R. C.,
- [12] Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709-734.
- [13] Kenworthy, J. B., & Jones, J. (2009). The roles of group importance and anxiety in predicting depersonalized ingroup trust. *Group Processes & Intergroup Relations*, 12(2), 227-239.
- [14] Collins, N. L., & Read, S. J. (1990). Adult attachment, working models, and relationship quality in dating couples. *Journal of Personality and Social Psychology*, 58(4), 644-663.
- [15] Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, 86(2), 307-324.
- [16] Das, T. K., & Teng, B. S. (2004). The risk-based view of trust: A conceptual framework. *Journal of Business and Psychology*, 19(1), 85-116.
- [17] Evans, A. M., & Krueger, J. I. (2011). Elements of trust: Risk and perspective-taking. *Journal of Experimental Social Psychology*, 47(1), 171-177.
- [18] Patra, G. R., & Mohanty, M. N. (2022). Price Prediction of Cryptocurrency Using a Multi-Layer Gated Recurrent Unit Network with Multi Features. *Computational Economics*, 1-20.
- [19] Shaik, A., Laxmi, P. S., Anusree, E., Abbas, S., Rajesh, S., & Teja, A. (2022). Forecast Bitcoin Price Prediction Using Time Series Analysis through Machine Learning. *Journal of Algebraic Statistics*, 13(3), 2113-2123.
- [20] Breaves, A. and Au, B. (2015) *Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin*, Stanford: CS224, pp. 1–6, 2015.
- [21] Sherman, A., Javani, F., Zhang, H., Golaszewski, E. (2019). On the Origins and Variations of Blockchain Technologies. *IEEE Security Privacy*. 17 (1): 72–77.
- [22] Narayanan, A., Bonneau, J., Felten, E., Miller, A. Goldfeder, S. (2016) *Bitcoin and cryptocurrency technologies: a comprehensive introduction*. Princeton: Princeton University Press.
- [23] Haber, S., Stornetta, W. (1991). How to timestamp a digital document. *Journal of Cryptology*. 3 (2): 99–111.
- [24] S. Nakamoto (2008) *Bitcoin: A Peer-to-Peer Electronic Cash System*, Tech. Rep., 2008.
- [25] Bakos, Y., Halaburda, H., Mueller-Bloch, C. (2021). When Permissioned Blockchains Deliver More Decentralization Than Permissionless. *Communications of the ACM*. 64 (2): 20–22.
- [26] Panetta, K. (2018). *Digital Business: CIO Agenda 2019: Exploit Transformational Technologies*.
- [27] Wegner, P. (1996). Interoperability. *ACM Computing Surveys*. 28: 285–287
- [28] Sun, C. (2008) *Stock Market Returns Predictability: Does Volatility Matter?*, Columbia University: QMSS, pp. 1-30, 2008.
- [29] Ciaian, P., Rajcaniova, M. and Kancs, D. (2016) *The Economics of Bitcoin Price Formation*, *Appl. Econ.*, vol. 48, no. 19, pp. 1799–1815, 2016.
- [30] McNally, S., Roche, J. and Caton, S. (2018) *Predicting the Price of Bitcoin Using Machine Learning*, 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), 2018, pp. 339-343, doi: 10.1109/PDP2018.2018.00060.
- [31] Gujarati, D. and Porter, D. *Essentials of Econometrics*. New York, NY, USA: McGraw-Hill, 1999.
- [32] Unknown, S. (2015, September 15). *Variance Inflation Factor*. <https://www.statisticshowto.datasciencecentral.com/variance-inflationfactor>
- [33] Berdik, D., Otoum, S., Schmidt, N., Porter, D., Jararweh, Y. *A Survey on Blockchain for Information Systems Management and Security*, *Information Processing & Management*, Volume 58, Issue 1, 2021, 102397, ISSN 0306-4573
- [34] Akaike, H. (1969) *Fitting autoregressive models for prediction*. *Ann Inst Stat Math* 21:243–247

- [35] Al-Khazali O, Bouri E, Roubaud D (2018) The impact of positive and negative macroeconomic news surprises: gold versus Bitcoin. *Econ Bull, AccessEcon* 38:373–382
- [36] Almeida, J., Tata, S., Moser, A. and Smit, V. (2015) Bitcoin Prediction Using ANN, *Neural Networks, Group 7*, pp. 1–12, June 2015.
- [37] Sun, C. (2008) *Stock Market Returns Predictability: Does Volatility Matter?* Columbia University: QMSS, pp. 1-30, 2008.
- [38] Ciaian, P., Rajcaniova, M. and Kancs, D. (2016) The Economics of Bitcoin Price Formation,” *Appl. Econ.*, vol. 48, no. 19, pp. 1799–1815, 2016.

AN INTELLIGENT VIDEO EDITING AUTOMATE FRAMEWORK USING AI AND COMPUTER VISION

Mei Yi Yang¹ and Yu Sun²

¹Southpointe Academy, 1900 56 St, Delta, BC V4L 2B1

²California State Polytechnic University,
Pomona, CA, 91768, Irvine, CA 92620

ABSTRACT

The issue to be resolved was videos may be a difficult and time-consuming process to edit, specifically with cropping videos [4]. The solution that was implemented was a mobile application that was capable of cropping videos using a body tracking solution called MediaPipe Pose [5]. Using a landmark model that labeled the body parts of a person, the application can recognize where the person is located in the video frame by saving the minimum and maximum x- and y- coordinates. In the image array, the rows and columns not included within those coordinates are deleted, which leaves only the area with the person inside. To prove the effectiveness of the application in daily life, a survey was performed on fifteen participants. Each participant was shown the same video demonstration of the application being used, then the participants answered questions regarding how useful the application would likely be in daily life and how convenient the application would be to use. Results indicate that the general public would be willing to use the application as a long-term solution for video cropping [7].

KEYWORDS

Video Editor, Dancing, Video Cropping, Flutter.

1. INTRODUCTION

The idea of this application sparked when dancers faced the problem of not being able to edit their dance videos efficiently and aesthetically to post online. Users had to conduct a lot of research over a period of time to find all the methods available. Because they change positions when dancing, they have to film the frame fairly wide so it covers my entire range of motion. However, this makes the body really small in the video and very hard to see. When looking for apps that have the functions of having the frame follow the person in a video but there were no such products. Over the past few months, we have experimented with similar functioning applications. The popular video sharing platform TikTok has a body zoom filter which can be used when filming on the app [8]. However the flaw in this method is that the filter is exclusively used for filming in the application, meaning that a video cannot be uploaded to then apply the filter. Another flaw was that it zoomed very far in, which captured the full range of motion however is occasionally aesthetically displeasing. Another editing app that has a similar idea to what we were looking for was CapCut. This editing software allows users to upload videos and edit by hand. They have a feature that allows users to hand edit each frame they wish to zoom in or out and left or right. This allows the users creative freedom in how they wish the frame to track their body, however the major flaw in this is that it is very time consuming. The benefits of

this app is that it is able to be used to film videos with the app that will automatically edit the video for the user as well as upload and process filmed videos. This product should be extremely relevant in solving many dancers' problems that are similar to mine.

Some existing video editors attempt to make video cropping easier. Video editors exist in which cropping is performed by typing in the minimum and maximum x- and y-coordinates of the video that should still be kept in [9]. However, this may be too much effort for an ordinary user, as the user will have to measure the resolution of the video, then estimate how many pixels they would have to crop out from each side of the video. By having a designated crop button and allowing the user to manually stretch and resize the region that should be kept in, the user may find this method of cropping videos more intuitive than the previous method.

One of the biggest issues with these general-purpose video editors, however, is that they are so feature-heavy that it can become difficult to navigate through them all. With so many buttons on the screen, these editors can be incredibly daunting for those who are new or inexperienced with video editing. Furthermore, more heavy-duty video editors generally require a strong machine to use, which can lead to those with older devices either running the video editor very slowly or not being able to run the video editor at all. The ideal video cropping solution is a lightweight application that is capable of video cropping with very little effort needed from the user [10].

A method that has been tested before is "smart" video cropping, which examines the video for what parts should be kept in the video before cropping automatically. Much of this was done to resolve the issue of unusual aspect ratios, which may have resulted in awkward cropping. For example, the video could be completely centered and cut off a person that is on the very side of the screen, but this new method could crop so the person shows up on the screen. However, many of these implementations were made more than ten years ago, and there may be newer technologies today that may be more accurate.

The tool that was created to tackle the issue of video cropping being a difficult and daunting process is a mobile application that automatically crops the video for the user. The mobile application was made using Flutter for the front-end code and Python for the back-end code. The main feature of this application is smart video cropping. The user of the application will simply be able to input a video and be returned a video that is cropped to have the person in the image be the focus of the video. This feature was intended for dancing videos, but it can also be applied to other videos that require the person of the video to be the primary focus. As some people can be too far away from the camera at the time of recording, such an application can refine videos and enhance the viewing experience of those who watch them. Compared to general-purpose video editors, this application is much more lightweight. However, the application lacks the multitude of features that come with full video editors. Since the application was designed for only one primary purpose, the lack of features is not too detrimental to the application. As this application was made with recent technologies, they may be more consistently accurate than implementations developed in the past.

In order to prove that the application would be effective and would provide benefits to the general public if it was widely released, a survey was conducted on Google Forms. Each participant would watch the same video demonstration of a person using the application and showing off each of the features. Then, the participants would answer questions that ask how useful the application is in daily life and how convenient the application would be to use. They will be provided with a scale from one to ten to answer each question, in which one is the worst rating and ten is the best rating. In case the participants have more thoughts they would like to share, they would be able to do so inside an optional free-response feedback section at the bottom of the survey.

The results of the survey could indicate if there are any particular aspects of the application that could use more work. They could also gauge how successful the application would be and how much of a positive impact it would potentially make on society. Convenience and usefulness are both very valuable qualities for an application to have. If the application is too inconvenient to use or requires too much time and effort to deliver the desired result, people will likely try to find alternatives. If the application provides no benefit or practical use in daily life, users may not be intrigued enough to download the application in the first place. Only when both qualities are done well in the application would users install the application and use it long-term.

The remainder of the paper will be divided into five sections, labeled from 2 through 6. Section 2 describes the hurdles that had to be overcome when coming up with ideas for the application and implementing the application. Section 3 offers an explanation for how the application as a whole, as well as specific sections of the application, was implemented. Section 4 analyzes how effective the application is using experiments that test the functionality of the application and the feedback from the general public, and Section 5 introduces several related works and how they compare to this work. The final section, Section 6, provides a conclusion that includes the summary of the application, some current limitations to the application, and what can be done in the future to resolve these limitations.

2. CHALLENGES

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

2.1. Coming up with an Idea for What the Application Should Do

The first challenge was coming up with an idea for what the application should do. The topic that was to be addressed was how dancing videos had the dancer too far away from the camera, which may detract from the quality or viewing experience of those who watch the video. At first, one idea that was considered was that the application should let the user choose a specific box in the video, and the application would crop out everything else in the video except the specified box. However, that would still take much effort from the users of the application, and they would likely not use the application again if it was too inconvenient for them to use. To address this, the video cropping idea was still utilized, but what was cropped out of the video could be decided based on an AI that detected where the person was in the video.

2.2. Implementing the Application itself

Another obstacle was implementing the application itself. While the ideas and concepts were already confirmed, applying those ideas and concepts to code and software was a much more challenging feat. While creating the front-end with Flutter took a relatively short amount of time, coding with Python was much more difficult, as it involved using a model that would be able to track the human body quickly. The model that was settled on was a detection model from MediaPipe Pose, which would use landmarks to predict where the specific body parts of a person were inside video frames. Using the detection model for the first time was confusing and frustrating, and there were not many examples of working Python code involving MediaPipe Pose that could be easily found online [11]. After much trial and error with the code, a working implementation was finally created. By retrieving the x- and y-coordinates of where the predicted landmarks were in the image array, a box could be created and the rest of the rows and columns in the image array could be deleted.

2.3. Coming up with an Experiment that could Effectively Test the Application for Relevant Information

A challenge that was also encountered was coming up with an experiment that could effectively test the application for relevant information. The first step in doing so was determining exactly what would be considered as “relevant information” [12]. Anything that would improve the chances of the application succeeding if it were to be widely released to the general public, such as practicality and ease of use, would be very important to gauge. In order to do so, a survey was used by gathering participants to view a video demo and asking them to complete a Google Form. The questions that the survey would have should ask for relevant information, which would allow the participants to share their thoughts regarding the convenience and usefulness of the application. However, there is the possibility that the participants have other feedback they would like to share that is not covered by the previous questions. Therefore, an optional free-response section was added to the survey as well.

3. SOLUTION

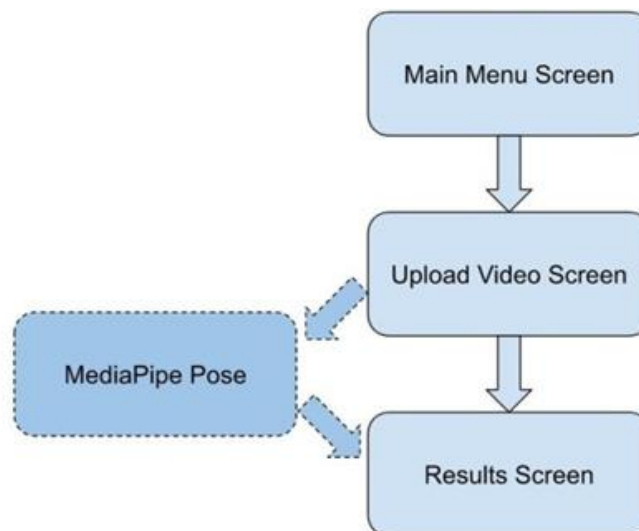


Figure 1. Overview of the solution

The application is comprised of three primary screens in Flutter. The first screen is the main menu screen, which is what the users of the application see when they first open up the application. On this screen, the words “Zoom in Dancing” are displayed as the title, and an image of a dancer is located below the title. Below the image is a single button labeled “Start”. The second screen is the upload video screen, which can be accessed by tapping the start button on the main menu screen. On this screen, the users are provided with two options as to how they will input their video into the application. The first option is to select a video file from their library, which allows the user to choose a file that is saved in the device’s storage. The second option is to access the camera from the application, in which users will have the opportunity to take a video of themselves dancing without having to exit the application and do so from the camera app. Once a video file has been selected or a video has been taken using the camera, the application reaches its third and final screen. This screen is responsible for displaying the resulting video, which will be cropped to place more emphasis on the person in the video. Such a feat is performed using Python back-end code, specifically with MediaPipe Pose, which is a body pose tracking solution that can detect human bodies in video frames.

The application was created using Flutter for the front-end code and Python for the back-end code. To access the different screens in the application, certain triggers were used [15]. To transition from the main menu screen to the upload file screen, the user would have to press the start button. The button is programmed so that when it is pressed, Flutter calls `Navigator.push()` which routes the application to a separate screen. In the case of transitioning from the upload video screen to the results screen, this same method is used, but it only triggers when the application has had a proper video file inputted into it, whether that be through selecting a file from the device or recording a video within the application using the device camera.

The functionality related to the cropping of the input video in the application comes from the Python back-end code. One of the most significant components of the application is MediaPipe Pose. MediaPipe Pose tracks body poses in videos using the landmark model. The landmark model inside MediaPipe Pose performs a prediction on thirty-three “landmarks” of the human body [6]. These landmarks range from areas on the limbs such as the elbows and knees to more specific body parts, including the pinkies and thumbs of each hand. Facial features are even included as some of the landmarks, such as the corners of the mouth, inner eyes, outer eyes, nose, and ears. Within the Python code, a minimum detecting and tracking confidence level of 0.5 is needed to determine the pose in the inputted video. Since MediaPipe Pose’s models were designed with real-time detection in mind, the majority of modern-day smart devices possess the capabilities to utilize these models.

As for how the MediaPipe Pose is implemented into the application, the landmarks that the landmark model finds are processed and stored in a variable. Each of the landmarks is looped through. For the x-coordinate of each landmark, it is checked with the current minimum and maximum x-values and updated if the x-coordinate is lower than the current minimum or higher than the current maximum. The same procedure happens for the y-coordinates of the landmarks. This is mainly done to determine where the person is located in the video. With this information, the application will know what can safely be cropped out of the video. The minimum and maximum x- and y-coordinates that MediaPipe has detected are multiplied by the width and height of the screen to take ratios into account. Using built-in functions in MediaPipe, the landmarks that the landmark model has recognized are drawn out and indicated directly on the output video file. The outputted video file is created by taking each frame of the inputted file and saving the frame as an image array. When the time comes for the video to be cropped, the image arrays are retrieved from a list and the image arrays have only a portion of their rows and columns saved, based on the final screen height and width that MediaPipe has determined.

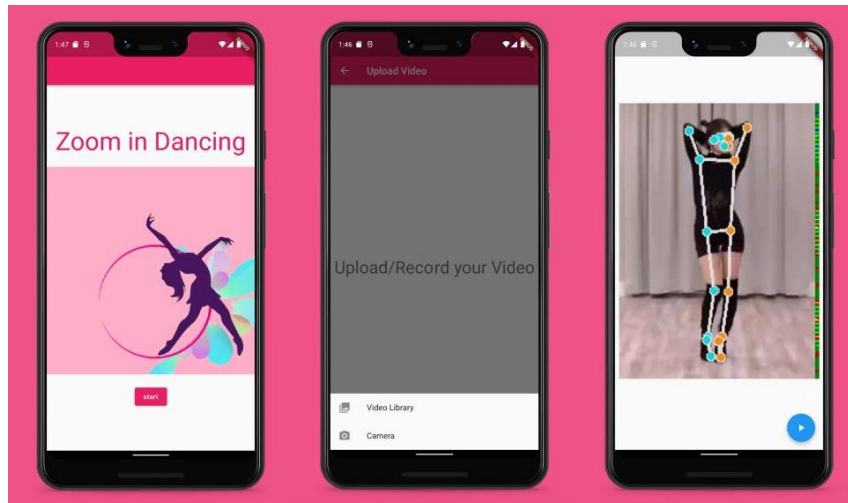


Figure 2. Screenshots from the video cropping application

```

class _VideoAppState extends State<VideoApp> {
  late VideoPlayerController _controller;

  @override
  void initState() {
    super.initState();
    _controller = VideoPlayerController.network(widget.videoUrl)..initialize().then(() {
      // Ensure the first frame is shown after the video is initialized, even before the play button has been pressed.
      setState(() {});
    }).catchError((e) {
      print("Failed to init the video: ");
      print(e);
    });
  }
}

max_x = 0
min_x = float('inf')
max_y = 0
min_y = float('inf')

if results.pose_landmarks != None:
    for mark in results.pose_landmarks.landmark:
        #print(mark.x, mark.y)
        if mark.x > max_x:
            max_x = mark.x
        if mark.x < min_x:
            min_x = mark.x
        if mark.y > max_y:
            max_y = mark.y
        if mark.y < min_y:
            min_y = mark.y
    # print("min_x", min_x, "min_y", min_y, "max_x", max_x, "max_y", max_y)
    #print coordinates of ratios above
    min_x_pos = int(min_x*width)
    min_y_pos = int(min_y*height)
    max_x_pos = int(max_x*width)
    max_y_pos = int(max_y*height)

uploadFileToServer(image) async {
  print(image?.path);
  print(image?.name);

  // read the file data
  var fileBytes = await image?.readAsBytes();
  // upload the video to the server
  // var url =
  //   'http://ec2-54-176-5-252.us-west-1.compute.amazonaws.com:5000/video';
  var url = 'http://10.0.2.2:5000/video';

  http.MultipartRequest request =
    http.MultipartRequest('GET', Uri.parse('$url'));
  request.files.add(
    await http.MultipartFile.fromBytes(
      'file',
      fileBytes!,
      filename: image?.name,
      contentType: MediaType('audio', 'midi'),
    ),
  );
}

for i in range(len(img_arr)):
  print("writing", frame_size[i])
  # Flip the image horizontally for a selfie-view display.
  cropped_image = img_arr[i]
  int(frame_size[i][0][1] - (screenHeight - (frame_size[i][1][1] - frame_size[i][0][1])) / 2)
  : int(frame_size[i][1][1] + (screenHeight - (frame_size[i][1][1] - frame_size[i][0][1])) / 2),
  int(frame_size[i][0][0] - (screenWidth - (frame_size[i][1][0] - frame_size[i][0][0])) / 2)
  : int(frame_size[i][1][0] + (
    screenWidth - (frame_size[i][1][0] - frame_size[i][0][0])) / 2) # crop the image
  out.write(cropped_image)

```

Figure 3. Screenshots of the video cropping application code

4. EXPERIMENT

4.1. Experiment 1

The experiment to test the effectiveness of the video cropping application is a survey created in Google Forms. Fifteen participants were gathered to take the survey, which is a large enough sample size to mitigate the effects of any variability. First, each participant was asked to watch a video demo of the application and how it would affect an inputted video. Because every participant was given the same video demo, there are fewer confounding variables in the experiment. Then, the participants would be given the link to a Google Forms survey. The first question asks how practical and useful the application seemed on a scale from one to ten.

Participant Number	Usefulness Rating of Application
1	7
2	7
3	9
4	7
5	8
6	6
7	8
8	10
9	8
10	8
11	8
12	9
13	7
14	8
15	6
Average	7.73

Figure 4. Table of participant number and usefulness rating of applicant

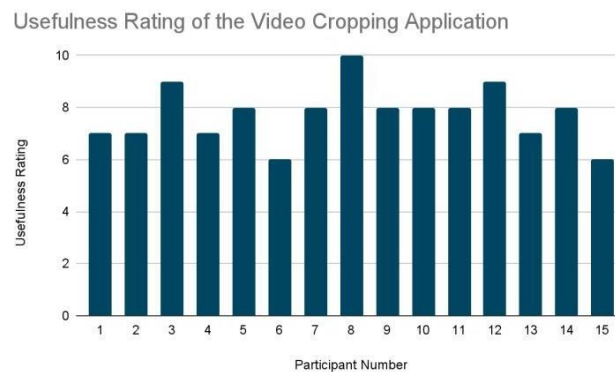


Figure 5. Graph of usefulness rating of the video cropping application

The responses from the first question were generally positive. All of the usefulness ratings from the participants were 6 or higher, and the average of the ratings was a comfortable 7.73. This could indicate that if the application were to be widely released, the general public may perceive the application as a tool that could reliably crop videos in a meaningful way that enhances the viewing experience of those who watch the refined videos. With the video demo that was . In the feedback that was provided by the participants, many said that this would be a helpful tool. Even those who claimed they were proficient and had experience in video editing admitted that this application could save a lot of time and effort in certain situations. However, one participant

expressed concern about how the application may not always be accurate when it comes to cropping the video in the right areas, and manually cropping the videos may be a more consistent choice.

4.2. Experiment 2

The second question from the Google Forms survey asked whether the application would be convenient and easy to use. Just like the previous question, this question would use a scale from one to ten for the participants to choose from. For both questions, a score of one indicates that the application completely lacked that specific quality, while a score of ten indicates that specific quality was done extremely well. Because an optional free-response section would be placed at the end of the survey, participants will have more freedom in expressing their thoughts and opinions, as the first two questions are quite limited in gathering the participants' full thoughts.

Participant Number	Convenience Rating of Application
1	6
2	7
3	10
4	8
5	6
6	7
7	8
8	10
9	9
10	7
11	8
12	7
13	8
14	8
15	6
Average	7.67

Figure 6. Table of participant number and convenience rating of applicant

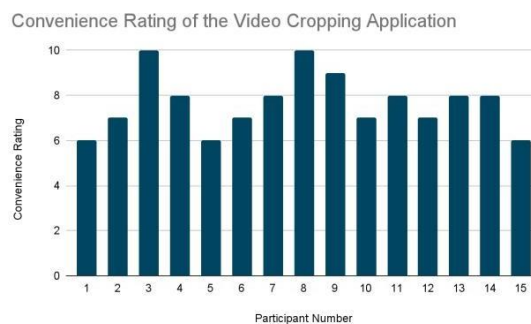


Figure 7. Graph of convenience rating of the video cropping applicant

From the results of the second question, participants were generally pleased with how intuitive the application was to use. The convenience ratings were similar to the usefulness ratings of the application, since all the ratings were 6 and above and the average was at a fairly high 7.67. However, the convenience of the application may have been reduced by the addition of too many features. The feature that emphasized the points where the application recognized the human body appeared to be particularly polarizing among the participants. While some stated in their feedback that they simply thought of such a feature as a nifty little bonus, others thought that it provided too much unnecessary information that the average user of the application simply had

no use for, as it only seemed to get in the way and people would normally already be able to tell where a person is in the video when watching it.

According to the results of the first experiment, the application can be a great help to many people around the world. The participants are regarded as a small sample of the general public, which means that the general public can make use of the application, particularly those that make dance videos or other similar videos that require cropping. A fairly high score on the usefulness of the application was to be expected, as the primary purpose of creating the application in the first place was to improve the quality of life for people who have recorded videos that could be improved through an automatic cropping system.

The results of the second experiment indicate that application will be able to be used by many without any steep learning curves and without much effort. This is significant for the success of the application, as users will likely not spend the time to learn how to use a complicated application and would prefer to use an application that requires minimal interaction and effort to reach a final product. This was also an expected result, as the general public was kept in mind during the development of the application.

5. RELATED WORK

One related work highlights how video editing can be a difficult process, and a new video editor named Silver was developed to address that. Research indicates that the use of metadata can help video editor tools become more helpful to the users, such as resolving inconsistencies in audio and video boundaries [1]. The related work is similar to this work in that tools were created in the hopes of making a complicated-seeming process like video editing become less daunting and more convenient to use. However, while the related work places more focus on video editing as a whole and dual channels of video and audio, this work focuses more on the idea of video cropping using a body detection model.

Another related work presents a way to convert video files to different aspect ratios. This would be particularly useful for devices that have screens with uncommon aspect ratios. A trained scoring algorithm is used to figure out what the screen should focus on most when cropping a video [2]. This related work is very similar to this work in that smart video cropping is performed using the help of advanced back-end code. The related work provides a large emphasis on varying aspect ratios fitting cleanly on a screen. On the other hand, this work presents an application that chooses to crop in order to focus on a person in the video.

A third related work also focuses on different aspect ratios and how to prevent the issue of awkward video crops. The solution that is presented is a video retargeting method that makes use of critical regions to determine areas of the video that should not be cropped out [3]. This related work also shares a similarity with this work in that a video cropping method is used to include important portions of the video. The related work focuses on critical regions to figure out what should stay inside the final video, while this work utilizes a person detection model involving landmarks.

6. CONCLUSIONS

To tackle the issue of people not being able to easily crop videos, an application was created that can automatically detect the person inside a video and crop out the rest of the video so that the person is the primary focus of the video. This application was mainly intended for dancing videos, but it can also be applied to other types of videos as well. The application was built using

Flutter and Python, and the application made use of a person detection model from MediaPipe Pose that tracks the “landmarks”, or body parts, of a person in video frames [13]. Using the coordinates of the landmarks, the application deletes the unneeded rows and columns of the image arrays to create the final video product.

The experiment that was performed on the application was a survey. Fifteen participants watched a video demonstration of someone using the application and showing off the features in the application. Then, the participants were asked to fill out the survey, which had two main questions. The first question asked how useful the application seems in daily life, and the second question asked how convenient using the application would be. According to the results, the participants seem to generally agree that the application would be useful and convenient based on the video demo, which can indicate that the application would likely be used by people long-term. This could also solve the issue of cropping videos in the future, and an application like this may potentially lead the way for more applications implementing a tool or feature similar to this in the future.

A major limitation in this program is that it has difficulty in processing videos containing more than one person. Running any video with multiple people can result in a lot of errors because the software is only built to acknowledge one body at a time. If multiple are present, the video product tends to jump from one person to another randomly. Another aspect that could be improved is video quality and processing speed. Additionally to improving the two, options for each could be implemented so the user could select the size of the file they wish to export.

With more development, the application could have more features to be implemented to appeal to the dance audience specifically such as special editing features, filters, effects and sharing features. For example, the user could choose to crop their video to an ideal size for TikTok, Instagram reels, Youtube, and other platforms [14].

REFERENCES

- [1] Casares, J., Long, A. C., Myers, B. A., Bhatnagar, R., Stevens, S. M., Dabbish, L., Yocum, D., & Corbett, A. (2002). Simplifying video editing using metadata. *Proceedings of the Conference on Designing Interactive Systems Processes, Practices, Methods, and Techniques - DIS '02*, 157 – 166. <https://doi.org/10.1145/778712.778737>
- [2] Deselaers, T., Dreuw, P., & Ney, H. (2008). Pan, Zoom, Scan - Time-coherent, trained automatic video cropping. *2008 IEEE Conference on Computer Vision and Pattern Recognition*. <https://doi.org/10.1109/cvpr.2008.4587729>
- [3] Wang, Y.-S., Lin, H.-C., Sorkine, O., & Lee, T.-Y. (2010). Motion-based video retargeting with optimized crop-and-warp. *SIGGRAPH '10*, (90), 1–9. <https://doi.org/10.1145/1833349.1778827>
- [4] Kopf, Stephan, et al. "Automatic scaling and cropping of videos for devices with limited screen resolution." *Proceedings of the 14th ACM international conference on Multimedia*. 2006.
- [5] Lugaresi, Camillo, et al. "Mediapipe: A framework for building perception pipelines." *arXiv preprint arXiv:1906.08172* (2019).
- [6] Zhang, Fan, et al. "Mediapipe hands: On-device real-time hand tracking." *arXiv preprint arXiv:2006.10214* (2020).
- [7] El-Alfy, Hazem, David Jacobs, and Larry Davis. "Multi-scale video cropping." *Proceedings of the 15th ACM international conference on Multimedia*. 2007.
- [8] Montag, Christian, Haibo Yang, and Jon D. Elhai. "On the psychology of TikTok use: A first glimpse from empirical findings." *Frontiers in public health* 9 (2021): 641673.
- [9] Song, Hwa-Sun. "Job analysis of video editors based on the DACUM method." *The Journal of the Korea Contents Association* 7.12 (2007): 95-104.
- [10] Deselaers, Thomas, Philippe Dreuw, and Hermann Ney. "Pan, zoom, scan— time-coherent, trained automatic video cropping." *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008.

- [11] Denning, Dorothy E. "An intrusion-detection model." *IEEE Transactions on software engineering* 2 (1987): 222-232.
- [12] Ehrlich, Danuta, et al. "Postdecision exposure to relevant information." *The journal of abnormal and social psychology* 54.1 (1957): 98.
- [13] Michon, Pierre-Emmanuel, and Michel Denis. "When and why are visual landmarks used in giving directions?." *International conference on spatial information theory*. Springer, Berlin, Heidelberg, 2001.
- [14] Che, Xianhui, Barry Ip, and Ling Lin. "A survey of current YouTube video characteristics." *IEEE MultiMedia* 22.2 (2015): 56-63.
- [15] Morrison, Alison. "Entrepreneurship: what triggers it?." *International Journal of Entrepreneurial Behavior & Research* (2000).

© 2022 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.

A SINGLE LEVEL DETECTION MODEL FOR TRAFFIC SIGN DETECTION USING CHANNEL SHUFFLE RESIDUAL STRUCTURE

Yuanzhi Luo and Jie Hao

College of Computer Science and Technology,
Nanjing University of Aeronautics and Astronautics, China

ABSTRACT

Traffic sign recognition (TSR) is a challenging task for unmanned systems, especially because the traffic signs are small in the road view image. In order to ensure the real-time and robustness of traffic sign detection in automated driving systems, we present a single level detection model for TSR which consists of three core components. The first is we use channel shuffle residual network structure to ensure the real-time performance of the system, which mainly uses low-level features to enhance the representation of small target feature information. Secondly, we use dilated convolution residual block to enhance the receptive field to detect multi-scale targets. Thirdly, we propose a dynamic and adaptive matching method for the anchor frame selection problem of small traffic signs. The experimental surface on Tsinghua-Tencent 100k Dataset and Chinese Traffic Sign Dataset benchmark has better accuracy and robustness compared with existing detection networks. With an image size of 800×800 , the proposed model achieves 92.9 running at 120 FPS on 2080Ti.

KEYWORDS

Computer Vision, Traffic Sign Detection, Convolutional Neural Networks, Label Assignment.

1. INTRODUCTION

Traffic sign recognition (TSR) is a critical component of Automated Driving Systems (ADS) [1] and High-Definition Maps (HD Map) [2], which can provide TSR aims to help make drivers more aware and able to make better safer driving decisions. Therefore, TSR must fulfill the requirements of high precision and real time. However, TSR still faces big challenges, on the one hand, the complex road conditions and natural environments that appear in the images [3], and on the other hand, most traffic signs in the Tsinghua-Tencent 100k dataset (TT100K) [4] are smaller than 32×32 pixels, which means that most traffic signs account for less than 0.3% of the image and detecting small objects is more challenging than large ones.

As a kind of object detection problem, TSR usually shares the same detection algorithms based on convolutional neural networks. On one hand, two-stage approaches such as Fast RCNN [5], Faster RCNN [6], Mask RCNN [7] and Cascade RCNN [8] use region proposals to detect objects. Although these neural networks are superior in terms of accuracy, the low per-frame detection speed limits their application in real-time TSR. On the other hand, one-stage detectors such as SSD [9], YOLO series [10]-[11] [12] [13], YOLOX [14], etc. possess good speedups. Classic detectors still play a great role in TSR. Faster R-CNN [15] and SSD [16] are adapted to reduce the computational complexity for TSR. Yuan [17] et al. extracted regions of interest by adding attention module to CNN to refine the feature extraction of traffic signs in complex

backgrounds. Zhang [18] et al. introduced image enhancement and spatial pyramid pooling (SPP) modules based on YOLOv3 to effectively fuse low-level and high-level features. These efforts have had some effect, but it is still not enough to satisfy the demands of TSR. This has inspired a series of works targeting lightweight architecture design and better speed-accuracy trade-offs, including MobileNet [19], MobileNetV2 [20], ShuffleNet [21], ShuffleNet V2 [22], and so on. Although efficient, they are still not sufficient for low-dimensional feature extraction. In order to achieve real-time, high accuracy, and we present a single level detection model for TSR which can extract the low-dimensional feature more quickly and efficiently, as shown in Figure 1.

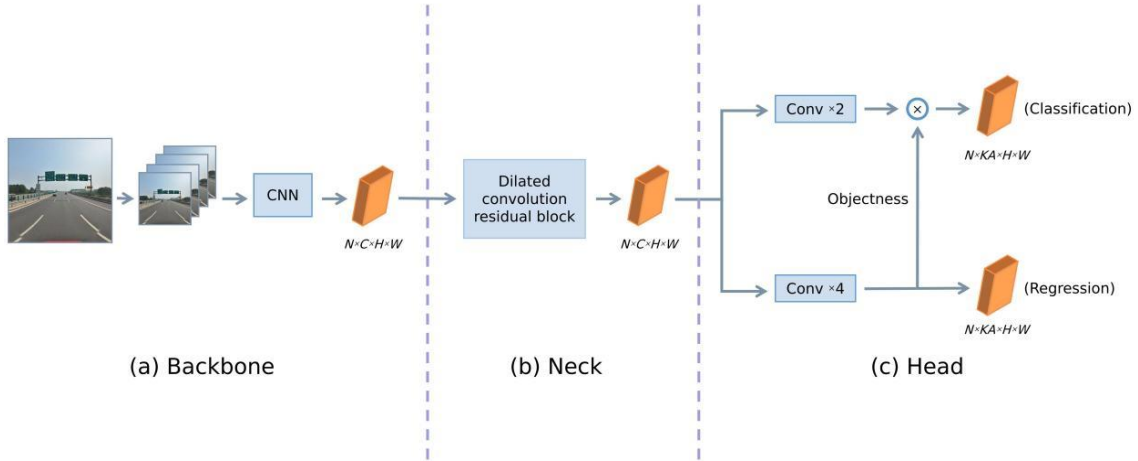


Figure 1. The network structure of TSD.

The main contributions of this paper are as follows:

- 1) We propose an effective TSR detection model, which uses channel shuffle residual structure for low-level feature extraction, which can effectively reduce the computational overhead of the model.
- 2) We introduce dilated convolution to increase the receptive field [23] and enhance the feature representation using an attention mechanism to improve the robustness of the model.
- 3) We design an adaptive matching training samples algorithm to solve the problem of unbalanced distribution of small-scale traffic sign training samples, which can dynamically assign positive anchors according to the targets of different scales to improve the detection effect and robustness of traffic signs.
- 4) We evaluate our model on the Tsinghua-Tencent 100k dataset (TT100K) [4] and the Chinese traffic sign dataset (CTSD) [24], and the results show that our model achieves outstanding performance in terms of faster speed and higher accuracy compared with state-of-the-art work.

2. PREPARATION

As a typical small object detection problem, the ability to detect small targets is critical for a TSR model [25]. CNNs use multi-level convolution and pooling operations to obtain deeper semantic features. These operations result in small objects existing only at shallow layers, but shallow features are not powerful enough in complex traffic scenes due to the lack of deep semantic

information. Therefore, the researchers have devoted their efforts to improving the small object feature representation capacity, including designing multiscale feature fusion backbones, adding feature pyramids to neural networks [26], and exploiting attention mechanisms [27].

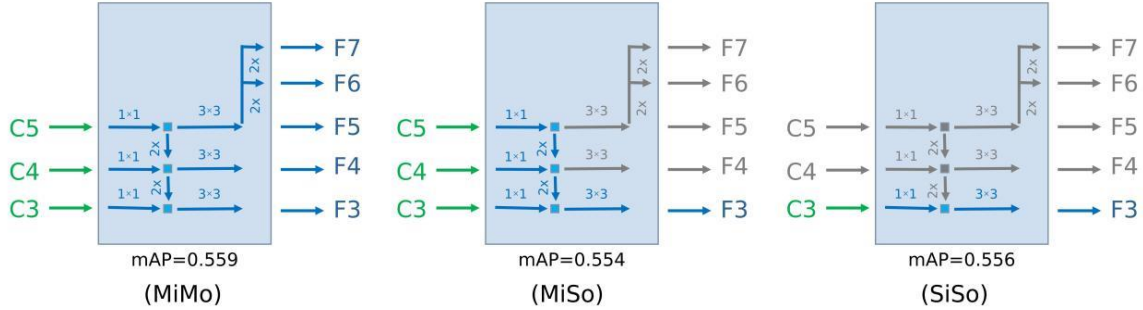


Figure 2. Three structures of FPN in the RetinaNet network. C3~C5 are the input feature layers with downsampling multiplier 8, 16, 32, respectively, and F3~F7 are the output layers after FPN, respectively. The gray lines denote unused channels.

Among all, FPN [26] is a popular module as it can rise the detection accuracy by fusing multi-scale features, in which small scale features contributes the most to TSR. Therefore, before we go to the elaboration of our detector, we first evaluate to what extent of the contribution of FPN in TSR, which we hope can inspire our design of TSR detector. Unfortunately, we found that the depth pyramid level plays a secondary role in TSR. Figure 2 shows the three structures of FPN we've evaluated, and Table. 1 shows the results of FPNs on RetinaNet [28] network and TT100K dataset.

Table 1. The result under TT100K dataset with 800×800 input.

model	mAP	AP ₅₀	AP _s	AP _m	AP _l
MiMo	0.559	0.841	0.266	0.690	0.778
MiSo	0.554	0.841	0.283	0.698	0.720
SiSo	0.556	0.847	0.285	0.698	0.750

From Table. 1, we found that MiMo, MiSo, and SiSo achieve comparable performance in TSR, and even SiSo structure is the best for small targets. We speculate that the small target information on C3 feature layer is detailed enough without using the complex information on C4 and C5 layers. Table. 1 demonstrates that FPN is of limited help in TSR. This is also claimed by YOLOF [29]. Motivated by this observation, we design a detector based on the SiSo structure and remove the deep layer to reduce the inference time. Also we redesign the network structure to extract traffic sign features efficiently, and design an adaptive allocation strategy in order to balance the problem of under-allocation of positive samples for small traffic sign training.

3. METHODOLOGY

This section proposes a real-time yet efficient single-stage detection framework based on SiSo structure for TSR, denoted as TSD hereafter. Its schematic diagram is shown in Figure 1. The backbone is used for feature extraction, the neck further processes the features and distributes them to the head, and the head performs the classification and regression tasks and generates the final prediction result. In this section, we will elaborate the three components of TSD respectively.

3.1. Backbone

The residual structure [30] can effectively solve the problem of gradient disappearance and gradient explosion caused by too deep network structure. However, a larger network structure also brings a boosting number of parameters and computational overhead. In order to better achieve the accuracy and real-time performance of TSR, we designed a new residual structure as shown in Figure 3. We replaced the conventional residual structure with CS residual structure, and the parameters were reduced by 36.7% with comparable performance.

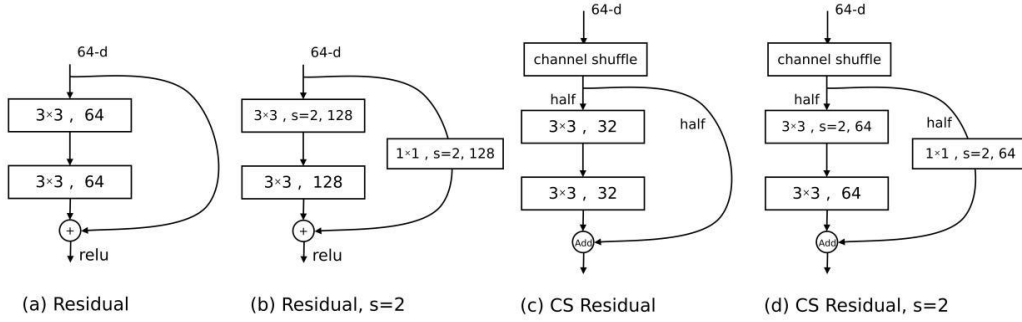


Figure 3. Residual structures. (a),(b) are traditional residual structures, c and d are CS residual structures we designed. where 64-d denotes the input dimension, Channel shuffle randomly disrupts the channel, which allows information exchange between channels of the same group.

The network parameters of the entire backbone are shown in Table 2. The Focus module can obtain a two-fold downsampling feature map without information loss, and the CBR module consists of Convolutional, Batchnorm [31], and ReLU activation functions [32].

Table 2. Structure of Backbone.

layer-name	Parametric	Output-size
Input	-	800×800×3
Focus	$\begin{bmatrix} \text{Slice} \\ \text{CBR } 1 \times 1 \end{bmatrix}$	400×400×32
CBR	$\begin{bmatrix} \text{conv } 3 \times 3, s = 1 \\ \text{Batchnorm} \\ \text{ReLU} \end{bmatrix}$	400×400×32
CS Residual, s=2	$\begin{bmatrix} \text{half1: CBR } 3 \times 3, s = 2 \\ \text{half2: CBR } 1 \times 1, s = 2 \\ \text{half1: CBR } 3 \times 3 \end{bmatrix}$	200×200×64
CS Residual	$\begin{bmatrix} \text{half1: CBR } 3 \times 3 \\ \text{half1: CBR } 3 \times 3 \end{bmatrix}$	200×200×64
CS Residual, s=2	$\begin{bmatrix} \text{half1: CBR } 3 \times 3, s = 2 \\ \text{half2: CBR } 1 \times 1, s = 2 \\ \text{half1: CBR } 3 \times 3 \end{bmatrix}$	100×100×128
CS Residual	$\begin{bmatrix} \text{half1: CBR } 3 \times 3 \\ \text{half1: CBR } 3 \times 3 \end{bmatrix}$	100×100×128

3.2. Neck

The input of the neck is the output of backbone, which we first simply process using a 1×1 convolution and a 3×3 convolution to obtain a 128-channel feature layer. Then, in order to make the output features of neck cover all the targets on various scales, we use the dilated convolution [33] to form the residual block, whose rate is 2, 3, 5 in order, and the channel of the dilated convolution is set to half of the input channel, and we add SELayer [34] to enhance the feature expression after the convolution, and the overall structure of the neck is shown in Figure 4.

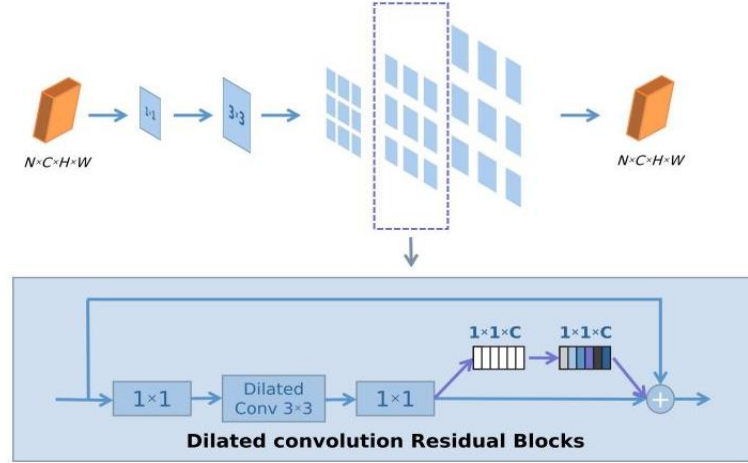


Figure 4. The neck structure. 1×1 and 3×3 represent the size of the convolution kernel, and three consecutive dilated convolution residual blocks can increase the receptive field range, and in the residual blocks we add an attention mechanism to enhance the feature representation, and all convolutional layers are followed by a Batchnorm layer and a ReLU layer.

3.3. Head

For the head, we use two parallel task-specific head compositions: a classification head and a detection head. We follow the design of the FFN in DETR [35], so that the two heads have different number of convolution layers. The classification head has a 1×1 convolution and a 3×3 convolution, and the regression head has a 1×1 convolution and three 3×3 convolutions, followed by batch normalization layer and ReLU layer. Also, we follow the design in Auto assign [36] by adding an implicit objectness prediction (without direct supervision) for each anchor on the regression head. The final classification scores for all predictions are generated by multiplying the classification output with the corresponding implied objectivity.

The loss function consists of categorical loss and regression loss, where we use Focal Loss as the category loss function and GIoU Loss as the location loss function. The total loss is shown in Equation 1.

$$\text{Loss} = \gamma_1 \frac{1}{N_{\text{pos}}} \sum_i L_{\text{cls}}^i + \gamma_2 \frac{1}{N_{\text{pos}}} \sum_j L_{\text{reg}}^j \quad (1)$$

Where N_{pos} denotes the number of positive samples, i denotes all positive and negative samples, j denotes all positive samples, and γ_1 and γ_2 are the values of the weights learned by the network.

3.4. Label Assignment

The definition of a positive sample is crucial for the optimization of the target detection problem, and in the anchor-based approach, the definition of a positive sample is based on the IoU between the anchor and the GT frame. For example, in RetinaNet [37], an anchor is a positive sample if the maximum IoU between it and the GT frame is greater than a given threshold value of 0.5. This strategy is called Max-IoU matching. When facing small-scale targets and large scale gaps between targets, this methods have certain drawbacks, the large target will match more positive anchors than the small target, which will lead to an unbalanced distribution of positive and negative anchors. Small targets may match few or no positive anchors, which makes the detector pay more attention to the large ground-truth boxes and ignore the small ground-truth boxes during training, resulting in poor detection. Thus we propose an adaptive matching strategy.

Based on the pre-generated anchors, we additionally use the predicted bbox as supplementary information and propose an adaptive matching strategy: calculate the IoU value of the predefined anchor and GT, select the IOU value greater than 0.15 At the same time, we use the predicted bbox with the IoU value of GT greater than 0.15 as a supplementary candidate positive sample. Then we calculate the mean and standard deviation based on the total candidate positive samples to obtain a threshold, and then use the threshold to filter the final set of positive samples. The specific algorithm is as follows:

Algorithm 1 Adaptive matching training samples	
Input:	G is a set of ground-truth boxes on the image A is a set of all anchor boxes P is a set of all Prediction boxes Y is the initial IoU threshold with a default value of 0.15
Output:	O is a set of positive samples N is a set of negative samples I is a set of ignore samples
Detail:	<pre> 1: for each ground-truth $g \in G$ do 2: build an empty set for candidate positive samples of the ground-truth g: C_g; 3: for each $a_i \in A$ do 4: compute IoU between g and a_i: $A_g = \text{IoU}(a_i, g)$; 5: if $A_g = \text{IoU}(a_i, g) > Y$ then 6: $C_g = C_g \cup a_i$; 7: end if 8: end for 9: for each $p_i \in P$ do 10: compute IoU between g and p_i: $P_g = \text{IoU}(p_i, g)$; 11: if $P_g = \text{IoU}(p_i, g) > Y$ then 12: $C_g = C_g \cup p_i$; 13: end if 14: end for 15: compute IoU between C_g and g: $D_g = \text{IoU}(C_g, g)$; 16: compute mean of D_g: $m_g = \text{Mean}(D_g)$; 17: compute standard deviation of D_g: $v_g = \text{Std}(D_g)$; </pre>


```

18:   compute IoU threshold for ground-truth  $g$ :  $t_g = m_g + v_g$ ;
19:   for each candidate  $c \in C_g$  do
20:       if  $\text{IoU}(c, g) \geq t_g$  then
21:            $P_2 = P_2 \cup c$ ;
22:       elif  $l \leq \text{IoU}(c, g) \leq t_g$  then
23:            $I = I \cup c$ ;
24:       end if
25:   end for
26: end for
26:  $N = A - P_2 - I$ ;
27: return  $P_2, N, I$ ;

```

4. EXPERIMENTS AND RESULTS

We evaluated our TSD on TT100k and CTSD, using the COCO benchmark [38]. Most of the category in the TT100K are less than 100, and to make full use of the annotation information, we divided both datasets into three categories (danger, prohibitory, mandatory). We divide all traffic signs into small, medium, and large groups, which can compare the performance of traffic signs at different scales in more detail. We compare with some currently popular lightweight detection frameworks, such as YOLOv4-Tiny [39]. A description of the datasets and specific experimental details are given below.

4.1. Datasets

CTSD has 700 annotated images in the training set and 400 in the test set, and the traffic signs are classified into three categories (danger, prohibitory, mandatory). The training set of TT100K includes 6105 labeled images and the test set has 3071 images, with a total of 232 categories. Considering many of them have very few, and in order to make full use of their labeling information, we process them into the same three categories as CTSD. As shown in Figure 5. For TT100K, we divided the traffic signs into three categories: small, medium, and large according to their bounding box sizes, with those smaller than 32×32 being small targets, those with sizes ranging from 32×32 to 96×96 being medium targets, and those larger than 96×96 being large targets. The original annotation of the training set of TT100K consists of 51.43% small targets, 41.45% medium targets, and 7.12% of large targets, as shown in Figure 6. When we resize it to an image size of 800×800 for training, the percentages of small, medium, and large targets are 89.28%: 10.69%: 0.04%, respectively. Table 3 contains the image size and annotation size for both datasets.

Table 3. Details of TT100K and CTSD.

Dataset	train images	test images	image size(px)	sign size(px)
TT100K	6105	3071	2048×2048	8×9~285×343
CTSD	700	400	768×1024 ~720×1280	20×20 ~ 380×378

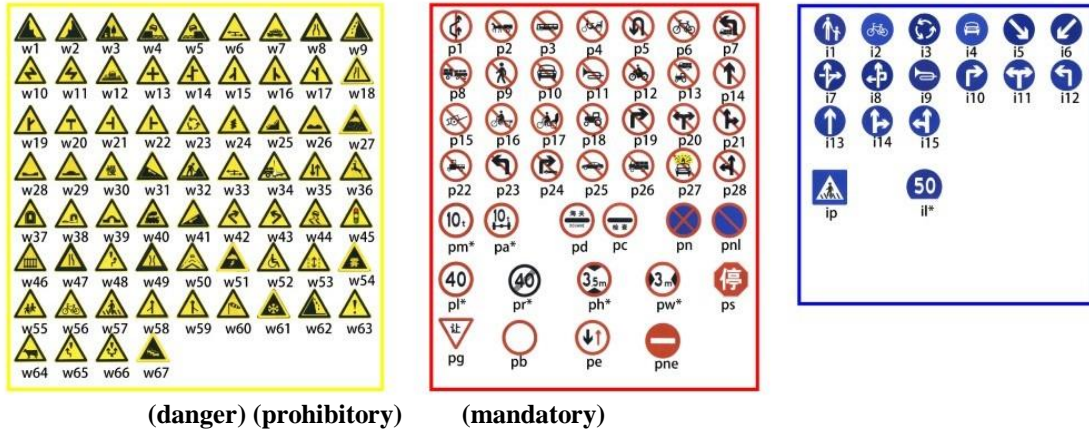


Figure 5. Traffic sign categories in TT100K.

4.2. Experiment Environment

All experiments in this paper are conducted on Ubuntu 20.04LTs system with GeForce RTX 2080Ti, based on MMDetection [40] framework, except for SSDLite where the input image size is 320×320 . The rest of the image preprocessing is exactly the same, each image is resized to 800×800 , RandomFlip, RandomShift, Normalize. RandomFlip, RandomShift, Normalize. Batchsize is set to 8, and we adopt the learning rate setting in DETR [28], the initial learning rate is set to 0.02, and a smaller learning rate is set in the backbone network, i.e., $1/3$ of the initial learning rate to stabilize the training at the beginning. We set the warmup times as 1500 times. For model inference, we post-processed the results using an NMS with a threshold of 0.5.

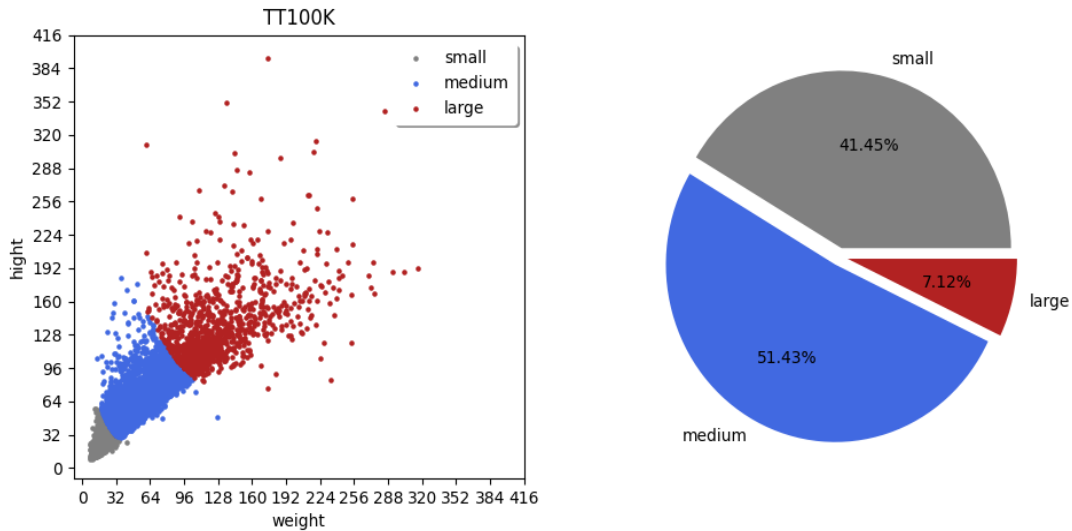


Figure 6. Sample distribution in the used dataset. The left figure shows the two-dimensional distribution of large, medium and small targets in the original annotation of TT100K training set, and the right one shows the percentage of large, medium and small targets in the training set.

4.3. Comparison With Previous Work

We used AP (Average Precision) (AP_s , AP_m , AP_l), AP_{50} , and AP_{75} which were introduced by MS COCO [38] benchmark to evaluate the accuracy and the number of images processed per second (FPS) to evaluate the inference speed. Since the image size in CTSD is not uniform, FPS metric is not evaluated. The results are shown in Table 4 and Table 5. The readers can also refer to Figure 7 for a more intuitive comparison result.

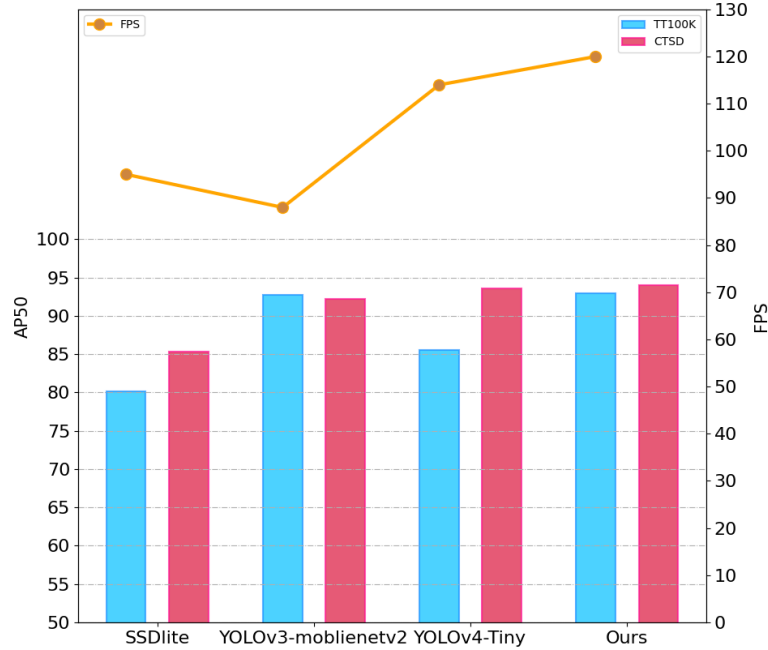


Figure 7. Compare our method with the previous method on TT100K and CTSD.

We can see that our detector can achieve the best results in most cases. Both YOLOv3-MobileNetV2 and YOLOv4-Tiny used the Kmeans method [41] to obtain the anchor sizes, and the anchor settings for TT100K and CTSD were $[(73, 73), (110, 111), (157, 154)]$, $[(31, 31), (42, 41), (56, 55)]$, $[(16, 16), (20, 20), (25, 25)]$ and $[(23, 25), (30, 32), (43, 45)]$, $[(12, 14), (15, 16), (19, 20)]$, $[(6, 7), (8, 9), (10, 11)]$.

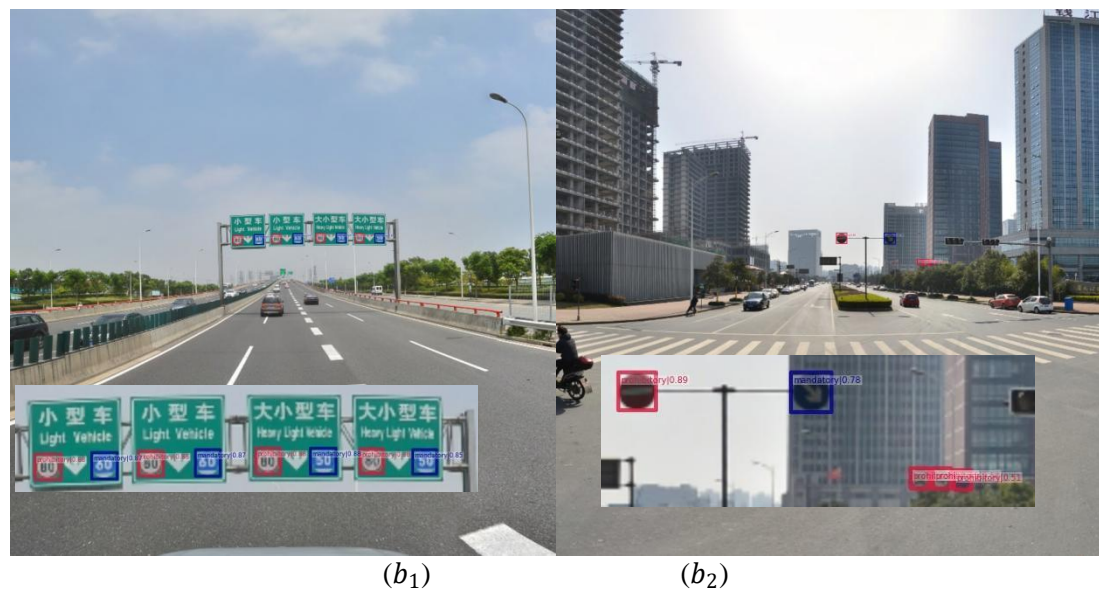
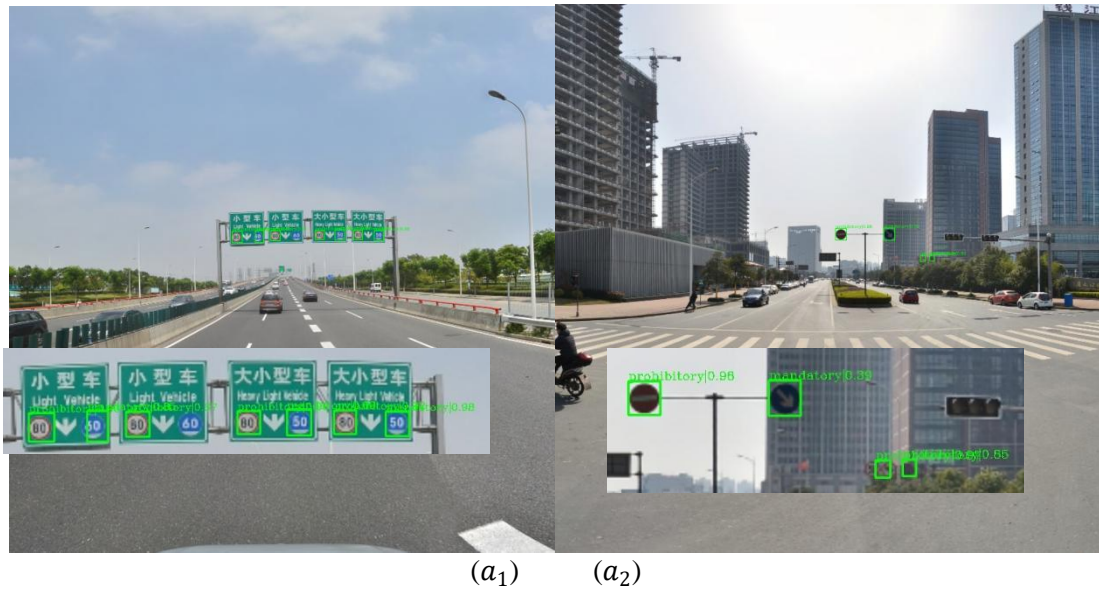
Table 4. Comparison of our method with other methods on TT100K.

method	mAP	AP_{50}	AP_s	AP_m	AP_l	FPS
SSDLite	0.460	0.802	0.264	0.582	0.635	95
YOLOv3-MobileNetV2	0.545	0.927	0.455	0.597	0.602	88
YOLOv4-Tiny	0.437	0.855	0.264	0.530	0.612	114
TSD (Ours)	0.643	0.929	0.498	0.717	0.746	120

In terms of accuracy, our method achieves the highest AP metric of 92.9% on TT100K, which is 7.4% higher than YOLOv4-Tiny, which indicates that our method has good results for small object detection. In the CTSD, YOLOv4-Tiny is comparable to our method. Please note, TSD is able to reach 94.3% if inputting 1024×1024 images for inference.

Table 5. Comparison of our method with other methods on CTSD.

method	mAP	AP ₅₀	AP _s	AP _m	AP _l
SSDLite	0.538	0.853	0.268	0.576	0.563
YOLOv3-MobileNetV2	0.644	0.922	0.411	0.687	0.532
YOLOv4-Tiny	0.722	0.936	0.430	0.767	0.765
TSD (Ours)	0.736	0.940	0.556	0.771	0.731

Figure 8. Partial visualization results, a₁ and a₂ are the visualization results of Yolov4-Tiny, b₁ and b₂ are our results.

In terms of inference speed, we test on TT100k with input image size of 800×800 for inference, our method has FPS of 120, which is slightly faster than YOLOv4-Tiny and fully satisfies the real-time requirement.

To show the results of our experiments more visually, Figure 8 shows the visualization of some tested samples on YOLOv4-Tiny and TSD.

4.4. Ablation Experiment

We investigated the effectiveness of the backbone and neck components of TSD, and all experiments were performed on TT100K.

In Table 6, we replace the CS Residual structure of the backbone part with the conventional residual structure, and we can see that CS Residual structure achieves comparable performance with the conventional residual structure, but CS Residual structure parameters were reduced by 70% and FLOPs by 61.8% when the output channel of the backbone is 128.

Table 6. backbone uses different residual structures for comparison.

Backbone	AP ₅₀	Params(M)	FLOPs (GFLOPs)
Residual	0.929	0.667	12.101
CS Residual(ours)	0.929	0.2 (70% ↓)	4.621 (61.8%↓)

We tested different choices of dilation for the expansion convolution in the neck part, and the results are shown in Table 7. We can see that the difference in model results is not significant for different choices of dilation, and we guess the reason is that the target scale of TT100K is small, and the small receptive field is enough to cover most of the targets in the image. We learned that the grid effect occurs when the dilation of the dilated convolution is greater than 1 [33]. To ensure the robustness of the model, we set the dilation to 2, 3, 5.

Table 7. Different dilations, RF means receptive field.

Dilation	AP ₅₀	AP _s	AP _m	AP _l	RF
1,2,3	0.926	0.490	0.713	0.748	13x13
2,2,2	0.925	0.492	0.711	0.751	13x13
1,2,5	0.924	0.486	0.712	0.761	17x17
2,3,5	0.929	0.498	0.717	0.746	21x21
2,4,6	0.921	0.479	0.710	0.754	25x25

5. CONCLUSIONS

In this work, we aim to improve the speed and accuracy of small traffic sign detection, and we find that FPN have limited improvement in detection capability for small targets. We propose an effective framework TSD for small TSR, detailing various optimization strategies, including for proposing CS Residual structure, Dilated convolution blocks and adaptive matching training sample methods. Experiments show that our method outperforms some lightweight detection algorithms on TT100K and CTSD.

Our model has some limitations when facing the detection of very large targets, because we focus more on the feature information within a certain range. If you need to, we suggest that you can further increase the range of receptive field or create another branch to detect large objects.

For future work, we will try to optimize the detection algorithm using the anchor-free approach, which can eliminate our customization work for anchor and enhance the robustness of the model.

ACKNOWLEDGEMENTS

This work is supported in part by the National Key R&D Program of China under Grant 2019YFB2102000, and in part by the Collaborative Innovation Center of Novel Software Technology and Industrialization.

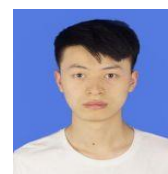
REFERENCES

- [1] A H U , A N K , B S Y . A Safety Knowledge Representation of the Automatic Driving System[J]. *Procedia Computer Science*, 2016, 96:869-878.
- [2] Takeuchi E , Yoshihara Y , Ninomiya Y . Blind Area Traffic Prediction Using High Definition Maps and LiDAR for Safe Driving Assist[C]// *IEEE International Conference on Intelligent Transportation Systems*. IEEE, 2015.
- [3] Muhammad K , Ullah A , Lloret J , et al. Deep Learning for Safe Autonomous Driving: Current Challenges and Future Directions[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, PP(99):1-21.
- [4] Zhe Z , Liang D , Zhang S , et al. Traffic-Sign Detection and Classification in the Wild[C]// *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016.
- [5] Girshick R. Fast R-CNN[J]. *Computer Science*, 2015.
- [6] Ren S , He K , Girshick R , et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39(6):1137-1149.
- [7] He K , Gkioxari G , P Dollár , et al. Mask R-CNN[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017.
- [8] Cai Z , Vasconcelos N . Cascade R-CNN: Delving into High Quality Object Detection[J]. 2017.
- [9] Liu W , Anguelov D , Erhan D , et al. SSD: Single Shot MultiBox Detector[J]. 2015.
- [10] Redmon J , Divvala S , Girshick R , et al. You Only Look Once: Unified, Real-Time Object Detection[J]. IEEE, 2016.
- [11] Redmon J , Farhadi A . YOLO9000: Better, Faster, Stronger[J]. IEEE, 2017:6517-6525.
- [12] Redmon J , Farhadi A . YOLOv3: An Incremental Improvement[J]. *arXiv e-prints*, 2018.
- [13] Bochkovskiy A , Wang C Y , Liao H . YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. 2020.
- [14] Ge Z , Liu S , Wang F , et al. YOLOX: Exceeding YOLO Series in 2021[J]. 2021.
- [15] Zuo Z , Kai Y , Qiao Z , et al. Traffic Signs Detection Based on Faster R-CNN[C]// *2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW)*. IEEE, 2017.
- [16] You S , Bi Q , Ji Y , et al. Traffic Sign Detection Method Based on Improved SSD[J]. *Information*, 2020, 11.
- [17] Yuan Y , Xiong Z , Wang Q . VSSA-NET: Vertical Spatial Sequence Attention Network for Traffic Sign Detection[J]. *IEEE Transactions on Image Processing*, 2019:1-1.
- [18] Zhang H , Qin L , Li J , et al. Real-Time Detection Method for Small Traffic Signs Based on Yolov3[J]. *IEEE Access*, 2020, PP(99):1-1.
- [19] Howard A G , Zhu M , Chen B , et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. 2017.
- [20] Sandler M , Howard A , Zhu M , et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks[J]. IEEE, 2018.
- [21] Zhang X , Zhou X , Lin M , et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices[J]. 2017.

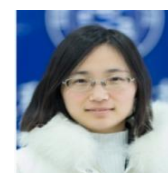
- [22] Ma N , Zhang X , Zheng H T , et al. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design[J]. Springer, Cham, 2018.
- [23] Li Y , Chen Y , Wang N , et al. Scale-Aware Trident Networks for Object Detection[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019.
- [24] Yi Y , Luo H , Xu H , et al. Towards Real-Time Traffic Sign Detection and Classification[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(7):2022-2031.
- [25] Deng C . A Review on the Extraction of Region of Interest in Traffic Sign Recognition System[C]// 2020 International Conference on Computing and Data Science (CDS). 2020.
- [26] Lin T Y , Dollar P , Girshick R , et al. Feature Pyramid Networks for Object Detection[J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- [27] Vaswani A , Shazeer N , Parmar N , et al. Attention Is All You Need[C]// arXiv. arXiv, 2017.
- [28] Lin T Y , Goyal P , Girshick R , et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):2999-3007.
- [29] Chen Q , Wang Y , Yang T , et al. You Only Look One-level Feature[J]. 2021.
- [30] He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.
- [31] Ioffe S , Szegedy C . Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. JMLR.org, 2015.
- [32] Glorot X , Bor De S A , Bengio Y . Deep Sparse Rectifier Neural Networks[C]// Journal of Machine Learning Research. 2011:315-323.
- [33] Yu F , Koltun V , Funkhouser T . Dilated Residual Networks[J]. IEEE Computer Society, 2017.
- [34] Jie H , Li S , Gang S , et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, PP(99).
- [35] Carion N , Massa F , Synnaeve G , et al. End-to-End Object Detection with Transformers[J]. 2020.
- [36] Zhu B , Wang J , Jiang Z , et al. AutoAssign: Differentiable Label Assignment for Dense Object Detection[J]. 2020.
- [37] Lin T Y , Goyal P , Girshick R , et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):2999-3007.
- [38] Lin T Y , Maire M , Belongie S , et al. Microsoft COCO: Common Objects in Context[C]// European Conference on Computer Vision. Springer International Publishing, 2014.
- [39] Bochkovskiy A , Wang C Y , Liao H . YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. 2020.
- [40] Chen K , Wang J , Pang J , et al. MMDetection: Open MMLab Detection Toolbox and Benchmark[J]. 2019.
- [41] Bahmani B , Moseley B , Vattani A , et al. Scalable K-Means++[J]. Proceedings of the VLDB Endowment, 2012, 5(7):622-633.

AUTHORS

Yuanzhi Luo is a graduate student at Nanjing University of Aeronautics and Astronautics, College of Computer Science and Technology. His research interests include deep learning, image processing, and computer vision.



Jie Hao received her BS degree from Beijing University of Posts and Telecommunications, China, in 2007, and the Ph.D. degree from University of Chinese Academy of Sciences, China, in 2014. From 2014 to 2015, she has worked as post-doctoral research fellow in the School of Computer Engineering, Nanyang Technological University, Singapore. She is currently an Assistant Professor at College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China. Her research interests are ubiquitous perception, multimodal data fusion and Internet of Things.



HYPER-PARAMETER TUNING IN DEEP NEURAL NETWORK LEARNING

Tiffany Zhan

USAOT, Las Vegas, Nevada, USA

ABSTRACT

Deep learning has been increasingly used in various applications such as image and video recognition, recommender systems, image classification, image segmentation, medical image analysis, natural language processing, brain-computer interfaces, and financial time series. In deep learning, a convolutional neural network (CNN) is regularized versions of multilayer perceptrons. Multilayer perceptrons usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The full connectivity of these networks makes them prone to overfitting data. Typical ways of regularization, or preventing overfitting, include penalizing parameters during training or trimming connectivity. CNNs use relatively little pre-processing compared to other image classification algorithms. Given the rise in popularity and use of deep neural network learning, the problem of tuning hyper-parameters is increasingly prominent tasks in constructing efficient deep neural networks. In this paper, the tuning of deep neural network learning (DNN) hyper-parameters is explored using an evolutionary based approach popularized for use in estimating solutions to problems where the problem space is too large to get an exact solution.

KEYWORDS

Deep Learning, Convolutional Neural Network, Deep Neural Network Learning, Hyper-Parameters.

1. DEEP LEARNING

In deep learning, a convolutional neural network (CNN) is a class of artificial neural network (ANN), most commonly applied to image and video recognition, recommender systems, image classification, image segmentation, medical image analysis, natural language processing, brain-computer interfaces, and financial time series [1-10]. CNNs are regularized versions of multilayer perceptrons. Multilayer perceptrons usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The "full connectivity" of these networks makes them prone to overfitting data. Typical ways of regularization, or preventing overfitting, include penalizing parameters during training (such as weight decay) or trimming connectivity (skipped connections, dropout, etc.) CNNs take a different approach towards regularization: they take advantage of the hierarchical pattern in data and assemble patterns of increasing complexity using smaller and simpler patterns embossed in their filters. Therefore, on a scale of connectivity and complexity, CNNs are on the lower extreme [11, 12].

Convolutional networks were inspired by biological processes with relatively little pre-processing compared to other image classification algorithms [13]. This means that the network learns to optimize the filters (or kernels) through automated learning, whereas in traditional algorithms

these filters are hand-engineered. This independence from prior knowledge and human intervention in feature extraction is a major advantage. A CNN consists of an input layer, hidden layers and an output layer. In any feed-forward neural network, any middle layers are called hidden because their inputs and outputs are masked by the activation function and final convolution. In a CNN, the input is a tensor with a shape: (number of inputs) \times (input height) \times (input width) \times (input channels). After passing through a convolutional layer, the image becomes abstracted to a feature map, also called an activation map, with shape: (number of inputs) \times (feature map height) \times (feature map width) \times (feature map channels). Convolutional layers convolve the input and pass its result to the next layer. This is similar to the response of a neuron in the visual cortex to a specific stimulus [14]. Each convolutional neuron processes data only for its receptive field. Convolutional networks may include local and/or global pooling layers along with traditional convolutional layers. Pooling layers reduce the dimensions of data by combining the outputs of neuron clusters at one layer into a single neuron in the next layer. Local pooling combines small clusters, tiling sizes such as 2×2 are commonly used. Global pooling acts on all the neurons of the feature map [15]. Each neuron in a neural network computes an output value by applying a specific function to the input values received from the receptive field in the previous layer. The function that is applied to the input values is determined by a vector of weights and a bias (typically real numbers). Learning consists of iteratively adjusting these biases and weights.

2. HYPER PARAMETERS

Constructing an efficient DNN for given applications is not a trivial task. It involves significant domain knowledge and efforts in exploring the properties of the data. For example, how sparse the data is, how many training or test samples are available, the definition of the data and its types, and the data representation power [16]. The goal of DNN is to be able to represent the available data as precisely as possible while avoiding the problem of over-fitting, thus constructing an efficient DNN should fit the data but not over-fit the data, which means that time must be spent determining when the data is being over-fit and when it is not. Fortunately, DNNs have several parameters that define their overall structure. These parameters are referred to as hyper-parameters in that they are used to define the structure of the DNN rather than are parameters to be used by the DNN [17].

There are many parameters to define DNNs, for example, the number of layers in the DNN, the number of nodes within a given layer, the algorithms used between layers, the overall algorithm used for the network, the optimization techniques [18] involved, the activation functions, the number of epochs, the size of batches, the number of folds, etc. Since there are variety of hyper-parameters, a critical question is what hyperparameters to be used for given problems. Exploring possible configurations of hyper-parameters will significantly impact the results. To complicate matters even further those parameters are often only reasonably transferable to another problem if the problems themselves can be mapped into the same problem space. Even if we may reasonably transfer some parameters to similar problems, there are often unique qualities within the datasets that differentiate them to a degree where one DNN configuration is not necessarily the most efficient for the next DNN configuration. The hyper-parameter tuning has been often based on intuition rather than scientific rationale because one experience with a set of parameters does not necessarily directly translate into another researcher's experience. This paper will explore the scientific evolutionary principles using genetic algorithms for hyper-parameter tuning.

3. GENETIC ALGORITHMS

Genetic algorithm (GA) is a metaheuristic inspired by the process of natural selection that belongs to the larger class of evolutionary algorithms. Genetic algorithms are commonly used to generate high-quality solutions to optimization and search problems by relying on biologically inspired operators such as mutation, crossover and selection [19, 20]. In a genetic algorithm, a population of candidate solutions (called phenotypes) to an optimization problem is evolved toward better solutions. Each candidate solution has a set of properties (its genotype) which can be mutated and altered; traditionally, solutions are represented in binary as strings of 0s and 1s, but other encodings are also possible [20, 21]. The evolution usually starts from a population of randomly generated individuals, and is an iterative process, with the population in each iteration called a generation. In each generation, the fitness of every individual in the population is evaluated; the fitness is usually the value of the objective function in the optimization problem being solved. The more fit individuals are stochastically selected from the current population, and each individual's genome is recombined or possibly randomly mutated to form a new generation. The new generation of candidate solutions is then used in the next iteration of the algorithm. Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population. A typical genetic algorithm involves a genetic representation of the solution domain, a fitness function to evaluate the solution domain. Once the genetic representation and the fitness function are defined, a GA proceeds to initialize a population of solutions and then to improve it through repetitive application of the mutation, crossover, inversion and selection operators.

The initial population in GA is generated randomly, allowing the entire range of possible solutions (the search space). During each successive generation, a portion of the existing population is selected to breed a new generation. Individual solutions are selected through a fitness-based process, where fitter solutions (as measured by a fitness function) are typically more likely to be selected. The fitness function is defined over the genetic representation and measures the quality of the represented solution. Genetic operators include Crossover (genetic algorithm) and Mutation (genetic algorithm). Although crossover and mutation are known as the main genetic operators, it is possible to use other operators such as regrouping, colonization-extinction, or migration in genetic algorithms [22, 23, 24, 25, 26]. This generational process is repeated until a termination condition has been reached. Common terminating conditions include a solution is found that satisfies minimum criteria, fixed number of generations reached, allocated budget (computation time/money) reached, the highest-ranking solution's fitness is reaching or has reached a plateau such that successive iterations no longer produce better results.

The best individual is kept and by the end of the GA run the best individual is the one with the optimal result. In this case optimal refers to the fitness function of the individual, which is the function that defines how good an individual is. Extrapolating this towards DNN we can then say that if we can express an individual as the set of hyper-parameters of a DNN then the optimal DNN configuration is the one which evaluates to the best outcome. The best outcome for a DNN is the one that achieves the best accuracy and the best learning rate. By treating the parameters that define a DNN as an optimization problem we map this problem to a Genetic Algorithm to find the best possible result within a given time.

4. RESEARCH PROBLEM

During the confirmation of DNNs, we need to determine which parameters to use. In this context, I capture a small subset of the parameters in a DNN and attempt to optimize them. It is possible

to model a DNN with much greater complexity but in order to verify the results it will take much longer than the computational resources available to complete.

The research problem is defined as follows: using a DNN and given datasets, can a GA be used to generate better results in a reasonable time period? Here a reasonable period refers to under one hour of time and the parameters explored are the batch sizes, number of layers, nodes within layers, and activations used within layers.

5. ALGORITHMIC SOLUTION

Since the number of possible combinations of these parameters is extremely large, an upper bound has been placed on these values in order to comply with the requirement that a result can be found within a reasonable time period. The activations used are those available within the Python Scikit-Learn framework. Although some activations are probably unnecessary there was no filtering done to prevent this. Instead, the GA identifies these as being poor performers and filter them out of the results. The batch sizes could potentially rise to half of the entire dataset but these instead were reduced to a more reasonable range. The reasonable range was determined experimentally ahead of time by noting that extremely large values would often result in extremely poor accuracy while at the same time extremely smaller batch sizes would run slowly and result in poor accuracy.

Algorithm A shows the procedure to build and run a GA. Algorithm B shows the procedure of creating and running a model.

Algorithm A

Population Initialization

```

while generations < 25 do
  for all individuals in population
    do Calculate fitness
  end for
  select N best fitness individuals to create population
  for all individuals in best do
    if random < Mutation Probability then
      mutate individual
    end if
  end for
  for all individual A, individual B in best do
    if random < Crossover Probability then
      crossover individual A, individual B
    end if
  end for
  generations++
end while
output best individual

```

Algorithm B

```

CREATEMODEL (individual)
model = Sequential
add input layer to model
for all layers in individual do
  if random < Layer Probability then

```

```

        add random layer to model
    end if
end for
add output layer to model
return model

RUNMODEL (individual)
    train, test  data.split()
    compile model from train, test
    model  CreateModel(individual)
    run model
    return model accuracy

```

6. EXPERIMENTS

The DNN and GA are implemented using Python Scikit-Learn framework on Windows. I used the Distributed Evolutionary Algorithms in Python (DEAP) which is an evolutionary computation framework for rapid prototyping and testing of ideas [27]. It incorporates the data structures and tools required to implement most common evolutionary computation techniques such as genetic algorithm, genetic programming, evolution strategies, particle swarm optimization, differential evolution, traffic flow [27, 28, 29, 30] and estimation of distribution algorithm. The benchmarks dataset MNIST and CIFAR-10 are used, which make use of categorical cross entropy as the loss function as well as the adam optimizer. For the GA, the population was limited to 25 in order to limit the computational running time.

Both datasets show a clear advantage to using the GA as the best accuracy achieved was better than on a basic CNN with no hidden layer. The best performing GA had around batch size of 100 whereas the worse performing had a batch size of 500. This indicates that the batch size is closely related to the data itself rather than a generic value and is a great candidate for hyper-parameter tuning.

7. CONCLUSION

The genetic algorithm is beneficial to be used for the purposes of hyper-parameters tuning in deep neural network learning. It can enhance the chances to identify an efficient architecture as well as optimize the performance to discover the best parameters for a given problem. Instead of researching what the current best solution is, it is wise to allow the genetic algorithm to do the work instead. Rather than a human applying an educated guess at what will perform better, the genetic algorithm will do the same thing without requiring additional effort to review the results. It is therefore a good approach to tuning hyper-parameters, particularly in a domain where there is not sufficient intuitive knowledge. However, the genetic algorithm suffers from high computational costs and in some cases is not realistic if computational resources are limited.

REFERENCES

- [1] Valueva, M., Nagornov, N., Lyakhov, P., Valuev, G., Chervyakov, N. (2020). Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. *Mathematics and Computers in Simulation*. Elsevier BV. 177: 232–243.
- [2] Zhang, W. (1988). Shift-invariant pattern recognition neural network and its optical architecture. *Proceedings of Annual Conference of the Japan Society of Applied Physics*.

- [3] Avilov, O., Rimbert, S., Popov, A. Bougrain, L. (2020). Deep Learning Techniques to Improve Intraoperative Awareness Detection from Electroencephalographic Signals. The 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Montreal, QC, Canada: IEEE. 2020: 142–145.
- [4] Tsantekidis, A., Passalis, N., Tefas, A., Kannianen, J., Gabbouj, M., Iosifidis, A. (2017). Forecasting Stock Prices from the Limit Order Book Using Convolutional Neural Networks. 2017 IEEE 19th Conference on Business Informatics (CBI). Thessaloniki, Greece: IEEE: 7–12. doi:10.1109/CBI.2017.23. ISBN 978-1-5386-3035-8. S2CID 4950757.
- [5] Matusugu, M., Katsuhiko, M., Yusuke, M., Yuji, K. (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network (PDF). *Neural Networks*. 16 (5): 555–559. doi:10.1016/S0893-6080(03)00115-1. PMID 12850007. Retrieved 17 November 2013.
- [6] Goodfellow, I., Bengio, Y., Courville, A. (2016). *Deep Learning*. MIT Press. p. 326.
- [7] Habibi, H. (2017). *Guide to convolutional neural networks: a practical application to traffic-sign detection and classification*. Heravi, Elnaz Jahani. Cham, Switzerland. ISBN 9783319575490. OCLC 987790957.
- [8] Venkatesan, R., & Li, B. (2017). *Convolutional Neural Networks in Visual Computing: A Concise Guide*. CRC Press. ISBN 978-1-351-65032-8.
- [9] Cireşan, D., Meier, U., Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. 2012 IEEE Conference on Computer Vision and Pattern Recognition. New York, NY: Institute of Electrical and Electronics Engineers (IEEE). pp. 3642–3649.
- [10] LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*. 521 (7553): 436–444. Bibcode:2015 Natur.521..436L. doi:10.1038/nature14539. PMID 26017442. S2CID 3074096.
- [11] LeCun, Y., Bengio, Y. (1995). Convolutional networks for images, speech, and time series. In Arbib, Michael A. (ed.). *The handbook of brain theory and neural networks* (Second ed.). The MIT press. pp. 276–278.
- [12] Patrick, L. Viard-Gaudin, C., Barba, D. (2006). A Convolutional Neural Network Approach for Objective Video Quality Assessment (PDF). *IEEE Transactions on Neural Networks*. 17 (5): 1316–1327. doi:10.1109/TNN.2006.879766. PMID 17001990. S2CID 221185563. Retrieved 17 November 2013.
- [13] Viebke, A., Memeti, S., Pillana, S., Abraham, A. (2019). CHAOS: a parallelization scheme for training convolutional neural networks on Intel Xeon Phi. *The Journal of Supercomputing*. 75 (1): 197–227. arXiv:1702.07908. doi:10.1007/s11227-017-1994-x. S2CID 14135321.
- [14] Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. 1: 1097–1105 – via ACM.
- [15] Haotian, J., Zhong, L., Qianxiao, L. (2021). Approximation Theory of Convolutional Architectures for Time Series Modelling. *International Conference on Machine Learning*. arXiv:2107.09355.
- [16] Passos, D., & Mishra, P. (2022). A tutorial on automatic hyperparameter tuning of deep spectral modelling for regression and classification tasks. *Chemometrics and Intelligent Laboratory Systems*, 104520.
- [17] Gonzales-Martínez, R., Machacuay, J., Rotta, P., & Chinguel, C. (2022). Hyperparameters Tuning of Faster R-Cnn Deep Learning Transfer for Persistent Object Detection in Radar Images. *IEEE Latin America Transactions*, 20(4), 677-685.
- [18] Thavasimani, K. & Srinath, N. (2022). Optimal Hyperparameter Tuning using custom genetic algorithm on deep learning to detect twitter bots. *Journal of Engineering Science and Technology*, 17(2), 1532-1549.
- [19] Gerges, F., Zouein, G., Azar, D. (2018). Genetic Algorithms with Local Optima Handling to Solve Sudoku Puzzles. *Proceedings of the 2018 International Conference on Computing and Artificial Intelligence*. ICCAI 2018. New York, NY, USA: Association for Computing Machinery: 19–22. doi:10.1145/3194452.3194463. ISBN 978-1-4503-6419-5. S2CID 44152535.
- [20] Ting, C. (2005). On the Mean Convergence Time of Multi-parent Genetic Algorithms Without Selection. *Advances in Artificial Life*: 403–412. ISBN 978-3-540-28848-0.
- [21] Deb, K. & Spears, W. (1997). C6.2: Speciation methods. *Handbook of Evolutionary Computation*. Institute of Physics Publishing. S2CID 3547258.

- [22] Patrascu, M., Stancu, A.F., Pop, F. (2014). HELGA: a heterogeneous encoding lifelike genetic algorithm for population evolution modeling and simulation. *Soft Computing*. 18 (12): 2565–2576. doi:10.1007/s00500-014-1401-y. S2CID 29821873.
- [23] Srinivas, M., Patnaik, L. (1994). Adaptive probabilities of crossover and mutation in genetic algorithms (PDF). *IEEE Transactions on System, Man and Cybernetics*. 24 (4): 656–667. doi:10.1109/21.286385.
- [24] Zhang, J., Chung, H., Lo, W. L. (2007). Clustering-Based Adaptive Crossover and Mutation Probabilities for Genetic Algorithms. *IEEE Transactions on Evolutionary Computation*. 11 (3): 326–335. doi:10.1109/TEVC.2006.880727. S2CID 2625150.
- [25] Fraser, A. & Burnell, D. (1970). *Computer Models in Genetics*. New York: McGraw-Hill. ISBN 978-0-07-021904-5.
- [26] Fogel, D. B., ed. (1998). *Evolutionary Computation: The Fossil Record*. New York: IEEE Press. ISBN 978-0-7803-3481-6.
- [27] Fortin, F., De Rainville, F., Gardner, F., Gagné, C., Parizeau, M. (2012). DEAP: Evolutionary Algorithms Made Easy. *Journal of Machine Learning Research*. 13: 2171–2175.
- [28] Gonzales-Martínez, R., Machacuay, J., Rotta, P., & Chinguel, C. (2022). Hyperparameters Tuning of Faster R-Cnn Deep Learning Transfer for Persistent Object Detection in Radar Images. *IEEE Latin America Transactions*, 20(4), 677-685.
- [29] Shankar, K., Kumar, S., Dutta, A. K., Alkhayyat, A., Jawad, A. J. A. M., Abbas, A. H., & Yousif, Y. K. (2022). An Automated Hyperparameter Tuning Recurrent Neural Network Model for Fruit Classification. *Mathematics*, 10(13), 2358.
- [30] Elhoseny, M., Metawa, N., Sztano, G., & El-Hasnony, I. M. (2022). Deep Learning-Based Model for Financial Distress Prediction. *Annals of Operations Research*, 1-23.

ANOMALY DETECTION BASED ON ALARMS DATA

Michel Kamel, Anis Hoayek and Mireille Batton-Hubert

Mathematics and industrial engineering department, Ecole des Mines de Saint-
Etienne, University Clermont Auvergne, CNRS UMR 6158 LIMOS

ABSTRACT

Alarms data is a very important source of information for network operation center (NOC) teams to aggregate and display alarming events occurring within a network element. However, on a large network, a long list of alarms is generated almost continuously. Intelligent analytical reporting of these alarms is needed to help the NOC team to eliminate noise and focus on primary events. Hence, there is a need for an anomaly detection model to learn from and use historical alarms data to achieve this. It is also important to indicate the root cause of anomalies so that immediate corrective action can be taken. In this paper, we aim to design an anomaly detection model in the context of alarms data (categorical data) in the field of telecommunication and that can be used as a first step for further root cause analysis. To do this, we introduce a new algorithm to derive four features based on historical data and aggregate them to generate a final score that is optimized through supervised labels for greater accuracy. These four features reflect the likelihood of occurrence of events, the sequence of events and the importance of relatively new events not seen in the historical data. Certain assumptions are tested on the data using the relevant statistical tests. After validating these assumptions, we measure the accuracy on labelled data, revealing that the proposed algorithm performs with a high anomaly detection accuracy

KEYWORDS

Alarms, Anomaly detection, Events data, Probabilistic scoring distribution.

1. INTRODUCTION

Anomaly detection is an aspect of data mining that has been the subject of research in many fields, such as telecommunications, information technology and finance.

There are several definitions of anomaly in the literature. Hawkins [1] defines an anomaly/outlier as an observation, which deviates considerably from the remaining observations, as if generated by a different process. Dunning and Friedman [2] state that anomaly detection involves modelling what is normal in order to discover what is not. In general, anomalies are events with a special behaviour that is dissimilar to that of normal events, and it is expected that this behaviour would be detected by analysing underlying data. Therefore, there is an urgent need for intelligent algorithms to identify such abnormal behaviour.

Anomaly detection improves data quality by deleting or replacing abnormal data. However, in certain cases, anomalies reflect an extreme event and can provide useful new knowledge. For example, the detection of such anomalies can prevent material damage and encourage predictive maintenance in the industrial field. It also has applications in several other areas such as health [3], cybersecurity [4], finance [5], natural disaster [6], and telecommunication [7].

Several methods have been proposed for detecting anomalies, each of which has its own strengths and weaknesses. Patcha and Park [8] reviewed all the known methods used for anomaly detection. Additionally, an overview of existing techniques covering several approaches is presented in [9] and [10].

Despite the large volume of literature on anomaly detection for numeric data e.g., time series, there is limited knowledge on the problem of abnormal behaviour in the context of categorical and structured textual data.

In this paper, we aim to design an anomaly detection algorithm in the context of alarms data (categorical data) in the field of telecommunication. In other words, in a given period of time, each network element of a telecommunication network generates a set of Key Performance Indicators (KPIs) and alarms that describe its behaviour. Alarms are typically categorical data with different characteristics (i.e., name, description, severity of the event, start time, end time), triggered to indicate a certain event occurring on the network element. Based on this information, those intervals of time are detected that have a high probability/score of displaying abnormal behaviour. Alarms data is important in a real-world context when KPIs are unavailable and cannot be calculated or extracted. It should be noted that alarms are events that can start popping up on a certain network element at any time. Therefore, each alarm can be considered to be equivalent to a binary random variable that can appear at any time with a certain probability.

Here, we propose an approach that introduces two new, innovative aspects. First, four features are calculated and aggregated to define events data during a certain interval of time; this includes the number of alarms, occurrence time, inter arrival time, transition frequency (Markovian model) and historical frequency. By combining this information, we compute an abnormality score which is, to the best of our knowledge, the first time that an anomaly detection algorithm has incorporated all the attributes of an event. In fact, in the majority of prior influential studies only a few of the previously cited attributes were considered. [11] consider just the Markovian component; in [12], a feature selection step is proposed prior to anomaly detection, which is a process that is associated with a high risk of loss of key information and requires significant effort for data labelling; [13] consider categorical data to be textual and vectorize it before the anomaly detection phase which is also associated with a high risk of loss of information. Second, the proposed algorithm enables users to extract local and focused information about one of the previously discussed features which may provide greater insight into the root cause of the anomaly (also known as anomaly fingerprint).

In this paper, we first describe the methodology used to build the abnormality score. We then present an application of the algorithm and analyse the results.

2. METHODOLOGY

We propose a semi-parametric scoring system that reflects the different behavioural aspects of a component of a network during a given interval of time using alarms data generated for that component. These aspects are (2.1) the number of alarms, (2.2) the inter-arrival time between alarms, (2.3) the transition probability of two consecutive alarms, and (2.4) the historical frequency of an alarm. The calculation of the final score is demonstrated in Subsection 2.5 and the optimization of the model weights is shown in Subsection 2.6. Because alarms are generated from each network component, of which there are different types, we group these components by type when drawing inferences from the data to reduce volatility and heterogeneity in the calculated statistics.

2.1. Number of Alarms

It is a common practice in parametric statistics to assume a Poisson distribution while modelling the number of occurrences of a certain event during a fixed period. Therefore, under this assumption, we begin by estimating the rate parameter λ of the Poisson distribution by calculating the arithmetic average of the number of alarms across all the intervals for each different component type of the network. Therefore, if we have L different types of components in the network, L different rate parameters $\lambda_1, \dots, \lambda_L$ are estimated.

Now, let $N_l, l = 1, \dots, L$ denote the random variables (r.v.) indicating the number of alarms generated by a component of type $l \in \{1, \dots, L\}$ over an interval of time. Based on the previous assumption, N_l follows a Poisson distribution with rate parameter λ_l . Note that $\mathbb{E}[N_l] = \lambda_l$ and it can easily be shown that the proposed estimator is a minimum variance unbiased estimator (MVUE) of λ . Hence, if n denotes the observed number of alarms in a fixed interval for a component of type l , the associated probability can be computed as shown in Equation (1).

$$\mathcal{P}_1^l = \mathbb{P}[N_l = n] = \frac{e^{-\lambda_l} \lambda_l^n}{n!} \quad (1)$$

Hence, in order to standardize this probability and transform it into a score that reflects the number of alarms, and the fact that a higher-than-average score indicates a higher probability of abnormal behaviour, S_1^l can be defined as:

$$S_1^l = \begin{cases} \frac{\mathbb{P}[N_l = \text{int}(\lambda_l)] - \mathcal{P}_1^l}{\mathbb{P}[N_l = \text{int}(\lambda_l)] - \min_{\text{over all data}} (\mathcal{P}_1^l)}, & \text{if } n \geq \text{int}(\lambda_l) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where $\text{int}(\cdot)$ denotes the integer part of a real number. Then, a value of S_1^l close to one implies that the number of alarms indicates abnormal behaviour in the specified interval. Note that $\text{int}(\lambda_l)$ represents the mode of a Poisson distribution of parameter λ_l .

2.2. Inter-Arrival Time

Using the same logic, we consider the intervals of time during which at least two alarms were detected for each type of component L . We also define the r.v. Y_l representing the time between two consecutive alarms that occurred within the same interval of time. It is common to model such r.v. by an exponential distribution with rate parameter μ_l , which is estimated by calculating the inverse of the arithmetic average of the time between two consecutive alarms during all the intervals for each different component type of the network. Note that under the previous assumption, $\mathbb{E}[Y_l] = 1/\mu_l$. Hence, if the number of alarms during an interval for a component of type $l \in \{1, \dots, L\}$ is $n \geq 2$, an associated probability can be computed as shown in Equation (3).

$$\mathcal{P}_2^l = \mathbb{P} \left[Y_l \leq \frac{\sum_{j=1}^{n-1} y_j}{n-1} \right] = 1 - e^{-(1/\mu_l) \frac{\sum_{j=1}^{n-1} y_j}{n-1}} \quad (3)$$

where $y_j, j = 1, \dots, n-1$ denotes the time between alarms j and $j+1$. Similarly, this probability can be standardized and transformed into a score to reflect that alarms occurring consecutively

within a very short span of time are more likely to indicate abnormal behaviour, as shown in Equation (4).

$$S_2^l = \frac{\max_{\text{over all data}} (\mathcal{P}_2^l) - \mathcal{P}_2^l}{\max_{\text{over all data}} (\mathcal{P}_2^l) - \min_{\text{over all data}} (\mathcal{P}_2^l)} \quad (4)$$

Here, a value of S_2^l close to one implies that the time between consecutive alarms indicates abnormal behaviour in the specified interval.

2.3. Transition Probability

In the same context as that of the inter-arrival time score, and based on all the observed alarms during all the intervals for a component of type l , we define the state space of alarms as $E^l = \{a_1, \dots, a_K\}$, where K denotes the number of unique observed alarms in component l . Subsequently, we empirically compute the transition probabilities, $\forall i, j \in \{1, \dots, K\}$, as shown in Equation (5).

$$p_{a_i a_j} = \text{probability of observing } a_j \text{ after } a_i \quad (5)$$

Hence, we obtain a transition matrix in a similar manner to a Markov chain, that summarizes all the historical transitions that have occurred for each type of component. Then, to highlight abnormal behaviour during a given interval, we identify the occurrence of transitions that are historically uncommon. Practically, if the number of alarms during an interval for a component of type l is $n \geq 2$, where these alarms are elements of the state space E^l denoted by x_1, \dots, x_n , an associated probability can be computed as shown in Equation (6).

$$S_3^l = \frac{\max_{\text{over all data}} (\mathcal{P}_3^l) - \mathcal{P}_3^l}{\max_{\text{over all data}} (\mathcal{P}_3^l) - \min_{\text{over all data}} (\mathcal{P}_3^l)} \quad (7)$$

As described previously, the probability is standardized and transformed into a score to reflect that the alarms that occur consecutively and that have not occurred one after the other frequently in the past are more likely to be displaying abnormal behaviour. This score is obtained as shown in Equation (7).

$$S_3^l = \frac{\max_{\text{over all data}} (\mathcal{P}_3^l) - \mathcal{P}_3^l}{\max_{\text{over all data}} (\mathcal{P}_3^l) - \min_{\text{over all data}} (\mathcal{P}_3^l)} \quad (7)$$

Here, a value of S_3^l close to one implies that during this interval, a non-frequent transition is occurring, which is likely to be abnormal behaviour.

2.4. Historical Frequency

Now, we consider the historical frequency of the alarms occurring during an interval. In other words, an alarm of a certain type that is historically infrequent is considered to be more critical and should be highlighted. In real world scenarios, given that access to big data can be limited, this attribute helps in identifying infrequent or non-occurring events in the network, especially

high impact events that occur rarely. Then, for a component of type l we consider the state space of alarms $E^l = \{a_1, \dots, a_K\}$, and the historical frequency of each of these alarms is computed and denoted by $f_{il} \in \{1, \dots, K\}$.

Further, to highlight abnormal behaviour during a given interval, we focus on the alarm with the lowest historical frequency among those that occurred during this interval, which are denoted by $x_1, \dots, x_n \in E^l$, with $n \geq 1$. P_4^l is first defined as shown in equation (8)

$$P_4^l = \max_{k=1, \dots, n} \frac{1}{f_k} \quad (8)$$

This is derived using all the available historical intervals data. This is followed by standardization, where P_4^l is transformed into a score quantity as shown in Equation (9).

$$S_4^l = \frac{P_4^l - \min_{\text{over all data}} (P_4^l)}{\max_{\text{over all data}} (P_4^l) - \min_{\text{over all data}} (P_4^l)} \quad (9)$$

Here, a value of S_4^l close to one implies that a non-frequent alarm occurs during this interval, which indicates abnormal behaviour.

2.5. Final score and individual contributions

To obtain a final abnormality score for a given interval of time and for a particular component of the network of type $l \in \{1, \dots, L\}$, the previously computed scores are aggregated as weighted average measures as shown in Equation (10).

$$S^l = \sum_{i=1}^4 w_i S_j^l \quad (10)$$

with $0 < w_i < 1$ and $\sum_{i=1}^4 w_i = 1$.

The values of different weights are determined based on interactions with subject matter experts (SMEs) and a supervised grid search approach that will be discussed later.

A value of S^l close to one indicates that abnormal behaviour is being displayed during the specified interval and addressing this should be considered to be a priority for the SMEs. In addition, diagnostic information can be extracted from the four individual scores which may provide a starting point for the SMEs to analyse the root cause of the detected anomalies. This additional information is used to explain the derived score by specifying which alarms occurred, which interarrivals are low, which transitions are rare and which alarms have the lowest occurrence historically.

2.6. Validation And Optimization

After assigning a score to each of the intervals across all the components of the network, we validate the results by comparing our labels to the ones given by the SMEs (labels are determined by manual inspection of the data to identify occurrences of anomalies). From the scores obtained

in Subsection 2.5, the labels are determined based on a predefined fixed threshold denoted by s , such that:

$$S_{\text{labeled}}^l = \begin{cases} 1 & \text{if } S^l > s \\ 0 & \text{Otherwise} \end{cases} \quad (11)$$

The values of the weights in Subsection 2.5 and the value of the threshold s are determined based on a grid search process [14], where several scenarios/combinations of the underlying parameters are considered. The selected combination is the one with the best performance based on the accuracy of the confusion matrix that shows the degree of similarity between our labels and the SMEs labels, and the value of the Area Under the ROC Curve (AUC) [15]. Such optimization makes the algorithm similar to supervised ML models with the aim of maximizing the correlation between labels and features. This is a unique supervision method to replicate and learn human decisions.

3. APPLICATION

The methodology described in Section 2 is applied on real world data obtained from a virtual telecommunication network to identify intervals with a high probability of displaying abnormal behaviour. Data description, results and analyses, and the advantages of the proposed algorithm will be presented in Subsections 3.1, 3.2 and 3.3 respectively.

3.1. Data Description

From a virtual telecommunication network, and for a given period, we consider alarms occurring on the different network elements over 30-minute intervals with a sliding window step of 5 minutes. The concept of the sliding window is introduced to consider events (i.e., alarms) that overlap between two consecutive intervals. In addition, to assure that the different aspects of the methodology of Section 2 are applicable, we estimate the parameters of the underlying distributions—Poisson for counting alarms and exponential for inter-arrival time—separately on the three types of components that are present in the network and are indexed as $l \in \{1,2,3\}$.

These alarms (categorical data) with their different levels of severity, e.g., critical, minor and major, occurring during a given interval indicate the occurrence of abnormal behaviour. Based on the abnormality score that is computed by the proposed algorithm, SMEs should prioritize intervention in such cases.

3.2. Results And Analysis

We first estimate the parameters of the Poisson distribution λ_l , $l \in \{1,2,3\}$ and the exponential distribution μ_l , for each type of component by applying the methods described in Subsections 2.1 and 2.2. For each type of component, all the available alarm occurrences across all the intervals are used. Note that only those intervals with at least two alarms are considered for the estimation of μ_l because the exponential distribution models the time between two consecutive alarms (in minutes). Moreover, goodness-of-fit tests are conducted to test the feasibility of the assumption that the number of alarms and the time between two consecutive alarms are governed respectively by Poisson and exponential distributions. Pearson's chi-squared test [16] is used for the goodness-of-fit test. The estimation results and the p-values of the statistical tests are represented in Tables 1 and 2 respectively.

Table 1. Parameter estimation.

Network element components	λ_l	μ_l
i=1	0.312	0.166
i=2	0.158	0.201
i=3	7.055	0.191

Table 2. Goodness of fit tests.

Network element components	p-values	
	Poisson	Exponential
i=1	0.98	0.84
i=2	0.231	0.279
i=3	0.785	0.871

Table 2 shows that when the different types of components are considered separately, the assumption about the underlying distribution appears to be reasonable. Therefore, the parametric approach defined in Subsections 2.1 and 2.2 can be relied upon to compute scores $S1_l$ and $S2_l$ for each interval for the different network elements.

The third type of score is based on a transition matrix of the probabilities of different descriptions of alarms for a given type of component. To compute such a matrix, the empirical approach described in Subsection 2.3 is applied. Table 3 shows the transition matrix of alarm descriptions for the component of type $l = 3$.

Table 3. Transition matrix.

Alarm Description	3(a)	3(b)	3(c)	3(d)	3(e)	3(f)	3(g)	3(h)
3(a)	1	0	0	0	0	0	0	0
3(b)	0	0.2	0	0	0	0	0	0.8
3(c)	0	0	1	0	0	0	0	0
3(d)	0	0	0	1	0	0	0	0
3(e)	0	0	0	0	0	0	1	0
3(f)	0	0	0	0	0	1	0	0
3(g)	0	0	0	0	0	0	1	0
3(h)	0	0.3	0	0	0	0	0	0.7

As an example of how to read Table 3, we can say that for network element of type $l = 3$, when at least two alarms occur during an interval, alarms of description 3(b) are followed by alarms of the same description in 20% of cases and alarms of description 3(h) in 80% of cases.

Using these matrices and applying the method described in Subsection 2.3, one can obtain the third score S_3^l for each interval for the different network elements.

The next step is to compute the historical frequency of each alarm description for a given type of component and to use these frequencies, as described in Subsection 2.4, to calculate the fourth score S_4^l for each interval for different network elements. An example of these frequencies is shown in Table 4 for the network element of type $l = 1$. Then, when an alarm of description 1(b) occurs during an interval and is observed to have a low historical frequency, then the interval is suspected to be displaying abnormal behaviour.

Table 4. Frequency table.

Alarm Description	Frequency
1(a)	2248
1(b)	10
1(c)	8608
1(d)	2324
1(e)	862
1(f)	17684
1(g)	29
1(h)	253
1(i)	441
1(j)	1348

Now, for each component type $l \in \{1,2,3\}$ and for all the intervals, the final abnormality score S^l can be computed by applying Equation (10) and by setting the initial values for the weights $w_i, i = 1,2,3,4$ (e.g., $w_i = \frac{1}{4} \forall i$). In addition, to apply Equation (11), a threshold s needs to be determined to label all the intervals with 1 if abnormal behaviour is taking place and 0 otherwise.

To optimize the choice of the underlying weights and threshold, the SMEs label a parallel and independent abnormal behaviour based on the same data for the same intervals. Then, based on a random grid search process, the parameters of the algorithm are optimized, for each component type, on two levels. First, among all the tested combinations of weights w_i verifying $0 < w_i < 1$ and $\sum_{i=1}^4 w_i = 1$ we select the one with the highest AUC. Second, among all the threshold used to draw the optimal ROC curve, we select the one with the highest accuracy in terms of true positives and true negatives, i.e., we maximize the sum of the diagonal terms of the confusion matrix. Hence, we begin the grid search process by considering the following equation:

$$w^* = \underset{w \in \mathcal{D}}{\operatorname{argmax}} AUC(w) \quad (12)$$

where w is a vector of weights $w_i, i = 1,2,3,4$ and \mathfrak{D} is the set of all the considered combinations of weights during the first level of the grid search process. Once the optimal combination of weights w^* is selected, we select optimal threshold by applying the following equation:

$$s^* = \underset{s \in \mathcal{T}}{\operatorname{argmax}}(TP(s) + TN(s)) \quad (13)$$

where $TP(s)$ and $TN(s)$ denote the true positive and true negative labels respectively when the threshold s is fixed. \mathcal{T} represents the set of all the considered thresholds during the second level of the grid search process. The construction of the sets \mathfrak{D} and \mathcal{T} is done with collaboration and validation by the SMEs. Furthermore, we are not concerned by the phenomenon of overfitting because we are using the latter grid search process merely to optimize the selection of the underlying weights of different scores and the threshold based on a matching method with labels fixed by the SMEs.

Considering network element of type $l = 1$, Fig. 1 shows the optimal ROC curve with a maximum AUC of 0.975 corresponding to the vector of weights $w^* = (w_1^* = 0.41, w_2^* = 0.29, w_3^* = 0.2, w_4^* = 0.1)$ for intervals with at least two alarms (i.e., S_2^l and S_3^l are computable) and a vector $w^* = (w_1^* = 0.8, w_2^* = 0, w_3^* = 0, w_4^* = 0.2)$ for intervals with less than two alarms.

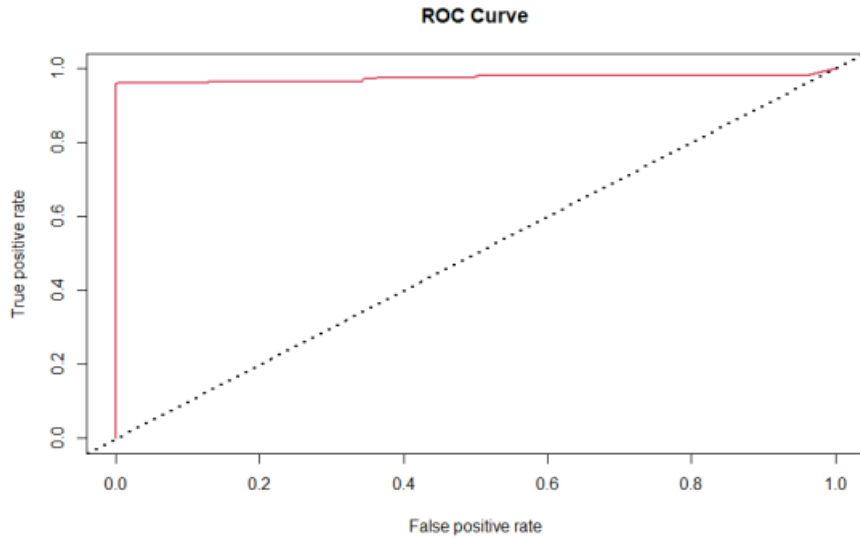


Figure 1. Optimal ROC curve

Table 5 shows the optimal confusion matrix corresponding to a threshold of 0.58. In other words, for components of type $l = 1$, we will apply the following rule to label anomalous behaviour across all the intervals:

$$S_{\text{labeled}}^l = \begin{cases} 1 & \text{if } S^l > 0.58 \\ 0 & \text{Otherwise} \end{cases}$$

In the following certain interpretations and metrics are discussed based on the confusion matrix:

- True positive: 2521 intervals; False positive: 531 intervals.
- True negative: 199513 intervals; False negative: 103 intervals.

- Sensitivity: proportion of true positive among SMEs abnormal intervals:
 $2521 / (2521 + 103) = 0.961$.
- Specificity: proportion of true negative among SMEs non abnormal intervals:
 $199513 / (199513 + 531) = 0.997$.
- Accuracy:
 $(2521 + 199513) / (2521 + 531 + 103 + 199513) = 0.997$.

Table 5. Confusion matrix.

		SMEs Labels	
		1	0
Predicted Labels	1	2521	531
	0	103	199513

Based on all the previous metrics computed after interactions with the SMEs, it is evident that the proposed algorithm is performing well with a high accuracy, and that we can rely on it to detect abnormal behaviour in future intervals. Furthermore, the algorithm is applied on online arriving alarms data and intervals that have been declared as anomalous and validated by experts. Here, approximately 1% of the intervals under control were behaving in an abnormal way, which is very reasonable in practice and is commonly encountered by SMEs. In addition, the algorithm presented in this paper has several advantages when compared to the classical anomaly detection approach. Most of these advantages will be enumerated in the next subsection.

3.3. Advantages Of The Proposed Algorithm

Compared to popular anomaly detection models, the proposed algorithm has four main advantages:

- Our algorithm is already adapted to be an online anomaly detection model applied directly to new arrivals for the purpose of highlighting abnormal behaviour. Hence, there is no need to train such a model on a sample and to test it on another because such a model has no risk of overfitting.
- To the best of our knowledge, this is the first time that an anomaly detection algorithm, based solely on alarms categorical data, has successfully extracted diagnostic information from the four different components of the global score (i.e., S_1^l, S_2^l, S_3^l and S_4^l) to help SMEs initiate root cause analysis of the detected anomalies.
- The proposed algorithm has the ability to generate abnormality scores based solely on alarms data without any additional information about numeric KPIs, which is uncommon in the field of anomaly detection for telecommunication networks.
- The interpretability of this model adds great value and is important for both developers and users.

3.4. Test Of Independence Between Different Type of Alarms

The algorithm proposed in this paper assumes the existence of one family of alarms. In fact, if other families of alarms are available, our model can easily be generalized by proposing a weighted anomaly score for the different families of alarms. Additionally, we can consider the same optimisation process proposed in Subsections 2.6 and 3.2 to determine the values of the different weights.

Further, to ensure the statistical independence between different families of alarms in terms of occurrence time we suggest an independence test. This is essentially a uniform distribution goodness-of-fit test using classical chi-squared test. Therefore, for alarms of family \mathcal{A} , we test whether the occurrence times of such alarms, between two alarms of another family \mathcal{B} , are uniformly distributed. It is important to note that in order to apply such an approach, the intervals of time separating the occurrence of two alarms of the same family need to be normalized.

4. CONCLUSION AND PERSPECTIVES

In this paper, an innovative anomaly detection algorithm that solely uses structured alarms (categorical data) has been presented. The proposed model takes into consideration four different attributes extracted from alarms occurrence data to compute a global anomaly score. This can then be used to extract diagnostic information that helps SMEs in performing root cause analysis. Our algorithm is shown to be more advantageous than other existing anomaly detection models, when applied in the same context.

Moreover, we applied the algorithm to real data in the field of telecommunication. The results were then validated by SMEs who provided positive feedback and found that our algorithm outperforms the previously used classical approaches. Users of such a model are also convinced by its output because it relies on the behaviour of historical data and generates real-time ranking of events occurring on a network component in terms of abnormality.

A first perspective of this work is to mathematically formalize a model/algorithm using the extracted information from different sub-scores in order to enhance existing root cause analysis methods based solely on alarms data. A second perspective is to combine alarms data with other type of non-numeric features, e.g., textual data, to build a more complete anomaly detection approach that covers novel aspects that have not been addressed before. Such pioneering work can be initiated by drawing inspiration from [17].

ACKNOWLEDGMENTS

The authors thank the data science team of B-Yond and the SMEs for useful discussions and for their support.

REFERENCES

- [1] D. M. Hawkins, (1980) Identification of outliers, vol. 11, Springer.
- [2] T. Dunning & E. Friedman (2014) Practical machine learning: a new look at anomaly detection, O'Reilly Media Inc.
- [3] A. Ukil, S. Bandyopadhyay, C. Puri, & A. Pal, (2016) "IoT healthcare analytics: The importance of anomaly detection", IEEE 30th international Conference on advanced information networking and applications (AINA), pp 994-997.
- [4] D. A. Bierbrauer, A. Chang, W. Kritzer, & N. D. Bastian, (2021) "Anomaly detection in cybersecurity: Unsupervised, graph-based and supervised learning methods in adversarial environments", arXiv preprint arXiv:2105.06742.

- [5] M. Sekar,(2022) “Fraud and anomaly detection”, *Machine Learning for Auditors*, pp. 193–202, Springer.
- [6] S. Miao & W.-H. Hung,(2020) “River flooding forecasting and anomaly detection based on deep learning”, *IEEE Access*, vol. 8, pp. 198384–198402.
- [7] M. Kamel, A. Hoayek & M. B. Hubert, (2022) “Probabilistic approach for anomaly detection with geometric dynamics”, unpublished.
- [8] A. Patcha&J.-M. Park,(2007) “An overview of anomaly detection techniques: Existing solutions and latest technological trends”, *Computer networks*, vol. 51, no. 12, pp. 3448–3470.
- [9] C. C. Aggarwal,(2017) “An introduction to outlier analysis”, *Outlier analysis*, pp. 1–34, Springer.
- [10] V. Chandola, A. Banerjee, &V. Kumar,(2009) “Anomaly detection: A survey”,*ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58.
- [11] H. Ren, Z. Ye, & Z. Li, (2017) “Anomaly detection based on a dynamic Markov model”, *Information Sciences*, vol. 411, pp. 52–65.
- [12] Y. Liu, H. Xu, H. Yi, Z. Lin, J. Kang, W. Xia, Q. Shi, Y. Liao, &Y. Ying, (2017) “Network anomaly detection based on dynamic hierarchical clustering of cross domain data”,*IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, pp. 200–204.
- [13] B. Nie, J. Xu, J. Alter, H. Chen, &E. Smirni, (2020) “Mining multivariate discrete event sequences for knowledge discovery and anomaly detection”, *50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pp. 552–563.
- [14] M. Claesen & B. De Moor,(2015) “Hyperparameter search in machine learning”, *arXiv preprint arXiv:1502.02127*.
- [15] T. Fawcett,(2006) “An introduction to roc analysis”, *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874.
- [16] K. Pearson,(1900) “X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling”, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 50, no. 302, pp. 157–175.
- [17] K. Ezukwoke, H. Toubakh, A. Hoayek, M. Batton-Hubert, X. Boucher, &P. Gounet, (2021) “Intelligent fault analysis decision flow in semiconductor industry 4.0 using natural language processing with deep clustering”, *IEEE 17th International Conference on Automation Science and Engineering (CASE)*, pp. 429–436.

AUTHORS

Michel KAMEL, twelve years of experience in data science and machine learning model development and currently working for B-yond handling the data science function and practices, in parallel actively researching in the field of anomaly detection within the telecommunication industry. PhD student at Ecole des Mines de SaintEtienne.

Anis HOAYEK, Associate professor in the field of probability and statistics at Ecole des Mines de Saint-Etienne. Member of mathematics and industrial engineering department.

Mireille BATTON-HUBERT, Full Professor in the field of probability and statistics at Ecole des Mines de Saint Etienne. Head of mathematics and industrial engineering department.

MACHINE LEARNING BASED TO PREDICT B-CELL EPITOPE REGION UTILIZING PROTEIN FEATURES

Fatema Nafa and Ryan Kanoff

Department of Computer Science, Salem State University, Salem, USA

ABSTRACT

Considering the current state of Covid-19 pandemic, vaccine research and production is more important than ever. Antibodies recognize epitopes, which are immunogenic regions of antigen, in a very specific manner, to trigger an immune response. It is extremely difficult to predict such locations, yet they have substantial implications for complex humoral immunogenicity pathways. This paper presents a machine learning epitope prediction model. The research creates several models to test the accuracy of B-cell epitope prediction based solely on protein features. The goal is to establish a quantitative comparison of the accuracy of three machine learning models, XGBoost, CatBoost, and LightGbM. Our results found similar accuracy between the XGBoost and LightGbM models with the CatBoost model having the highest accuracy of 82%. Though this accuracy is not high enough to be considered reliable it does warrant further research on the subject.

KEYWORDS

machine learning models; data exploratory techniques; B-cell epitope prediction.

1. INTRODUCTION

Given the current rate of globalization, we can expect with near certainty that we will see events similar to the Covid-19 pandemic in the coming years. This makes vaccine research and development, especially the speed and efficiency of these processes, more relevant than ever. An essential aspect of vaccine research and development is B-cell epitope prediction. B-cell epitope regions are areas of antigen proteins that when recognized by B-cells produce large amounts of antigen-specific antibodies. Historically, B-cell epitope predication has relied primarily on sequence data and not protein features. By utilizing these protein features we can potentially increase the speed of B-cell epitope prediction which may help expediate the research and development of new vaccines.

For the purpose of this paper, we used several different machine learning models to test the relationship between protein features and B-cell epitope regions in an attempt to find this highest accuracy possible. The input to our algorithm is a dataset consisting of protein/amino acid sequences and their associated protein features and target behavior (anti-body inducing behavior). Three different machine learning models applied to this dataset. All of these models utilize boosting algorithms, specifically gradient boosting. These algorithms function by first creating weak learners and converting them to strong learners [1]. XGBoost or extreme gradient boost, is a gradient boosting decision tree (GBDT) model that focuses on speed and performance [2]. In recent years it has gained significant popularity through its ability to yield highly accurate results from large datasets. LightBGM is another gradient boosting tree-based model. Alongside

David C. Wyld et al. (Eds): AI, AIMLNET, BIOS, BINLP, CSTY, MaVaS, SIGI - 2022

pp. 115-122, 2022. CS & IT - CSCP 2022

DOI: 10.5121/csit.2022.121811

the expected speed and performance of a gradient boosting model, LightBGM offers reduced memory usage which can help facilitate large-scale data processing. The final model, CatBoost, is another GBDT model. It behaves much like similar models; however, it specializes in data processing speed, with the potential to be notably faster than the previously mentioned models [3]. Each of these models will output an accuracy for the prediction of the target behaviour based on the features.

2. RELATED WORK

Machine learning algorithms have become one of the foremost methodologies within the field of bioinformatics and more specifically, immunoinformatic, for the purpose of predicting B-cell epitope regions. The following section will detail a number of papers regarding B-cell epitope prediction to provide a basic survey of the topic. Three of the most methodologies used include Naïve Bayes Classification, Support Vector Machine and Artificial Neural Networks [2]. Many papers that explore B-cell epitope region prediction will include at least one of these as part of their methodology. J Chen et al used the amino acid pair (AAP) antigenicity scale to predict B-cell epitopes. They saw significantly improved performance with the AAP antigenicity scale by utilizing support vector machine rather than existing scales [4].

Tao Lui et al utilized deep learning, a type of neural network with many hidden nodes and hidden layers, to create a predictive model for linear B-cell epitopes. They obtained sample peptides from the IEDB database and used this data to build a feed forward deep neural network. This ensemble prediction model was named DLBEpitope and performed better than current major models [5].

A great deal of success has also been made less conventional models such as BERT-based epitope prediction.

While BERT models are typically used for natural language processing, Minjun Park et al were able to create a Bert-based model, EpiBERTope. They pre-trained it with the Swiss-Prot protein database so that it could predict linear and structural epitope while exclusively relying on protein sequences. This model outperformed all the classic benchmark models include random forest, gradient boosting, naïve bayes and support vector machine [6].

3. PROBLEM DEFINATION

The main motivation behind this work is, to highlight the work done in B-cell epitope prediction using machine learning models, the limitation and the strength of each work presented in this work. Also, introducing machine learning methods to predict the B-cell epitope, in additionally, to provide future directions in terms of the limitation of the proposed method.

4. METHODOLOGY

This section covers pre-processing steps, and the methodology used in this research.

4.1. Data Pre-Processing

In this work Python [7], [8] is used for performing the prediction. Python is a general-purpose programming language with most data analysis features provided by NumPy and pandas. It has efficient high-level data structures and an object-oriented programming technique that is simple but effective.

This dataset was obtained from IEDB and UniProt available for public [9]. This dataset consists of three csv files, input_bcell, input Sars and input covid. Input_bcell is by far the largest file with dimensions 14387 x 14. All three of these files have the same columns excluding input Sars which does not have the target column. The first few columns are identifying pieces for the protein sequences and are not utilized by this paper. These include the parent protein id, protein sequence, peptide start position, peptide end position and peptide sequence. All of these columns are dropped in pre-processing because they are not needed when exclusively examining the protein features.

4.2. ML Models

This work implements Boosting algorithms [2], [10]. Gradient boosting is an algorithm that stands out for its predictability and speed, especially when dealing with big and complicated datasets[11], [12]. Gradient Boosting algorithm has three main components, loss function, weak learner, and additive model. The main goals of Gradient Boosting algorithms are to improve the prediction power by converting a number of weak learners to strong learners. Boosting algorithms work on the idea of first building a model on the training dataset, then building a second model to correct the errors in the first model. This process is repeated until the errors are minimized and the dataset is accurately predicted. Boosting algorithms are divided into three main categories, AdaBoost algorithm, Gradient algorithm, and Extreme Gradient Boosting, or XGBoost [2]. Gradient boosting algorithms can be Regressors (for predicting continuous target variables) or Classifiers (predicting categorical target variables). In this paper, XGBoost is used for predicting categorical target variables.

XGBoost stands for eXtreme Gradient Boosting. XGBoost is a powerful machine learning algorithm. it is a supervised learning algorithm used optimized loss function and applied several regularization techniques[2]. XGBoost is a more regularized form of Gradient Boosting. XGBoost uses advanced regularization which improves model generalization capabilities. XGBoost can be used for a variety of applications, including regression and classification problems [f]. XGBoost is a faster algorithm because of its parallel and distributed computing. It has a deep consideration in terms of systems optimization in machine learning [1]. The adoption of XGBoost is mostly due to its execution speed and model performance. It uses ensemble learning methods, which means it combines several different algorithms to produce a single model. This method allows for parallel and distributed processing while maximizing memory utilization. The steps of XGBoost shown in Algorithm 1.

Algorithm 1: XGboost (XB) Algorithm

Input: The n-dimensional data, $X \in \mathbb{R}^n$ and target outcome, $Y \in \mathcal{R}$
Output: The posterior probability, $P \in [0, 1]$ of unseen test data, x ,
 where $C = 2$, C_1 (Represent Covid-19) and C_2 (No Covid-19)

Data: Dataset

- 1 Initialize the model with constant value
 $F_0(X) = \text{argmin}_{\gamma} \sum_{i=1}^N L(Y_i, \gamma)$ where, $\sum_{i=1}^N L(Y_i, \gamma)$ is the differentiable lossfunction and N is the number of sample.
- 2 for $t \leq T$ (nclassifier) do
- 3 a) Train a weak learner using distribution D_t .
- 4 b) Select a weak hypothesis, $h_t : \mathbb{R}^n \rightarrow \mathbb{R}$ with low weight error,
 $\xi_t = \text{Pr}_{i \sim D_t}[h_t(x_i) \neq Y]$
- 5 Choose $\alpha_t = \frac{1}{2} \ln \frac{1 - \xi_t}{\xi_t}$
- 6 Update the Model
- 7

LightGBM is a tree-based learning optimization technique. This approach, as an optimization technique, minimizes both information loss and memory space usage. Furthermore, unlike traditional learning tree algorithms that develop at the level of the node with the largest share of information loss, Light GBM accelerates the learning process by

CatBoost, like the Light GBM method, is part of the GBDT model, which seeks to handle categorical features [3], [13], [14]. This approach is frequently used in search, system suggestions, personal help, self-driving automobiles, and weather forecasting. This optimization approach is notable for its data processing speed, which may be up to 60 times quicker than Light GBM in some situations [1].

The main motivation of choosing these models is as following:

- ✦ Faster training speed and higher efficiency.
- ✦ Lower memory usage.
- ✦ Better accuracy.
- ✦ Support of parallel and GPU learning.
- ✦ Capable of handling large-scale data.

implementations of the Gradient Boosted Trees algorithm, a supervised learning method that is based on function approximation by optimizing specific loss functions as well as applying several regularization techniques. Gradient boosting is a machine learning technique that can be used for a variety of applications, including regression and classification. It returns a prediction model in the form of an ensemble of weak prediction models, most commonly decision trees. The resulting approach is called gradient-boosted trees when a decision tree is the weak learner; it usually outperforms random forest. A gradient-boosted trees model is constructed in the same stage-wise manner as other boosting approaches, but it differs in that it allows optimization of any differentiable loss function [15], [16].

5. RESULTS

In this section, we present experimental results conducted on the dataset.

5.1. Dataset Description

The experimental data in this paper are derived from dataset developed during a research process obtained from IEDB and UniProt [9]. The dataset consists of 15 features, and the result in the data were taken from 14907 people. General statistical information about the dataset shown in Table 1.

Table 1. Statistical Information about the Dataset.

Dataset statistics		Variable types	
Number of variables	15	NUM	11
Number of observations	14907	CAT	3
Missing cells	0	BOOL	1
Missing cells (%)	0.0%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	1.7 MiB		
Average record size in memory	120.0 B		

The first few features are identifying pieces for the protein sequences and are not utilized by this paper. These include the parent protein id, protein sequence, peptide start position, peptide end position and peptide sequence. All of these features are dropped in pre-processing because they

are not needed when exclusively examining the protein features. The features and its abbreviation are shown in Table 2.

Table 2. Dataset and Feature Abbreviation.

Features	Abbreviation of Features
parent_protein_id	Parent Protein
protein_seq	parent protein sequence
start_position	start position of peptide
end_position	end position of peptide
peptide_seq	peptide sequence
chou_fasman	peptide feature, β turn
emini	peptide feature, relative surface accessibility
kolaskar_tongaonkar	peptide feature, antigenicity
parker	peptide feature, hydrophobicity
isoelectric_point	protein feature

5.2. Data Correlation

After data pre-processing, we have to learn whether or not the data is processed properly and how much the correlation between the data is. The correlation coefficient can be used to reflect the close relationship between the features. The correlation coefficient is calculated by the difference method. It is also based on the dispersion of the two features and their respective averages. The two differences are multiplied to reflect the degree of correlation between the two features. The linear single correlation coefficient is studied as shown in Fig.1.

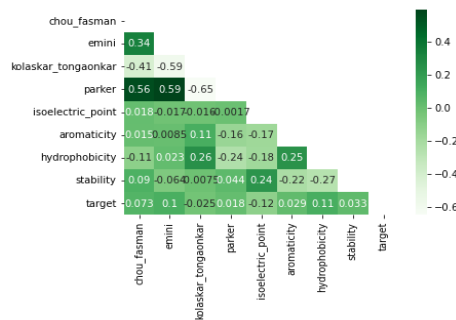


Figure 1. Correlation coefficient matrix

Diving more into the features, we have to learn whether or not the data is processed properly and how much the correlation between the data is. Finding the distribution of each feature is important. It is clear from Fig.2, some variables having skewed distribution. Most of machine learning models assume that the data is normally distributed those variables may need scaling.

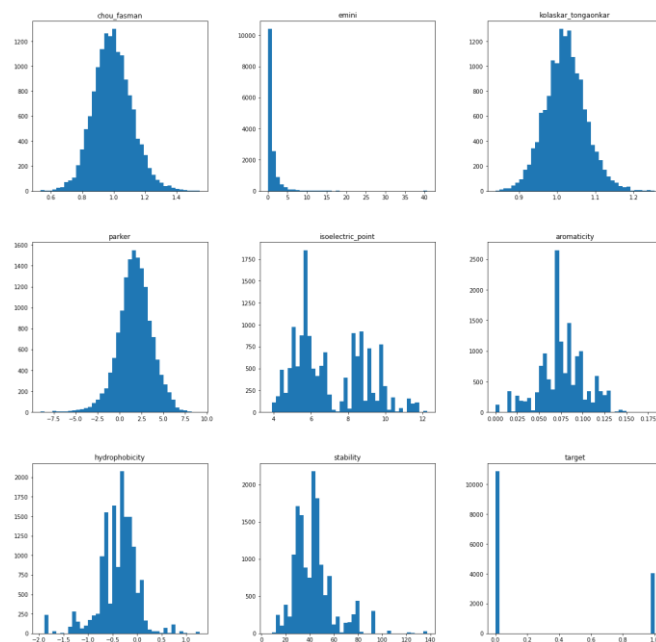


Figure 2. The population distribution of all attributes.

From correlation coefficient matrix, we can figure out whether there is a linear correlation between target and various parameters, such as the emini is 0.34.

5.3. Performance Metrics/Confusion Matrix

XGBoost, **LightGBM** and **CatBoost** Tree algorithms are used in this research work. Experiments are performed using internal cross-validation 10-folds [17], [18]. Accuracy, F-Measure, Recall, Precision and ROC (Receiver Operating Curve) measures are used for the prediction of this work. Table.3.3 defines accuracy measures below.

Table 3. List of Accuracy Measures

Measures	Definitions	Formula
1. Accuracy (A)	Accuracy determines the accuracy of the algorithm in predicting instances.	$A = (TP + TN) / (\text{Total no of samples})$
2. Precision (P)	Accuracy is measured by Precision.	$P = TP / (TP + FP)$
3. Recall (R)	To measure the classifier's completeness or sensitivity, Recall is used.	$R = TP / (TP + FN)$
4. F-Measure	F-Measure is the weighted average of precision and recall.	$F = 2 * (P * R) / (P + R)$
5. ROC	ROC (Receiver Operating Curve) curves are used to compare the usefulness of tests.	

5.4. Result Analysis

By the iterated training and adjustment parameters of the training set, the comparison results are showed between XGBoost, CatBoost, and LightGbm in Table.4.

Table 4. Performance comparison for dataset.

Prediction Models	Precision	Recall	Accuracy%
XGBoost	0.708	0.532	81.83
CatBoost	0.737	0.514	82.01
LightGBM	0.720	0.531	81.83

While carrying out B-cell epitope prediction, the model can also give the order of importance feature for improving the accuracy of the prediction model through the tree model mechanism. These important features can also be used as some clinical reference value for doctors.

Finally, the results of an attribute in all the models are weighted and summed, and then averaged to obtain the importance score.

As shown in Fig.3.[19]

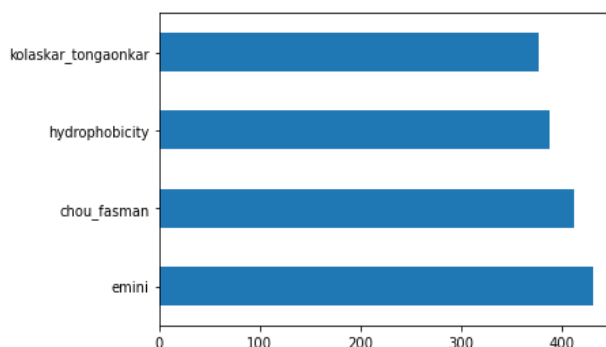


Figure 3. Feature importance of the Machine Learning Algorithm

As can be seen from Fig.3. we can clearly see that feature (peptide) is the most important feature which has contributed towards the prediction of the results. Followed by feature (kolaskar_tongaonkar) and feature (emini). The least important feature is feature (antigenicity). The model can also give the order of importance feature for improving the accuracy of the prediction model through the tree model mechanism. These important features can also be used as some clinical reference value for doctors.

6. CONCLUSION

Data preparation is a necessary step in ensuring model accuracy and speeding up the process. When the models are compared, it is evident that CatBoost outperforms several classic approaches. We suggested a better feature combination approach based on CatBoost after comparing it to the integrated algorithms. The experiment findings suggest that using the CatBoost model, it is possible to develop a B-cell epitope prediction model with high prediction accuracy, stability, and speed. However, there were no missing data in the datasets used in this experiment. As a result, the same conclusions cannot be predicted from a dataset with missing values and noisy data. As a result, we will test the accuracy and speed of the model's using datasets that include a variety of data types as well as missing data in the future. In addition, as we are entering the BigData era, we would attempt running the models in-memory distributed computing and design UGI for it to make it easy to use.

REFERENCES

- [1] R. B. Sundaram, "Gradient Boosting Algorithm: A Complete Guide for Beginners," analyticsvidhya, 2021.
- [2] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785–794.
- [3] G. Huang et al., "Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions," *J. Hydrol.*, vol. 574, pp. 1029–1041, 2019.
- [4] J. Chen, H. Liu, J. Yang, and K.-C. Chou, "Prediction of linear B-cell epitopes using amino acid pair antigenicity scale," *Amino Acids*, vol. 33, no. 3, pp. 423–428, 2007.
- [5] T. Liu, K. Shi, and W. Li, "Deep learning methods improve linear B-cell epitope prediction," *BioData Min.*, vol. 13, no. 1, pp. 1–13, 2020.
- [6] M. Park, S. Seo, E. Park, and J. Kim, "EpiBERTope: a sequence-based pre-trained BERT model improves linear and structural epitope prediction by learning long-distance protein interactions effectively," *bioRxiv*, 2022.
- [7] J. Faouzi and H. Janati, "pyts: A Python Package for Time Series Classification.," *J Mach Learn Res*, vol. 21, pp. 46–1, 2020.
- [8] J. Hao and T. K. Ho, "Machine learning made easy: a review of scikit-learn package in python programming language," *J. Educ. Behav. Stat.*, vol. 44, no. 3, pp. 348–361, 2019.
- [9] R. Vita et al., "The immune epitope database (IEDB): 2018 update," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D339–D343, 2019.
- [10] R. E. Schapire, "The boosting approach to machine learning: An overview," *Nonlinear Estim. Classif.*, pp. 149–171, 2003.
- [11] K. M. Ting, "A comparative study of cost-sensitive boosting algorithms," 2000.
- [12] A. J. Ferreira and M. A. Figueiredo, "Boosting algorithms: A review of methods, theory, and applications," *Ensemble Mach. Learn.*, pp. 35–85, 2012.
- [13] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: gradient boosting with categorical features support," *ArXiv Prepr. ArXiv181011363*, 2018.
- [14] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," *J. Big Data*, vol. 7, no. 1, pp. 1–45, 2020.
- [15] S. Neelakandan and D. Paulraj, "A gradient boosted decision tree-based sentiment classification of twitter data," *Int. J. Wavelets Multiresolution Inf. Process.*, vol. 18, no. 04, p. 2050027, 2020.
- [16] T. Pinto, I. Praça, Z. Vale, and J. Silva, "Ensemble learning for electricity consumption forecasting in office buildings," *Neurocomputing*, vol. 423, pp. 747–755, 2021.
- [17] S. De Bruyne and F. Plastria, "2-class Internal Cross-validation Pruned Eigen Transformation Classification Trees," *Optim. Online Httpwww Optim. OrgDB HTML2008051971 Html*.
- [18] H. Van Hasselt, "Estimating the maximum expected value: an analysis of (nested) cross validation and the maximum sample average," *ArXiv Prepr. ArXiv13027175*, 2013.
- [19] F. Nafa, A. Babour, and A. Melton, "Prerequisite Relations among Knowledge Units: A Case Study of Computer Science Domain," *Comput. Model. Eng. Sci.*, doi: 10.32604/cmcs.2022.020084.

TENSOR-BASED MULTI-MODALITY FEATURE SELECTION AND REGRESSION FOR ALZHEIMER'S DISEASE DIAGNOSIS

Jun Yu¹, Zhaoming Kong¹, Liang Zhan², Li Shen³ and Lifang He¹

¹Department of Computer Science and Engineering,
Lehigh University, Bethlehem, Pennsylvania, USA

²Department of Electrical and Computer Engineering,
University of Pittsburgh, Pittsburgh, Pennsylvania, USA

³Department of Biostatistics, Epidemiology and Informatics,
The Perelman School of Medicine, University of Pennsylvania,
Philadelphia, Pennsylvania, USA

ABSTRACT

The assessment of Alzheimer's Disease (AD) and Mild Cognitive Impairment (MCI) associated with brain changes remains a challenging task. Recent studies have demonstrated that combination of multi-modality imaging techniques can better reflect pathological characteristics and contribute to more accurate diagnosis of AD and MCI. In this paper, we propose a novel tensor-based multi-modality feature selection and regression method for diagnosis and biomarker identification of AD and MCI from normal controls. Specifically, we leverage the tensor structure to exploit high-level correlation information inherent in the multi-modality data, and investigate tensor-level sparsity in the multilinear regression model. We present the practical advantages of our method for the analysis of ADNI data using three imaging modalities (VBM-MRI, FDG-PET and AV45-PET) with clinical parameters of disease severity and cognitive scores. The experimental results demonstrate the superior performance of our proposed method against the state-of-the-art for the disease diagnosis and the identification of disease-specific regions and modality-related differences. The code for this work is publicly available at <https://github.com/junfish/BIOS22>.

KEYWORDS

Alzheimer's disease, multi-modality imaging, brain network, tensor, feature selection, regression.

1. INTRODUCTION

Alzheimer's Disease (AD) is one of the most common and incurable neurodegenerative diseases [1], which can result in progressive cognitive decline and behavioral impairment, and even cause death in severe cases. A recent report shows that 26.6 million AD patients exist in the world and 1 out of 85 people will be living with AD by 2050 [2]. Thus, for timely therapy that might be effective to slow the disease progression, it is important for early diagnosis of AD and its prodromal stages like Mild Cognitive Impairment (MCI). In particular, MCI is attractive because it represents a transitional stage between normal aging and dementia [3]. Evidence shows that patients with MCI have a high risk to convert into AD, about 12% per year, while healthy controls convert at a 1–2% rate [4].

Recent studies have showed great promise for the use of multi-modality imaging in predicting the progression of AD markers and clinical diagnosis [5–7], such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET). The results indicate that multi-modality imaging data contain complementary information of clinical importance and thus have the potential to improve the performance of AD diagnosis tasks and provide further insight into the pathophysiology of AD [8]. However, multi-modality imaging data contain noisy and heterogeneous information, making it challenging to effectively incorporate them into quantitative models. It is also challenging to analyze all voxels in the whole brain, as the number of whole-brain voxels usually far exceeds the number of observations available in practice, leading to overfitting.

Recently, various machine learning methods have been proposed for the analysis of multi-modality images in AD studies [6, 9–13]. Feature selection, which searches for a subset of significant features from the original set of features, is critical for effective machine learning, as spurious features can harm the learning process, especially in the presence of multi-modality and high dimensionality data. Furthermore, unlike feature extraction techniques such as principal component analysis (PCA) [14], which project raw data onto a new low-dimensional space, feature selection tends to preserve the interpretability of raw data, as the selected features have clear connections to the original ones. This can help experts to understand which features are relevant to certain disease states.

Feature selection methods are either classifier dependent or classifier independent. The classifier dependent method can potentially provide better performance as it directly makes use of the interaction between features and accuracy [15]. In this respect, several approaches at different levels of complexity have been developed and applied for multi-modality AD diagnosis and risk factor analysis (e.g., [6, 7, 10]). However, existing methods mainly adopt vector space models for multi-modality fusion, which may lead to the loss of valuable information due to vector quantization, and thus degrade the prediction performance.

In this paper, we propose a novel tensor-based method to model the inherent relationship between the features of multi-modality data and perform joint feature selection and tensor regression for AD diagnosis, in which tensor-structured sparsity and low-rankness properties are jointly exploited to learn the interpretable coefficients. We test the feasibility of this method on a large neuroimaging data set from the ADNI cohort using disease severity score and cognitive scores of MMSE and ADAS-13 across three different imaging modalities, including VBM-MRI, FDG-PET, and AV45-PET. Extensive experimental results demonstrate that the proposed method has better prediction performance compared to the vector space models, as well as good interpretability for diagnostic results from selected discriminative regions and modality-related differences.

2. MATERIALS AND METHODS

2.1. Data Acquisition and Preprocessing

In this work, a total of 692 non-Hispanic Caucasian participants in the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database [16] that passed quality control were used for our analysis, including 163 cognitively normal controls (CN), 73 normal controls with significant memory concern (SMC), 214 patients with early MCI (EMCI), 149 patients with late MCI (LMCI), and 93 AD patients. Each subject has three modalities of imaging data, including structural Magnetic Resonance Imaging (VBM-MRI), 18 F-fluorodeoxyglucose Positron Emission Tomography (FDG-PET) and 18 F-florbetapir PET (AV45-PET).

The multi-modality imaging data were aligned to each participant's same visit. The structural MRI scans were preprocessed with voxel-based morphometry (VBM) using the SPM software [17]. Generally, all scans were aligned to a T1-weighted template image, segmented into gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) maps, normalized to the standard Montreal Neurological Institute (MNI) space as $2 \times 2 \times 2$ mm³ voxels, and were smoothed with an 8 mm FWHM kernel. The FDG-PET and AV45-PET scans were also registered to the same MNI space by SPM. The MarsBaR toolbox¹ was used to group voxels into 116 ROIs defined by Automated Anatomical Labeling (AAL) [18]. ROI-level measures were calculated by averaging all the voxel-level measures within each ROI.

To assess the level of AD's development, we consider three clinical scores as our prediction responses: (1) Disease Severity Score (DSS), in which we treat different stages of AD progression as an index of disease severity (1-CN, 2-SMC, 3-EMCI, 4-LMCI, and 5-AD), (2) AD Assessment Scale–Cognitive 13-item (ADAS-13) [19], with a total score of 85, improving the responsiveness of classic ADAS-Cog [20] for MCI by considering more candidate tasks related to additional cognitive domains, and (3) Mini-Mental State Examination (MMSE) score [21], a widely used screening test of cognitive function among the elderly, with a maximum score of 30 points. We normalize these three prediction scores to the range of [0, 1] so that they have the same scale, and the higher scores indicate the greater severity of the cognitive impairments.

2.2. Tensor Construction

Tensors are higher-order generalizations of matrices to multiple indices that can be used to represent multi-dimensional and multi-relational data. To model the inherent relationship and connectivity between the ROIs from multi-modality data for AD/MCI assessment, we investigate three sizes of tensor representations as follows.

- **116 × 3.** We concatenate 116 feature values from all three imaging modalities along an additional modality dimension in the order of VBM, FDG and AV45. Each modality contains 116 ROI-based features, thus resulting in a 2D tensor of size 116×3 for each subject.
- **116 × 116.** We construct an ROI-to-ROI connectivity matrix based on the pairwise similarity of ROIs, thus obtaining a 2D tensor of size 116×116 for each subject. For each ROI, we concatenate features from three modalities into a vector of 3 dimensions, denoted as $r_i, i = 1, 2, \dots, 116$. Then we use the k -Nearest Neighbor (k NN) graph [22] to construct the connectivity matrix via the Gaussian similarity function, i.e., $X_{i,j} = \exp(-\|r_i - r_j\|^2 / \sigma^2)$, where σ is a user defined parameter specifying width. For simplicity, we set $\sigma = 1$ and consider $k = 1, 2, \dots, 116$ in our paper. In particular, when $k = 116$, it is a fully connected matrix, as each ROI is connected to other ROIs.
- **116 × 116 × 3.** We construct an ROI-to-ROI connectivity matrix in each modality using the k NN graph as above, and concatenate them along an additional modality dimension to generate a 3D tensor of size $116 \times 116 \times 3$ for each subject.

2.3. Joint Feature Selection and Tensor Regression

Figure 1 provides an overview of our proposed method, which uses tensor data as input features to perform feature selection and regression simultaneously.

¹<https://imaging.mrc-cbu.cam.ac.uk/imaging/MarsBaR>

Objective Function. Assume that we have a dataset containing N subjects, each subject is represented by an M th-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_M}$ and is associated with a regression label y indicating disease status. Similar to linear regression, the tensor regression model can be formulated as follows:

$$y = \langle \mathcal{W}, \mathcal{X} \rangle + \varepsilon, \quad (1)$$

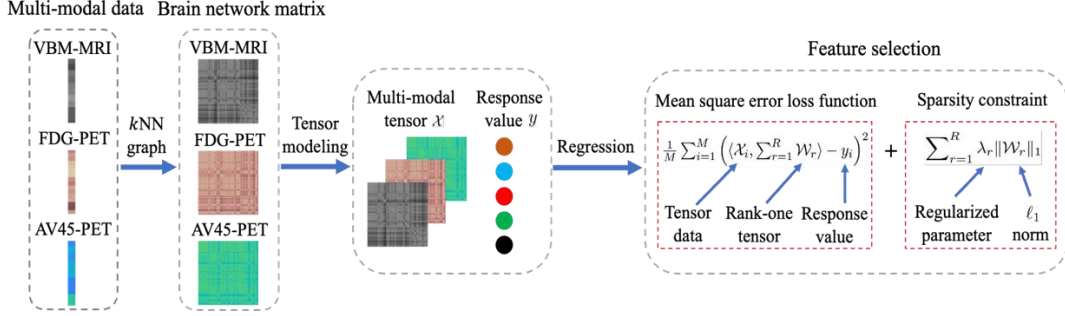


Figure 1. The framework of the proposed method. The input data contains three modalities: VBM-MRI, FDG-PET, and AV45-PET, then k NN graph is applied to construct tensor data representation. Finally, we regress the response values by using the proposed method.

where \mathcal{W} is the coefficient tensor, ε is the bias error, and $\langle \cdot, \cdot \rangle$ denotes the inner product operator. To exploit the high-dimensional structure and correlation in the tensor representation, we employ the following sparse and low-rank tensor regression model to solve Eq. (1) inspired by [23].

$$\min_{\mathcal{W}_r} \frac{1}{N} \sum_{i=1}^N \left(\langle \sum_{r=1}^R \mathcal{W}_r, \mathcal{X}_i \rangle - y_i \right)^2 + \sum_{r=1}^R \lambda_r \|\mathcal{W}_r\|_1, \quad \text{s.t. CP rank}(\mathcal{W}_r) \leq 1. \quad (2)$$

Where $\mathcal{W}_r = \mathbf{w}_r^{(1)} \otimes \dots \otimes \mathbf{w}_r^{(N)}$ is a unit-rank tensor defined upon the CP rank [24], \otimes denotes the outer product operator, and λ_r is the regularized parameter that contributes to the sparsity.

Due to the use of a ℓ_1 -norm regularizer in unit-rank tensors, after finding the optimal solution in Eq. (2), we have many zero elements in $\sum_{r=1}^R \mathcal{W}_r$, whose corresponding features are not useful in prediction of disease status. Furthermore, by simple arithmetic, we have $\|\mathcal{W}_r\|_1 = \prod_{j=1}^N \|\mathbf{w}_r^{(j)}\|_1$, i.e., the sparsity of a unit-rank tensor directly leads to the sparsity of its components. This allows us to produce a set of sparse factor components $\mathbf{w}_r^{(j)}$ for $j = 1, \dots, N$ simultaneously. Using this we can examine how each ROI behaves and how each modality contributes to the prediction.

Optimization. A common way to solve Eq. (2) is to utilize the alternating least squares (ALS) [25]. However, in this way it is difficult to estimate R and λ_r . Here we adopt a divide-and-conquer strategy to sequentially solve the following sparse unit-rank estimation problems instead based on the fast stagewise unit-rank tensor factorization (SURF) algorithm [23], which can automatically estimate λ_r and is easy to tune the parameter R .

$$\min_{\mathcal{W}_r} \frac{1}{N} \sum_{i=1}^N \left(\langle \mathcal{W}_r, \mathcal{X}_i \rangle - y_i \right)^2 + \lambda_r \|\mathcal{W}_r\|_1, \quad \text{s.t. CP rank}(\mathcal{W}_r) \leq 1 \quad (3)$$

where r is the sequential number of the unit-rank terms and y_i^r is the current residue of response with

$$y_i^r := \begin{cases} y_i & \text{if } r = 1 \\ y_i^{r-1} - \langle \mathcal{W}_{r-1}, \mathcal{X}_i \rangle, & \text{otherwise,} \end{cases} \quad (4)$$

Where \mathcal{W}_{r-1} is the estimated unit-rank tensor in the $(r - 1)$ -th step. The final estimator can be obtained as $\mathcal{W} = \sum_{r=1}^R \mathcal{W}_r$.

3. RESULTS

3.1. Experimental Settings

Competing Methods. To show the validity of the proposed method, we compared our method with the PCA feature extraction method followed by linear regression (PCA+LR), and three representative feature selection methods—Lasso [25], Elastic Net (ENet) [26], and Group Lasso (gLasso) [27].

Evaluation Metrics. For the quantitative performance evaluation, we employed two metrics: (1) root mean squared prediction error (RMSE), measuring the deviation between the ground truth response and the predicted values, and (2) sparsity of coefficients, defined as the percentage (%) of the number of zero elements to the total number of elements in the coefficients.

Implementation Details. For model validation, subjects were randomly split into training and test sets in the ratio of 5:1. The hyperparameters of all methods were optimized using 5-fold cross validation on the training set. Specifically, we traversed all possible percentage of the principal components to realize the best results in PCA+LR. The regularizer parameter of Lasso was selected from $\{0.1, 0.2, \dots, 1\}$. For the ENet, the weight of lasso (ℓ_1) versus ridge (ℓ_2) optimization was learned from $\{\{0.1, 0.2, \dots, 1\}\}$. For the gLasso, the features are grouped by modalities or ROIs, and the parameter-wise and group-wise regularisation penalty are selected via a grid search ranging from $\{10^{-6}, 10^{-5}, \dots, 10^1\}$, then applying a more fine-grained searching (i.e., $\{0.1, 0.2, \dots, 1\}$) after the magnitude is determined. In our proposed method, the rank R is incrementally learned from 1 to 70 with step size 1. To avoid randomness from affecting the experiments, five trails were conducted to record mean value and standard deviation of final results for each method.

3.2. Experimental Results

Table 1 shows the performance comparison of five methods. It is clear that the proposed method consistently outperforms the competing methods in terms of both RMSE and Sparsity. Specifically, we observe the following results.

- The prediction performance of our method improves consistently when the feature sizes or dimensions of input data grow, which proves the connectivity information among ROIs in the brain is helpful to AD assessment. The gLasso outperforms the other three baselines and performs better with the higher-order tensor data, because the group variables can consider the structure information to some extent. In contrast, the Lasso and ENet struggle to handle higher dimensional data since vectorization may lead to a certain loss of structural information. These results indicate not only the importance of considering multi-modality information and constructing connectivity among ROIs in the brain data, but also the effectiveness of maintaining the dimensionality in tensor space during the learning process.

- PCA+LR is more consistent to achieve better results with higher dimensionality of input data than lasso-based baselines, because the k NN graph-based data construction introduces the redundant and noisy information with effective connectivity information and PCA is effective in filtering out unwanted information. However, PCA is a feature extraction method that destroys the structure of original feature space and turns it into principal components, making it difficult to identify significant ROIs of clinical importance. It further validates that our method is robust via seeking a low-rank and sparse tensor space that is helpful to feature selection and noise removal [23].
- Due to the usage of joint sparsity constraint in our objective function, the proposed method can produce the best predictive values and achieve the sparsest solution simultaneously, which increases the interpretability of our model for diagnostic results. More detailed medical significance will be shown in the following qualitative analysis section.

Table 1. Performance comparison over different tensor structures on ADNI dataset. The results of mean values and standard deviation (mean \pm std) are calculated across 5 trails. \downarrow means the lower the better, and \uparrow means the higher the better.

Features	Scores	Metrics	PCA+LR	Lasso	ENet	gLasso	Proposed
116×3	DSS	RMSE \downarrow	0.39 ± 0.02	0.37 ± 0.01	0.36 ± 0.02	0.33 ± 0.01	0.31 ± 0.01
		Sparsity \uparrow	0.00 ± 0.00	85.12 ± 4.14	77.41 ± 7.54	88.39 ± 0.56	97.16 ± 0.43
	ADAS-13	RMSE \downarrow	0.20 ± 0.02	0.23 ± 0.01	0.23 ± 0.01	0.18 ± 0.01	0.17 ± 0.00
		Sparsity \uparrow	0.00 ± 0.00	85.17 ± 2.76	82.82 ± 2.52	91.21 ± 0.88	98.56 ± 0.41
	MMSE	RMSE \downarrow	0.27 ± 0.01	0.26 ± 0.20	0.26 ± 0.02	0.24 ± 0.02	0.22 ± 0.02
		Sparsity \uparrow	0.00 ± 0.00	87.70 ± 6.05	85.23 ± 5.46	91.95 ± 1.06	96.67 ± 0.16
116×116	DSS	RMSE \downarrow	0.36 ± 0.03	0.38 ± 0.01	0.38 ± 0.01	0.31 ± 0.02	0.29 ± 0.01
		Sparsity \uparrow	0.00 ± 0.00	98.34 ± 0.55	97.39 ± 1.74	97.69 ± 0.47	99.75 ± 0.04
	ADAS-13	RMSE \downarrow	0.23 ± 0.01	0.23 ± 0.01	0.23 ± 0.01	0.18 ± 0.01	0.16 ± 0.01
		Sparsity \uparrow	0.00 ± 0.00	98.77 ± 0.09	97.99 ± 0.80	98.86 ± 0.30	99.89 ± 0.02
	MMSE	RMSE \downarrow	0.29 ± 0.02	0.27 ± 0.01	0.27 ± 0.01	0.22 ± 0.02	0.20 ± 0.02
		Sparsity \uparrow	0.00 ± 0.00	99.10 ± 0.24	98.38 ± 0.86	98.84 ± 0.26	99.64 ± 0.03
$116 \times 116 \times 3$	DSS	RMSE \downarrow	0.34 ± 0.02	0.37 ± 0.02	0.36 ± 0.02	0.32 ± 0.01	0.28 ± 0.01
		Sparsity \uparrow	0.00 ± 0.00	99.38 ± 0.18	98.42 ± 0.76	98.49 ± 0.25	99.97 ± 0.01
	ADAS-13	RMSE \downarrow	0.19 ± 0.02	0.23 ± 0.01	0.22 ± 0.01	0.18 ± 0.01	0.14 ± 0.01
		Sparsity \uparrow	0.00 ± 0.00	99.55 ± 0.06	99.35 ± 0.17	98.85 ± 0.25	99.97 ± 0.00
	MMSE	RMSE \downarrow	0.22 ± 0.02	0.26 ± 0.02	0.26 ± 0.02	0.21 ± 0.02	0.18 ± 0.02
		Sparsity \uparrow	0.00 ± 0.00	99.77 ± 0.05	99.62 ± 0.12	98.70 ± 0.44	99.97 ± 0.01

3.3. Explanation Analysis for Feature Selection

Heatmaps of Discriminative ROIs and Modalities. To understand the effectiveness of feature selection, we explore the most discriminative regions using features identified by compared methods on the original data of 116×3 . We locate the most discriminative regions based on the selected frequency of each region. **Figure 2** illustrates the average coefficient weights of four feature selection methods across five trials. It is obvious that our method achieves more sparse solutions and pays higher attention to the VBM modality as it is a useful approach for investigating neurostructural brain changes in dementia. The top 10 selected brain regions by our method are: Temporal-Pole-Sup-L, Hippocampus-L, Vermis-10, Hippocampus-R, Temporal-Pole-Mid-L, Angular-L, Cerebellum-10-L, Pallidum-R, Cerebellum-10-R, Caudate-L, Cingulum-Post-L. Most of these selected regions are known to be highly related to AD and MCI in previous studies [28-32]. It is evident that only our method can identify the areas of Hippocampus and

Temporal Pole in VBM (as marked in **Figure 2**) which are two critical regions relevant to AD pathology. Furthermore, we also use the factor components learned by our method to measure the contributions of each modality for prediction. According to the average absolute value of $w_r^{(2)}$ for $r = 1, 2, \dots, 70$, we find that the three modalities are ranked as follows: VBM > AV45 > FDG, which coincide with our results in **Figure 2** and the previous findings [33].

Colormaps for Physical Brain Regions. We use the BrainNet Viewer [34] to visualize the brain structure and highlight the regions that the proposed method used to make the predictions against other compared methods, see **Figure 3**. It can be seen that our method is more sparse and uses more relevant ROIs to assess the progression of AD (as marked in **Figure 3** (d)).

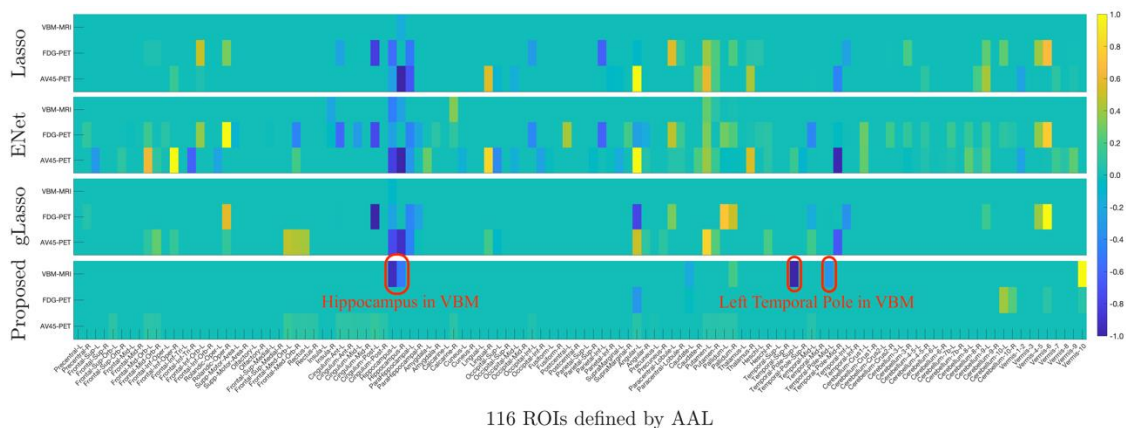


Figure 2. Comparison of coefficient weights in terms of each imaging modality across five trials. Each row corresponds to a feature selection method: Lasso, ENet, gLasso, and our proposed method (from top to bottom). Within each panel, there are three rows corresponding to three imaging modalities, i.e., VBM, FDG, and AV45.

3.4. Hyperparameter Analysis

We further analyse the influence of two important hyperparameters in our method: k and R , where k controls the neighborhood information in the k NN graph, and R controls the number of rank-one tensors that are required to approximate the tensor coefficient. We firstly vary the number of neighbours k in k NN graph data construction to explore the robustness of our method to the data variation. As shown in **Figure 4** (a), it can be noticed that 1) the higher dimensional data (red line) always outperforms low dimensional data (blue line) across nearly all of possible k values, and 2) the best results are consistently achieved in fully connected graphs (i.e., $k = 116$). The results indicate that our method can make full use of high-order relations among ROIs and modalities. Furthermore, we traverse CP-rank R via a step size 1 to investigate its influence on our model performance, as it plays an important role for feature selection. As shown in **Figure 4**(b), we can observe that 1) the proposed model consistently performs better with the increase of R and tends to be stable with ~ 60 , and 2) the sparsity of coefficient is reduced when R is increased. The reasons behind these results are because a higher value of R implies that more non-zero values of coefficients are included, and the most relevant features are selected in the beginning itself and the later attributes do not contribute much to the prediction performance. This shows that users of our model can adaptively make a balance between the accuracy and sparsity by controlling R .

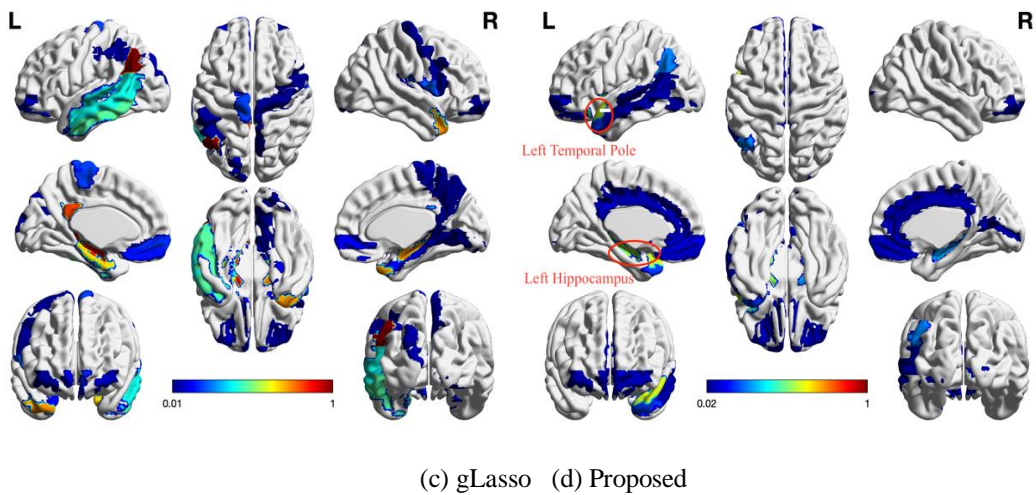
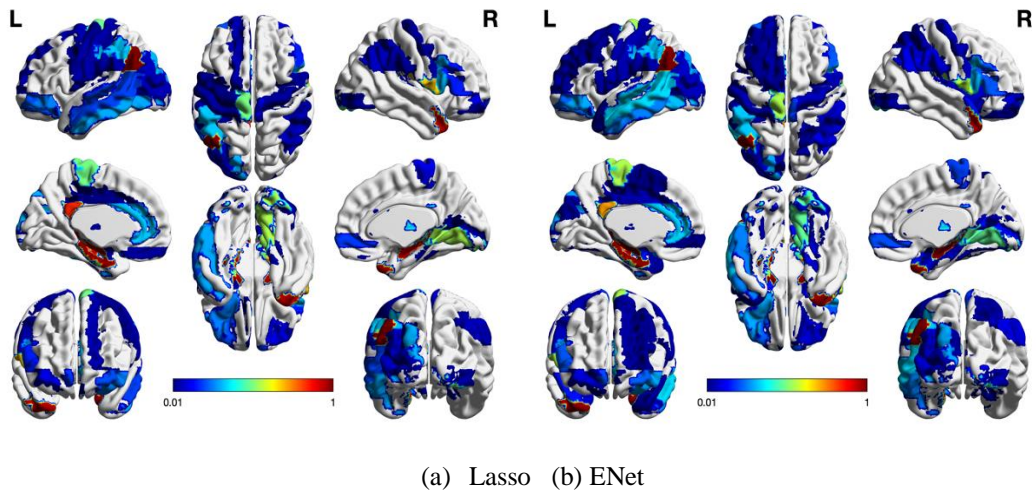


Figure 3. The colormaps of 116 ROIs on the physical brains to the corresponding sparse solutions of each feature selection method. Each method shows the full eight-brain views, in which the first row from left to right are lateral view of left hemisphere, topside, lateral view of right hemisphere, the second row from left to right are medial view of left hemisphere, bottom side, medial view of right hemisphere, and the third row are frontal side and backside.

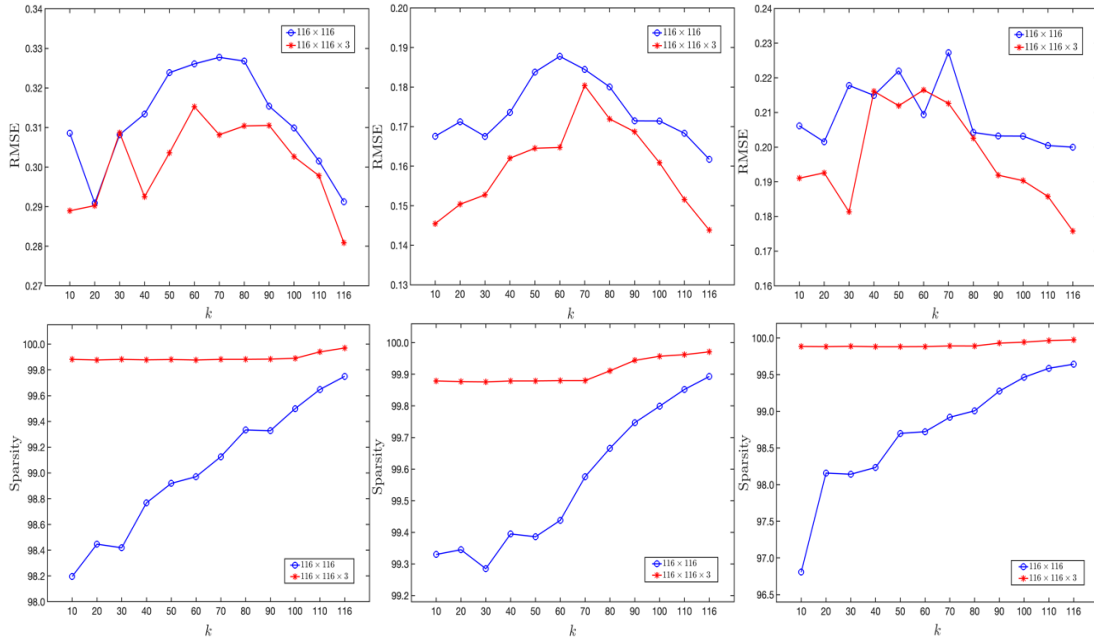
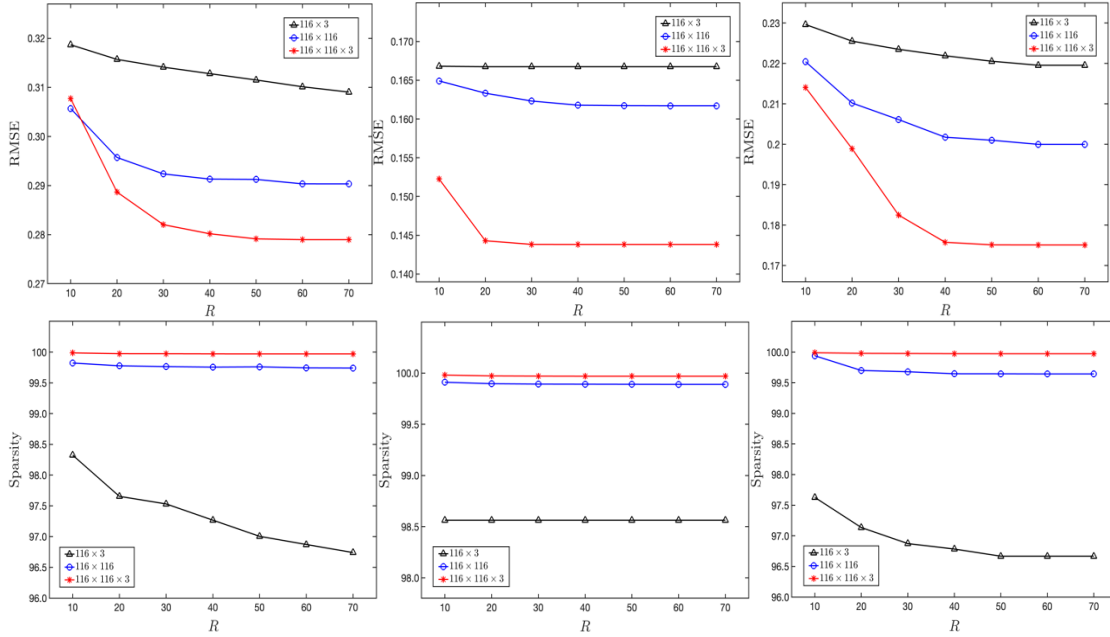
(a) Influence of k on RMSE and Sparsity of DSS, ADAS-13 and MMSE ($R = 60$).(b) Influence of R on RMSE and Sparsity of DSS, ADAS-13 and MMSE ($k = 116$).

Figure 4. Influence of different hyperparameters on model performance associated to DSS, ADAS-13, and MMSE (from left to right). (a) The k value in k NN graph to construct the brain network matrix; (b) the tensor CP-rank R to control the low-rank and sparse property of coefficient weights.

4. CONCLUSIONS

In this paper, we proposed a joint feature selection and tensor regression model for the prediction of AD-related clinical scores and corresponding biomarker identification, in which tensor-

structured sparsity and low-rankness properties are simultaneously exploited to learn the interpretable coefficients. We investigated three different tensor representations to model multi-modality imaging data based on the ROI-level measures with the ADNI database. Our extensive experimental results validated that the proposed method can successfully identify biomarkers related to AD and achieve higher predictive performance than traditional sparse regression methods. Our approach is of wide general interest as it can be applied to other diseases when multi-modality data are available. In the future work, we will extend this approach to simultaneous multiple regression analysis for jointly modelling multiple responses and identifying the importance of regions with multiple clinical scores at the same time.

ACKNOWLEDGEMENTS

This work is supported in part by NIH grants R01AG071243, R01MH125928 and U01AG068057, NSF grants IIS 2045848 and IIS 1837956, ONR grant N00014-18-1-2009 and Lehigh's accelerator grant S00010293.

REFERENCES

- [1] Alzheimer's Association, "2019 alzheimer's disease facts and figures," *Alzheimer's & dementia*, vol. 15, no. 3, pp. 321–387, 2019.
- [2] Ron Brookmeyer, Elizabeth Johnson, Kathryn Ziegler-Graham, and H Michael Arrighi, "Forecasting the global burden of alzheimer's disease," *Alzheimer's & dementia*, vol. 3, no. 3, pp. 186–191, 2007.
- [3] Francesco Angelucci, Gianfranco Spalletta, F d Iulio, Antonio Ciaramella, Francesca Salani, AE Varsi, W Gianni, G Sancesario, C Caltagirone, and P Bossu. 2010. Alzheimer's disease (AD) and Mild Cognitive Impairment (MCI) patients are characterized by increased BDNF serum levels. *Current Alzheimer Research* 7, 1 (2010), 15–20.
- [4] Juan P Amezcua-Sanchez, Anahita Adeli, and Hojjat Adeli. 2016. A new methodology for automated diagnosis of mild cognitive impairment (MCI) using magnetoencephalography (MEG). *Behavioural brain research* 305 (2016), 174–180.
- [5] Biao Jie, Daoqiang Zhang, Bo Cheng, and Dinggang Shen, "Manifold regularized multi-task feature selection for multi-modality classification in alzheimer's disease," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 275–283.
- [6] Daoqiang Zhang, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al., "Multimodal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease," *Neuroimage*, vol. 59, no. 2, pp.895–907, 2012.
- [7] Claudia Plant, Stefan J Teipel, Annahita Oswald, Christian B'ohm, Thomas Meindl, Janaina Mourao-Miranda, Arun W Bokde, Harald Hampel, and Michael Ewers, "Automated detection of brain atrophy patterns based on mri for the prediction of alzheimer's disease," *Neuroimage*, vol. 50, no. 1, pp. 162–174, 2010.
- [8] Angelo Del Sole, Simona Malaspina, and Alberto Magenta Biasina. 2016. Magnetic resonance imaging and positron emission tomography in the diagnosis of neurodegenerative dementias. *Functional neurology* 31, 4 (2016), 205.
- [9] Jiayu Zhou, Jun Liu, Vaibhav A Narayan, Jieping Ye, Alzheimer's Disease Neuroimaging Initiative, et al., "Modeling disease progression via multi-task learning," *Neuroimage*, vol. 78, pp. 233–248, 2013.
- [10] Xiaofeng Zhu, Heung-Il Suk, and Dinggang Shen, "Matrix-similarity based loss function and feature selection for alzheimer's disease diagnosis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3089–3096.
- [11] Morshedul Bari Antor, AHM Jamil, Maliha Mamtaz, Mohammad Monirujjaman Khan, Sultan Aljhdali, Manjit Kaur, Parminder Singh, and Mehedi Masud, "A comparative analysis of machine learning algorithms to predict alzheimer's disease," *Journal of Healthcare Engineering*, vol. 2021, 2021.
- [12] Haifeng Chen, Weikai Li, Xiaoning Sheng, Qing Ye, Hui Zhao, Yun Xu, and Feng Bai, "Machine learning based on the multimodal connectome can predict the preclinical stage of alzheimer's disease: a preliminary study," *European Radiology*, vol. 32, no. 1, pp. 448–459, 2022.

- [13] Jinhua Sheng, Yu Xin, Qiao Zhang, Luyun Wang, Ze Yang, and Jie Yin, "Predictive classification of alzheimer's disease using brain imaging and genetic data," *Scientific Reports*, vol. 12, no. 1, pp. 1–9, 2022.
- [14] Svante Wold, Kim Esbensen, and Paul Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.
- [15] Sanmay Das, "Filters, wrappers and a boosting-based hybrid for feature selection," in *International Conference on Machine Learning*, 2001, pp. 74–81.
- [16] Susanne Mueller, Michael Weiner, Leon Thal, Ronald Petersen, Clifford Jack, William Jagust, John Trojanowski, Arthur Toga, and Laurel Beckett, "The alzheimer's disease neuroimaging initiative," *Neuroimaging Clin. N. Am.*, vol. 15, no. 4, pp. 869, 2005.
- [17] John Ashburner and Karl J Friston. 2000. Voxel-based morphometry—the methods. *Neuroimage* 11, 6 (2000), 805–821.
- [18] Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 1 (2002), 273–289.
- [19] Richard C Mohs, David Knopman, Ronald C Petersen, Steven H Ferris, Chris Ernesto, Michael Grundman, Mary Sano, Linas Bieliauskas, David Geldmacher, Chris Clark, et al . 1997. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. *Alzheimer disease and associated disorders* (1997).
- [20] Wilma G Rosen, Richard C Mohs, and Kenneth L Davis. 1984. A new rating scale for Alzheimer's disease. *The American journal of psychiatry* (1984).
- [21] Tom N Tombaugh and Nancy J McIntyre. 1992. The mini-mental state examination: a comprehensive review. *Journal of the American Geriatrics Society* 40, 9 (1992), 922–935.
- [22] Ulrike Von Luxburg, "A tutorial on spectral clustering," *Stat. Comput.*, vol. 17, no. 4, pp. 395– 416, 2007.
- [23] Lifang He, Kun Chen, Wanwan Xu, Jiayu Zhou, and Fei Wang, "Boosted sparse and low-rank tensor regression," *Advances in Neural Information Processing Systems*, 2018, pp. 1017–1026.
- [24] Tamara G Kolda and Brett W Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [25] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [26] Hui Zou and Trevor Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.
- [27] Ming Yuan and Yi Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.
- [28] Judith Miklossy, "Alzheimer's disease-a neurospirochetosis. analysis of the evidence following koch's and hill's criteria," *Journal of neuroinflammation*, vol. 8, no. 1, pp. 1–16, 2011.
- [29] Emma J Bubb, Claudia Metzler-Baddeley, and John P Aggleton, "The cingulum bundle: anatomy, function, and dysfunction," *Neuroscience & Biobehavioral Reviews*, vol. 92, pp. 104–127, 2018.
- [30] Paul J Mattis, Martin Niethammer, Wataru Sako, Chris C Tang, Amir Nazem, Marc L Gordon, Vicky Brandt, Vijay Dhawan, and David Eidelberg, "Distinct brain networks underlie cognitivedysfunction in parkinson and alzheimer diseases," *Neurology*, vol. 87, no. 18, pp. 1925–1933, 2016.
- [31] Andrei G Vlassenko, Tammie LS Benzinger, and John C Morris, "Pet amyloid-beta imaging in preclinical alzheimer's disease," *Biochimica et Biophysica Acta*, vol. 1822, no. 3, pp. 370–379, 2012.
- [32] Vahram Haroutunian, Pavel Katsel, and James Schmeidler, "Transcriptional vulnerability of brain regions in alzheimer's disease and dementia," *Neurobiology of aging*, vol. 30, no. 4, pp. 561–573, 2009.
- [33] Marion Ortner, Ren é Drost, Dennis Hedderich, Oliver Goldhardt, Felix Müller-Sarnowski, Janine Diehl-Schmid, Hans Förstl, Igor Yakushev, and Timo Grimmer, "Amyloid pet, fdg-pet or mri?-the power of different imaging biomarkers to detect progression of early alzheimer's disease," *BMC neurology*, vol. 19, no. 1, pp. 1–6, 2019.
- [34] Mingrui Xia, Jinhui Wang, and Yong He, "Brainnet viewer: a network visualization tool for human brain connectomics," *PloS one*, vol. 8, no. 7, pp. e68910, 2013.

AUTHORS

Jun Yu is a third-year Ph.D. student in the Department of Computer Science and Engineering at Lehigh University, PA, USA. He received his Bachelor's degree and Master's degree from Shandong University in Software Engineering in 2017 and in Computer Science and Technology in 2020, respectively. His research interests primarily focus on tensor techniques, multi-task learning, and deep learning methods for multi-dimensional data analysis and image computing.



Zhaoming Kong is a fourth-year Ph.D. student in the Department of Computer Science and Engineering at Lehigh University, PA, USA. He received his Bachelor's degree and Master's degree from South China University of Technology in Applied Mathematics in 2016 and in Software Engineering in 2019, respectively. His research interests primarily focus on tensor analysis, image processing and deeplearning for multidimensional data analysis.



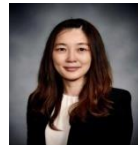
Liang Zhan is currently an Assistant Professor in the Departments of Electrical and Computer Engineering and Bioengineering at the University of Pittsburgh, PA, USA. He received his Ph.D. degree in Biomedical Engineering from University of California, Los Angeles (UCLA). His research interests include computational neuroimaging, brain connectomics, machine learning and bioinformatics.



Li Shen is a Professor of Informatics in Biostatistics and Epidemiology. His research interests include medical image computing, bioinformatics, machine learning, network science, visual analytics, and big data science in biomedicine. He received his BS degree (Computer Science) at Xi'an Jiao Tong University in 1993, MS degree (Computer Science) at Shanghai Jiao Tong University in 1996, and PhD degree (Computer Science) at Dartmouth College in 2004.



Lifang He is currently an Assistant Professor in the Department of Computer Science and Engineering at Lehigh University, PA, USA. She received her Ph.D. degree in Computer Science from South China University of Technology in 2014. Her research interests primarily focus on machine learning, data mining, tensor analysis, with major applications in biomedical data and neuroscience. She has designed several tensor powerful processing and advanced machine learning methods for a range of real-world problems. Her work have been featured by major publications such as WIREs Computational Molecular Science, Neural Networks, TKDE, TIP, NeurIPS, and ICML.



MACHINE LEARNING GUI BASED FOR DETECTING ALZHEIMER'S

Fatema Nafa¹, Evelyn RodriguezArgueta¹,
Annie Dequit¹ and Changqing Chen²

¹Department of Computer Science, Salem State University, Salem, MA

²Department of Chemistry and Physics, Salem State University, Salem, MA

ABSTRACT

Alzheimer's disease (AD), a kind of dementia, is marked by progressive cognitive and behavioural problems that appear in middle or late life. Alzheimer's disease must be detected early in order to create more effective therapies. Dr. Alois Alzheimer was the first doctor in the medical field to notice an unusual state of change in the brains of his deceased patients with mental illness, which marked the start of Alzheimer's study. Machine learning (ML) techniques nowadays employ a variety of probabilistic and optimization strategies to allow computers to learn from vast and complex datasets. Because of the limited number of labelled data and the prevalence of outliers in the current datasets, accurate dementia prediction is extremely difficult. In this research, we propose a sustainable framework for dementia prediction based on ML techniques such as Support Vector Machine, Decision Tree, AdaBoost, Random Forest, and XGmodel. All the experiments, in this literature, were conducted under the same experimental conditions using the longitudinal MRI Dataset.

KEYWORDS

Machine learning, Alzheimer's disease, Feature selection, Biomechanical parameters.

1. INTRODUCTION

The earliest research into the Alzheimer's field began with Dr. Alois Alzheimer who was the first doctor within the medical field to begin to notice an unusual state of change of the brain from his deceased patients with mental illness. This article helped share some light on how the evolution of Alzheimer's research has changed drastically throughout the years. Early onset research focused on the initial connection linked with the genetic traits that were present for the older members of families that had AD and those younger members that over a course of time began showing symptoms or for some generations would never get to that point of progression with the disease [1]. As technology began making its strides forward, we followed its lead and used this new technology to make the discovery that there was a consistent behaviour seen with the high production of amyloid beta and individuals with Alzheimer's disease.

Human based detection of this disease is a time-consuming and expensive process that requires a large amount of data and the involvement of an experienced clinician.

Automated systems are not prone to human mistake, they are more accurate than human assessments and can be employed in medical decision support systems. Several studies on

Alzheimer's disease diagnostics have been completed, with the focus recently shifting to the accurate prediction of the disease's early stages.

Furthermore, the current rapid advancement of machine learning technology, which uses AI to anticipate various result, has had a substantial impact on medicine [2], [3]. Recently, several machine learning modelled has been used for AD detection. such as SVM (support vector machines), KNN (K-nearest neighbour), NN (Neural Network) [4]– [6]. However, it remains a major challenge to select best parameters for each model.

In this study, we will build a more effective prediction GUI system for Alzheimer's detection at an earlier stage by using machine learning models, such as Support Vector Machine, Decision Tree, AdaBoost, Random Forest, and XGmodel. The models are compared, CatBoost model clearly outperforms various other models. The results of the experiment reveal that using the CatBoost model for a B-cell epitope prediction with high prediction accuracy, stability, and speed may be developed.

2. RELATED WORK

Many scholars have advocated a comprehensive study on the classification and diagnosis of Alzheimer's disease. A quick review of the relevant work is included in this section.

The major drivers of advanced data analysis model are artificial intelligence (AI) and machine learning (ML). ML approaches are built on the basis of data representations and discover significant insights from the data. The most important step in creating such models is extracting features from the data. Models can be produced manually automatically based on the data representations [2]. The authors [3] utilized a semi-supervised learning model to create a MRI biomarker of MCI-to-AD conversion. Mild cognitive impairment (MCI) is a transitional stage between age-related cognitive decline and Alzheimer's disease (AD). The authors [7] suggested a new eigenbrain-based computer-aided diagnostic (CAD) system for MRI brain imaging. They used kernel support vector machines with multiple kernels to generate the discriminant areas that separate AD from NC, MIE coefficients with values higher than 0.98 were highlighted. Sharma et al. in [8] introduced and tested a support vector machine (SVM) model for discriminating between patients with Alzheimer's disease (AD) and older control subjects of whole-brain anatomical magnetic resonance imaging. Another work by [4] investigated Alzheimer's infections. Using support vector machine (SVM) The researchers studied three main segments: Sagittal, Axial, and Frontal regions of the cerebrum.

The most relevant work is featured in this section, but there are a few others that have investigated the same problem. The scope of this work is limited for biomarkers produced from MRI images. In the above discussed stateof-art techniques are bounded by their performance over many limitations. The authors of this study proposed a machine learning models, Support vector machine, Decision Tree, AdaBoost, Random Forest, and XGmodel to accurately predict progress of a patient from mild cognitive impairment to dementia.

3. PROBLEM DEFINITION

To help clinicians and therapies predicting an early dementia of a patient we propose to develop a machine learning models that can accurately predict progress of a patient from mild cognitive impairment to dementia. Also, the models have a friendly GUI for non-programmers, users do not need advanced knowledge to take full advantage of these models.

4. METHODOLOGY

This section focuses on methodology used for this study. User Graphical Interface (GUI) was developed using Python. GUI designed to predict the dementia. GUI employs easy architecture to be used by non-programmer users. Users can upload their dataset; clean the data using data pre-processing tool; and perform the prediction using multiple machine learning models. The data pre-processing tool provides a set of GUI icons that allow user to read and clean the dataset. Then prediction tool provides icons that can help user to easily use the models to help predict dementia with controlling of different parameters. The section will explain the dataset, data pre-processing, crossfold validation, and Machine Learning models respectively.

4.1. Dataset

The ML models were trained and tested on publicly available longitudinal MRI data[9]. The dataset consists of 373 in total, there are around 39% demented cases in the dataset i.e., majority of the data is of non-Demented cases while 10% of the data is of Converted as shown in Fig.1. The descriptions of the attributes and brief statistical summary are shown in Table. 1.

The paper used the dataset to investigate the following:

- To assist physicians and therapies in detecting a patient's early dementia
- what biomarkers (or variables) are associated with dementia?
- Is the data of the different datasets normally distributed?
- How would you build and evaluate a predictive model, for dementia on these data?

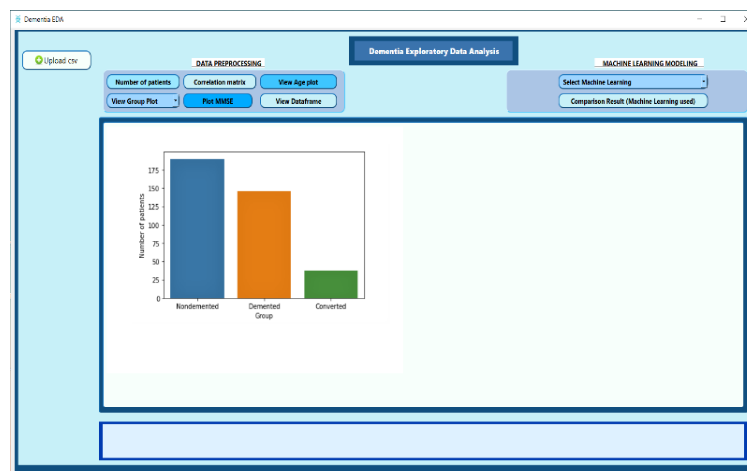


Figure 1. Distrbuation of Dementia Group.

Table 1. The Overview of the Dementia Patient Dataset.

Attribute	Description	Mean±Std
1. Group	Patient grouped as Converted (Previously Normal but developed dementia later), Demented and Nondemented (Normal Patients)	1.412±0.6644
2. M/F	Gender	0.571±0.495
3. Age	Age in years	17.134±7.641
4. EDUC	Years of education	5.611±2.59
5. SES	Socioeconomic status as assessed by the Hollingshead Index of Social Position and classified into categories from 1 (highest status) to 5 (lowest status)	1.93±1.55
6. MMSE	Mini-Mental State Examination score (range is from 0 = worst to 30 = best)	15.059±3.84
7. CDR	Clinical Dementia Rating (0 = no dementia, 0.5 = very mild AD, 1 = mild AD, 2 = moderate AD)	0.58±0.72
8. eTIV	Estimated Total Intracranial Volume	142.92± 78.14
9. nWBV	Normalize Whole Brain Volume	65.83± 32.94
10.ASF	Atlas Scaling Factor	132.22 ±71.0056

4.2. Data Pre-Processing

This step enhances data quality by detecting and eliminating mistakes and inconsistencies. GUI in Fig.2. showing the main interface of the system. The left side include the pre-processing part. This consist of checking the overall distribution of categorical and numerical columns, the missing values, outlier rejection, and feature selection of the attribute. The data pre-processing described as follows:

Figure 2: The main GUI of the system.

Outlier detection [10], [11] outliers can negatively affect the training process of the model resulting in lower accuracy. there are different techniques can be used to detect the outliers. In this paper IQR (Inter Quartile Range) Score been used to detect the outliers [12], [13].

Missing values, it is important to identify and manage missing values effectively during data preparation; it led to drawing incorrect conclusions and inferences from the data[14]. There are two approaches to deal with missing data, deleting the rows that has missing values and calculating the mean or median. In this study, calculating the median method is used. Some features have float value cannot impute a float value of mean in place however median imputed, and median is the most representative value of the features in this scenario. Calculating the mean values of the attributes can be formulated as in (1).

$$M(x) = \begin{cases} \text{mean}(x) & \text{if } x = \text{null} \\ x & \text{otherwise} \end{cases}$$

The standardization the practice of rescaling attributes to achieve a standard normal distribution with zero mean and unit variance is known as standardization or Z-score normalization. As illustrated in Fig.3. standardization (R) reduces the skewness of the data distribution.

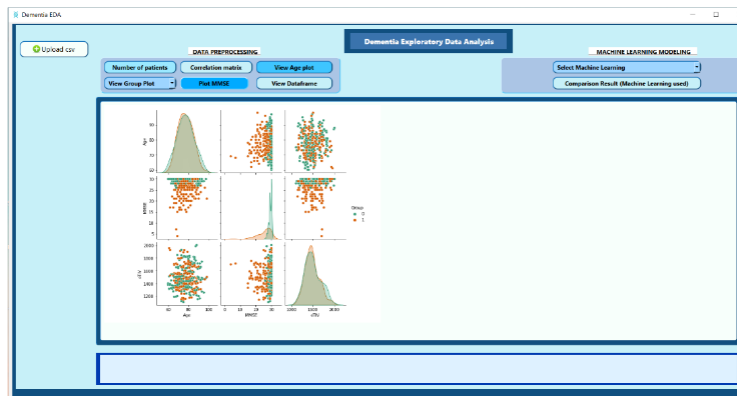


Figure 3. The Population Distribution of all Features.

Feature Engineering, there are approximately 168 Nondemented cases in the dataset, with 34 Converted cases. There are also 97 cases of Dementia. For further information, we investigate another key characteristic called Clinical Dementia Rating. There is a scoring reference used to help doctors determine proper ratings, according to [15], [16]. The score (Normal=0, Very Mild Dementia=0.5, Mild Dementia=1, Moderate Dementia=2, Severe Dementia=3) can be used to characterize and track a patient's level of impairment or dementia.

4.3. Cross-Fold Validation

The K-fold Cross-validation (KCV) technique is one of the most widely used approaches by practitioners for model selection and error estimation of classifiers [15]. The k-fold cross-validation procedure in this study implemented using the scikit-learn machine learning library in Python.

4.4. Machine Learning Models

This is the most crucial step, which includes the development of a dementia prediction model. Various machine learning techniques for dementia prediction have been implemented. The

models are Support Vector machine, Decision Tree, Random Forest, Ada Boost model, and Gradient Boost model. After feeding the input dataset, the model will predict using machine learning techniques and deliver the best result in the form of a comparison to predict the best accuracy for detecting dementia. In this experiment, the best result has been delivered using random forest model. A random forest model is made up of tree-structured [17], [18]. The model adds objects from an array of input to each forest tree. Every single tree vote for classification of the elements of the unit vector independently. The forest selects out the classifications with the greatest votes.

5. RESULT

5.1. Experiment

This section discusses the dataset used in the experiment and the technical requirements. The technical requirements are Python programming language. Then import the necessary libraries and import the dataset to the Jupyter notebook. An experiment was conducted on dataset has been used around 90 blood biomarkers and other demographic data, the total size was 43.8 KB. Table 2. provides a descriptive statistic to describe and summarize the data. It uses quantitative approach describes and summarizes data numerically. We can observe that the data set contain 373 rows and 15 columns. The missing values are 21.

To get a further insight into the data, correlation values were calculated to know how much an attribute affects the dementia attribute (Outcome) or if other attributes are affected by it. Correlation values were calculated using the Pearson (product-moment) correlation coefficient equation. It computes the ratio of the covariance of both features to the product of their standard deviations consequently finding the measure of the linear relationship between those two features. Correlation values are shown in Fig.4.

Table 2. Statistical Information about Dementia Dataset.

Dataset statistics		Variable types	
Number of variables	15	Categorical	
Number of observations	373	Numeric	
Missing cells	21		
Missing cells (%)	0.4%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	43.8 KiB		
Average record size in memory	120.3 B		

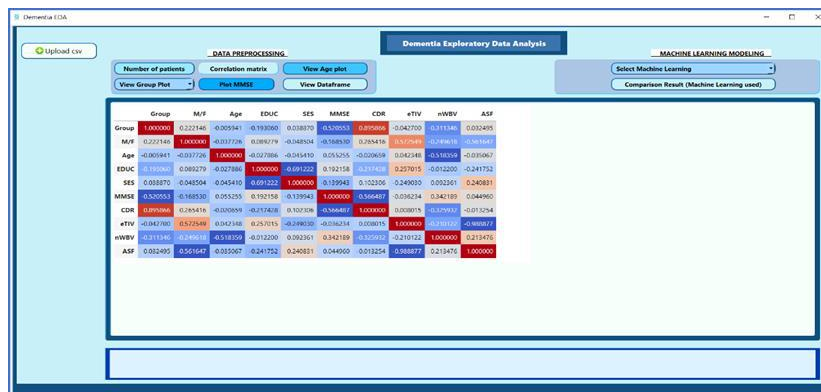


Figure 4. The confusion matrix of the attribute's correlation with the outcome.

Histograms were also created to provide a better visual representation of the data, as seen in Fig.5. Aside from the improved presentation that histograms provide, the figures can make it easier to see potential outliers that could harm the proposed model.

5.2. Evaluation Metrics

The metrics used for measuring the performance of the system are precision, recall, and Accuracy. Here follow the definitions of precision, recall, and Accuracy. see Formulas 1.1,1.2 ,and 1.3 respectively [19], [20].

Precision: Precision is calculated by multiplying the number of TP by the number of TP 'Plus' the number of FP. False positives occur when a model is wrongly classified as positive when it is negative.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall: the number of true TP separated by the TP '+' FN used to calculate the recall.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Accuracy: another metric is accuracy, which is defined as the percentage of true cases (both positive and negative) among all examples retrieved

$$\text{Accuracy} = \text{AUC} = \frac{tp + tn}{tp + tn + fp + fn}$$

5.3. Results Comparison

As it shown in Fig.6. Decision Tree model is very good when we have no idea on the data. Even with unstructured and semi structured data like text, images, and trees Decision Tree algorithm works well. The drawback of the Decision Tree model is that to achieve the best prediction results for any given problem, several key parameters are needed to be set correctly.

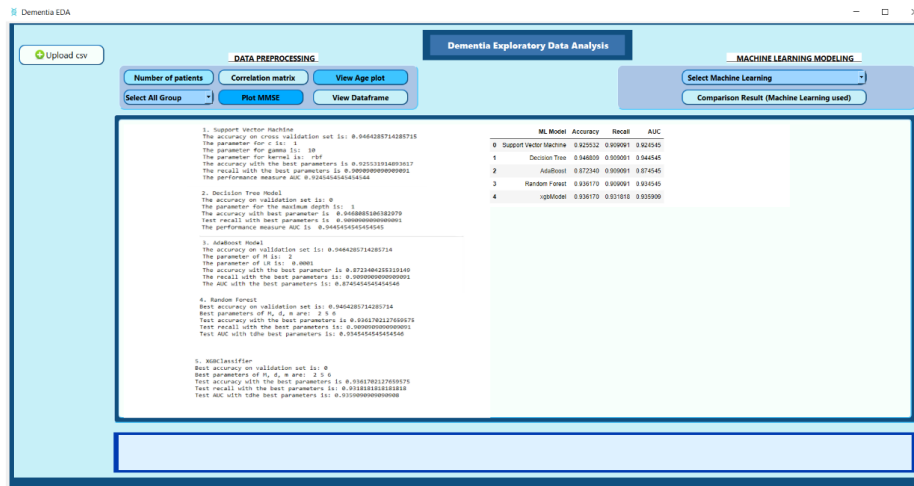


Figure 5. Five Machine learning Models and their Comparison Accuracy of Result.

6. CONCLUSION AND DISCUSSION

One of the real-world medical challenges that must be solved is the early detection of dementia. Throughout this research, deliberate attempts have been made toward developing a framework that will eventually be used to predict illnesses such as dementia. Four machine learning models used; Decision Tree algorithm was explored and assessed on several measures during this project, Decision Tree model achieved the highest accuracy. The goal of this project is to create a system that can predict dementia in a patient earlier and more accurately utilizing machine learning techniques that provide advance support for dementia prediction accuracy.

REFERENCES

- [1] L. Caly, J. D. Druce, M. G. Catton, D. A. Jans, and K. M. Wagstaff, "The FDA-approved drug ivermectin inhibits the replication of SARS-CoV-2 in vitro," *Antiviral Res.*, vol. 178, p. 104787, 2020.
- [2] S. Sharma and P. K. Mandal, "A Comprehensive Report on Machine Learning-based Early Detection of Alzheimer's Disease using Multi-modal Neuroimaging Data," *ACM Comput. Surv. CSUR*, vol. 55, no. 2, pp. 1–44, 2022.
- [3] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, J. Tohka, and A. D. N. Initiative, "Machine learning framework for early MRIbased Alzheimer's conversion prediction in MCI subjects," *Neuroimage*, vol. 104, pp. 398–412, 2015.
- [4] A. Sharma, S. Kaur, N. Memon, A. J. Fathima, S. Ray, and M. W. Bhatt, "Alzheimer's patients detection using support vector machine (SVM) with quantitative analysis," *Neurosci. Inform.*, vol. 1, no. 3, p. 100012, 2021.
- [5] T. A. Assegie, "Support Vector Machine And K-Nearest Neighbor Based Liver Disease Classification Model," *Indones. J. Electron. Electromed. Eng. Med. Inform.*, vol. 3, no. 1, pp. 9–14, 2021.
- [6] M. A. Ibrahim, M. U. G. Khan, F. Mehmood, M. N. Asim, and W. Mahmood, "GHS-NET a generic hybridized shallow neural network for multi-label biomedical text classification," *J. Biomed. Inform.*, vol. 116, p. 103699, 2021.
- [7] Y. Zhang et al., "Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning," *Front. Comput. Neurosci.*, vol. 9, p. 66, 2015.
- [8] B. Magnin et al., "Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI," *Neuroradiology*, vol. 51, no. 2, pp. 73–83, 2009.
- [9] J. H. Kramer et al., "Longitudinal MRI and cognitive change in healthy elderly.," *Neuropsychology*, vol. 21, no. 4, pp. 412–418, 2007, doi: 10.1037/0894-4105.21.4.412.

- [10] I. Ben-Gal, "Outlier detection," in *Data mining and knowledge discovery handbook*, Springer, 2005, pp. 131–146.
- [11] V. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artif. Intell. Rev.*, vol. 22, no. 2, pp. 85–126, 2004.
- [12] S. Walfish, "A review of statistical outlier methods," *Pharm. Technol.*, vol. 30, no. 11, p. 82, 2006.
- [13] C. Andreou and V. Karathanassi, "Estimation of the number of endmembers using robust outlier detection method," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 1, pp. 247–256, 2013.
- [14] W. Z. Liu, A. P. White, S. G. Thompson, and M. A. Bramer, "Techniques for dealing with missing values in classification," in *International Symposium on Intelligent Data Analysis*, 1997, pp. 527–536.
- [15] J. C. Morris, "Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type," *Int. Psychogeriatr.*, vol. 9, no. S1, pp. 173–176, 1997.
- [16] W. J. Burke et al., "Reliability of the Washington University clinical dementia rating," *Arch. Neurol.*, vol. 45, no. 1, pp. 31–32, 1988.
- [17] C. Brokamp, R. Jandarov, M. Hossain, and P. Ryan, "Predicting daily urban fine particulate matter concentrations using a random forest model," *Environ. Sci. Technol.*, vol. 52, no. 7, pp. 4173–4179, 2018.
- [18] S. J. Rigatti, "Random forest," *J. Insur. Med.*, vol. 47, no. 1, pp. 31–39, 2017.
- [19] H. Dalianis, "Evaluation metrics and evaluation," in *Clinical text mining*, Springer, 2018, pp. 45–53.
- [20] J.-O. Palacio-Niño and F. Berzal, "Evaluation metrics for unsupervised learning algorithms," *ArXiv Prepr. ArXiv190505667*, 2019.

LABBUDDY: A GAME-BASED INTERACTIVE AND IMMERSIVE EDUCATIONAL PLATFORM FOR PHYSICS LAB LEARNING USING ARTIFICIAL INTELLIGENCE AND 3D GAME ENGINE

Yuxing Ji¹, Mingze Gao² and Yu Sun³

¹Oaks Christian School, 31749 la tienda rd, westlake village, CA, 91362

²University of California, Irvine, Irvine, CA 92697

³California State Polytechnic University,
Pomona, CA, 91768, Irvine, CA 92620

ABSTRACT

The concepts of physics play an important role in many fields of people's lives, and physics learning is abstract, challenging, and sometimes intimidating [1]. However, how to motivate students to learn physics in a fun way becomes a question. This paper develops a gamic, interactive, educational application to allow students to learn abstract physics in an illustrative way. We have implemented a visual physics lab by using a 3D game engine supporting the immersive environment of visualization and providing a playful learning tool for physics experiments at the same time [2].

KEYWORDS

Physics, Virtualization, Artificial Intelligence, 3D Game Engine.

1. INTRODUCTION

The learning of physics has become a mandatory class in education, and the experiments and labs are the most helpful tools to understand the concepts of physics. In most cases, students can only do experiments within the school labs; however, this limits them from re-do the experiments outside of the school, which results in students not fully interacting with the learning concepts, and it may cause students to have a lack of understanding and memorization. Learning the knowledge of physics only takes half part of the procedure, the other half is the operating step where students combine experiments in order to find out the relationship between physics functions and principles [3]. Therefore, the opportunity of doing experiments and labs in the learning of physics becomes an important issue that needs to be considered. For example, I had a deep impression about writing a report in my physics class. My group and I did an experiment during school time, but we were working on the lab report at home but we had a lack of memory on one of the experiment steps. We wanted to redo the lab but there is no place for us. I was really distressed at that time because the report was due on the same night. Therefore, I was thinking why couldn't we do it virtually? This idea guided me to do a lot of research and readings, and I finally decided to create a virtual physics lab that could benefit more students in learning physics.

There are many physics experiment simulators that have been published in public, which allow users to choose a specific lab or experiment. Users are free to put any values such as speed, gravity, direction, etc. However, these simulators are formed by a 2D platform where the users are only allowed to observe one side of the experiment [4]. Their implementations are limited in the experience of virtualization causing the users hard to see the changes on other sides during the experiment. In addition, the environment around the experiment is very shabby because the 2D platform method only provides a single background. A practical problem is that users are not able to immerse themselves in the experiment and it is hard to focus on details happening in the experiment.

Our goal is to create a virtual physics lab that allows users to choose any specific lab with appropriate inputs. However, our simulator is different compared to the others where the simulator is implemented as a 3D game engine instead of a 2D platform [5]. Within the virtual lab, users are free to walk around the lab, and there are different stations of labs placed in the lab room, and users can choose to play any of them. There are some good features of the virtual lab room. First, the experiment provides multi-angle observation that helps users to catch more details in the experiment. Second, the lab room is created as a virtual reality that makes an immersive feeling of doing the labs. Third, a short quiz will be generated to test the concepts of the experiment based on observation and operation. By doing this, it draws users' attention to recall the images of the experiment in order to have a deep understanding of the details. Therefore, we believe that our immersive educational platform of physics labs can effectively help users to get success in learning physics. In the early stages of the implementation, we have planned three different physics labs which are Magnetic Drop, Floating Egg, and Hot Water Pump.

Since the application is designed as a 3D platform, the virtual experience and virtualization of the labs are the most important keys of the application. We will create a checklist to test various features and functionalities of the virtual lab. The corresponding screenshots will be added along with the checklist in order to support the expected user experience. For example, the checklist will test the movements of the first-person perspective and make sure that the character can move freely within the virtual lab room. In addition, the experiment needs a different angle of observation in order to find out more details such as the lab of Magnetic Drop, which requires a top view to observe the moving speed of the object [6]. It is also necessary to include an observation angle testing in the checklist. We will carefully list all the testing areas in the checklist and test with a real machine demonstration to record how users could interact with our virtual 3D physic lab. In addition, the lab results will also be compared with the real experiments that will be finished in real life in order to evaluate the accuracy of the results. We will record the results of both happening in the application and the real life on the same table so that we can have a clear mind while comparing the results.

The rest of the paper is organized as follows: Section 2 gives the details on the challenges that we met during the experiment and designing the sample; Section 3 focuses on the details of our solutions corresponding to the challenges that we mentioned in Section 2; Section 4 presents the relevant details about the experiment we did, following by presenting the related work in Section 5. Finally, Section 6 gives the conclusion remarks, as well as pointing out the future work of this project.

2. CHALLENGES

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

2.1. Learning process of programming language and Unity

The first challenge was the learning process of programming language and Unity since the idea was to create a 3D virtual lab, Unity would be the first choice to accomplish it. However, this was the first time that I started to learn code, and I had no idea how I could implement code with Unity in order to reach my goal. I did a lot of research and watched different videos of Unity, then I decided to implement C# with Unity to do this project [7]. At the beginning of learning C#, it was a tough time since I was really new to the program, and the structure made me confused all the time. However, I was working hard and always researched when I had questions. I also set a schedule for each day in order to keep on track. The process turned out really well after I familiarized myself with C#, and I started to implement C# with Unity by doing a lot of exercises and testing.

2.2. Environment designing process for the virtual lab

The second challenge was the environment designing process for the virtual lab, and it took me a long time to figure out how to fit everything that I want in the virtual lab room. I first sketched a draft on paper and started to set up the environment on Unity by referring to the sketch; However, the Unity was not as easy as I had expected, the dimensions and scales were really hard to handle, and the things that were placed in the lab room looked messy. However, I took some extra time carefully calculating the dimensions and re-set the sizes of the objects. By doing this, it promised a learning environment in the virtual lab that provided immersive feelings for the users.

2.3. Converting the necessary physics functions into C#

The third challenge was converting the necessary physics functions into C# in order to keep the accuracy of the results showing in the virtual lab. The physics functions looked really easy to write by hand; however, it was hard for me to implement the functions in a programming language because some values need to be saved in different places. For example, the value x needed to be saved outside of the function, but the y value needed to be saved inside the function. I did deep research by watching different tutorials and videos to fully understand different variable types in C# in order to implement physics functions correctly in the experiment.

3. SOLUTION

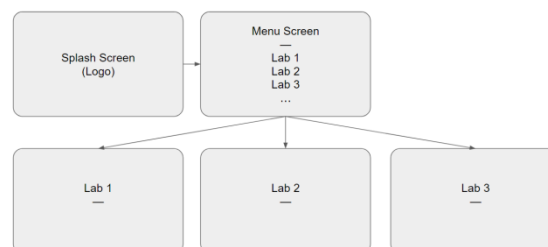


Figure 1. Overview of the Virtual Lab

The immersive educational platform for physics lab is designed by using C# with Unity where the C# supports all the features and operations happening in the lab, and the application of Unity creates a 3D virtual world to pull users into the immersive learning environment. The whole design of the virtual physics lab is broken into three big sections (Figure 1). The first section is the interface when users open the application where they will be required to press the open button

to join the menu screen for the lab selection. The second section is the main menu where users are free to choose any lab to start with, the current implementation plan consists of three different labs. Once users have selected one of the three labs, they will join the virtual lab room to test and operate the corresponding experiment. The third section is the lab room which contains all the labs where users have the ability to switch between the labs. For example, after the users have finished the first lab experiment, they are free to move the character around the lab room. Each lab has its own desk to do the experiment, so users can walk to the next desk to do the other labs. It saves the extra steps of going back to the main menu to do the selection. Instead, it provides an immersive learning experiment where users have the same experience as the actual physics lab and make flexible choices while doing the labs. We are also planning to make quizzes for each lab which provide a short quiz when users have finished doing the lab [8]. By doing the quiz after the lab, it helps users to recall their memories while doing the experiment, making a deep impression of the details that happened in the lab. It is also an extra opportunity to practice and discover the weaknesses.

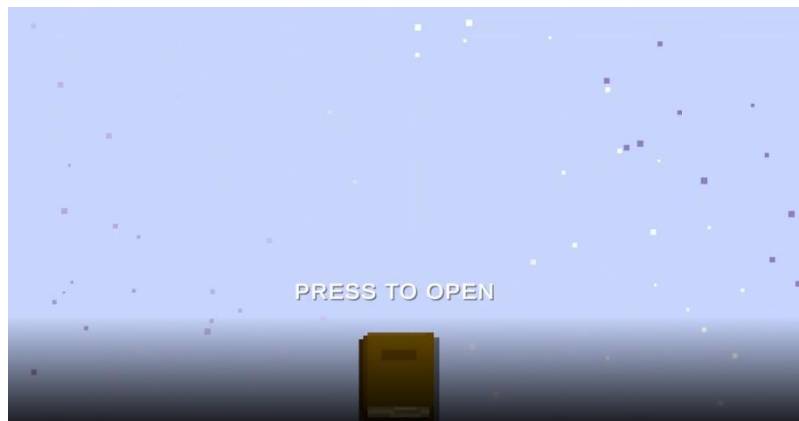


Figure 2. Starting Interface

First, the start of the user interface is designed as a magic style that sets off the theme of physics (Figure 2). The background is implemented as a dynamic background, and a book icon is placed at the center of the bottom. Once the book is pressed, it will lead users to the lab selection interface.



Figure 3. Lab Selection Interface

After pressing to open the book, the interface will change to lab selection where the background contains physical elements (Figure 3). There are two labs that have been implemented so far, users are free to choose any of them to start the lab. Each choice has an experiment picture, lab

number, and lab name in order to help users to distinguish and make the correct choice. In the future, there will be more labs coming into the list, we are also planning to set different categories or folders to group labs based on topics. There will also be a search button or filter to support quick access to specific labs.



Figure 4. First View of Magnetic Drop Lab

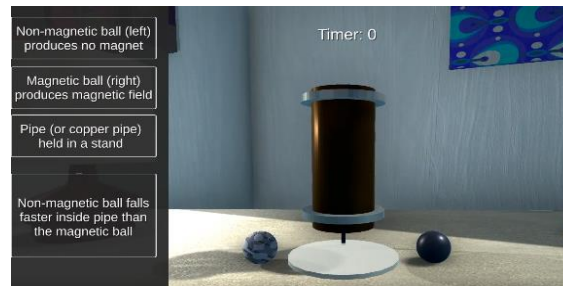


Figure 5. The Explanation of Magnetic Drop Lab

For example, if the user has selected “Lab 1: Magnetic Drop” to start the experiment, the user will be brought into a virtual lab room where he or she is free to move the camera and walk around the room. A piece of short instruction information will be provided on the screen that asks the user to “Press E to activate Magnetic Drop Lab” (Figure 4). After that, the screen will display an explanation of the materials and results of the experiment in order to provide a clear understanding of the goals for the current lab (Figure 5).

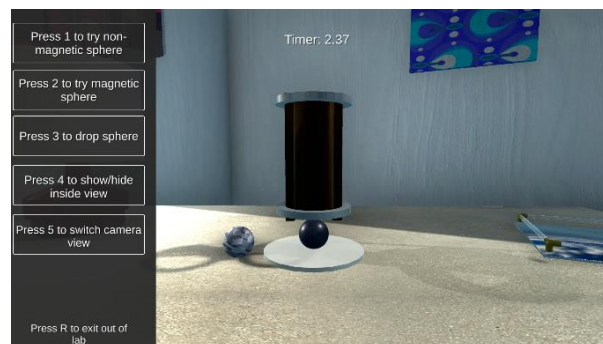


Figure 6. Front View of Magnetic Drop

Once the experiment starts, a list of operations is displayed on the left side of the screen (Figure 6). The user can press “1” to try a non-magnetic sphere. “2” to try the magnetic sphere, “3” to drop the sphere, “4” to show or hide inside view, “5” to switch camera view, and “R” to exit out of the current lab. When each drop has started, the timer will start to record the time

that sphere will fall from top to bottom. Once a new sphere is switched, the timer will reset to 0 and start counting for a new round. This lab also supports different observations of the experiment where the inside view of the sphere can be shown or hidden. It provides a flexible experiment for observing the movement of the sphere within the object. In addition, a different angle of observation is also provided to the lab.

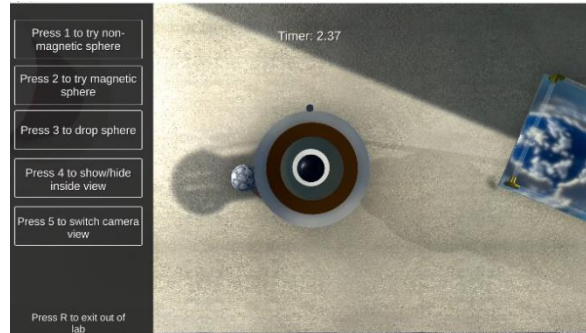


Figure 7. Top View of Magnetic Drop

After pressing “5” to switch the camera view, the observation angle will change to the top where the user can see the movement of the sphere vertically in the object (Figure 7). The interface is similar to the front view, and the user still has access to the operation list from the left of the screen. A timer is also shown on the top of the screen to record the time when the sphere reaches the bottom. By providing a different view of the lab, it supports users to enjoy the immersive learning experience where they are able to learn the experiment by observing it from different angles in order to discover any changes and patterns in the results.



Figure 8. View of the Lab Room

Once the user has finished the current lab, he or she is free to press “R” to exit the current lab. The user can move to the other lab desk for a new experiment (Figure 8). When they walk closer to the lab desk, the instruction information will display on the screen.

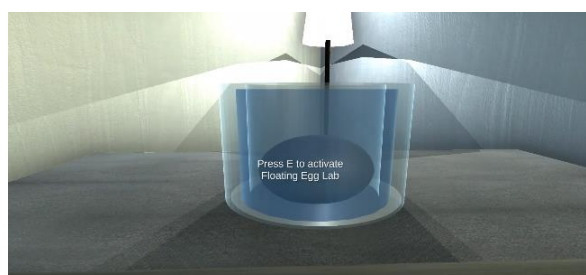


Figure 9. First View of Floating Egg

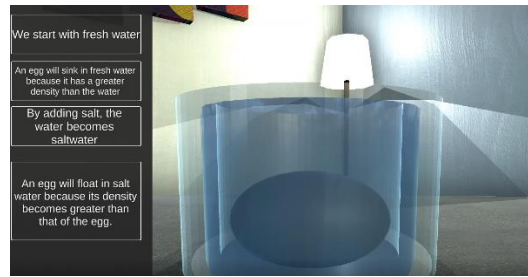


Figure 10. The Explanation of Floating Egg

When the user wants to do another lab in the lab room, they can walk close to the lab desk and press “E” to activate the lab. For example, if a user has finished Magnetic Drop Lab and wants to do Floating Egg Lab, he or she can start it by pressing “R” (Figure 9). After that, the explanation and instruction will display to the left of the screen in order to provide a clear mind of how the experiment would work (Figure 10).

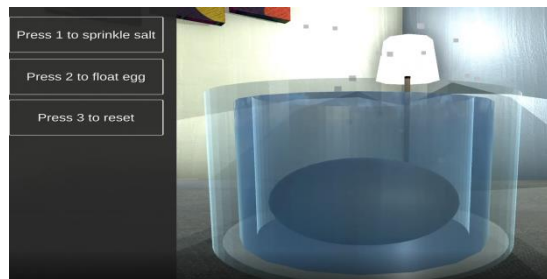


Figure 11. The Front View of Floating Egg #1

Once the Floating Egg experiment starts, it will provide an operation list from the left of the screen such as press “1” to sprinkle salt, press “2” to float the egg, and press “3” to reset. The action of adding salt is implemented as a dynamic action where the little white spots are the salts added to the water (Figure 11).

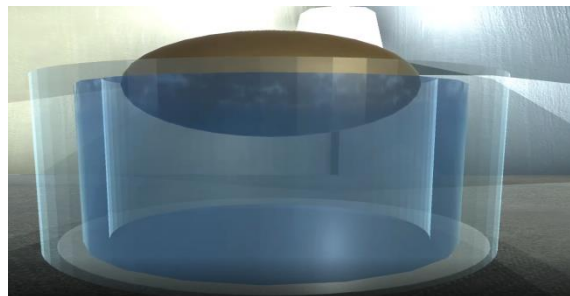


Figure 12. The Front View of Floating Egg #2

Because the salt is added to the water, the egg will float in the salted water since the density becomes greater than that of the egg (Figure 12). Instead, if the water is not salted, the egg will stay under the water all the time. The whole experiment is designed by dynamic reactions so that users are easier capture the changes in the lab.

4. EXPERIMENT

4.1. Experiment 1

In order to check the user experience within the virtual lab room, we have created a checklist that covers all the major operations that could happen while using the application. We will join the virtual lab as a user to test all the tasks that are listed in the checklist, making sure the features work in the good condition. By looking at Table 1 of the checklist, we have self-tested the camera movement, character movement, lab instruction, timer, different angle of observation, exit button, switch between labs and change mouse sensitivity. The major operations and functionalities work great and reflect good feedback while testing the whole application. However, we also realize a few issues with the beginning instructions which might cause new users not able to get familiar with the controls. For example, once the user has successfully joined the virtual lab room, there are no tips or hints about how they can interact with the character. By discussing this issue, we have decided to add a list of buttons with their functionalities along to the side of the screen such as buttons “W” moving forward, “S” moving backward, “A” moving to the left, and “D” moving to the right. In this case, not only the new users will have a clear mind of how they could control the virtual character, but also it reminds the returned users of how they could interact with the application.

Table 1. Experiment 1 - Check List

Operations	Results
Camera Movement	✓
Character Movement	✓
Lab Instruction Displayed	✓
Timer	✓
Different Angle of Observation	✓
Press “R” to Exit Current Lab	✓
Switch Lab	✓
Change Mouse Sensitive	✓

By considering the accuracy of the results from the virtual lab, we have created the second experiment to evaluate and compare both labs from the virtual lab and real life. We set two separate teams Team 1 recorded lab results from the virtual physics lab, and Team 2 recorded lab results from the real-life lab. By doing this, it supports the virtual labs to have the same authenticity and accuracy as people doing the labs in the real lab. First, it proves the results are trustable while doing the labs in a virtual way. Second, it gains feedback from the results in order to determine any mistakes in the calculation or logic implementation. Both Team 1 and Team 2 tested two labs which were Magnetic Drop and Floating Egg, and we compared the results on the same table (Table 2) [10]. By looking at the table, when we compare the results of Magnetic Drop from Team 1 and Team 2, the result looks accurate since the Non-magnetic Sphere takes a shorter time than the Magnetic Sphere. Although the time between Team 1 and Team 2 has a deviation of 0.01-0.02s, it does not affect the result of the experiment where the logic of the lab is still correct. The lab of Floating Egg also has the accurate result where Team 1 and Team 2 both test the condition of an egg in salted water and clean water. The egg is floated with salted water in both virtual and real-life lab, and the egg with clear water is not floated. Comparing the results

from the virtual lab and real-life lab plays an important role to prove the consistency of the results, and it's also a good behavior to check the deviation between the virtual lab and real life. This kind of evaluation could help us to minimize deviation and mistakes while implementing new labs in the future [9].

Table 2. Team 1 vs Team 2

Tasks	Team 1: Virtual Lab	Team 2: Real Life Lab
Magnetic Drop: Non-magnetic Sphere	Time: 0.58s	Time: 0.56s
Magnetic Drop: Magnetic Sphere	Time: 2.37s	Time: 2.36s
Floating Egg: With Salted Water	Egg Floated	Egg Floated
Floating Egg: With Water	Egg Not Floated	Egg Not Floated

5. RELATED WORK

Hamed, G. et al mentioned an important key concept of using the virtual learning lab where students could have a flexible learning environment [11]. For example, students are free to do experiments at home, school, computer lab, etc. The virtual learning environment brings benefits to both time management and flexibility. It is also the main reason why we designed this virtual physics lab to support students doing or testing labs at anyplace and anytime. Therefore, the immersive feeling becomes a necessary demand for students to enjoy learning.

M Erfan et al introduced an Android application of Physics Virtual Lab which provides a mobile platform for students [12]. This idea solved the issue when students do not have the laptop or desktop but they can still do physics labs on smartphones. By learning the idea from this research, it makes me have the action to implement our application as an App in the future in order to provide a wide and convenient user experience. Because the current system is only playable on laptop or desktop, it restricts some students who are willing to do virtual physics labs without a computer.

Nindha Ayu B. et al researched on how virtual physics labs affect students' study by collecting information from students using surveys [13]. It's a really good way to receive feedback from users, and the feedback would be the best information to improve the application in different aspects such as adding a new physics experiment, implementing a new feature, and changing the user interaction. By reading this research, it reminds me the importance of making quizzes for each lab. Since the quizzes section is still in progress, we might need to reconsider the format of the quiz after each lab. For example, the quiz can contain 10 questions, the first 9 questions ask about the knowledge from the lab, and the last question is a feedback question where users are allowed to put any feedback about the lab or any improvements that they recommend to us.

6. CONCLUSIONS

In this application, we have designed a game-based interactive and immersive educational platform for physics lab learning. We have designed the whole lab by using C# with a 3D Game Engine [14]. It provides an immersive feeling like doing the labs in real life. Users are being taken into a virtual physics lab where different experiments are implemented to the application. A virtual character is controlled by users to move around the lab room where each lab has its own desk to do the experiment. Users will first see an overall instruction for that specific lab, and they can start observing the experiment by following the different inputs or commands. Some of the

experiments are supporting different angle observations so that it provides a great experience to do the observation in a special view in order to discover more details of it. Based on the evaluations, the virtual lab provides the same accurate results as real-life, and it is more effective for users who want to use the labs outside of the school or lab room. They can practice and exercise in the labs at any place and any time, and they can refer to the labs while doing the reports or homework. Since the quiz section is still in progress, it might limit users who want to ask questions right after the experiments. Once the quiz section is ready for users, the virtual physics lab will be more comprehensive as a practice tool, and users can check themselves by answering the quiz questions [15]. It not only helps them to recall the images of the lab but also supports them to remember details in mind. In the future, we will keep working hard to implement more labs into the application in order to provide a diverse lab to all grades of students. Since the current application only supports a laptop or desktop, our next plan is to publish the application as a small application as well so that users are more flexible in using the application without computer limitations.

REFERENCES

- [1] Johnson, Paul E. "Associative meaning of concepts in physics." *Journal of Educational Psychology* 55.2 (1964): 84.
- [2] Shiratuddin, Mohd Fairuz, and Walid Thabet. "Utilizing a 3D game engine to develop a virtual design review system." (2011).
- [3] Zitzewitz, Paul W., and Physics Textbook Review Committee. "Physics: Principles and problems." *The Physics Teacher* 37.5 (1999): 292-293.
- [4] Fan, Yuancheng, et al. "Graphene plasmonics: a platform for 2D optics." *Advanced optical materials* 7.3 (2019): 1800537.
- [5] Cheah, Thomas CS, and K-W. Ng. "A practical implementation of a 3D game engine." *International Conference on Computer Graphics, Imaging and Visualization (CGIV'05)*. IEEE, 2005.
- [6] Sero-Guillaume, O. E., et al. "The shape of a magnetic liquid drop." *Journal of Fluid Mechanics* 241 (1992): 215-232.
- [7] Juliani, Arthur, et al. "Unity: A general platform for intelligent agents." *arXiv preprint arXiv:1809.02627* (2018).
- [8] Narloch, Rodger, Calvin P. Garbin, and Kimberly D. Turnage. "Benefits of prelecture quizzes." *Teaching of Psychology* 33.2 (2006): 109-112.
- [9] Lemert, Edwin. "Primary and secondary deviation." *Crime. Critical concepts in sociology* 3 (1951): 603-607.
- [10] Brancher, J. P., and D. Zouaoui. "Equilibrium of a magnetic liquid drop." *Journal of magnetism and Magnetic Materials* 65.2-3 (1987): 311-314.
- [11] Hamed, Ghadeer, and Ahmad Aljanazrah. "The effectiveness if using virtual experiments on students' learning in the general physics lab." (2020).
- [12] Erfan, M., et al. "Utilization of the Android physics virtual lab application to improve understanding of light and optics concepts." *IOP Conference Series: Materials Science and Engineering*. Vol. 1088. No. 1. IOP Publishing, 2021.
- [13] Berlianti, Nindha Ayu, Nur Hayati, and Lina Arifah Fitriyah. "STUDENTS' PERCEPTIONS AND ATTITUDES: IMPLEMENTATION OF VIRTUAL LABORATORY PHYSICS APPLICATIONS (PVL) DURING THE COVID-19 PANDEMIC." *Jurnal Pena Sains* Vol 8.2 (2021).
- [14] Benton, Nick, Luca Cardelli, and Cédric Fournet. "Modern concurrency abstractions for C#." *ACM Transactions on Programming Languages and Systems (TOPLAS)* 26.5 (2004): 769-804.
- [15] Pols, Freek. "A Physics Lab Course in Times of COVID-19." *Electronic Journal for Research in Science & Mathematics Education* 24.2 (2020): 172-178.

AN APPROACH USING MACHINE LEARNING MODEL FOR BREAST CANCER PREDICTION

Fatema Nafa, Enoc Gonzalez and Gurpreet Kaur

Department of Computer Science, Salem State University, Salem, MA, USA

ABSTRACT

Breast cancer is one of the most common diseases that causes the death of several women around the world. So, early detection is required to help decrease breast cancer mortality rates and save the lives of cancer patients. Hence early detection is a significant process to have a healthy lifestyle. Machine learning provides the greatest support to detect breast cancer in the early stage, since it cannot be cured and brings great complications to our health system. In this paper, novel models are generated for prediction of breast cancer using Gaussian Naive Bayes (GNB), Neighbour's Classifier, Support Vector Classifier (SVC) and Decision Tree Classifier (CART). This paper presents a comparative machine learning study based to detect breast cancer by employing four different Machine Learning models. In this paper, experiment analysis carried out on a Wisconsin Breast Cancer dataset to evaluate the performance for the models. The computation of the model is simple; hence enabling an efficient process for prediction. The best overall accuracy for breast cancer detection is achieved equal to 94% using Gaussian Naive Bayes.

KEYWORDS

Machine Learning, Breast Cancer, Representation Learning, Gene Embeddings.

1. INTRODUCTION

Cancer that initiates in the cells of the breasts is called breast cancer, and it occurs more in women and rarely in men. Statistics indicate that breast cancer-related complications are the top causes of death among women, and the significant cases of breast cancer are attributed to a shortage of information. In America, breast cancer is the leading cause of cancer-related deaths, and the mortality rate is relatively high compared to the neighbouring countries in America.

According to [1], about 1 in 8 women in the United States is predicted to develop breast cancer during her lifetime. Statistically, this number is very high, and it shows the importance of studying and analysing the factors associated with breast cancer. Understanding breast cancer and what causes it has helped increase the survival rates, with the deaths associated with it declining.

According to [2], In the United States, breast cancer has become the second leading cause of cancer death in women, only second to lung cancer. The older the person, the higher the risk of getting breast cancer, with women being more likely to develop breast cancer than men. Factors such as family genetics also play a big part in breast cancer. People with close relatives older the person, the higher the risk for breast cancers who have been diagnosed with breast cancer are more likely to develop the disease at some point in their life.

The symptoms and warnings of breast cancer people can experience vary from everyone. What this means for people is that in some cases an individual might not experience what is considered “typical” symptoms, and this makes breast cancer screenings the most important. According to [3], 99% of breast cancer occurs in women, with the lifetime of breast cancer risk for men being 1 in 1000. This means that while women are more likely to get breast cancer, men are also victims of this.

Although science has done a great job at saving so many lives from breast cancer, there’s still much more to do, and detecting breast cancer in very early stages is crucial in order to increase the chances of survival of both women and men affected by the disease.

2. RELATED WORK

There has been a growing demand for machine learning models in the biomedical domain, each model aims at addressing challenges associated with a particular set of factors. The models included but not limited to, clustering, classification, neural networks, and associated rule mining [4]– [7] Many of them show good accuracy for the result. There are two kinds of machine learning models: predictive models and descriptive models [8]. A predictive model is to predict unknown variables of interest and it is applied to supervised learning. On the other hand, descriptive models are used to discover patterns in data, and it is applied to unsupervised learning [9]. In this work supervised learning models have been used for prediction.

Many scientists designate themselves to develop appropriate approaches for the detection of breast cancer. However, different machine learning algorithms have been used with different breast cancer dataset and the result varies based on the algorithm and the dataset used by different researchers. The research associated with the prediction task is outlined in brief as follows. Authors in [10], authors used several machine learning models to predict breast cancer and found that logistic regression performed best. Work by [11] proposed using three machine learning algorithms, Naïve Bayes, random forest and K-Nearest Neighbour for breast cancer prediction. In the result, K-Nearest Neighbour (KNN) has better performance.

Another study [12] tested different machine learning models such as SVM, KNN, DT, Logistic Regression, and Random Forest for breast cancer prediction. The highest effectiveness was determined to be 89 percent for random forest. Nikita Rane et al [13] tested six different machine learning algorithms for breast cancer prediction. Naive Bayes, Random Forest, Artificial Neural Network, Nearest Neighbour, Support Vector Machine, and Decision Tree. They classified the cancers as benign and malignant.

According to [14] authors surveyed machine learning techniques based on the Scopus database. Examples of the techniques are Support Vector Machine, Logistic Regression, and Decision Tree model, it used for breast cancer classification task. Also, ML techniques widely used for developing CAD systems are Decision Tree (DT), Naive Bayes, nearest neighbour, Artificial Neural Network (ANN), Support Vector Machines, and set Classifiers have been used for breast cancer classification.

Using the background and related work, a missing element to existing research was to assess the sensitivity, specificity, and accuracy of the performance and quality of the proposed models. This work presents a comparative study of the efficiency for four classifiers: SVM, Naive Bayes, Neighbour’s, and Decision Tree, which are the most popular machine learning techniques. Also, the optimization of these models has been investigated for breast cancer detection.

3. PROBLEM DEFINITION

The dataset provides the attribute of women and their incidence of breast cancer (diagnosis). Performing exploratory data analysis on the dataset and come up with insights on the factors that cause breast cancer. Also, in this paper we'll analyze the data and evaluate different Machine Learning models to predict whether a specific set of symptoms will be high risk breast cancer or not which is predicting breast cancer at early stage.

4. METHODS

This section will introduce the proposed models to solve the problem. Overview of the system presented in Figure 1.1. The first step that needed to be completed is the preprocessing of the dataset. Different python libraries have been used in order to handle different issues. In this step, exploratory data analysis level has been applied to clean and prepare the data. A dummy variable (diagnosis), handle the missing values, visualize the data, and create heatmaps and a scatter matrix. In additionally, a Multicollinearity has been checked because of one of the assumptions of the proposed model is that there isn't any Perfect multicollinearity. Multicollinearity is where one of the variables is highly correlated with another explanatory variable. To check Multicollinearity, a correlation matrix using the corr() python function used. The function creates a matrix with each variable having its correlation calculated for all the other variables. The visualization for the matrix created using heatmaps. if you travel diagonally down the matrix all the associations should be one, as it is calculating the correlation of the variable with itself. heatmaps can quickly help identify the highly correlated variables, by just looking for the darker colors. So, the correlation between diagnosis other variables was found. More details can be found in the result section.

Second step, another level of in-depth exploration of the data before building the models, in this step the data get explored to see how the data is distributed and if there are any outliers. Third step, build the model, Random Forest Regression model is used. Then 4 machine learning models imported, the best accuracy gained by this model. last step, evaluating the performance of the model using Confusion Matrix. and explored the data we can proceed to the next part, building the model.in additionally, the interpretation of the prediction result needs a discussion to see if the result make sense.

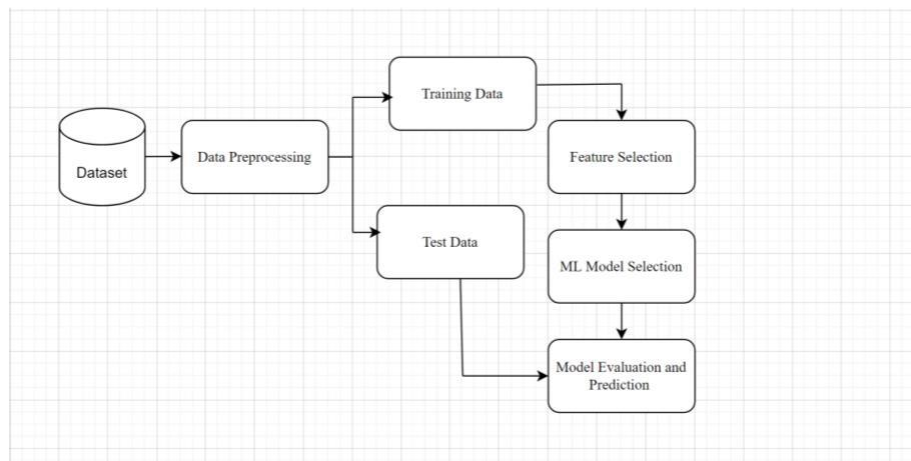


Figure 1. Overview of the proposed Model.

4.1. Model Inputs

Supervised Learning algorithms, it is a machine learning algorithm that learn to anticipate new results based on previous learning by learning patterns from pre-existing data. Existing data is identified using machine learning methods such as probability-based, function-based, rule-based, tree-based, instance-based, and so on.

Naïve bayes is a machine learning model based on based on Bayes' Theorem with an assumption of independence among predictors. In other words, the attributes are independent of each other. Naïve Bayes predicts datasets with the assumption that attributes belonging to a class that is independent of each other. This study uses Gaussian Naïve Bayes algorithm which works well with both continuous and discrete datasets. And Bayes refers to the statistician and philosopher Thomas Bayes theorem [15]. The NB theorem can be expressed mathematically as follows:

$$P(A/B) = (P(B/A) P(A)) / (P(B)) \quad (1)$$

$P(A / B)$: Probability of occurrence of event A given the event B is true.

$P(A)$ and $P(B)$: Probabilities of the occurrence of event A and B respectively.

$P(B / A)$: Probability of the occurrence of event B given the event A is true.

Naïve bayes has been used in different application such as Real time Prediction, Text classification, Spam Filtering and Sentiment Analysis [6], [16], [17].

5. RESULTS

5.1. Dataset

An experiment was conducted on dataset. The dataset is Wisconsin Breast Cancer dataset from UCI Machine

Learning Repository. There are around 30 numeric attributes of features in the dataset. The total size was 156.7 KB. Table 1.1 provides a descriptive statistic to describe and summarize the data. It uses quantitative approach describes and summarizes data numerically. We can observe that the data set contain 569 rows and 31 columns. The missing values are none.

Table 1. Descriptive Statistics about the Dataset.

Dataset statistics		Variable types	
Number of variables	31	Categorical	1
Number of observations	569	Numeric	30
Missing cells	0		
Missing cells (%)	0.0%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	165.7 KB		
Average record size in memory	298.2 B		

we should analyze and “get to know” the dataset. in other words, get familiarize with the dataset, to gain some understanding of the potential features and to see if data cleaning is needed. After the statical information has been presented. The pre-processing step needed to be applied. In this dataset, reordering for some columns has been performed, delete ID column. The diagnosis

feature represents the number of Benign and Malignant cases. It has been replaced with Benign = 0 and 'Malignant = 1 to make all the data numerical type.

Using descriptive method that generates descriptive statistics that summarize the central tendency, dispersion, and shape of a dataset's distribution, excluding NaN values. This method tells us a lot of things about a dataset. Table1.2 shows the statistics that are generated by the describe () method: count tells us the number of non-empty rows in a feature. mean tells us the mean value of that feature. std tells us the Standard Deviation Value of that feature. min tells us the minimum value of that feature. 25%, 50%, and 75% are the percentile/quartile of each feature. This quartile information Max tells the maximum value of that feature Correlation is a statistical technique that can show whether and how strongly pairs of variables are related/interdependent [18], [19]. Figure 3. show Pearson correlation coefficient. Pearson correlation coefficient is a measure of the strength linear association between two variables. Looking at the heatmap along with the correlation matrix we can identify a few highly correlated variables as in Figure 3. This is an extremely high correlation and marks it as a candidate to be removed. Logically it makes sense that these two are highly correlated.

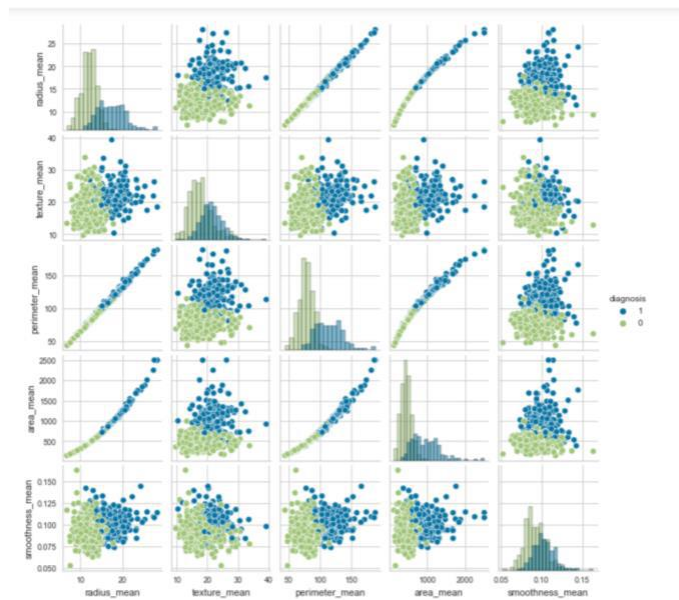


Figure 2. Correlation between Variables.

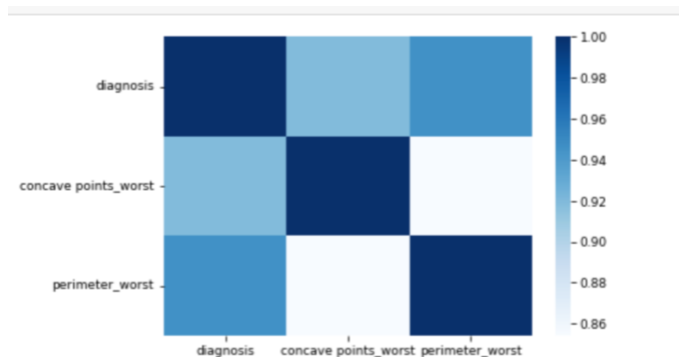


Figure 3. Correlation between concave points worst, perimeter worst and diagnosis.

5.2. Performance Metrics/Confusion Matrix

The performance metrics of the classification model were calculated based on precision, recall, and accuracy and are presented in Table 2. TP and TN specify the numbers of diabetes and normal patients that were correctly classified, respectively, while FN and FP specify the numbers of normal and diabetes patients that were incorrectly classified, respectively. 10-fold cross-validation was used to train and test the dataset for the entire classification model.

6. CONCLUSION AND DISCUSSION

As a result, Gaussian Naive Bayes (NB) has been used effectively to predict whether a patient will have breastcancer based on some features. The model has achieved an accuracy of 93% as shown in Table 2. The most important features are the one with high values. Based on the result, it looks that variable6 level has the most significant influence on the model. the second highest value is variable9 and the last one is variable8. The three features have a positive influence on the prediction their higher values are correlated with person being diabetes. I noticed a lot of information needed to be discussed with medical expertise to make sure that it is correct.

Table 2. The Accuracy for the Prediction model.

Algorithm	Recall	precision	F1-Score	Accuracy	Run Time
DecisionTreeClassifier	90.00%	99.00%	97.00%	92.00%	0.059s
Support Vector Machine	87.00%	99.00%	90.00%	92.00%	0.056s
Gaussian Naive Bayes (NB).	93%	100%	90%	94.00%	0.016s
KNeighborsClassifier	91.00%	91.00%	92.00%	91.00%	0.028s

REFERENCES

- [1] "Breast Cancer Statistics | How Common Is Breast Cancer?" <https://www.cancer.org/cancer/breast-cancer/about/how-common-isbreast-cancer.html> (accessed May 13, 2022).
- [2] J. A. Ajani et al., "Gastric cancer, version 2.2022, NCCN clinical practice guidelines in oncology," *J. Natl. Compr. Canc. Netw.*, vol. 20, no. 2, pp. 167–192, 2022.
- [3] P. A. McElfish et al., "Diabetes and hypertension in Marshallese adults: results from faith-based health screenings," *J. Racial Ethn. Health Disparities*, vol. 4, no. 6, pp. 1042–1050, 2017.
- [4] R. Khan, Y. Qian, and S. Naeem, "Extractive based Text Summarization Using K-Means and TF-IDF.," *Int. J. Inf. Eng. Electron. Bus.*, vol. 11, no. 3, 2019.
- [5] M. A. Ibrahim, M. U. G. Khan, F. Mehmood, M. N. Asim, and W. Mahmood, "GHS-NET a generic hybridized shallow neural network for multi-label biomedical text classification," *J. Biomed. Inform.*, vol. 116, p. 103699, 2021.
- [6] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: A comprehensive survey," *Expert Syst. Appl.*, vol. 165, p. 113679, 2021.
- [7] N. Alami, M. Meknassi, N. En-nahnahi, Y. El Adlouni, and O. Ammor, "Unsupervised neural networks for automatic Arabic text summarization using document clustering and topic modeling," *Expert Syst. Appl.*, vol. 172, p. 114652, 2021.
- [8] M. Kantardzic, *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons, 2011.
- [9] B. Lantz, *Machine learning with R: expert techniques for predictive modeling*. Packt publishing ltd, 2019.
- [10] T. A. Assegie, "An optimized K-Nearest Neighbor based breast cancer detection," *J. Robot. Control JRC*, vol. 2, no. 3, pp. 115–118, 2021.
- [11] T. A. Assegie, "An optimized K-Nearest Neighbor based breast cancer detection," *J. Robot. Control JRC*, vol. 2, no. 3, pp. 115–118, 2021.

- [12] R. Rawal, "Breast cancer prediction using machine learning," *J. Emerg. Technol. Innov. Res. JETIR*, vol. 13, no. 24, p. 7, 2020.
- [13] R. Hazra, M. Banerjee, and L. Badia, "Machine learning for breast cancer classification with ann and decision tree," in *2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2020, pp. 0522–0527.
- [14] E. H. Houssein, M. M. Emam, A. A. Ali, and P. N. Suganthan, "Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review," *Expert Syst. Appl.*, vol. 167, p. 114161, 2021.
- [15] T. Bayes and D. Hume, "BAYES'S THEOREM," in *Proceedings of the British Academy*, 1763, vol. 113, pp. 91–109.
- [16] E. Ezpeleta, U. Zurutuza, and J. M. Gómez Hidalgo, "Does sentiment analysis help in bayesian spam filtering?," in *International Conference on Hybrid Artificial Intelligence Systems*, 2016, pp. 79–90.
- [17] R. Mallik and A. K. Sahoo, "A novel approach to spam filtering using semantic based naive bayesian classifier in text analytics," in *Emerging technologies in data mining and information security*, Springer, 2019, pp. 301–309.
- [18] M. Ezekiel and K. A. Fox, "Methods of correlation and regression analysis: linear and curvilinear," 1959. [19] S. L. Crawford, "Correlation and regression," *Circulation*, vol. 114, no. 19, pp. 2083–2088, 2006.

CONVERTING REAL HUMAN AVATAR TO CARTOON AVATAR USING CYCLEGAN

Wenxin Tian

Graduate School of Information Sciences and Arts,
Toyo University, Kawagoe, Saitama, Japan
Dept. of Information Sciences and Arts,
Toyo University, Kawagoe, Saitama, Japan

ABSTRACT

Cartoons are an important art style, which not only has a unique drawing effect but also reflects the character itself, which is gradually loved by people. With the development of image processing technology, people's research on image research is no longer limited to image recognition, target detection, and tracking, but also images. In this paper, we use deep learning based image processing to generate cartoon caricatures of human faces. Therefore, this paper investigates the use of deep learning-based methods to learn face features and convert image styles while preserving the original content features, to automatically generate natural cartoon avatars. In this paper, we study a face cartoon generation method based on content invariance. In the task of image style conversion, the content is fused with different style features based on the invariance of content information, to achieve the style conversion.

KEYWORDS

Deep learning, CNN, Style transfer, Cartoon style.

1. INTRODUCTION

Cartoon faces appear in animations, comics, and games. They are widely used as profile pictures in social media platforms, such as Facebook and Instagram. Drawing a cartoon face is labor intensive. Not only it requires professional skills, but also it is difficult to resemble unique appearance of each person's face. Through style transfer, which can express the picture effect more perfect and be able to achieve the desired effect. It can be done without complex PS retouching and does not require particularly good drawing skills to complete the corresponding task. In the film production or webcast, it can make the image performance more involved, more vivid image special effects to do more abstract perfection. Image stylization originated from the research of Gatys and others, who found that although today's style migration has achieved good results, there are still some areas for improvement. The first thing to solve is the time consuming problem, even if you choose the optimal solution, it takes a long time to train a model, obviously there is still a lot of room for improvement, and there is still a lot of room for optimizing the time problem for selfie images.

In recent years some social networking services have been popular such as TikTok. Photo-to-cartoon style transfer for face can be useful for the services especially when the users do not want to show their own faces. And due to COVID-19, many schools have adopted online classes to prevent the expansion of the infection. Teachers want to know how well their students understand, what they learned and how well the students focus on. What the teachers said from nonverbal

information such as facial expression, facial pose, eye-gaze, etc. On the other hand many students do not want to show their faces. In this case photo-to-cartoon style transfer can be useful, because it can keep facial expression, facial pose, and eye-gaze, while it converts real photo style into cartoon style. The purpose of this paper is to construct a model which can convert real face images into cartoon face images using deep neural networks.

2. BACKGROUND AND SIGNIFICANCE

In today's hot deep learning, image processing based on deep learning has also been more researched, such as super-resolution reconstruction, repair of missing images, colorization of black and white images, and AI face replacement. In addition to the above applications of deep learning in images, research on image stylization has also emerged in recent years, with the main idea of turning a common image into an artistic style painting. Gatys et al. used convolutional neural networks to synthesize the texture of image features and found that the feature image extracted by convolutional neural networks can show the style characteristics of an image and also the content characteristics of an image, and the feature map is a representation of the deeper features of an image. The idea proposed by Gatys et al. is to input the image into the convolutional neural network, reconstruct the image with different convolutional layers, extract the features of different dimensions of the image with different convolutional layers, calculate the style features with the Gram matrix, simulate the texture features of different style maps, and reconstruct the final result by fusing the content map and style map to create a beautiful painting with the content and artistic style of the photograph. The final result is created by fusing the content map and the style map to create a beautiful painting with photo content and artistic style.

Nowadays, image stylized migration is also used in various fields, such as video broadcasting, movie special effects, etc. Stylized photos are often sought after and loved by young people on various social networks. Although the stylized migration of images has made a good effect in some fields, most of the artistic style migration nowadays is in oil paintings with obvious textures, and line images such as cartoons and sketches are seldom involved.

The influence of manga in our lives is huge, for example, the Japanese classic anime such as Black Deacon and Cherry Puff; the American Disney and Marvel anime are still loved by countless fans all over the world. The other world created by anime and manga is a place that many people dream of, and reading comics can relax their mood, improve their imagination, find their ideals and trust in comics, and see the hope for the future. After the above analysis, I think it is still necessary to continue to improve the existing research and apply the neuro art algorithm to the creation of comic style rendering.

3. CURRENT STATUS OF RESEARCH ON THE TOPIC

Neural network based image style migration was proposed in 2015 by Gatys et al. in two papers. In the literature of Gatys et al. in 2015, a texture model based on the feature space of convolutional neural network is proposed, in this model, the texture is represented according to the interrelationship between the feature maps in each layer of the network, and the texture extraction increasingly captures the stylistic content features of natural images while making the object information more and more clear. As shown in the figure, the original image is input to the neural network on the left, and the texture analysis is obtained by extracting features in different convolutional layers and calculating the Gram matrix, while the white noise image is input on the right, and the loss function of the texture in different layers is calculated to synthesize the texture. In another paper, the texture synthesis method of the first paper is used to perform the image style migration in the oil painting style. The authors use the intermediate neural network layers to

reconstruct the content, keep the largest pixel value part, and then fuse it with the extracted texture image to get the final image containing different oil painting art styles.

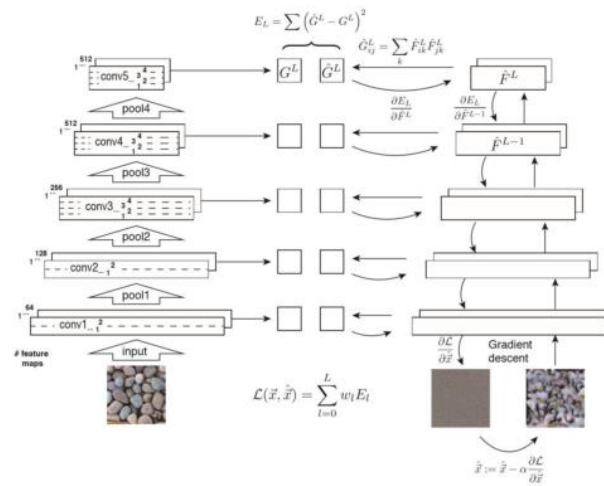


Figure 1. Image texture extraction based on CNN

Li used traditional methods for the artistic rendering of images and videos. Ulyanov D et al. used a feed-forward convolutional network to generate multiple samples of the same texture of arbitrary size and transfer the artistic style from a given image to any other image, the network was hundreds of times faster than the results of Gatys et al. The main component of the style conversion method is a block-matching-based operation for fixing layers, constructing target activities in a given style and content image, a process called "Style-swap" by Chen et al. By replacing the content image with a patch of the style image, this method performs the relevant processing on one layer only. The core of the paper is the proposal of AdaIN, a new adaptive instance normalization that aligns the mean and variance of content features with the mean and variance of style features. Liao et al. proposed a deep image analogy technique in which two images, image A and image B, are given in the paper. The two images are mutual content map and style map, and the conversion needs to result in images with mutual style in their original content, the two images of this method must have high similarity, or use the network of VGG for the extraction of features of the image, the high-level convolution will extract to the texture of the image, so the paper believes that A and A^B after the extraction of features by VGG19, the coarse-grained output of the top convolutional layer of feature maps should be very similar, i.e., the feature maps of A and A^B are basically the same. If they are considered the same, we can reconstruct the content of A^B by deconvolution of the top layer feature map of A, and then fuse it with the features of image B to get the final image A^B , while image B^A is the same. A new project by Chinese students from Cornell University and engineers from Adobe is based on image style migration for high-definition image style migration, which is more inclined to style migration between two photos, detail and clarity are the characteristics of this article, the input of the style image is a high quality photo, the result can change day to night, is a high-definition photo to high-definition photo Different styles of conversion, using to the style image is no longer an artistic painting. The team introduced two main core innovations in the style migration. In the optimization process, a set of realistic regularized loss functions are added to the loss function to prevent distortion of the generated images. The paper introduces semantic segmentation by transferring the style features of the lawn to the lawn, the style features of the sky to the sky, etc., which can avoid the mismatch of the content of the style migration and make the output image more realistic.

From the endless image processing work in recent years, people have more and more research and ideas in image processing, and there are new breakthroughs in image stylization, image recognition, video stylization and so on. Especially in the area of image style migration, deep learning has been hot in recent years and has brought more new ideas and better results for image processing. Nowadays, people's standard of living is getting higher and higher, and the pursuit of art is also getting higher and higher, so the image stylization algorithm based on deep learning has become an area worth exploring more and more, and there is a high possibility that more excellent image stylization results can be achieved by deep learning and other related algorithms in the future.

4. RELATED RESEARCH

4.1. Unpaired Image-to-Image translation

When paired images such as a real face image and the corresponding cartoon face image are needed to train photo-to-cartoon style transfer model, obtaining strictly paired images is very difficult. So the model that do not rely on paired images are of great practical importance.

There is a gap between paired and unpaired picture training that cannot be eliminated. Nevertheless, in many cases, it is still feasible enough to use unpaired data exclusively. Zhu et al. [2] expand the range of possible uses of "unsupervised" configurations. To some extent, it solves the problem of deep learning: too little labeled data, difficulty in finding paired data, and using unpaired datasets for training.

They proposed a method to learn from the source data domain X to the target data domain Y in the absence of paired data. Its goal is to use an adversarial loss function to learn the mapping $G: X \rightarrow Y$, making it difficult for the discriminator to distinguish the picture $G(X)$ from the picture Y . Since such a mapping is subject to huge limitations, an opposite mapping $F: Y \rightarrow X$ is added to the mapping G to make them pairwise, and a cyclic consistency loss function is added to ensure that $F(G(X)) \approx X$.

4.2. Landmark assisted CycleGAN

Wu et al. [3] proposed a method to generate cartoon faces based on input human faces by utilizing unpaired training data.

The process is divided into three main steps. First, the generator generates a rough cartoon face based on CycleGAN; afterwards, the model generates a pre-trained regression volume to predict the facial landmark based on the image generated in the first step, which marks the key points of the face. Finally, with both local and global discriminators, the researchers refine the face features in the cartoon image and the corresponding real image. In this stage, the consistency of the landmark is emphasized, so the final generated results are realistic and recognizable. Consequently, landmark Assisted CycleGAN is proposed to define consistency loss using facial landmark features to guide the training of local discriminators in CycleGAN.

4.3. Unpaired Photo-to-Caricature translation

Cao et al. [4] proposed a learning-based approach to solve the conversion from ordinary photographs to cartoons. A two-way model with a coarse-distinctive and a fine-distinctive discriminator is designed in order to take into account both local statistics and global structure during the conversion. For the generator, perceptual loss, adversarial loss, and consistency loss

are utilized to achieve representation to learn in two different domains. Moreover, an auxiliary noise input can be used to understand the style.

It also presents a generative adversarial network (GAN) for photo-to-comic transformation without paired training datasets: the CariGAN. It uses two modules to explicitly model geometric exaggeration and appearance stylization, one is CariGeoGAN the other is CariStyGAN. In this way, a difficult cross-domain transformation problem is decomposed into two simpler tasks. Compared to advanced methods, CariGAN generates caricatures that are closer to hand-drawn while better maintaining the personality characteristics of the original face. In addition, the user is allowed to control the degree of exaggeration and variation of the shapes, or to give example caricatures to generate the corresponding styles.

4.4. Cartoon adversarial generation network

Researchers at Tsinghua University have proposed CartoonGAN [5], a comic style-based generative adversarial network that can train a comic style migration model. Previous image styling algorithms based on generative adversarial networks often require two sets of corresponding images to be trained to obtain better results, such as CycleGAN, which also makes the training data difficult to obtain. The paper proposes a GAN network architecture dedicated to cartoon style migration and two simple and effective loss functions.

The cartoon generative adversarial network mainly proposes a cartoon stylized framework for generative adversarial networks, which directly trains the captured images with the cartoon images without matching them one by one and is easy to use. And the authors propose two kinds of losses, one is semantic content loss, which is a sparse regularization constructed in the high-level feature graph of VGG network to cope with the large number of style variations between photos and cartoons; the other is an edge-promoting adversarial loss to maintain clear edges. To improve the convergence of the network to the target, an initialization phase is further introduced in this paper. The GAN framework consists of two CNNs: a generator, which is trained to produce outputs that make the discriminator indistinguishable, and a discriminator D, which classifies whether the image is from a real target or a synthetic one. The generator uses a convolutional layer for downsampling, follows a layout of eight residual blocks, and finally upsamples the image by microstrip convolution. The discriminator network D is used to determine whether the input image is a real cartoon image or not. The discriminator network has shallow layers, in fact, the discriminator network is a classification network, the discriminator mainly discriminates whether there are obvious boundary lines, the structure is two convolutional layers for downsampling, and then the convolutional layers return the classification results.

5. RESEARCH METHOD

5.1. Convolutional neural networks

Convolutional neural networks are the most widely used of all kinds of deep neural networks and have achieved good results in many problems of machine vision, in addition to its successful applications in natural language processing, computer graphics, and other fields. In 1989, LeCun [6] proposed a convolutional neural network that is quite efficient for handwritten character recognition, which is also the origin of many convolutional neural networks nowadays.

After nearly two decades of neural network coldness, the AlexNet network was proposed in 2012, which won the ImageNet competition at that time. the parameter scale of the AlexNet network became larger, the convolutional layers of the network became deeper, with a total of five

convolutional layers, and the maximum pooling layer was added to the first, second, and fifth convolutional layers to reduce the computation, and finally, the fully connected The network divides the convolutional layers into two parallel networks, which can effectively reduce the computation and improve the computational efficiency.

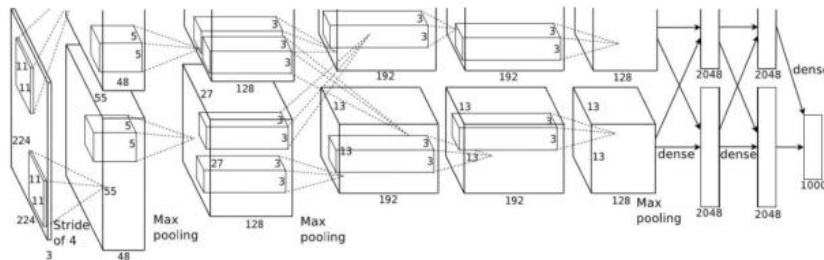


Figure 2. AlexNet Network Structure Diagram

GoogLeNet was proposed two years after AlexNet, the key in this network is the Inception block, that is, the input image is extracted with different scales of features, this mechanism change reduces the number of parameters to one twelfth of AlexNet. The network integrates multi-scale convolutional kernels and pooling layers, which effectively reduces network parameters, prevents overfitting and reduces computational effort to improve efficiency.

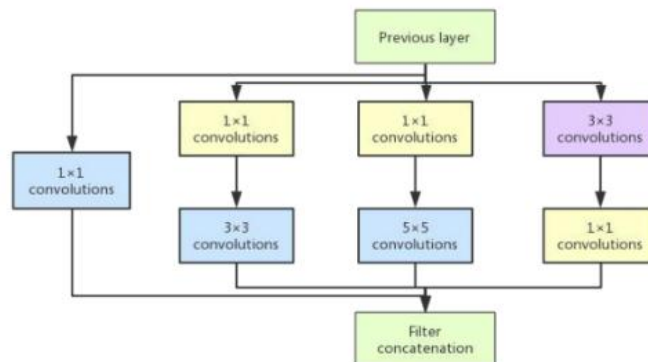


Figure 3. GoogLeNet Inception Modular

ResNet residual neural network was proposed by four Chinese including KaimingHe from Microsoft Research [9], the network structure of ResNet can pass down all the previous parameters, and the accuracy of the network is greatly improved. However, when the number of network layers is increased to 1202, the result decreases due to the overfitting caused by the deep number of network layers. The main core of the residual network is to pass the information of each layer directly to the output, which will not lead to degradation problems due to the increase of the number of network layers, and the accuracy rate can be increased based on the increase of the number of network layers to ensure the integrity of the extracted features.

5.2. CycleGAN

The field of image transfer is the domain of GAN networks, and recently many people have applied CycleGAN networks to the field of image style transfer.

Figure 4 shows structure of general GAN which is composed of generator G and discriminator D. The generator G generates data $G(z)$ from a random input z , and makes the generated data as close to the real data as possible. On the other hand the discriminator D tries to distinguish the real data from the generated data $G(z)$ as much as possible. The two networks are always playing a game, in which G gradually gains the upper hand and the generated data is no different from the real data.



Figure 4. General GAN

The goal is to realize the data migration of two domains, to realize the translation between images with the help of GAN, as shown in Figure 5. There should be two discriminators of domains, and each discriminator will judge separately whether the data of their respective domains are real data. As for the generator, the image translation needs to turn the image of domain A into the image of domain B. Therefore the generator is somewhat like the autoencoder structure, except that the output of the decoder is not the image of domain A, but the image of domain B. To make full use of the two discriminators, there should also be a translation back, which means there is another generator that translates the data from domain B to domain A.

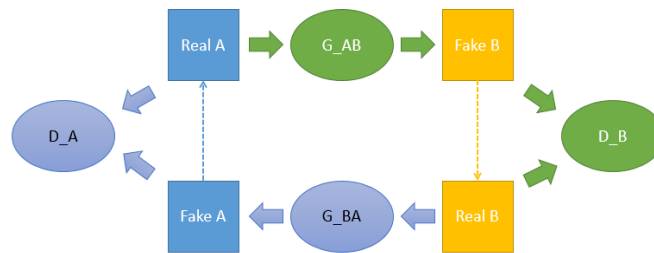


Figure 5. Data migration of two domains

The dashed arrow in Figure 5 indicates ‘treat the image at the beginning of the arrow as the image at the end of the arrow and continue according to the flow chart’. It means that for Real A, the complete process is like this: $A_{real} \rightarrow B_{fake} \rightarrow A_{fake}$; for Real B, the process is like this: $B_{real} \rightarrow A_{fake} \rightarrow B_{fake}$. It can be seen that the whole process is a cycle for both domain A and domain B images, so it is called CycleGAN. The whole cycle can be seen as an autoencoder, the two generators are seen as encoder and decoder, and the two discriminators are criteria.

In general, the two generators are designed in such a way like:

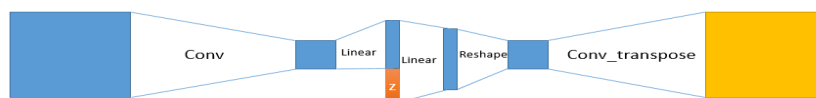


Figure 6. The two generators

Z controls some properties in G such that the generated results are not unique and can be diverse [12]. The process of CycleGAN is clarified to write its objective function as follows.

For discriminator A: $L_{D_A} = E_{x \in P_A} \log D_A(x) + E_{x \in P_{B2A}} \log(1 - D_A(x))$

For discriminator B: $L_{D_B} = E_{x \in P_B} \log D_B(x) + E_{x \in P_{A2B}} \log(1 - D_B(x))$

For generator BA: $L_{G_{BA}} = E_{x \in P_{B2A}} \log D_A(x) + \lambda E_{x \in P_A} \|x - G_{BA}(G_{AB}(x))\|_1$

For generator AB: $L_{G_{AB}} = E_{x \in P_{A2B}} \log D_B(x) + \lambda E_{x \in P_B} \|x - G_{AB}(G_{BA}(x))\|_1$

Adding refactoring error terms for generators, like pairwise learning, can guide the two generators to better perform the task of encoding and decoding. In turn, the two Ds serve to correct the encoding result to conform to the style of a certain domain. Not only does the structure is simple and effective but also the data of the pair is not required. Cycle consistency loss has been proposed which makes generic unpaired image-to-image translation possible. Given only two domains of image collections, CycleGAN can explore the collection-level supervised information and realize image transfer.

5.3. Generative networks based on content invariance

This paper studies the method of generating face cartoon drawings in the absence of paired experimental data. In the absence of paired data, the content features of the images need to be constrained by indirect means, and a cyclic generative adversarial network was first proposed in the literature [14] to achieve the interconversion of the two styles, which ensures the invariance of the content by reconstructing the generated image out of the original image. The same principle is used in the literature [15] and [16], with the difference that the encoding network is divided into a style encoder and a content encoder, which encodes the image to be converted and the style image, and then performs the fusion to go through a decoding network of a specific style.

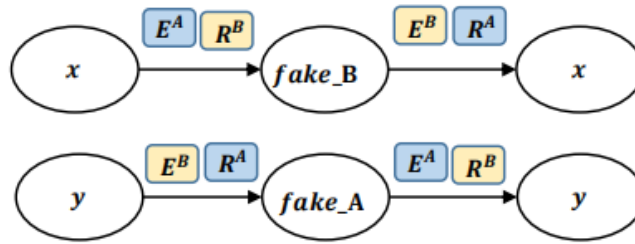


Figure 7. Network Structure Comparison [14]

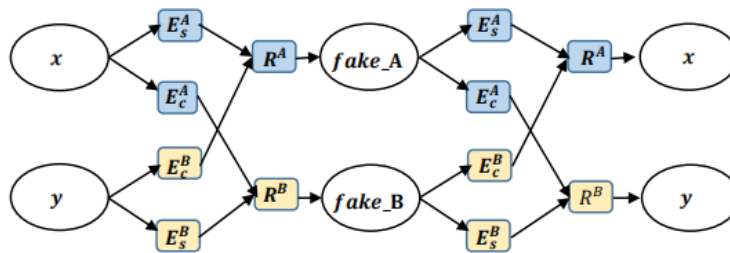


Figure 8. Network Structure Comparison [16]

The above figure shows the network structure of the literature [14] and the literature [16], and the encoders or decoders used by both methods are designed for different styles. The two styles are noted as style A and style B; $x \in A, y \in B$. where the network used for the different styles is distinguished by two colors in the figure, blue indicates encoding or decoding of images or features of style A, and yellow indicates encoding or decoding of images of style B. The encoder is denoted as E and the decoder is denoted as R, and the style is distinguished by superscripts A and B. The content and style encoders are distinguished by subscripts c and s.

The method in Figure (a) directly converts the input image to the corresponding style image by using different style encoders and decoders.

Since the style conversion is essentially a combination of the content information of the image to be converted and the style information of the reference image. Therefore, the method in Figure (b) extracts the content x and the y for the two input images x and y , respectively style features, and then the target images are obtained by the decoding network of corresponding styles. Both methods above restore the input image x or y by encoding and decoding the generated image, a process called image reconstruction.

In the style conversion task, the purpose of image reconstruction is to ensure the content information of the original image, while the confrontation is to make the generated image with a specific style, a certain balance needs to be maintained between the two, if the content information is protected more, the network will choose to ignore the style information; on the contrary, the content of the generated image will not be guaranteed.

However, training the interconversion of the two styles requires facing a problem in that the network needs to build additional models. for different styles of encoding and decoding networks and discriminative networks, which means that additional memory is occupied during the training phase space in the training phase. In the testing phase, only the A to B conversion needs to be implemented, and the extra models should be minimized. This study aims to find a method that can be trained in only one direction to achieve the generation of face cartoons, so it is necessary to find another reconstruction method that can be used to guarantee the invariance of the content and at the same time be able to reach a balanced state with the styles.

Derived from the idea of the literature [16], the desired content features and style features can be extracted by constructing a content encoder and a style encoder separately, and to make the encoding network general, the same network is used for the extraction of content or style features for both style images. The two main reasons are as follows.

Firstly the so-called content features mainly include the shape structure information of the face, for the extraction of the image content features should not be affected by the image style, i.e. the content coding network should be general for all face images. Whether it is a real face image or a style image, the content features should be able to be extracted correctly.

Secondly, in the style migration task, style can be used as an attribute, and for the same content feature, based on different style attributes, it should be possible to obtain images with different styles and keep the content unchanged. Further, in the style conversion, the style features are used as the condition of conversion, and the style features are fused with the content features, and then the decoder is used to get images of different styles. Although using different style encoders for different styles of images can get more style features, it increases the complexity of the network. In deep networks, the classification of images is performed by extracting images with classification is performed by extracting discriminative features from images, and for style determination, it can also be The discriminable features are extracted from the image to

determine the style of the image, and such discriminable features can be used as the style features of the image. can be used as the style feature of the image.

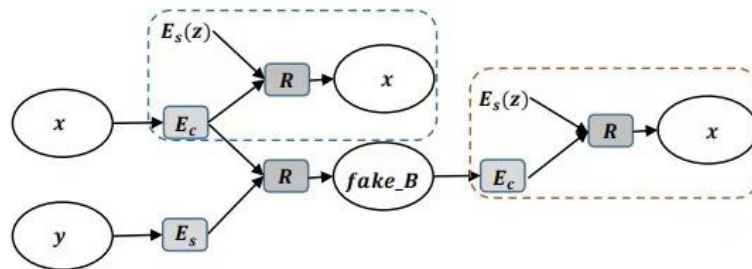


Figure 9. Generate network structure

Since the network only needs to be trained to convert from style A to style B, the reconstruction of the image also only needs to reconstruct the input image. To ensure content invariance, the content encoding of the generated image should be as similar as possible to the content encoding of the input image, so the content encoding is performed on the generated image, and then the style features of the real face are fused to recreate the original image, i.e., the orange dashed box part in the figure. The reconstructions all use the style features of the real face, and theoretically, the style features of different images of the same style should be as similar as possible. Therefore, during the training process, for each input image to be transformed, a random image of the same style is input to extract its style features, the so-called feature input for reconstruction.

6. EXPERIMENT

I have already trained a generative adversarial network to generate cartoon faces after showing pictures of many real faces. Most of them here are from the CycleGAN implementation in PyTorch.

I use the CelebFaces Attributes Dataset as training data, which is a large-scale face attributes dataset with more than 200K celebrity images, each with 40 attribute annotations. The images in this dataset cover large pose variations and background clutter, which means CelebA has large diversities, large quantities, and rich annotations.

I have done several experiments to change the transformation effect of the images by changing the parameters, here are a few examples. In this time I changed the learning rate from 0.0002 to 0.0016:

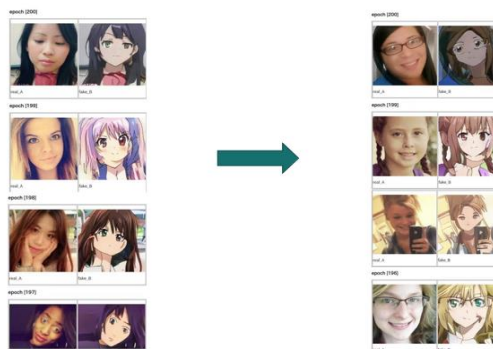


Figure 10. Experiment1

This time I changed the lambda identity from 0.5 to 2.5:

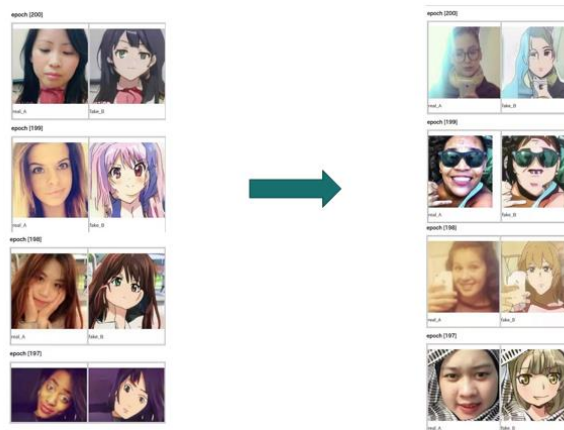


Figure 11. Experiment2

7. CONCLUSIONS

The influence of cartoons in life is enormous, such as Japan's Black Deacon, Cherry Puff and other classic anime; the U.S. Disney, etc. is still being loved by countless fans around the world. Since the cartoon style is different from the oil painting style in style transfer, the cartoon needs a neat and clear border to avoid a large number of uneven color blocks.

Style transfer algorithms are increasingly studied. I have studied the common neural networks of deep learning, learned the principles and practice for the basic style transfer algorithm to lay the foundation for the next step of research. In addition, I read some papers about style transfer, studied and analyzed the cartoon translation style, to come up with a suitable model for cartoon style transfer. Based on the existing model, different parameters were experimentally analyzed to select the parameters suitable for cartoon style transfer, the dataset mentioned in this paper was used for training, which means the deep learning based cartoon style transfer algorithm was implemented. It is possible to generate cartoon style avatars with clear lines and simple character features, the results of different parameters are compared, with the aim of improving the algorithm and making the generated cartoon style avatar better.

At present, there is some progress in cartoon style transfer, but there is still a difference for real cartoon avatars, the details of the transfer are not very good, and the details of the cartoon style are still lacking. The next work will focus on the following aspects: first of all, more research on the details to achieve a more realistic cartoon style transfer effect, secondly adjust the parameters to make the characters retain more features. Hopefully, after the details are improved, it can be applied to daily life, saving more time and creating more value.

ACKNOWLEDGEMENTS

At this point in the writing, although there is reluctance, also had to say goodbye to the past. When I look back, I realize that what I want to say most is I would like to thank you. First of all, I would like to thank my supervisor, Prof. Murakami Since I entered his lab as a graduate student, I have benefited a lot from his rigorous and serious research attitude and practical research style; I thank Prof. Murakami for his guidance and help in my research so that I can overcome one

problem after another; I thank Prof. Murakami for his tireless teaching so that I always keep a positive attitude and try my best to accomplish everything.

Thank you to my labmates. During my graduate studies, my classmates provided me with a lot of help, and I learned from them the excellent qualities of aggressiveness, diligence, optimism, and perseverance. Finally, I would like to thank my family for their support, whether, in study or in life, the warmth of my family is the motivation for my efforts, and I will live actively to repay their love.

REFERENCES

- [1] Leon A. Gatys, Alexander S. Ecker, Matthias Bethge; Image Style Transfer Using Convolutional Neural Networks; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2414-2423
- [2] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros; Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks; Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2223-2232
- [3] Ruizheng Wu, Xiaodong Gu, Xin Tao, Xiaoyong Shen, Yu-Wing Tai, J iaya Jia; Landmark Assisted CycleGAN for Cartoon Face Generation
- [4] Kaidi Cao, Jing Liao, Lu Yuan; CariGANs:Unpaired Photo-to-Caricature Translation; ACM Transactions on Graphics, Vol. 37, No. 6, Article 244. Publication date: November 2018
- [5] Chen Yang, Lai YuKun, Liu YongJin. CartoonGAN: Generative Adversarial Networks for Photo Cartoonization [C], IEEE Computer Society Conference on Computer Vision and Pattern Recognition.Salt Lake City:IEEE Computer Society,2018:9465 - 9474
- [6] Furlanetto R W, Underwood L E, Van Wyk J J, et al. Estimation of somatomedin-C levels in normals and patients with pituitary disease by radioimmunoassay[J]. Journal of Clinical Investigation,1977, 90(3):648-657.
- [7] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C], Advances in Neural Information Processing Systems, Lake Tahoe:Neural information processing systems foundation,2012:1097-1105
- [8] Szegedy c, Liu W, Jia Y, et al. Going deeper with convolutions[C], IEEE Conference on Computer Vision and Pattern Recognition, Boston:IEEE Computer Society,2015:1-9
- [9] He K, Zhang x, Ren S, et al. Deep Residual Learning for Image Recognition[C], IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas:IEEE Computer Society, 2016:770-778.
- [10] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large- -Scale Image Recognition[J]. Computer Science, 2014.5(2):230-237.
- [11] Goodfellow I J,Pouget-Abadie J,Mirza M,et al.Generative Confersarial Nets[C], International Conference on Neural Infonmation Processing Systems, 2014:2672-2680.
- [12] Zhu J Y, Park T, Isola P, et al. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks[J]. arXiv preprint arXiv:1703.10593, 2017.
- [13] Gao H, Zhuang L, Laurens M, et al. Densely Connected Convolutional Networks[C], IEEE Conference on Computer Vision and Pattern Recognition,Hawaii:Institute of Electrical and Electronics Engineers Inc.,2019:2261 - 2269
- [14] J. Y. Zhu, T. Park, P. Isola, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]. IEEE International Conference on Computer Vision. 2017: 2223-2232.
- [15] Y. Lu, Y. W. Tai, C. K. Tang. Attribute-Guided Face Generation Using Conditional CycleGAN[C]. European Conference on Computer Vision (ECCV). 2018: 282-297.
- [16] H. Kazemi, F. Taherkhani, N. M. Nasrabadi. Unsupervised Facial Geometry Learning for Sketch to Photo Synthesis[C]. 2018 International Conference of the Biometrics Special Interest Group (BIOSIG). IEEE, 2018: 1-5.rs.

AUTHOR

My name is Wenxin Tian. I'm currently studying for my master's degree at the Graduate School of Information Sciences and Arts, Toyo University.



© 2022 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.

AN INTERACTIVE AND SCENARIO- BASED SIMULATION GAMING SYSTEM FOR BUSINESS EDUCATION USING GAME ENGINE AND MACHINE LEARNING

Yingzhi Ma¹ and Yu Sun²

¹Crean Lutheran High School, 12500 Sand Canyon Ave, Irvine, CA 92618

²California State Polytechnic University, Pomona,
CA, 91768, Irvine, CA 92620

ABSTRACT

Technology has become increasingly vital in society. The COVID-19 pandemic demonstrated how useful technology was in keeping society running, especially education [15]. One major trend is the use of simulations as a tool for education. Business is one of the fields that could benefit massively from the implementation of new educational simulations. For this study, a survey was conducted to gauge their prior educational experience and interest in fields such as business and computer science. Additionally, the survey participants were questioned on their previous experiences with using interactive simulations. The study had fifty-one participants both complete the survey and give consent to have their data shared in this research paper. These participants were given an additional survey to either test a business simulation or watch a video of one and respond whether they learned from it. The results indicate that although most people would want to play a game that taught more about business, only roughly 45% of individuals expressed interest in the topic of business. Furthermore, the survey responses also indicated that a large majority of individuals would also prefer more interactive educational simulations for other topics. The reception to the business simulation was mostly positive, and participants indicated that it was effective at helping them learn business. Overall, it was concluded that there is not enough access to business simulations to meet the public's interest, and either more should be created or existing ones should be made better known.

KEYWORDS

Machine Learning, Game development, business education.

1. INTRODUCTION

In a world that is becoming more reliant on technology, schools and other institutions of education are transforming technology into effective learning tools. Computer simulations have been implemented into education to stimulate class discussions and enhance student learning (Flake 55) [2]. When testing the effectiveness of computer simulations in topics such as evolution (in Biology), results indicate that the simulations have a positive effect on students' learning and understanding of the concepts (Hongsermeier et al. 130) [1]. In fields such as nursing science, simulations can provide students with a virtual environment that is similar to their expected working environment and help them gain a more vivid experience. Another upside of computer simulations is that they allow students to learn in more fun and engaging ways (Koukourikos et al. 16-17) [3].

Simulations are primarily used as a supplement to traditional teaching techniques, rather than as a complete replacement for them. In addition to new concepts, they can teach students how to analyze information and develop critical thinking skills. Teachers can benefit from using technology as well since they can develop creative and interesting in-class simulation activities. Unlike a regular lecture-style course where a teacher writes on a board while the students sit and watch, involving students directly in the learning process will make them more likely to stay motivated to learn. Computer simulations can also teach students how to work in a group, as many of these simulations involve team-based activities and decision-making (Coffman 5) [4].

The purpose of business is to maximize production efficiency in a variety of sectors. From the development of finance long ago until today, the goal is to continue increasing the efficiency and success rate of diverse businesses. Business profits off human motivation to provide goods and services. Unless human nature changes and everyone can actively provide the goods required by society, business will remain a necessary part of society. Business has already existed for a long time, about 20,000 years ago. The first known trade occurred in New Guinea in around 17,000 BCE, where one obsidian was traded for other goods. Over time, business changed from a barter economy to one which used currency, usually gold or silver, starting with the Lydian Kingdoms in the ancient world (Weatherford 31-32) [5]. By the 19th century, modern capitalism arose in the world in Europe alongside the Industrial Revolution. This economic development created prosperity throughout the world through capitalism's ability to promote innovation and new technologies (Sachs 97-98) [6]. The creation of digital technologies and computers led to the third industrial revolution, which society is currently in. This revolution has shown the power of these new technologies to transform the economy and society (Dosi and Galambos 28-30) [7]. Because of the importance of the digital economy and technology to business, expanding simulations in business education is an important goal.

This paper seeks to study the general public's interest in business and the effectiveness and prevalence of simulations and games in their education, as well as their response to testing or observing a business simulation, to determine whether simulations are effective and needed in society. The development and maintenance of businesses involve direct communication and good foresight; hence, business simulations are necessary for individuals to build upon their understanding of the business world. Through using a survey that examines the interests and experiences of individuals regarding simulations, this goal is achievable.

2. CHALLENGES

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

2.1. Gap in the Research

Current research on business education indicates that not much is known about the audience of business simulations and what effect the characteristics of this demographic will have on learning—the audience of business simulations can be regarded as both students of entrepreneurship education and members of the general public who are not students of entrepreneurship education. For example, there is a lack of past research regarding the willingness and opinions of the general public to participate in a business simulation (Ewijk 9) [8]. Therefore, steps can be taken to identify the general public's perception of business and interactive simulations and broaden the field of business education research.

2.2. Factors of Simulation-Based Learning in Business Education

When learning business, students should focus on their direct communication with others as well how they can apply their business knowledge to their situation; simulations can effectively help students improve on these business skills. According to Professor Kevin Hindle, now CEO of the Mentor Entrepreneurship Group (MENTREG), three main factors are measured when judging simulations related to entrepreneurship [9]. One factor is verification, which is the confirmation of whether the simulation is functioning and portraying a scenario as intended. Verification, therefore, allows a simulation to be more reliable with less prevalent or major glitches. Verification also involves how simple the user interface is to work with and how well-trained the instructor or trainer is with the simulation. Overall, in existing studies of using simulations related to business education, the ability of a simulation to perform consistently well is a crucial factor in terms of keeping students engaged in classroom settings (Hindle 238) [10].

While simulations would engage students in the classroom, simulations are essential to exposing students to situations that would bring a realistic business environment for experience. Entrepreneurship Professor Jeffery McMullen and Dr. Dean A. Shepherd in the Mendoza College of Business detail the next factor, validation, and how it refers to the ability of the simulation to properly reflect how a given scenario would play out in real life. A simulation with strong validation would allow the users to have an accurate representation of real-life processes. However, measuring validation in a simulation can be very complicated. A scenario can occur in which results would be uncertain due to unexpected and inventive actions taken by the user. To address this issue, a simulation can choose to balance an accurate modeling of reality with the assessment of the user's skills and abilities (McMullen and Shepherd 134) [11].

Simulations have the ability to involve students in realistic scenarios that would improve their ability to assess and analyze. Dr. Joseph Fox, a professional researcher of entrepreneurship and instructional technology, weighs in the final factor of fidelity, which measures how realistic the simulation is. Unlike the other two factors, in which more of a factor means better, too much fidelity in a simulation designed for entrepreneurial education can potentially hurt the experience of its users. For instance, providing the users with too much information or details can leave the users overwhelmed and unable to focus on the simulation's main points. As a result, those who create simulations that involve business education would have to utilize a delicate balance of fidelity to maximize effectiveness. This can be done by creating believable yet simple scenarios that emphasize the most significant features of the business experience (Fox et al. 64) [12].

2.3. Benefits of Including Simulation-Based Learning in Business Education

The incorporation of simulations in a classroom poses a number of benefits for both teachers and students alike. According to Dr. Ernest Cadotte, the Emeritus Professor of Innovative Learning at the College of Tennessee, business simulations have been reported to have an overall positive effect on student learning. They have been proven to boost teacher and student enthusiasm as well as student performance measured by grades on assignments. Some other benefits of these simulations include increased flexibility and realism in learning experiences and involvement in the learning process. An explanation of why simulations are such an effective learning tool is the Experiential Learning Theory, which describes that learning is best understood to be a holistic process of creating knowledge rather than a set of outcomes. Furthermore, this theory states that learning is a result of forming transactions between an individual and the individual's environment (Cadotte 281-282) [13]. In this theory, involving the students in their education is a crucial part of developing their long-term retention of knowledge.

As teachers and students gain a positive experience from using simulations in the classroom, developers also recognize the resourcefulness of simulations. Celina Byers, instructional technology Professor at the Bloomsburg University of Pennsylvania, and Hugh M. Cannon, the Emeritus Professor of Marketing at Wayne State University, state the positive impacts that business simulations can have from the perspective of programmers. Business simulations contain clear results and feedback that indicates learning progress and can serve as a source of motivation and encouragement for the players to continue with the simulation. These business simulations work as groups of systems that provide the players with complex problems and allow them to apply both structural knowledge and strategic knowledge. Structural knowledge is defined as knowledge gained through developing hypotheses and testing these hypotheses. On the other hand, strategic knowledge is defined as selecting the appropriate approaches to solve a problem (Byers and Cannon 262-263) [14]. Due to the intricate planning regarding the creation and development of the simulations, the simulation's ability to stimulate learning can be optimized to be as effective as possible.

3. SOLUTION

This study required the development of a game to simulate driving a food truck to various areas and selling food to people to make a profit. The game was created to provide players with an interactive experience of business education, as players can learn the processes and costs behind running a food truck. Many well-known games that involve the food business tend to focus much more heavily on the food-making process rather than the business aspect. To bring more attention to this business aspect, the game intentionally limits the player's interaction by only allowing the player to select which location to drive to and which chefs to hire. While implementing the process of creating the foods in the game was considered, such an addition would likely cause the player to invest an unnecessary amount of attention and effort in an activity that is not related to the primary goal of the simulation. The game was programmed to focus on the overhead selection of employees because the hiring process is a crucial practice in business since it determines the efficiency of the business overall.



Figure 1. Food Truck Selection

This figure shows the first step of the player building their business which is establishing the main property.

The player of the game would start on a select screen, in which the player can choose what color they would like the food truck to be from three color options. Then, the player can choose from three places to visit: the city, the playground, and the amusement park. After choosing a place for the food truck to drive to, the player would then be able to select which chefs they would like to

hire from a list. Each chef has a different price to hire and a different selection of food items that they can create. The Michelin star rating of the chefs determined how much they cost and how many food items they could create. After a certain number of days is reached, the player is provided with the end screen, which displays to the player how much money has been earned. The player can then either choose to play again or quit the game.



Figure 2. Operating the Food Truck Business

This figure shows the player driving their selected food truck around the map and attempting to draw in customers.

The game was created in Unity, which makes use of scripts in the C# programming language. These scripts were implemented in almost all parts of the game, from calculating the amount of money earned in a day to displaying the food menu on the screen. The game also makes use of six scenes; one scene brings the player to the select screen in which the player can choose their food truck's color, another scene displays the end screen, a third scene displays a world map that the player can interact with by clicking, and the final three scenes animate and design the inside of a food truck for each of the three selectable colors.

```

public class OrderingFood : MonoBehaviour
{
    public GameObject foodMenuPanelPanel;
    public GameObject[] customerLocations;
    public List<GameObject> customers = new List<GameObject>();

    public void OrderFood()
    {
        Debug.Log("OrderFood");
        int size = foodMenuPanelPanel.transform.childCount;

        string[] foods = new string[size];
        for (int i = 0; i < size; i++)
            foods[i] = foodMenuPanelPanel.transform.GetChild(i).gameObject.GetComponent<F

        GameObject mapChoose = GameObject.Find("Canvas/MapChoose");
        int location = mapChoose.GetComponent<MapChoose>().intLocation;

        int numCustomers = customerLocations[location].transform.childCount;

        System.Random random = new System.Random();
        int index = random.Next(2, numCustomers);

        GameObject customer;

        for (int i = 0; i < numCustomers; i++)
        {
            if (i % index == 0)
            {
                customer = customerLocations[location].transform.GetChild(i).gameObject;
                customer.SetActive(true);
                customers.Add(customer);
            }
        }

        StartCoroutine(DisplayOrders(customers, foods));
    }

    IEnumerator DisplayOrders(List<GameObject> customers, string[] foods)
    {
        while (GetComponent<GameController>().orders.Any()==true)
        {

```

Figure 3. Code Base for Ordering Food Feature

This figure shows the C# code for the ordering food operation, which involves the creation of objects for each order as well as loops to keep track of the customers.

4. EXPERIMENT

4.1. Experiment

A business and computer science survey was conducted using Google Forms, including both multiple-choice and free-response questions. To ensure the participants' data could be used for analysis, the first question of the survey asked for their consent to have their answers collected for a research paper; respectively, participants who responded with "No" to this question would not have their data considered. The questions in this survey asked for the participants' learning preferences, knowledge in the topics of business and computer science, and thoughts regarding interactive simulations and games to enhance learning. To provide the participants with the flexibility to express themselves, multiple-choice questions had a range of possible answers to choose from. One instance of this is in the question that asks the participants whether they prefer to learn by reading or learn by doing. Participants were provided with five possible responses to this question: "Always prefer learning by reading," "Mostly prefer learning by reading," "No preference," "Mostly prefer learning by doing," and "Always prefer learning by doing". Rather than providing a "Yes" or "No" response, offering a spectrum of answers for participants to choose from ensures that they provide more accurate and meaningful responses.

Do you prefer learning by reading or learning by doing? (Ex: gaming, labs) *

Choose

- Always prefer learning by reading
- Mostly prefer learning by reading
- No preference
- Mostly prefer learning by doing
- Always prefer learning by doing

Figure 4. Learning Preference

This figure shows the question of what the participant prefers to learn and all the preferences available to answer.

A number of the questions were based on the questions beforehand. For example, one question asked the participants whether they have ever tried using an interactive simulation before. If a participant answered yes, they were then asked to list what simulations they have tried; otherwise, the participant would be asked to skip to the next question. Another question asked participants whether they have ever played a game that was beneficial to them in some way. If they answered yes, they would be asked to provide specifics in the following question. For the participants who answered that a game had not been beneficial to them before, they would simply be asked to instead move on to the next question.

Did you ever play a game that was beneficial for you? *

Yes

No

If yes to the previous question, can you describe why it was beneficial? If no to the previous question, skip to the next question.

Your answer _____

Figure 5. Benefits of Playing a Learning Game

This figure shows the questions about whether the participants have or have not gained a positive experience from playing a learning game and why.

The final type of question on the survey was a free-response question about what they felt they could learn from simulations and other interactive experiences as portrayed through images. The first image featured a rocket-building simulation, in which parts of a rocket could be moved around and adjusted. A weight and mass counter was also shown at the bottom of the image. The second image portrayed a flight simulation, in which a person was able to sit in a cockpit with realistic, physical controls and operate these controls. The screens, which are cleverly placed to look like the windows of a plane, allow the person to perform flight training in a safe yet

effective environment. The final image shows a game of Scrabble, which is a multiplayer word game about creating words by placing letter tiles on a board. Although Scrabble would not necessarily be considered a simulation, it is an interactive experience that allows participants to learn vocabulary and spelling by creating words with the randomized letter tiles they are provided with.



Figure 6. Survey Simulation Example

This figure shows an example of a simulation in aircraft and asks the participant if they can learn anything from observing it. The survey took between approximately five and ten minutes to fully complete, depending on how detailed their free-response answers were. The link to the Google Forms was distributed among approximately 70 individuals within several social networks, and 54 of them completed full responses to the survey. The responses to the survey are stored in the Google Forms as well as linked Google Sheets. Importantly, the linked Google Sheets provided a way to both check the timestamps of each survey response and view all of the responses more easily.

Timestamp	I consent to having my a	Do you prefer learning by	Do you believe that there	Optional: Do you have ar	Have you ever learned b	Do you find business inte
2/27/2022 16:19:30	Yes, I consent	Always prefer learning by	Yes	N/A	No	Not interesting
2/27/2022 16:40:02	Yes, I consent	Mostly prefer learning by	Yes		Yes	Interesting
2/27/2022 18:50:53	Yes, I consent	Mostly prefer learning by	Yes		Yes	Interesting
2/27/2022 19:47:50	Yes, I consent	Mostly prefer learning by	Yes		Yes	Neutral
2/27/2022 21:26:17	Yes, I consent	Always prefer learning by	Yes		No	Neutral
2/27/2022 21:28:36	Yes, I consent	Mostly prefer learning by	Yes		Yes	Neutral
3/1/2022 17:04:26	Yes, I consent	Always prefer learning by	Yes		No	Neutral
3/1/2022 17:28:32	Yes, I consent	Mostly prefer learning by	Yes	None	Yes	Interesting
3/3/2022 17:16:53	Yes, I consent	Mostly prefer learning by	Yes		No	Neutral
3/3/2022 17:17:17	Yes, I consent	Always prefer learning by	Yes		No	Neutral
3/3/2022 17:17:23	Yes, I consent	No preference	Yes		No	Interesting
3/3/2022 17:17:43	Yes, I consent	No preference	Yes		Yes	Interesting
3/3/2022 17:18:00	Yes, I consent	Mostly prefer learning by	Yes		No	Neutral
3/3/2022 17:18:50	Yes, I consent	No preference	Yes		No	Neutral
3/3/2022 17:19:08	Yes, I consent	Mostly prefer learning by	Yes		No	Neutral

Figure 7. Survey Results on Excel

This figure shows a list of the participants' answers for the survey; they are all organized on an excel sheet, making it resourceful to analyze.

Following the survey, participants attempted a demo gameplay of the Unity game. Each of the participants had the Unity game sent to them online as a zipped project folder, which they would try to unzip, import to Unity, and run. Participants had to directly make decisions regarding the starting budget they had with the business compared to hiring. Although their simulated food truck business would collect profits throughout the day, players had to consider the costs of both employees and facilities. After completing the simulation, the participants were asked to write and send a paragraph regarding their thorough thoughts on the simulation and whether they felt that they had learned from it.



Figure 8. Hiring Chef Selection

This figure shows the game’s hiring selection interface that the players will observe before selecting according to their business plan.

This survey collected fifty-four responses. As 3 of the 54 participants indicated in their responses that they did not consent to have their answers in the survey be used as data collection for a research paper, only 51 responses will be used. According to the results of the survey, only 51% of participants have previously studied computer science. As more high schools and other institutes of education push for more computer science education in recent times and it is becoming a more relevant subject, that almost half of the participants have never learned any computer science is an unexpected statistic (Yadav et al.). On the other hand, about 43.1% of participants have learned business before. As business is an integral part of society that most people are aware of, the majority of participants never learning about business before is also an unexpected statistic.

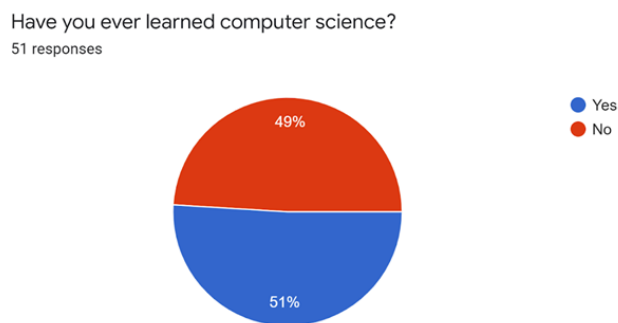


Figure 9. Participants' Computer Science Experience

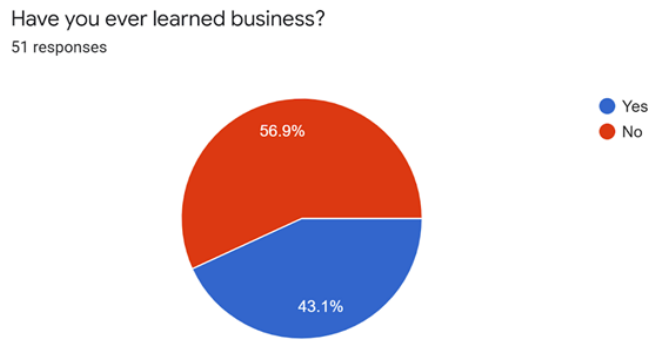


Figure 10. Participants' Business Experience

The data regarding the participants' prior education in computer science and business only partially matches the data regarding the participants' interest in these subjects. Among the participants, only 31.4% indicated that they found computer science interesting, while about 45.1% indicated that they found business interesting. The percentage of participants who found business interesting was relatively close to the percentage that had studied business before. In contrast, there was a significant difference between the percentage of participants who had learned business in school previously and the lower percentage of participants interested in business. This data could demonstrate the inaccessibility of business courses in high schools and other institutes of education. Alternatively, the data could also demonstrate the widespread accessibility of computer science education in the current world when compared to that of business. The interest in business is also emphasized through the responses to another question, in which 68.6% of participants indicated that they would play a game if it could help them gain a better understanding of business.

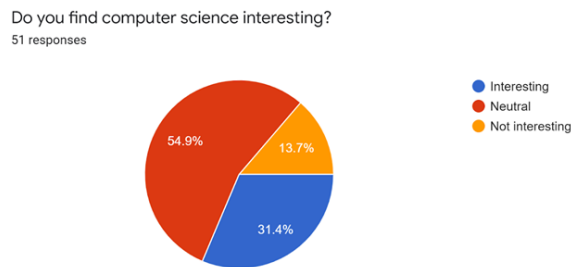


Figure 11. Participants' Computer Science Interest

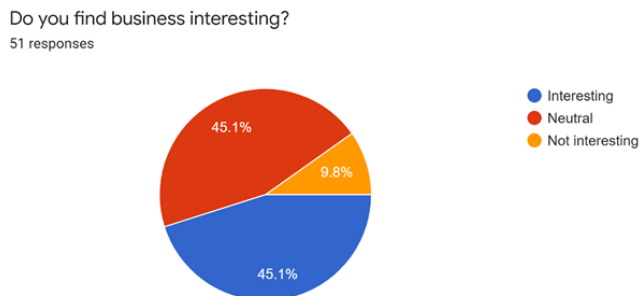


Figure 12. Participant's Business Interest

If there is a game which can help you understand business more, would you play it?
51 responses

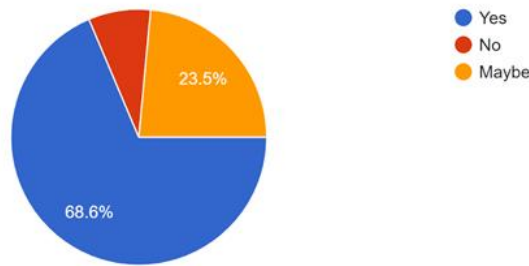


Figure 13. Participants’ Business Game Interest

According to the responses to the survey, the majority of participants seem to prefer learning by doing, with 23.5% indicating that they always prefer learning by doing and about 43.1% indicating that they mostly prefer learning by doing. In total, approximately 66.6% would rather learn by doing than learn by reading. These results represent how most individuals within the general public would be more enthusiastic to learn through a simulation, which could demonstrate a need for more simulations and interactive experiences to be incorporated into education.

Do you prefer learning by reading or learning by doing? (Ex: gaming, labs)
51 responses

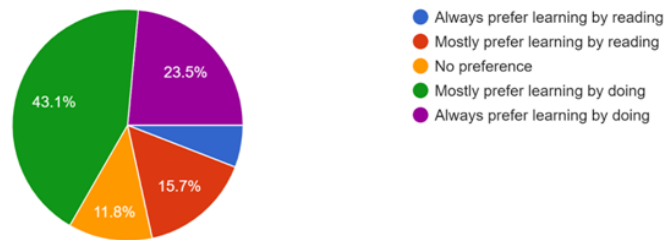


Figure 14. Participants’ Learning Preference

Similar to the results above, the interests of the participants also seem to contradict the participants’ experiences with learning by doing. Only 45.1% of participants have reported that they have used an interactive simulation before. From these results, either the participants have engaged in learning through other means, such as participating in a lab rather than in a simulation, or they have not had easy access to simulations that could boost their learning and/or enthusiasm about certain topics. The lack of easily accessible simulations may be emphasized in another question, which asks the participants whether there should be more interactive simulations for more topics. An overwhelming majority of 92.2% responded with “Yes” to this question. Among the list of simulations that the participants have tried before, the most popular responses were flight simulations, car simulations, and virtual lab simulations. These responses are an indicator of how simulations are already being widely used in educational environments.

Computer Science & Information Technology (CS & IT)

Have you ever tried using an interactive simulation before? (Ex: Flight simulator, car simulator)
51 responses

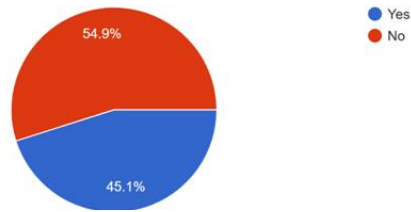


Figure 15. Participants' Interactive Simulation Experience

Do you believe that there should be more interactive simulations for other topics as well?
51 responses

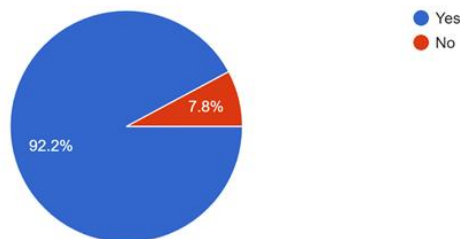


Figure 16. Participants' Interactive Simulation Opinion

When participants were asked whether they had played a game that was beneficial to them in some way, approximately 62.7% reported that they had. When asked to elaborate in the next question, many of the participants reported that the games were educational. A few responses indicated that games can improve reaction time, mood, critical thinking, and organizational skills. The genre of first-person shooter games (FPS) was most frequently associated with enhanced reaction time by the participants. Meanwhile, when bringing up enhanced critical thinking abilities, participants tended to mention strategy games more frequently. These patterns could indicate how certain genres of games can have greater benefits in certain areas.

Did you ever play a game that was beneficial for you?
51 responses

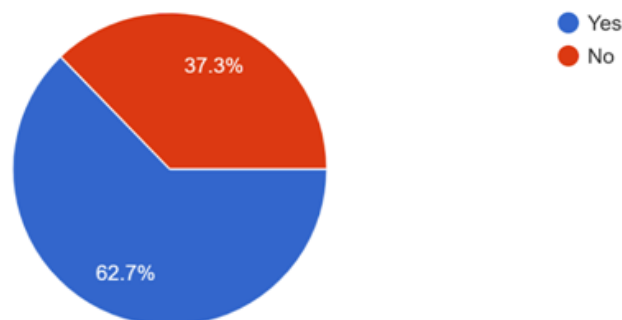


Figure 17. Participants' Experience with Beneficial Games

When asked what could be learned from the simulations and interactive experiences shown in the images, many of the participants gave only superficial responses such as they could learn just

what was specifically portrayed in the image. For example, in response to the image of a flight simulation, some participants stated that they could use such a simulation to learn how it looks and feels to fly a plane. However, other responses delved deeper, indicating in their responses that becoming a pilot requires handling complicated controls, much background knowledge, and strong multitasking ability. The majority of participants understood that what was being displayed in the picture was a word game, indicating in their responses that this game could help to expand vocabulary and test overall knowledge of English. Out of the 51 participants who had fully completed the survey and had given consent to their data being used in a research paper, only 11 of them were able to fully test the game and submit their responses. Among those who fully tested the game, every participant except one responded that they learned more about business and wished to see similar business simulations in the future. The participants noted that through the simulation, they better understood the layering relationship between profit and cost when it comes to business. Many of them also reported that they had become more mindful of how businesses operate and how complicated the inner workings of society can be.

Did you successfully run the game?
51 responses

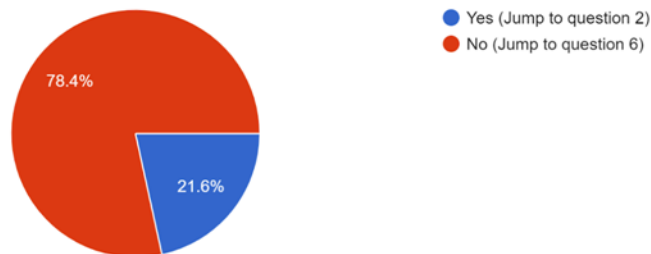


Figure 18. Participants' Game Installation Experience

Did you learn more about business from this game?
11 responses

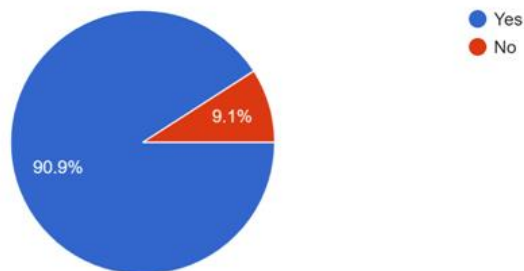


Figure 19. Participants' Learning Experience with the Game

Would you like to see more business simulations like this in the future?
11 responses

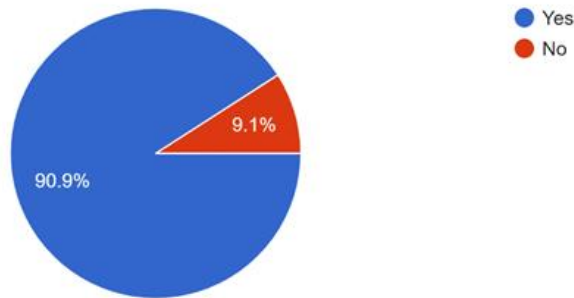


Figure 20. Participants' Interest in Future Business Simulations

Although most of the participants did not test the game, they were able to view a video of the game and provide their responses in the survey as well. Based on the video, 36 out of 40 participants who did not test the game for themselves stated that they were able to learn more about business from this simulation.

Did you learn more about business from this game?
40 responses

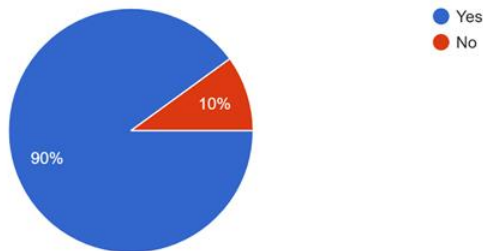


Figure 21. Participants' Learning Growth from Spectating the Gameplay

Would you like to see more business simulations like this in the future?
40 responses

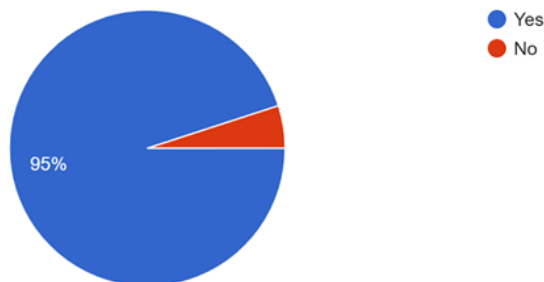


Figure 22. Participants' Interest in Future Business Simulations After Spectating

5. CONCLUSIONS

The conclusion that was drawn from this research is that not many people understand business and simulations are a significant and effective educational tool to help people learn about business. Therefore, there could be a need for more easily accessible and understandable simulations regarding education for business to be created and advertised. The participants who answered the surveys are treated as a subset of the general public. After analyzing the survey responses, the general public appears to have some level of interest in the topic of business and is generally willing to learn more about it, but either does not have access to any simulations regarding business or does not know where to access such simulations. The majority of the participants believed that the simulation was useful for learning business and wanted more to be created in the future. Furthermore, despite business playing such a significant and integral role in society, many of the participants reported not having previous education in or knowledge of business; this reinforces the need to make business simulations more prevalent and accessible to the public.

A limitation that was faced with the research was being unable to distribute the surveys to more individuals. Due to time constraints and the refusal of some individuals to either consent to have their data used in a research paper or take the surveys at all, the total sample size was 51. While a sample of 51 participants is large enough to account for variability, more participants would be preferable to reduce sample variability further. The sample size's randomness and representativeness could be improved as well, as the surveys were only distributed through some social networks. If more effort and planning had been directed towards the distribution of the surveys, the reach of the surveys could extend beyond these few social networks and gather results from those who belong to various demographics. A method to include a more diverse sample is distributing physical paper copies of the surveys. The surveys would need to be simpler and quicker to complete as a result; making this change to the surveys could also prevent non-responses, as some who take unnecessarily long and complicated surveys may give up on completing them partway through.

Another limitation of the research was the inability to reliably have the Unity game regarding business education tested and evaluated by the participants. This game was solely created to be included in this paper, and the research originally involved asking all the participants who took the survey to test the game as well. However, some of the participants lacked access to a computer with modern enough specifications to comfortably and smoothly run the game. Furthermore, the method of transferring the game to the participants to evaluate involved transferring an entire Unity project as a zipped folder. Many of the participants were unfamiliar with Unity and found it difficult to follow the process of downloading Unity, installing the correct Unity version for the game, unzipping the Unity project folder, importing the game through this project folder, and trying the game by clicking on the Select Screen scene before clicking Play. In the end, participants were only required to take the surveys instead of both taking the surveys and testing the game. Therefore, the game was not able to play as important of a role in the research as had been initially planned.

The most significant portion of the research to be added in the future is the testing and evaluation of the game by each of the previous survey participants. In the original study, an attempt was made to have participants test the game by transferring the entire Unity project as files over the internet, but some participants were unable to evaluate the game due to a lack of sufficient technology as well as an arduous and difficult-to-understand testing process. This obstacle can be overcome, however, by choosing a different method of testing. Currently, the best method would be to upload the game on a website, such as itch.io or [GitHub](https://github.com). Asking participants to simply click a link and play a game from a website is a much more convenient and reasonable testing process

than the previously attempted method. Furthermore, using a website to play the game requires fewer resources and is less hardware-intensive than testing the game directly from Unity, which can allow even those with older, lower-spec computers to smoothly test the game. The only possible issue would be internet connectivity, but there are enough locations providing internet access to overcome this. Finally, having the game available online would allow for participants to come from a variety of locations and interact with the game at different times because they would not need to come to a single physical location to play the game.

Since further research would involve the examination of the game, the research could also be improved upon by improving the game itself. Many features that could enhance the player's experience have yet to be added to the game, including the addition of a win condition of reaching a target amount of money and a losing condition of having a negative balance. A detailed page that states the player's total balance, along with the amount of money spent and gained in a day, would be crucial for the player to understand the scenario and can influence actions to make a higher profit the next day. Currently, the game is too static and unchanging, since replays of the game have only limited variation from each other. Therefore, adding potentially positive or negative random events and having the chefs' levels depend on location could both greatly improve replayability and test the player's ability to adapt. The game could also be polished by adding more animations, updating the map with improved graphics, introducing a pause menu along with an option to quit the game anytime, and adding more areas to the map.

The effectiveness of already existing business simulations is another aspect of this that can be further investigated. While some sections of the literature review were able to address this point to some extent, this paper never emphasized the effectiveness of simulations that were specifically related to business as a main point. The paper covered the possible benefits of using a simulation to learn business interactively and the three factors in a business simulation. Covering the effectiveness of simulations can allow the observation of how much of each of these factors were utilized in certain simulations and if the benefits of using these simulations that were described in other works were significant. Including this aspect in future research would also tie in well with the evaluation of the Unity game that was meant to be included more in this paper, as the effectiveness of that game would also be tested. To test the effectiveness of the game, the sample size can be split into 2 groups. The experimental group would play the game, while the control group would receive a standard lesson on business. Both groups would take a test and the results could be compared to form a conclusion. Additionally, the surveys could be further enhanced to discuss specific features of the gameplay or the specific participant's experiences. This information would be valuable in evaluating both how effective the simulation was and how participants responded to it compared to their other survey responses. If the simulation is proven to be an effective learning tool, policies may be created in the future for business education simulations to be applied to more high schools. Real-world applications of these simulations can enhance the financial literacy education of students.

REFERENCES

- [1] Byers, Celina, and Hugh M. Cannon. "The programming game: An exploratory collaboration between business simulation and instructional design." *Developments in Business Simulation and Experiential Learning: Proceedings of the Annual ABSEL conference*. Vol. 34. 2007.
- [2] Cadotte, Ernest. "The use of simulations in entrepreneurship education: Opportunities, challenges and outcomes." *Annals of entrepreneurship education and pedagogy-2014* (2014): 280-302.
- [3] Coffman, Teresa. "Using simulations to enhance teaching and learning." *Virginia Soc. Technol. Educ. J* 21.2 (2006): 1-6.
- [4] Dosi, Giovanni, Louis Galambos, and Luigi Orsanigo, eds. *The third industrial revolution in global business*. Cambridge University Press, 2013.

- [5] van Ewijk, Anne. "Persistence and acuteness of research gaps in entrepreneurship education: A systematic content analysis of previous reviews (1987-2017)." *International Journal of Entrepreneurship* 22.2 (2018): 1-18.
- [6] Flake, Janice L. "Interactive computer simulations for teacher education." *Educational Technology* 15.3 (1975): 54-57.
- [7] Fox, Joe, Luke Pittaway, and Ikenna Uzuegbunam. "Simulations in entrepreneurship education: Serious games and learning through play." *Entrepreneurship Education and Pedagogy* 1.1 (2018): 61-89.
- [8] Hindle, Kevin. "A grounded theory for teaching entrepreneurship using simulation games." *Simulation & Gaming* 33.2 (2002): 236-241.
- [9] Hongsermeier, Abby, Nealy F. Grandgenett, and Dawn M. Simon. "Modeling evolution in the classroom: an interactive LEGO simulation." *The American Biology Teacher* 79.2 (2017): 128-134.
- [10] Koukourikos, Konstantinos, et al. "Simulation in Clinical Nursing Education." *Acta Informatica Medica* 29.1 (2021): 15.
- [11] McMullen, Jeffery S., and Dean A. Shepherd. "Entrepreneurial action and the role of uncertainty in the theory of the entrepreneur." *Academy of Management review* 31.1 (2006): 132-152.
- [12] Sachs, Jeffrey D. "Twentieth-century political economy: A brief history of global capitalism." *Oxford Review of Economic Policy* 15.4 (1999): 90-101.
- [13] Wray, L. Randall. "Introduction to an alternative history of money." Levy Economics Institute, working paper 717 (2012).
- [14] Yadav, Aman, et al. "Expanding computer science education in schools: understanding teacher experiences and challenges." *Computer Science Education* 26.4 (2016): 235-254.
- [15] Ciotti, Marco, et al. "The COVID-19 pandemic." *Critical reviews in clinical laboratory sciences* 57.6 (2020): 365-388.

WORD PREDICTABILITY IS BASED ON CONTEXT - AND/OR FREQUENCY

Rodolfo Delmonte¹ and Nicolò Busetto²

¹Ca Foscari University, Venice (Italy)

²Accenture TTS Computational Linguistics

ABSTRACT

In this paper we present an experiment carried out with BERT on a small number of Italian sentences taken from two domains: newspapers and poetry domain. They represent two levels of increasing difficulty in the possibility to predict the masked word that we intended to test. The experiment is organized on the hypothesis of increasing difficulty in predictability at the three levels of linguistic complexity that we intend to monitor: lexical, syntactic and semantic level. To test this hypothesis we alternate canonical and non-canonical versions of the same sentence before processing them with the same DL model. The result shows that DL models are highly sensitive to presence of non-canonical structures and to local non-literal meaning compositional effect. However, DL are also very sensitive to word frequency by predicting preferentially function vs content words, collocates vs infrequent word phrases. To measure differences in performance we created a linguistically based “predictability parameter” which is highly correlated with a cosine based classification but produces better distinctions between classes.

KEYWORDS

Deep Learning Models, BERT, Masked Word Task, Word Embeddings, Canonical vs Non-canonical sentence structures, Frequency Ranking, Dictionary of Wordforms, Linguistic Similarity Measures, Predictability Parameter.

1. INTRODUCTION

In this paper we will discuss in detail the set up and results of an experiment carried out with a small dataset of Italian sentences, using the output of the first projection layer of a Deep Learning model, the raw word embeddings. We decided to work on Italian to highlight its difference from English in an extended number of relevant linguistic properties.

The underlying hypothesis aims at proving the ability of BERT [1] to predict masked words with increasing complex contexts. To verify this hypothesis we selected sentences that exhibit two important features of Italian texts, non-canonicity and presence of words with very low or rare frequency. To better evaluate the impact of these two factors on word predictability we created a word predictability measure which is based on a combination of scoring functions for context and word frequency of (co-)occurrence. The experiment uses BERT assuming that DNNs can be regarded capable of modeling the behaviour of the human brain in predicting a next word given a sentence and text corpus - but see the following section.

It is usually the case that paradigmatic and syntagmatic properties of words in a sentence are tested separately. In this experiment we decided to test them together by combining non-canonicity and infrequent word choice. Italian sentences are taken from two domains: newspapers

David C. Wyld et al. (Eds): AI, AIMLNET, BIOS, BINLP, CSTY, MaVaS, SIGI - 2022

pp. 195-211, 2022. CS & IT - CSCP 2022

DOI: 10.5121/csit.2022.121818

and poetry domain. Italian bureaucratic and newspaper language can be easily understood by sufficiently literate people. This is not so for people affected by return illiteracy, which constitute a good majority of middle aged people. The second set of sentences taken from Italian poetry of last century is fairly hard to understand. This is due both to choice of infrequent words and uncommon structures. The hypothesis is that predictable masked words would be more frequent in the first than in the second set of sentences. In addition we expect the canonical version of the dataset to be more predictable.

We decided to work with a small dataset which is made of 18 sentences with 150 content words which are then duplicated in the canonical structure thus summing up to 36 sentences and 300 words. This has been done in order to be able to comment the import of every single masked word and its role in the overall sentence structure from a linguistic point of view. This has allowed us to come to precise conclusions on the type of errors the encoding phase systematically makes. In particular, the experiment has allowed us to evaluate the bias of the model towards one of the domains, the newswire one, where the best results have been obtained. In the case of the poetry domain, errors were in general mostly due to the absence of word embeddings with an appropriate context for the input word and the consequent inability to predict the masked word.

The most important feature of the experiment is that all sentences are characterized by non-canonical structures. Italian is a language in which non-canonical structures are fairly common due to the weakly configurational nature of the language and to the existence of the pro-drop parameter that allows sentences to freely omit lexically expressed subjects [2]. We then operated on the dataset in two ways: at first we reformulated the text obtained modifying each sentence structure in order to make it canonical. The inclusion of sentences from poetry has been done in order to focus on the effects of context in conjunction with word level frequency effects - a thorough syntactic and semantic description of these sentences can be found in [3]. The reason for this choice is that poetry is the only domain where rare words are used consistently thus making available a full real context of use for (very) low frequency words. The combined effect of using rare words in a non-canonical syntactic configuration and then restructuring the same sentence with a canonical structure allowed us to make important comparisons.

Non-canonical sentences in Italian can be found in great number due to the pro-drop nature of the language which thus resembles Chinese and Japanese [4]. In addition, Italian is a morphologically rich language thus possessing a very large vocabulary of unique wordforms which, if compared to the total number of wordforms obtainable from the WordNet list of citation forms for English is an order of magnitude higher – from 500K to 5 million wordforms in Italian, only considering the corresponding number of grammatical categories [5]. We already discussed elsewhere [6] that languages like Italian, which have a rich morphology, need embeddings with higher dimensions and a vocabulary size more than doubled in order to account for the variety of semantically relevant wordforms.

When referring to context in BERT, the whole preceding sentence portion is included. BERT being bidirectional the context will apply to both the right and the left previous sequence of tokens. However, when referred to Distributional Semantic Models, the context is usually determined by the number (2 to 5) of co-occurring tokens to be considered when building vectors for word embeddings: if the masked word is the first word in the sentence only the right context will be available and this fact reduces the ability of prediction as shown by our data. The result of our experiment shows that DNNs are very sensitive to context and that frequency of occurrence is less relevant for word predictability – but see below.

The paper is organized as follows: in the following section, we introduce briefly state of the art on the problem of word predictability as seen from the cognitive point of view; in section three

we present the experimental setup and the typology of non-canonical structures contained in our dataset; section 4 presents the experimental results and discuss its import for the predictability parameter, then our conclusion. In the Appendix we reported the translated version of the sentences, while the detailed analysis is contained in the Supplemental Material.

2. WORD PREDICTABILITY IN COGNITIVE AND PSYCHOLINGUISTIC RESEARCH

Word prediction or predictive language processing has been a foundational topic for psycholinguistic research in the last 50 years or so for all that concerns human sentence processing and comprehension. In this paper we intend to exploit the hypothesis presented lately in a number of papers [7, 8] where human word predictivity is compared and tested by the performance of DNNs in next-word-prediction tasks. In particular, in their conclusion, Schrimpf et al. comment on the results of their findings defining them as an attempt to create a viable hypothesis for modeling predictive language processing in human brain by the use of predictive artificial neural networks, specifying that so-called “transformer” models - BERT - are best-performing models. In another paper (see [9]), they had already come to the conclusion that it is by the use of working memory as a whole that word predictivity works: i.e. the integration of all levels of language processing, lexico-semantic, syntax and knowledge of the world conspire to make word prediction viable in order to carry out the primary function of human language, “the extraction of meaning from spoken, written or signed words and sentences (see [8:2]).

The question of word frequency and their predictability is dealt with in great detail in a paper by [10]. Words which have high predictability scores are also those which are somehow more related to the prior context, and words which are more related to the prior context are also easier to integrate semantically. “...there is no such thing as an unexpected word; there are only words which are more or less expected.” (ibid. 309). In this approach, predictability changes from one word to the next due to syntactic and semantic constraints, eventually coming to the conclusion that speakers tend to choose words more likely to occur in a given context.

Estimating the level of difficulty or the “surprisal” or unpredictability - of a word in a given context is done by the negative log probability measure which counts as 1 words fully predictable and as 0 those unpredictable, where the former ones convey no additional information as opposed to the latter. Thus, in a serial-search model imagining lexical access in a frequency sorted lexicon, the 100th most frequent word would take twice as long to access as the 50th most frequent word. As a consequence, most frequent words are less informative and are easier to pronounce and to understand. However, this may only be regarded as a theoretically viable hypothesis since even when words are infrequent and unknown they may still serve to formulate some meaning related bit of information and help in understanding the content of the utterance. From the results obtained in our experiment based on BERT raw embeddings, both frequency and context conjure to establish word predictability. In some cases it is clearly the low frequency to prevent embeddings to be made available, but in other cases - see the example of the ambiguous word "ora"/now-hour below - even though the word and the local context is fairly typical, the word is not predicted.

A partly similar approach has been attempted by Pedinotti et al.[11], in a paper where they explore the ability of Transformer Models to predict transitive verb complements in typical predicate-argument contexts. Their results show clearly the inability to predict low frequency near synonyms, thus confirming the sensitivity of BERT-like models to frequency values. The experiment also included a version of the dataset where the surface syntactic structure of the sentences was modified in order to introduce non-canonical structures. In fact this was only

limited, though, to two cases: interrogative and cleft-structures. The second structure showed how the model suffered from non-recurrent word order by an important drop in performance (from 70 to 38% accuracy).

Another parameter which has loomed large in the cognitive literature is the relevance of the effort/time required to pronounce/read a word: a short word, both phonetically and as grapheme, is preferred and confirmed in an experiment based on semantic grounds by Mahowald et al. [12], where pairs of near synonym words inserted in frame sentences and user have consistently chosen the shortest ones as the most predictable. This seems to be confirmed by the well-known fact that the top range of frequency lists of wordforms are occupied by short words thus confirming the inverse correlation existing between word length and frequency. Most frequent words are not only the shortest but the ones with more senses as confirmed in a paper by Piantadosi et al. [13], hence the more frequent. To verify this we inspected the top 200 words in the frequency lists of ItWac for Italian and English and counted their number of syllables with the following results: Italian has 75 monosyllabic words and 125 words with more than one syllable; English has 149 monosyllabic words and 51 words with more syllables. The two languages have an opposite distribution as has also been documented in a previous paper [4]. In addition, English top 200 words contain only 30 content words, while Italian contains 61 content words, ten of which are morphological variants, English has only one morphological variant.

3. THE EXPERIMENTAL SETUP

We assume that word predictability can be characterized by two parameters: word (co-occurrence) frequency/ies and linguistic complexity measured by syntactic/semantic related scoring functions. We evaluate word co-occurrence frequencies by means of embeddings as the cosine value made available by BERT in its first projection layer, using pre-trained models and no fine-tuning. We produced the whole experiment leveraging the ability of the Huggingface implementation [7]. We used BERT – with the Italian model taken from UWAC corpus, Umberto-commoncrawl - and examined the output of the first or projection layers. In this way we intended to check the predicting ability of BERT on the masked word, by selecting in turn one content word at a time allowing BERT to use the rest of the sentence as a context to make appropriate predictions. Of course, we are aware of the fact that by training a DNN, its error rate may be reduced in cycles through back propagation. This involves comparing its predicted function value to the training data that we did not intend to use. Error reduction is done by computing the gradient of a cross entropy loss error function and proceeding by specified increments of the weights to an estimated optimal level, determined by stochastic gradient descent, which in the case of a test set, does not necessarily correspond to what has been learnt.

It is a fact that words are represented in a DNN by vectors of real numbers. Each element of the vector expresses a distributional feature of the word - in our case by cosine values. These features are the dimensions of the vectors, and they encode their co-occurrence patterns with other words in a training corpus. Word embeddings are generally compressed into low dimensional vectors (200-300 dimensions) that express similarity and proximity relations among the words in the vocabulary of a DNN model.

In the experiment we ran BERT by masking each content word and some function word, one at a time in order to be able to make a detailed error analysis and parameter evaluation. The text we use in the experiment has been organized to allow us to focus on the context: it is made up of 18 sentences, 11 belonging to the newswire domain and 7 sentences belonging to Italian poetry of last century. All sentences are fully commented and analysed in a previous paper where parsers of Italian have been used to parse them and have resulted in an accuracy value below 50%. In a section below are the description of the non-canonical features of the sentences we used for the experiment. The English translation is available in the Appendix. We signed every sentence with

letter A for those belonging to the poetry domain - 7, and letter B for newswire domain - 11. The newswire sentences are taken from the treebank of Italian – VIT, Venice Italian Treebank – available also under UD repositories at <https://universaldependencies.org>; the poetry set of sentences is taken from publicly available collections of poets of the first half of the nineteenth century which have already undergone specific analysis in previous work(see [2;3]).

In order to evaluate frequency values associated to each masked word, we cleaned the frequency list of Italian wordforms compiled on the basis of ItWaC - which contains approximately 388,000 documents from 1,067 different websites, for a total of about 250M tokens. All documents contained in the PAISA' corpus date back to Sept./Oct. 2010. The itWaC corpus is available at <https://wacky.sslmit.unibo.it/> accessed on October, 2021 -, deleting all numbers and websites. Then we created a list of 50000 most frequent wordforms to be used to check what words would be included by a model created on the basis of BERT tokenization module. Wordforms included are up to a frequency value of 1377. The remaining list is cut at frequency value 4, thus leaving out Rare words, made up of Trislegomena, Dislegomena and Hapaxlegomena, which is by far the longest list: it counts 1,642,949 entries. The inclusive List – the list that includes the 50000 plus the rest of wordforms down to and including words with frequency 4, is made up of 513,427 entries. Then, we divided the 50000 vocabulary into two halves: first half with “high” frequency words, including three segments - highest, high and middle frequency words down to 10000 -, second half from 10000 to 1377 we call “low” frequency words. We then consider as “very-low” frequency words those included in the so-called inclusive List - from 1377 down to 4 occurrences -, and the remaining long tail are classified simply as “Rare Words”. The final classification is then organized into four classes: High, Low, Very Low and Rare. To make frequencies more visible, we mark with one asterisk words belonging to “Low”, with two asterisks words belonging to “Very-Low”, and three asterisks “Rare” words.

4. THE DATASET AND NON-CANONICAL STRUCTURES

As said above, Italian is very rich in number and types of non-canonical structures. This is also due to its being a direct derivation from Latin, a free word-order language (see [4]). Our approach has been previously adopted by other researchers but with slightly different aims that we describe in what follows. The first work is by Paccosi et al. [15] where the authors present a new dataset of Italian based on “marked” sentences, which is then used to verify the performance of a neural parser of Italian (TINT) on the dataset. The result for LAS dependency structures is 77%, 3 points below the best results previously obtained on the UD corpus of Italian, which was 80% accuracy. This result confirms previous work documented also in [16] with a small dataset containing strongly marked sentences, which have been included in the text used in this paper, where the results were well below 50% accuracy. The authors make a detailed description of the type of marked structures they annotated in their treebank corpus. It is a list of seven structures - cleft, left dislocated, right dislocated, presentative “ci”, inverted subject, pseudo-clefts, hanging topic - with a majority of Cleft sentences and Left dislocated sentences. As said above, similar results are obtained by the experiment presented in the paper by Pedinotti et al. [11] where in Section IV they test the ability of Transformers - they use RoBERTa - on a small dataset with surface syntactic structures different from the recurrent word order. They modify the sentences to produce cleft and interrogative versions of the same sentences. The result for core semantic roles - this is what they are testing - is a dramatic drop of performance from 0.65 of correlation in canonical transitive versions down below 0.35.

When compared to the corpuses above, our dataset is smaller but it contains many more types of marked constructions, which makes it more difficult to come to terms with, and this is due mainly to presence of sentences from the poetry domain. We present now the structures contained in our dataset:

- ***complete argument inversion*** (the complement is fronted and the subject is in post verbal position) in sentence 7B - with copula deletion, and in sentence 17B with infinitival structure as subject;
- ***object fronting*** (the object comes before the subject at the beginning of the sentence) in sentence 2A and 5A;
- ***adjective extraction*** (the adjective is extracted and fronted from the noun phrase) in sentence 13A and 14A;
- ***PPadjunct preposing from participial clause*** in sentence 1B and 13A;
- ***lexical verb left extraction*** (the main verb - un- tensed non-finite - is positioned before the auxiliary/modal) in sentence 3A;
- ***subject right dislocation*** (the subject is positioned after the complements) in sentence 3A and 6B; ***subject and object fronting*** (the subject comes before the object and both are positioned before the main verb) in sentence 4A and 5A; ***PPspecification extraction from the noun phrase and fronted to the left*** in sentence 5A;
- ***clitic left dislocation*** in sentence 8B;
- ***object right dislocation*** (the object is positioned after the indirect object or the adjuncts) in sentence 10B;
- ***parenthetical insertion*** (a parenthetical is inserted after the subject before the main verb) in sentence 11B and 16B;
- ***adjective right extraction*** (the adjective is extracted from the noun phrase and positioned after the noun adjuncts) in sentence 11B and 14A; ***PPspecification right stranding*** - the PPof is stranded to the right out of the noun phrase in sentence 14B;
- ***lexical verb right extraction*** (the main verb - un- tensed non-finite - is positioned after the complements) in sentence 12A;
- ***double parenthetical insertions*** (after the subject and after the verb complex and before the complements) in sentence 15B and 16B;
- ***clitic left dislocation with subject fronted as hanging topic*** in sentence 18B.

5. EXPERIMENTAL RESULTS AND DISCUSSION

The evaluation has been carried out in two modalities: the first modality considered only cosine values, while the second one introduced linguistically based similarity scores. In the first modality, only cosine values were considered as they were made available by the first five candidates computed by BERT. Word predictability has been measured by BERT raw word embeddings and their cosine measure, by masking one content word at a time - and a few function words. In case the masked word was present we took its cosine value disregarding its position; when it was not correctly predicted in the first five candidates we selected the first candidate and its cosine value. We then added all the values found at sentence level. Then each content word has been searched in the frequency list made available by the ItWac frequency list in order to evaluate its frequency impact. Frequency contribution was computed simply by each word position in the frequency list, dividing very frequent words from low frequency ones.

We organized the experiment in three different configurations: on a first configuration, part of the sentences, the last 7 – are withheld with the aim to reduce the overall context at sentence level. This is done both for non-canonical and canonical structures. Then the last 7 sentences are added and the cosine values verified to see if predictions have been modified.

Then the second evaluation modality: the linguistically based one. In creating our linguistic evaluation scheme, we assumed that a better form of evaluation should account for gradable differences between predictions in which the actual word is not found but the ones predicted are very “similar”. The word “similar” is then decomposed into its various linguistic aspects and we

have devised a graduality which may be turned into scores according to simple linguistic criteria. Similarity may attain morphological, lexical, grammatical, syntactic, semantic criteria. Thus the more the choices are close to the actual meaning and the linguistic usage of the expected word, the higher the score will be which we assume will be a real value from 0 to 1.

Table 1. Graded Evaluation Scale for a Linguistically Based Similarity Scoring.

<ul style="list-style-type: none"> • Identical (first position) = 1 • Identical (second position) = 0.99 • Identical (third position) = 0.97 • Identical (fourth position) = 0.95 • Identical (fifth position) = 0.93 • Same word but different morphology = 0.8 • Same word but different grammatical category = 0.7 • Hyponym/Antonym/Meronym/Synonym word same morphology same grammatical category = 0.6 • Hyponym/Antonym/Meronym/Synonym word different morphology same grammatical category = 0.5 • Hyponym/Antonym/Meronym/Synonym word different morphology different grammatical category = 0.4 • Different word same grammatical category same morphology – no semantic similarity - 0.3 • Same grammatical category but different word – no semantic similarity = 0.2 • Different grammatical category different word = 0.1 • Punctuation, <ukn> = 0
--

We applied the scores reported in the table to the whole set of sentences and computed the results in the two tables below. In Table 2. we evaluate the seven sentences from the poetry domain, and in Table 3. the eleven sentences from the newswire domain. We computed three main parameters: in column 2, Number of Words masked with respect to total number of tokens; in columns 3 and 4 we list words correctly predicted with the identical corresponding word respectively in the Non Canonical and in the Canonical sentence structure; then in columns 5 and 6 we list the number of words with frequency values respectively Higher and Lower than a given threshold that we established at 10.000 occurrences. We also considered words that don't appear in the 50000 vocabulary and reported them after a slash: we assume their import should be valued double. Thus for instance, in the Poetry text, we found 5 such words and the total number of Low Frequency Words is thus increased by 10 points. Finally, in column 7, we reported the result of applying the scoring function described in Table 1.

Table 2. Evaluation of Poetry Sentences

Sent. No.	No. Words/ Masked	Non Canon. Words	Canon. Words	High Freq. Words	Low Freq. Words	Ling. Evaluat.
2.A	10/8	0	3	4	3/1	3.76
3.A	14/9	3	4	6	3	6.04
4.A	10/8	2	2	4	4	3.99
5.A	9/6	0	0	4	1/2	2
12.A	11/7	1	2	4	1	3.49
13.A	15/7	0	0	5	0/2	2.4
14.A	14/9	1	1	6	3/1	3.1
totals	83/54	7	12	33	15/6=27	24.78
ratios	0.65	0.583			0.818	0.4589

Table 3. Evaluation of Newswire Sentences

Sent. No.	No. Words/ Masked	Non Canon. Words	Canon. Words	High Freq. Words	Low Freq. Words	Ling. Evaluat.
1.B	14/8	3	5	8	0	5.97
6.B	6/5	2	3	5	0	3.84
7.B	5/4	0	0	3	1	2.4
8.B	10/7	1	2	6	1	2.37
9.B	7/4	1	1	4	1	2.99
10.B	12/9	1	1	7	2	4.79
11.B	15/10	2	4	10	0	6.17
15.B	25/10	7	7	8	2	8.23
16.B	22/10	4	4	8	2	7.2
17.B	15/9	6	6	10	0	7.1
18.B	22/10	4	4	9	0/1	5.7
totals	153/86	30	36	78	9/1=11	56.76
ratios	0.56	0.834			0.141	0.66

As can be easily noticed by comparing all parameters, poetry and news have opposite values. Quantities measured in column 2 show how the ratio of masked words is higher in poetry than in the news domain – 0.65 vs 0.56 -, the reason being that poetry text makes use of less grammatical or function words, like articles, clitics, prepositions which are highly predictable but are less informative. The first important parameter is the difference in number of masked words identified in Non-Canonical vs Canonical Sentences, and here again as can be easily noticed the newswire domain has a much higher score than the poetry domain – 0.834 vs 0.583. Then the second relevant parameter derived by the proportion of High Frequency words vs Low Frequency words and computed as a ratio between the sum of the absolute number of words plus a doubling of the number of very low frequency words. Here the scores show the opposite relation, Poetry domain has a much higher number of Low Frequency words than Newswire domain – 0.818 vs 0.141. Eventually, the linguistic evaluation of every single masked word on the basis of its cosine measure and the graded scoring scale reported in Table 1. Here we see again a much higher overall score for the Newswire than the Poetry domain – 0.66 vs 0.4589. The difference in scores is approximately 20 points and is strongly comparable to the difference found in “context”, i.e. Canonical vs Noncanonical.

The conclusion we can safely draw from these data is that in general the News domain has a higher linguistically and frequency-based evaluated prediction score:

- ✓ because it has a much lower number of Low Frequency words
- ✓ because it has a higher number of contextually predictable words in Non-canonical structures

In other words, the relevance of context varies according to the domain: in the Poetry domain it is both dependent on word frequency and context, i.e. word structural position, but context seems more relevant. Not so in the Newswire domain where context varies less and frequency plays a higher role.

One example is highly representative of the interplay between frequency and context and is the word "Ora", an ambiguous word with two homographs-homophones: one meaning "now", an adverbial contained in sentence n. 9 - the newswire domain; and another meaning "hour", a (temporal) noun, contained in sentence n. 5 - the poetry domain. Only the adverbial is predicted

in both structural versions. On the contrary, the noun is contained in a sentence belonging to the poetry domain where the overall context is not supportive for that word predictability.

Below, we list the words which have been assigned a cosine value higher than 0.5 in canonical and non-canonical structures. All cases of non-canonical structures are included in canonical ones where four additional words are present.

Table 4. Best cosine values for identically predicted masked words

Sent. No.	Masked Word	Cosine Value Non-Can.	Cosine Value Can.	Phrase Including	Lexical Type
1	miei	0.88233		miei colleghi	Function
1	più		0.55960	più acuta	Function
11	questo		0.76715	questo libro	Function
11	esempi	0.65383	0.73481	esempi di carità	Content
15	come	0.9186		come già	Function
15	ha	0.97755		ha voluto	Function
16	viene	0.79483		viene interrogato	Function
16	senatore	0.80796		senatore a vita	Content
16	vita	0.99582		senatore a vita	Content
17	fare	0.81857		intervento da fare	Content
17	questi	0.96136		questi giorni	Function
17	giorni	0.83000		questi giorni	Content
17	detto		0.55038	ha detto	Content
18	modo	0.79384		modo di	Content

As a general remark, the comparison of function and content words we see that function words have a much higher cosine score than content words – with the exception of the collocation or polirematic form: “senatore a vita”/life_long_senator, where both "senatore"/senator and “vita”/life receive a high cosine value, again confirming the relevance of the context, which in this case is as relevant as that one of function words and is the most important parameter to consider.

In Figure 1. below we show cosine values weighted by number of masked words - by choosing always the value associated with the first candidate - when compared with weighted Linguistic Parameter, by listing sentences in descending order according to their score. Correlation evaluation between our Linguistic Parameter and Cosine values is estimated at 0.8705 when computed on absolute values, but it goes down to 0.6349 when using weighted values. News texts have overall higher parameters in both evaluations: the descending trend is however much more linear for linguistic parameters than for the cosine ones.

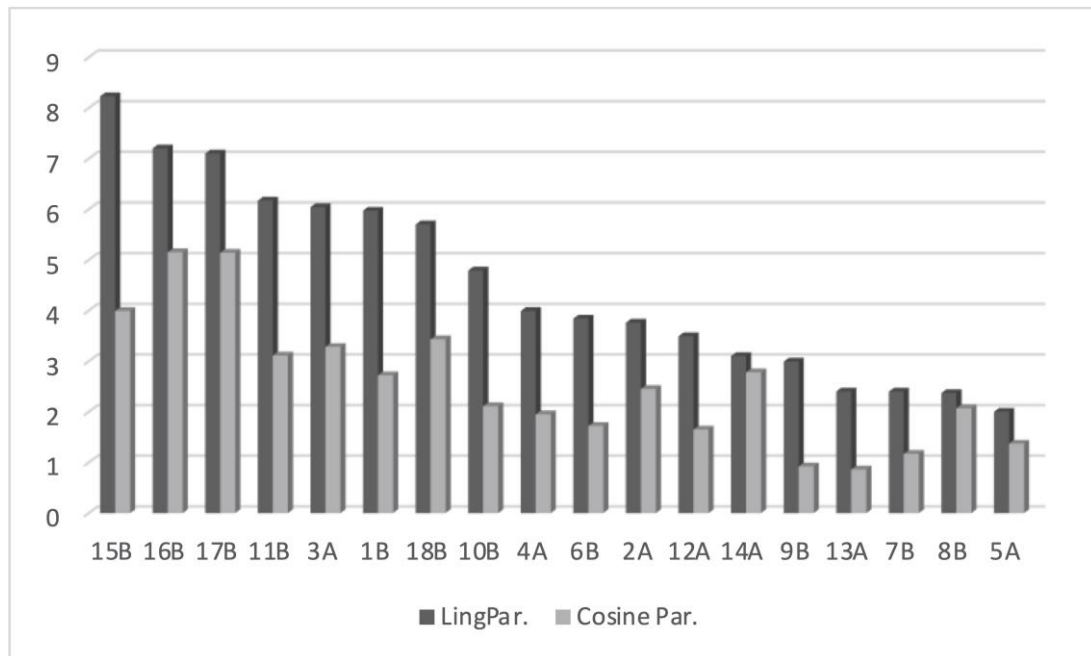


Figure 1. Comparing Sentence Level Cosine and Linguistic Weighted Parameters

The predictability score is a parameter that combines the linguistically weighted output of the masked task which is based on embeddings' cosine measure evaluation, and the frequency ranking of each word as reported in the ItWack vocabulary list. If we divide up the ratio of the evaluation score by the ratio of the frequency score we obtain the following predictability score: Poetry $0.4589 / 0.818 = 0.561$, and News $0.66 / 0.141 = 4.68$.

In sum, even though the poetry domain has a smaller number of sentences and almost half the number of words than the newswire domain, the three parameters we evaluated show the correctness of our hypothesis. In the poetry domain the two main parameters – word frequency and word context - conspire to reduce the predictability score. The context in poetry domain is characterized by metaphorical usage of word combination thus dramatically reducing the ability of BERT to find embeddings. Thus context has a double dimension: word combination aiming at producing metaphorical meaning is accompanied by constituent displacement and discontinuity contributing surprisal effects due to non-canonical structures. These two aspects are further constrained by the low frequency of some words thus justifying the low value of the overall predictability parameter. The opposite applies to the news domain: word linear combinations remain fairly literal in their semantic usage thus favouring the possibility for BERT to find embeddings even when words have low frequency values. Absolute frequency is thus less relevant in the Newswire than in the Poetry domain.

6. CONCLUSIONS AND FUTURE WORK

In this paper we have proposed a word predictability parameter based on linguistically motivated information that we have tested in a highly constrained context determined by the combination of three fundamental factors for a sentence meaning understanding perspective on the prediction task represented by BERT masked task: use of infrequent words - as measured against the ItWac frequency list - and their phrase level combination – word poetic usage for metaphors w.r.t possible semantic association -, and their larger sentential context in uncommon syntactic structures – non-canonical structures. In order to be able to evaluate the different impact of the

three adversarial factors on masked word prediction, we have included in the dataset a higher number of sentences from newswire domain showing the same structural syntactic properties but lacking both the usage of very infrequent words – with a few exceptions - and their uncommon combination to produce metaphors.

The results have clearly shown the ability of newswire sentences to receive an overall higher word predictability score thanks to the smaller effect of adversarial factors we investigated. The answer to the question: is frequency or context the determining factor for Transformer Language Models to predict the masked word, is both are, but their relevance depends on the domain. The news domain has less infrequent words and less uncommon non-canonical structures than the poetry domain, which is what explains the remarkable difference in final results.

In future work we intend to proceed in two directions: enlarging the dataset and completing the experiment using DNNs. We intend to use sentences contained in the treebank of Italian called VIT [3] - which is made up of 11,000 sentences - where some 30% of sentences have been manually classified as non-canonical. Using this dataset we will produce a set of experiments based on Machine Learning always using some variant of BERT, at first with a totally unsupervised approach, and finally a fully supervised approach also introducing syntactic information as has been done in a recent task we participated [17].

REFERENCES

- [1] Niki Parmar, Jakob Uszkoreit, Llion Jones Aidan, N Gomez, Lukasz Kaiser, Ashish Vaswani, Noam Shazeer, and Illia Polosukhin (2017) Attention is all you need. In *Advances in Neural Information Processing Systems*, pp. 6000–6010.
- [2] Rodolfo Delmonte, Antonella Bristot, and Sara Tonelli (2007) VIT - Venice Italian Treebank: Syntactic and quantitative features. In *Proc. Sixth International Workshop on Treebanks and Linguistic Theories*, volume 1, Nealt Proc. Series, pp. 43-54.
- [3] Rodolfo Delmonte (2009) Treebanking in VIT: from Phrase Structure to Dependency Representation, in S. Nirenburg (Ed.), *NATO Series, Language Engineering for Lesser Studied Languages*, Volume 21, IOS Press, Amsterdam, pp. 51–79.
- [4] Rodolfo Delmonte (2014) A computational approach to poetic structure, rhythm and rhyme. In *Proceedings of CLiC-it - The First Italian Conference on Computational Linguistics*, Volume 1, Pisa, Italy. Pisa University Press, pp. 144–150,.
- [5] Rodolfo Delmonte (2018) Syntax and semantics of Italian poetry in the first half of the 20th century. *Umanistica Digitale*, pp. 35–66.
- [6] Rodolfo Delmonte (2021) What’s wrong with deep learning for meaning understanding. In *Proceedings of the 2nd Italian Workshop on Explainable Artificial Intelligence (XAI.it 2021)*.
- [7] Adam Goodkind and Klinton Bicknell (2018) Predictive power of word surprisal for reading times is a linear function of language model quality. In *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018)*, pp. 10–18.
- [8] Martin Schrimpf, Idan Blank, Greta Tuckute, Carina Kauf, Eghbal A. Hosseini, Nancy Kanwisher, Joshua Tenenbaum, and Evelina Fedorenko (2021) The neural architecture of language: Integrative modeling converges on predictive processing. In *Proceedings of the National Academy of Science of the United States of America (PNAS) 2021*, Volume 118, pp. 1–12.
- [9] Evelina Fedorenko, Idan Blank, Matthew Siegelman, and Zachary Mineroff (2020) Lack of selectivity for syntax relative to word meanings throughout the language network. *Cognition*.
- [10] Smith N. and R. Levy (2013) The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3):302–319.
- [11] Paolo Pedinotti Giulia Rambelli, Emmanuele Chersoni, Enrico Santus, Alessandro Lenci, and Philippe Blache (2021) Did the cat drink the coffee? Challenging transformers with generalized event knowledge. In *Proceedings of the 10th Conference on Lexical and Computational Semantics*, pp. 1–11, Bangkok, Thailand. Association for Computational Linguistics.
- [12] K. Mahowald, E. Fedorenko, S.T. Piantadosi, and Edward Gibson (2012) Info/information theory: speakers choose shorter words in predictive contexts. *Cognition*, 126(2):313–318.

- [13] Steven T. Piantadosi, Harry Tily, and Edward Gibson (2012) The communicative function of ambiguity in language. *Cognition*, 122(3):280–291.
- [14] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, and Morgan Funtowicz et al. (2019) *Huggingface's transformers: State-of-the-art natural language processing*.
- [15] Teresa Paccosi, Alessio Palmero Aprosio, and Sara Tonelli (2022) It is markit that is new: An italian treebank of marked constructions. In *Proceedings of CLiC-It 2021 - Eighth Italian Conference on Computational Linguistics*.
- [16] Rodolfo Delmonte (2016) Syntactic and lexical complexity in italian noncanonical structures. In *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity*, Stroudsburg, PA, USA. ACL, pp. 67–78.
- [17] Rodolfo Delmonte (2020) Venses @ AcCompl-It: Computing Complexity vs Acceptability with a Constituent Trigram Model and Semantics, in Basile V., Croce D., Di Maro M., Passaro Lucia C. (eds). *Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian – EVALITA 2020*, Vol-2765, paper 103.

Appendix - English Version of the Canonical and Non-Canonical Text

1.B Today I thank for the courtesy on several occasions demonstrated to me and my colleagues. 2.A She alone maybe the cold dreamer would educate to the tender prodigy. 3.A I think of a green garden where with you resume can conversing the soul maiden. 4.A If spring my generous heart choked of deaf spasms. 5.A Neither the oblivious enchantment of the hour the iron-like beat grants. 6.B Becomes thus sharper the contradiction. 7.B Good instead overall the rest. 8.B An important decision Ghitti reserved after the holidays. 9.B The important thing is now to open it more. 10.B His information would also give to the guidelines of laique democracy greater boosts. 11.B In this book Maria Teresa, they explain at Mondadori's, will give examples of charities concrete. 12.A Said that they have his heart from inside the chest removed. 13.A The reluctant opinions and not ready and in the midst of executing works hampered. 14.A An echo of mature anguish reverdived to touch signs to the flesh dark of joy. 15.B The government, therefore, though giving up the absolute majority, has wanted, as already in IMI, focusing on a gradual privatization. 16.B At a conference in the Viminale the minister, when he is questioned on the senator to life, at first does not understand the name. 17.B First intervention to do, he said these days, is to implement the reform. 18.B I conceive the private as a work method, as work contracts, as a way to manage in short.

1.Bc Today I thank you for the courtesy demonstrated to me and my colleagues on several occasions. 2.Ac Maybe the cold dreamer educated her alone to the tender prodigy. 3.Ac I think of a green garden where the soul maid can resume conversing with you. 4.Ac Spring if you choked my generous heart of deaf spasms. 5.Ac Neither the iron-like beat of the hour grants the oblivious enchantment. 6.Bc The contradiction becomes thus sharper. 7.Bc Instead, overall the rest is good. 8.Bc Ghitti reserved an important decision after the holidays. 9.Bc Now it's important to open it more. 10.Bc His information would also give greater boosts to the guidelines of laique democracy. 11.Bc In this book Maria Teresa will give concrete examples of charities, they explain at Mondadori's. 12.Ac They said they took off his heart from the chest. 13.Ac The reluctant opinions and not ready works hampered in the middle of executing. 14.Ac An echo of mature anguish reverdressed to touch signs of joy obscure to the flesh. 15.Bc So the government wanted to focus on a gradual privatization while giving up the absolute majority as already in IMI. 16.Bc At a conference in the Viminale, when he is questioned on the senator to life at first the minister does not understand the name. 17.Bc To implement the reform is first intervention to do, he said these days. 18.Bc I conceive the private as a work method, such as work contracts, as a way to manage in short.

Supplemental Material

Sentence 1.B - *Oggi ringrazio della cortesia in più occasioni dimostrata a me e ai miei colleghi. 1.Bc Oggi ringrazio della cortesia dimostrata a me e ai miei colleghi in più occasioni.*

The sentence belongs to the newswire domain: it is computed best in the canonical form, with 5 words over 8 while the non-canonical version has only 3 words predicted correctly – only "più/more", "occasioni/chances" and "miei/my". Cosine values are not particularly high except for "miei/my" the possessive which being in its attributive position has a favourable predictive condition. "Oggi" is wrongly

predicted as being a separator with very high value, “<s> 0.99998”. It can be noted that “ringrazio” is partially predicted by “Grazie” in first position but very low value 0.14397. Now the canonical version: Ringrazio (0.0238), più (0.287), occasioni (0.545), dimostrata (0.165), miei (0.882). Interesting to note that the three words predicted in both structural versions have the same cosine values. When we add the remaining 7 sentences, another word is predicted, colleghi (0.076). No connection with frequency values of the missing words: they are all positioned in the high part of the frequency list – excluding “più” and “miei” which are grammatical words and are positioned close to the top.

Frequency List: °-più; °-miei; °-Oggi; °-colleghi; °-occasioni; °-ringrazio; °-dimostrata; °-cortesia

Sentence 2.A - *Lei sola forse il freddo sognatore educerebbe al tenero prodigio. 2.Ac Forse il freddo sognatore educerebbe lei sola al tenero prodigio.*

The second sentence belongs to the poetry domain. The original non-canonical version has no candidate found in the first 5 positions. This may be due to presence of a rather infrequent word like “educerebbe/would+educate” as main verb which only appears listed low only in the Upper List. On the contrary, the canonical form has three words predicted: first “Forse/Maybe”, second word “lei/She”, and third word “solo”/alone but with wrong masculine morphology. However, these words are correctly predicted with low cosine values - Forse (0.149), lei (0.0355) solo (0.0145). No version provides useful approximations of the meaning of the missing words even though “freddo/cold” is included in the high portion of the 50000 vocabulary. As to the remaining words, they are still included in the Vocabulary but in the lower portion. It is important to note that the lack of prediction can only be motivated just because by combining not so frequent words in unusual combination has produced metaphors like “cold dreamer”, “tender prodigy”, in association with a verb like “educate”.

Frequency List: °-solo; °-lei; °-Forse; °-freddo; *-tenero; *-prodigio; *-sognatore; **-educerebbe

Sentence 3.A - *Penso a un verde giardino ove con te riprendere può a conversare l'anima fanciulla. 3.Ac Penso a un verde giardino ove l'anima fanciulla può riprendere a conversare con te.*

The non-canonical version of this sentence has two words correctly predicted, giardino/garden, ove/where and a third word with different morphology, in slot 5, Pensa/Think(3rd+person+singular+present+indicative), rather than Penso(1st+person). In the canonical version we find correctly Penso/think in second slot, and another word is added può/can, the modal auxiliary that is now positioned correctly in front of its main verb "riprendere/restart", which is by itself a very frequent verb. As to cosine values, we have the following low values for the canonical version: Penso (0.085), giardino (0.194), ove (0.146), può (0.0865). The non-canonical version has a lower value for Penso but a higher value for giardino (0.291). In the longer context, the interesting fact is constituted by the substitution of “Pensa” with fino/until in the non-canonical version; while in the canonical version Penso/think is moved to a worse position from second slot to last slot, slot 5 and a lower cosine value (0.06112). As to the non-predicted noun modifier "fanciulla/maid", this is certainly an unusual combination even though the two words are highly frequent. The result of the combination is of course a beautiful metaphor which combines “primavera”/spring with “fanciulla”/maid and the garden. Notice the different position of Penso+1st+pers, with respect to Pensa+3rd+pers which is by far less frequent. Now consider the word conversare/conversing which receives the following list of non-word predicted candidates: erare/?? (0.4455), rare/rare?? (0.16737), lare/?? (0.0549), mare/sea?? (0.0479), scere/?? (0.03124). Apart from RARE and MARE which I don't regard being selected for their current meaning but just for being part of the list of subwords, the remaining segments are all meaningless and bear no semantically useful relation with the masked word CONVERSARE.

Frequency List: °-può; °-ove; °-anima; °-verde; °-Penso; °-riprendere; *-Pensa; *-fanciulla; *-conversare

Sentence 4.A - *Se primavera il mio cuor generoso soffocasti di spasimi sordi. 4.Ac Primavera, se soffocasti il mio cuor generoso di spasimi sordi.*

In this sentence only the phrase "mio cuor"/my heart is predicted in both structural versions. mio (0.291), cuor (0.394). The word “Primavera”, which is the first word in the canonical version, has no close prediction: as happens in all sentences, the prediction is totally missed whenever a content word appears in first position. In the non-canonical version, the word comes second, after the conjunction “Se”/If, which predicts the appearance of an auxiliary BE/HAVE in their correct morphological word form – fossi/were, avessi/had in both cases with first person morphology, but also fosse/were, and the last two: con/with and solo/alone. The version with the addition of the 7 sentences has the worsening effect of introducing a

subword in place of con/with, MMAI which I assume derives from the wrongly split SEMMAI/if+ever. The word has been wrongly split because the segment SE is wrongly – at least in the word SEMMAI - regarded as a legitimate segment due to its very high frequency. Again the problem seems the unusual combination of the remaining words which are fairly common, apart from soffocasti/choked which is not included in the frequent nor in the Rare wordform list; and spasmi/spasms which is only included in the Upper List. In other words, it's their metaphorical import that prevents the correct prediction. However, it is the position that produces the worst results: the adjective “sordi/deaf” in predicative position is predicted as a punctuation mark in both structural versions.

Frequency List: °-Se; °-mio; °-cuore; °-primavera; *-generoso; *-Primavera; *-sordi; **-spasmi

Sentence 5.A - *Né l'oblioso incanto dell'ora il ferreo battito concede. 5.Ac Né il ferreo battito dell'ora concede l'oblioso incanto.*

This sentence is the worst case of the poetry domain lot: it has no word predicted neither in the non-canonical nor in the canonical version. This may be due to the presence of a very infrequent word "obliosi/oblivious". However, we notice the presence of an unusual combination of the attributive metaphoric use of "ferreo/iron-like", a rather unusual word. But of course, it is just the combination of words used to build a powerful metaphor that prevents predictions to take place. It is worthwhile noting that "incanto"/enchantment is substituted by ten candidates semantically loosely related to the domains evoked by the masked word: temporal dimension (rhythm, stepping, passing, proceeding, beat), and a condition of the contemplating mind (silence, rest, meaning, thought, sound). Also another important remark regards the inability to predict the ambiguous word "ora"/hour, homograph with "ora"/now, thus clearly showing that context is the determining factor.

Frequency List: °-ora; °-Né; °-concede; °-incanto; *-battito; **-ferreo; **-oblioso

Sentence 6.B - *Diventa così più acuta la contraddizione. 6.Bc La contraddizione diventa così più acuta.*

This sentence has different predicted words in the two structural representations, Diventa/Becomes is present in both. Then "così/so" and "più/more" are predicted in the canonical sentence - diventa (0.215), così (0.0439), più (0.559); while in the non-canonical structure only acuta/sharp is predicted, acuta (0.0441), and the cosine value for "Diventa" is lower being in sentence first position. The canonical form has predicted the discourse marker "così/so" positioned in sentence center: not so in the non-canonical structure where we can again assume that it is the position right after the verb at the beginning of the sentence that does not allow the prediction, notwithstanding its high frequency. Now consider the high frequency of "contraddizione" which is not predicted presumably because of its position at the end of the sentence: the first candidate is the subword “mente” with cosine value (0.16536), followed by sensibilità/sensibility, coscienza/conscience, gioia/joy.

Frequency List: °-più; °-così; °-contraddizione; °-acuta; *-Diventa

Sentence 7.B - *Buono invece in complesso il resto. 7.Bc Invece in complesso il resto è buono.*

No word was predicted in either versions. In order to transform the original non-canonical version in the corresponding canonical one we added the copula "è" that is missing in the original sentence. This is predicted in the canonical version but since it has been added we do not count it for the actual predictive task. All the words are very frequent. As will be clarified further on, whenever the first word of the sentence coincides with a discourse marker or a conjunction the prediction is very close if not equal. This is the case for the canonical form of the sentence starting with “Invece”/Rather, which has the five following best predictions: “Ma”/But, “E”/And, “Però”/However, “Più”/More, “Ed”/And, all belonging to the same grammatical category and in two cases, also to the same semantic type (“Ma”, “Però”). Considering the status of the adjective “Buono”/Good which comes in first position in the non-canonical structure and in second position in the canonical one, one can clearly realize the importance of the respective position and the context on the ability of BERT to predict. In the first case, the word coming first position has no left context and there is no similarity, not even at a grammatical level: only conjunctions and verbs are predicted. On the contrary, in the canonical form, “buono” appears as predicate in a copulative structure and the predictions are very close: diverso/different, risolto/resolved, compiuto/achieved, secondario/secondary, positivo/positive.

Frequency List: °-invece; °-resto; °-complesso; *-Buono

Sentence 8.B - *Una decisione importante Ghitti l'ha riservata a dopo le feste. 8.Bc Ghitti ha riservato una decisione importante a dopo le feste.*

Only one word is predicted in both versions but it is not the same word. The canonical version predicts "importante/important", (0,0605), the non-canonical version predicts "dopo/after", (0,0152). As can be noticed, the cosine values are very low and again the frequency of occurrence of the words contained in the sentence is fairly high - excluding the proper name "Ghitti" which does not exist in the overall frequency list. The unexpected fact is constituted by the inability to predict the auxiliary "ha"/has in the non-canonical structure - as opposed to what happens in the canonical one -, and the association in fourth slot of a non-word like "vamteen", presumably a subword of some kind. The only explanation could be the presence of a past participle with feminine+singular ending which is only allowed by presence of the resumptive clitic "la" needed to construct the Clitic Left Dislocation of the object NP "Una decisione importante". As said above, the canonical version predicts the presence of the auxiliary HAVE in the correct form and also in two additional morphologically possible forms: "aveva"/had+3rd+pers and "avrebbe"/would+have+3rd+pers; final word predicted in the other auxiliary legal form "è"/is. Frequency List: °-dopo; °-importante; °-decisione; °-riservata; °-feste; ***<ukn>-Ghitti

Sentence 9.B - *L'importante ora è aprirlo di più. 9.Bc Ora è importante aprirlo di più.*

This sentence is perhaps too short and only function words are captured by BERT embeddings: ora/now (0.3825) più/more (0.0911). The ambiguous word "ora"/now is better predicted in the non-canonical structure - in first position - for the availability of right context - the canonical version predicts "Ora" in fourth position (0.0844). Again this is not relatable to a frequency problem but just structural problems, with the exception perhaps of the final word "aprirlo" which is only present in the very-low frequency list. In fact, in the canonical version, "aprirlo"/open+it is substituted by cliticized verbs - though semantically unrelated, however, showing that the morphology has been captured correctly. As to "importante"/important, it does not appear in the first five candidates, but it is predicted in sixth position (0.04902).

Frequency List: °-ora; *-aprirlo

Sentence 10.B - *Le sue informazioni darebbero anche agli orientamenti di democrazia laica maggiori spinte. 10.Bc Le sue informazioni darebbero maggiori spinte anche agli orientamenti di democrazia laica.*

This sentence has the same predicted word "maggiori/major" in both structural representations. As before, the words are all very frequent with the exception of "darebbero/+would+give, which is below the threshold and is only part of the "very+low" List. Now consider the word spinte/boosts: predicted masked words are as follows: certezze/certainties (0.0852), garanzie/guarantees (0.0824), informazioni/information (0.04183), taria/tary (0.04003), opportunità/opportunities (0.0383). The fourth slot contains a subword, in fact a non-word, which is assigned a score higher than the one assigned to "opportunities". The question is that the masked word is not frequent enough to be able to collect the co-occurrences required. As a result, even very low scored embeddings are considered. The non-word gets a slightly better score when the text is considered as a whole with the last 7 sentences added, up to (0.06002), but remains always in fourth position.

Frequency List: °-anche; °-informazioni; °-sue; °-maggiori; °-democrazia; °-orientamenti; °-laica; *-spinte; *-darebbero

Sentence 11.B - *In questo libro Maria Teresa, spiegano alla Mondadori, darà esempi di carità concreti. 11.Bc In questo libro Maria Teresa darà esempi di carità concreti, spiegano alla Mondadori.*

In this sentence there is a striking difference in prediction between the two structures. The non-canonical version has only two words predicted, "libro/book" and "esempi/examples", libro (0.0242), esempi (0.653). On the contrary, in the canonical version BERT manages to predict four words, "questo/this", "Maria/Mary", "Teresa/Therese", "esempi/examples", questo (0.767), Maria (0.283), Teresa (0.141), esempi (0.734). Strangely enough, the word "libro" does not figure in the first five candidates. Useless to say, the remaining words are all very frequent. The third run with a longer text including the following 7 sentences gives interesting results: "Teresa" now becomes first candidate substituting the previously chosen first candidate "ci"/us. The word "esempi"/examples, predicted as first candidate, in the text is followed by "carità"/charity which is not predicted in both version: in its place, the first candidate is again "esempi", thus certifying that predictions are made one word at a time disregarding the textual context. Now consider the adjective "concreti" which has been dislocated and is disjoined from its head, "esempi". The list of five candidates for the canonical version is the following: "cristiana+fem+sing"/Christian (0.1919), ' (0.0909), ' (0.0387), "civile+sing"/civil (0.0383), "esemplare+sing"/exemplar (0.0222). None of the candidates is plural in number as it should be, if the morphology of Italian has to be respected.

On the contrary, the first candidate agrees both in number and gender with the preceding word “carità+fem+sing”/charity, which is not to be considered the correct nominal head. The non-canonical version has one punctuation mark less and an additional adjective “pastorale+sing”/pastoral. \

Frequency List: °-questo; °-libro; °-esempi; °-carità; °-concreti; °-darà; °-spiegano; °-Mondadori

Sentence 12.A - *Disse che gli hanno il cor di mezzo il petto tolto. 12.Ac Disse che gli hanno tolto il cuore di mezzo il petto.*

This sentence from the poetry subset has only one word in common "cor/heart" and an additional word predicted in the canonical structure, "tolto/taken+off". The cosine values are all very low, cor-cuore (0.1019), for the non-canonical, and cor-cuore (0.0756), tolto (0.156) in the other structure. Interesting enough, when using the configuration with the whole text, also “mezzo/means” is predicted in second slot.

Frequency List: °-mezzo; °-cuore; °-petto; °-tolto; *-Disse

Sentence 13.A - *I ritrosi pareri e le non pronte e in mezzo a l'eseguire opere impedito. 13.Ac I ritrosi pareri e le opere non pronte e impedito in mezzo a l'eseguire.*

No prediction found by BERT in the two structural representations - with the exception of "mezzo"/means which however is only appearing in 8th position and not considered in this evaluation. However it is important to note that the previous seven predicted words are in fact only subwords, mostly meaningless, and some having a corresponding identical wordform with a totally different meaning. Here they are: "dotti"/learned+mas+plur, "dotte"/learned+fem+plur, "tente"/meaningless, "sistenti"/meaningless, "sistenza"/meaningless, "difficoltà"/difficulty, "fami"/meaningless. As to their frequency, words are mostly frequent but there are two missing words in the overall frequency lists: "ritrosi/reluctant" and "impedite/hampered". These two words may have been supplemented as subwords but with no useful context for the current analysis. The five candidates appearing are as follows: for “ritrosi” we have -suoi/his+hers, non/not, buoni/good+mas+plur, mal/bad(truncated), loro/their+them+they; and for “impedite” - ‘.’, buone/good+fem+plur, inutili/useless+plur, nuove/new+fem+plur, pubbliche/public+fem+plur. In all of these cases, even if the correct word has not been predicted, the morphology has been matched correctly.

Frequency List: °-mezzo; °-opere; °-pareri; °-eseguire; °-pronte; ***ritrosi; ***impedite

Sentence 14.A - *Un'eco di mature angosce rinverdiva a toccar segni alla carne oscuri di gioia. 14.Ac Un'eco di mature angosce rinverdiva a toccar segni di gioia oscuri alla carne.*

This is another sentence from poetry domain very hard to tackle and to understand. Both the canonical and the non-canonical analyses have just one word found, "eco/echo" (0.0984). Of course the main verb "rinverdiva" is not amongst the frequent words in the list: in fact, it is missing. The remaining words are frequent but they are organized in a peculiar structural configuration with the declared aim to produce metaphors. No changes or improvements when the sentence is analysed with the canonical version of the text. As we did for example 11, we now consider the discontinuous adjective “oscuri+mas+plur”/obscure and the morphology of the five candidates predicted. In the non-canonical version we have: “pieni+mas+plur”/full (0.5461), “piena+fem+sing”/full (0.0486), “e”/and, ‘.’, “pieno+mas+sing”/full (0.0216). Now the canonical version: “fino”/until (0.1139), “intorno”/around (0.1139), “dentro”/inside (0.1001), “sino”/until (0.0476), “vicino”/close (0.0437). As can be noticed, all of the predicted words for the non-canonical structure are function words and none – with the possible exclusion of the ambiguous “vicino+mas+sing” - is an adjective. The reason for this lack of grammatical match may be due to the presence of the articulated preposition “alle”/to the+fem+plur in the canonical version. In the non-canonical version the word “oscuri” was followed by a preposition “di” which is the most frequent wordform with 65 million occurrences.

Frequency List: °-alla; °-carne; °-gioia; °-segni; °-toccare; °-eco; *-oscuri; *-mature; *-angosce; ***rinverdiva

Sentence 15.B - *Il governo, quindi, pur rinunciando alla maggioranza assoluta, ha voluto, come già nell'IMI, puntare a una privatizzazione graduale. 15.Bc Quindi, il governo ha voluto puntare a una privatizzazione graduale pur rinunciando alla maggioranza assoluta come già nell'IMI.*

This long sentence belongs to the domain of the news and even in its non-canonical structure, it is more linear and thus more predictable. There are seven words predicted (over ten we masked) in the two versions: governo/government (0.304), maggioranza/majority (0.0377), assoluta/absolute (0.349), ha/has

(0.977), voluto/wanted (0.491), puntare/aim (0.0385). The proper name IMI is in the very low list. Strangely enough the function word come/like (0.1925/0.9186) is predicted as first candidate in its non-canonical position, as second position, but with a much lower cosine measure in canonical position.

Frequency List: °-governo; °-maggioranza; °-voluto; °-assoluta; °-puntare, °-privatizzazione; °-graduale; *-rinunciando; **-IMI

Sentence 16.B - *In una conferenza al Viminale il ministro, quando viene interrogato sul senatore a vita, sulle prime non capisce il nome. 16.Bc In una conferenza al Viminale, quando viene interrogato sul senatore a vita sulle prime il ministro non capisce il nome.*

There are four words predicted in this long sentence, again in the domain of the news, in the canonical and the non-canonical structures. They are: ministro/minister (0.497), viene (0.795), senatore/senator (0.808), vita/life (0.996). Again, most words are very frequent. An apparent difficulty is constituted by presence of a multiword: "sulle prime/at first" which may be hard to distinguish and differentiate on the basis of the context. In fact, in both structures, "prime" is substituted by riforme/reforms, banche/banks, dimissioni/resignation, pensioni/pensions, cose/things.

Frequency List: °-vita; °-viene; °-nome; °-ministro; °-prime; °-senatore; °-conferenza; °-capisce; *-interrogato; *-Viminale

Sentence 17.B - *Primo intervento da fare, ha detto in questi giorni, è di attuare la riforma. 17.Bc Primo intervento da fare è di attuare la riforma, ha detto in questi giorni.*

This is another fairly simple sentence which has the major number of predicted words in the whole set in relation to the total number in the sentence. There are six words predicted both in the canonical and the non-canonical version: "fare/do" (0.818), "ha/has" (0.283), questi/these (0.961), giorni/days (0.83), riforma/reform (0.194). The only difference being the slot assigned to riforma/reform, which has first slot in the canonical version and second slot in the non-canonical one, preceded by Costituzione/Constitution. Useless to say, the missing words are all very frequent.

Frequency List: °-fare; °-giorni; °-detto; °-intervento; °-riforma; °-Primo; °-attuare

Sentence 18.B - *Io il privato lo concepisco come un metodo di lavoro, come contratti di lavoro, come modo di gestire insomma. 18.Bc Io concepisco il privato come un metodo di lavoro, come contratti di lavoro, come modo di gestire insomma.*

In this final sentence again belonging to the newswire domain, there are four words predicted: metodo/method (0.0618), lavoro/work (0.214), lavoro/work (0.214), modo/way (0.794). Again very frequent missing words, apart from "concepisco/surmise" which is the only word present in the Rare-Words list. When analyzed with the canonical version of the text, the word lavoro/work moves from third to first slot, with a slightly improved cosine score.

Frequency List: °-lavoro; °-modo; °-Io; °-contratti; °-privato; °-metodo; °-insomma; °-gestire; ***-concepisco.

AN INTELLIGENT AND SOCIAL-ORIENTED SENTIMENT ANALYTICAL MODEL FOR STOCK MARKET PREDICTION USING MACHINE LEARNING AND BIG DATA ANALYSIS

Muqing Bai¹ and Yu Sun²

¹Brookfield Academy, 3215 N Brookfield Rd, Brookfield, WI 53045

²California State Polytechnic University, Pomona, CA, 91768, Irvine, CA 92620

ABSTRACT

In an era of machine learning, many fields outside of computer science have implemented machine learning as a tool [5]. In the financial world, a variety of machine learning models are used to predict the future prices of a stock in order to optimize profit. This paper proposes a stock prediction algorithm that focuses on the correlation between the price of a stock and its public sentiments shown on social media [6]. We trained different machine learning algorithms to find the best model at predicting stock prices given its sentiment. And for the public to access this model, a web-based server and a mobile application is created. We used Thinkable, a powerful no code platform, to produce our mobile application [7]. It allows anyone to check the predictions of stocks, helping people with their investment decisions.

KEYWORDS

Stock, Machine learning, Thinkable.

1. INTRODUCTION

Stock market plays an important role in the modern world [8]. It facilitates and efficiently allocates resources to the industries, or more specifically, companies who show promising outlook. There is no doubt that one can make a lot of money by investing in the stock market, which incentivize many to study the stock market, in hope of finding a better investment strategy [9]. Our application uses machine learning to figure out a stock's trend and then make a prediction on its price. The system can be run automatically, in comparison to analyzing a stock manually. This method is more advanced, efficient, and is friendly to the common users.

Many research papers have discussed the different factors that can affect a stock's prices [3], and the different methods to approach stock prediction [1]. Some focus on creating a more accurate sentiment analysis model. They implemented a "novel sentiment index", which weights each stock review differently and takes in consideration of many other potential effects on a stock's price [4]. However, these sources of data might not be good representations of the public, since the weight of each review is not equal.

As previous studies suggest, taking social sentiment into consideration can help improve stock predictions [3]. Our project follows the path of those who used sentiment analysis to predict stock

prices. The difference is that we do sentiment analysis on any user's tweet from Twitter.com. We believe that Twitter.com is able to offer more globalized opinions of everyday investors. We then use TextBlob, a Natural Language Processing library, to analyze the tweets [10]. The number of tweets who held positive, negative, or neutral outlook on the stock is recorded respectively in a realtime database. We then put the data into a machine learning model, which is hosted on a web-server, to get a prediction. The web server allowed us to make something new — a simple, easy to use mobile application that is designed for everyday investors. It communicates with the web server and displays the prediction of a stock's price change when its ticker symbol is entered into the application by the user. Our application aims to help users make their investment decisions.

For the above-mentioned machine learning model, specifically, we are using Scikit-learn's Linear Regression model. We will prove to you that this model performs better than polynomial regression and Bayesian regression on a dummy dataset. And we will show you the relevance of each factor in the machine learning process.

The rest of the paper is structured as follows: Section 2 list out the challenges we encountered while doing this research; Section 3 focuses on the details of our solutions corresponding to the challenges that are mentioned in Section 2; Section 4 explains the experiments we did, which also acts as a supplement to Section 3 as it details part of the solution to the aforementioned challenges; Section 5 presents related works; Section 6 provide the conclusion and end this paper with future works.

2. CHALLENGES

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

2.1. Performing Large-Scale Data Collection and Sentiment Analysis

In order to build and train an accurate machine learning algorithm, a large amount of data is needed, since if we are only taking the comments from a couple people regarding the stock, the overall sentiment can be highly biased because each person has a great weight to the overall sentiment [11]. And on top of that, we need to be able to consistently extract data, since the project is aimed to be updating on a daily basis.

When the data is gathered, it will then be analyzed using a Natural Language Processing algorithm. The difficult part is to figure out which algorithm to use. Since our data is specific to stock trading, it has special terminologies and phrases such as "AAPL to the moon!!!" These may be difficult to analyze.

2.2. Optimizing the Prediction Model with High Accuracy

We were in face of a big problem — which machine learning model should we use to make the prediction? Also, which variables are actually relevant to a stock? Is it the public sentiment? The price? or other variables?

We conducted a total of three experiments to figure out the aforementioned questions. Experiment 1 compares the result of three different types of machine learning model. Experiment 2 compares the result of 3 different configurations of the polynomial-regression. Experiment 3 compares the accuracy of the linear-regression model with different categories of input data cut off. These experiments will be explained in detail in Section 4.

2.3. Making a User-Friendly Interface to Display the Data

The goal of our project is to help casual investors invest in the stock market. So naturally, there has to be a way that the end-users can access our prediction results of the stocks. Thus we created a mobile application. We aim to have a simplistic design and a user-friendly interface for the application. This part is further discussed in the Solution, Section 3.2.4.

3. SOLUTION

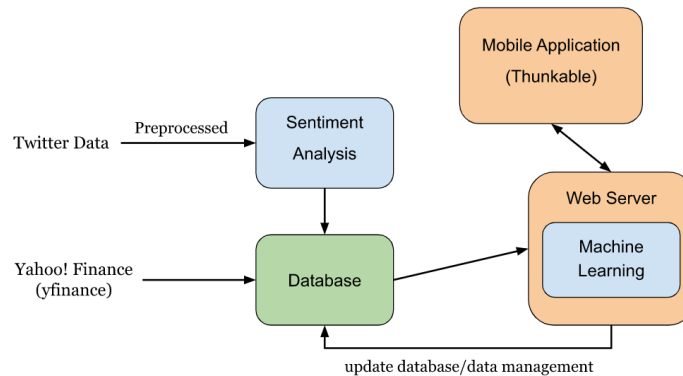


Figure 1. Overview of the solution

3.1. Web Scraping

In our case, Twitter.com is used to gather social sentiments. With Twitter's API, we can obtain the most recent 100 tweets about a particular stock. As shown in Figure 2, after authenticating using Twitter's API, the program passes the company's name or the company's ticker symbol as the keyword (query) to search for 100 matching tweets using the `.search_recent_tweets()` method.

```

38 # Twitter API Initialization
39 client = tweepy.Client(bearer_token=config.BEARER_TOKEN)
40 # Query
41 if 'coName' in locals():
42     response = client.search_recent_tweets(query=coName, max_results=100)
43 else:
44     response = client.search_recent_tweets(query=tickerS, max_results=100)
45
  
```

Figure 2. Screenshot of code 1

Alongside the tweets, the information about the stock, such as the market-closing price and the date, as well as the company's full name, are all collected through Yahoo! Finance's API. These data are stored in the database, which will be discussed in section 3.2.3.

3.2. Sentiment Analysis

After gathering the tweets, we need to analyze and distinguish them into 3 categories based on their polarity: positive, negative, and neutral. In order to achieve this, we need a Natural Language Processing (NLP) algorithm. Initially, we set up a model using TensorFlow, the configuration is as follows.

```
[ ] # building the neural nw model
model = tf.keras.Sequential([
    tf.keras.layers.Embedding(vocab_size, embedding_dim, input_length=max_len),
    tf.keras.layers.GlobalAveragePooling1D(),
    tf.keras.layers.Dense(24, activation='relu'),
    tf.keras.layers.Dense(1, activation='sigmoid')
])
model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
```

Figure 3. Screenshot of code 2

```
model.summary()
```

Layer (type)	Output Shape	Param #
embedding_16 (Embedding)	(None, 100, 256)	2560000
global_average_pooling1d_16 (GlobalAveragePooling1D)	(None, 256)	0
dense_32 (Dense)	(None, 24)	6168
dense_33 (Dense)	(None, 1)	25

Total params: 2,566,193
 Trainable params: 2,566,193
 Non-trainable params: 0

Figure 4. Screenshot of model summary

However, it performed poorly with merely 26% accuracy [12]. But do note that the dataset itself may not be accurate.

```
Epoch 26/50
135/135 - 4s - loss: -7.9707e+02 - accuracy: 0.2636 - val_loss: -1.3482e+02 - val_accuracy: 0.1993 - 4s/epoch - 30ms/step
Epoch 27/50
135/135 - 4s - loss: -8.8439e+02 - accuracy: 0.2657 - val_loss: -1.5342e+02 - val_accuracy: 0.2067 - 4s/epoch - 30ms/step
Epoch 28/50
135/135 - 4s - loss: -9.7694e+02 - accuracy: 0.2696 - val_loss: -1.6950e+02 - val_accuracy: 0.2073 - 4s/epoch - 30ms/step
Epoch 29/50
135/135 - 4s - loss: -1.0745e+03 - accuracy: 0.2674 - val_loss: -1.7475e+02 - val_accuracy: 0.1920 - 4s/epoch - 31ms/step
Epoch 30/50
135/135 - 4s - loss: -1.1769e+03 - accuracy: 0.2636 - val_loss: -1.9569e+02 - val_accuracy: 0.1973 - 4s/epoch - 30ms/step
Epoch 31/50
135/135 - 4s - loss: -1.2862e+03 - accuracy: 0.2594 - val_loss: -2.2550e+02 - val_accuracy: 0.2120 - 4s/epoch - 30ms/step
Epoch 32/50
135/135 - 4s - loss: -1.3975e+03 - accuracy: 0.2713 - val_loss: -2.3868e+02 - val_accuracy: 0.2053 - 4s/epoch - 31ms/step
Epoch 33/50
135/135 - 4s - loss: -1.5166e+03 - accuracy: 0.2597 - val_loss: -2.6920e+02 - val_accuracy: 0.2160 - 4s/epoch - 30ms/step
Epoch 34/50
135/135 - 4s - loss: -1.6378e+03 - accuracy: 0.2751 - val_loss: -2.7285e+02 - val_accuracy: 0.1987 - 4s/epoch - 30ms/step
Epoch 35/50
135/135 - 4s - loss: -1.7668e+03 - accuracy: 0.2662 - val_loss: -2.7654e+02 - val_accuracy: 0.1907 - 4s/epoch - 30ms/step
Epoch 36/50
135/135 - 4s - loss: -1.9017e+03 - accuracy: 0.2562 - val_loss: -3.1320e+02 - val_accuracy: 0.1993 - 4s/epoch - 30ms/step
Epoch 37/50
135/135 - 4s - loss: -2.0404e+03 - accuracy: 0.2702 - val_loss: -3.4411e+02 - val_accuracy: 0.2060 - 4s/epoch - 30ms/step
Epoch 38/50
135/135 - 4s - loss: -2.1822e+03 - accuracy: 0.2697 - val_loss: -3.6301e+02 - val_accuracy: 0.2020 - 4s/epoch - 30ms/step
Epoch 39/50
135/135 - 5s - loss: -2.3356e+03 - accuracy: 0.2613 - val_loss: -3.5802e+02 - val_accuracy: 0.1847 - 5s/epoch - 35ms/step
Epoch 40/50
135/135 - 4s - loss: -2.4907e+03 - accuracy: 0.2615 - val_loss: -3.9627e+02 - val_accuracy: 0.1920 - 4s/epoch - 30ms/step
Epoch 41/50
135/135 - 4s - loss: -2.6543e+03 - accuracy: 0.2699 - val_loss: -4.5797e+02 - val_accuracy: 0.2140 - 4s/epoch - 30ms/step
Epoch 42/50
135/135 - 4s - loss: -2.8206e+03 - accuracy: 0.2718 - val_loss: -4.4996e+02 - val_accuracy: 0.1920 - 4s/epoch - 30ms/step
Epoch 43/50
135/135 - 4s - loss: -2.9959e+03 - accuracy: 0.2683 - val_loss: -4.7724e+02 - val_accuracy: 0.1933 - 4s/epoch - 30ms/step
Epoch 44/50
135/135 - 4s - loss: -3.1782e+03 - accuracy: 0.2615 - val_loss: -5.4207e+02 - val_accuracy: 0.2107 - 4s/epoch - 30ms/step
Epoch 45/50
135/135 - 4s - loss: -3.3650e+03 - accuracy: 0.2669 - val_loss: -5.5043e+02 - val_accuracy: 0.2033 - 4s/epoch - 30ms/step
Epoch 46/50
135/135 - 4s - loss: -3.5617e+03 - accuracy: 0.2622 - val_loss: -5.8427e+02 - val_accuracy: 0.2027 - 4s/epoch - 30ms/step
Epoch 47/50
135/135 - 4s - loss: -3.7669e+03 - accuracy: 0.2583 - val_loss: -5.8690e+02 - val_accuracy: 0.1907 - 4s/epoch - 30ms/step
Epoch 48/50
135/135 - 4s - loss: -3.9742e+03 - accuracy: 0.2709 - val_loss: -5.9067e+02 - val_accuracy: 0.1780 - 4s/epoch - 33ms/step
Epoch 49/50
135/135 - 4s - loss: -4.1849e+03 - accuracy: 0.2555 - val_loss: -6.7279e+02 - val_accuracy: 0.2007 - 4s/epoch - 30ms/step
Epoch 50/50
135/135 - 4s - loss: -4.4108e+03 - accuracy: 0.2606 - val_loss: -7.3938e+02 - val_accuracy: 0.2107 - 4s/epoch - 30ms/step
```

Figure 5. Result accuracy

Then we tried using a RoBERTa model that was pre trained on 124 million tweets from January 2018 to December 2021. After testing it on a dataset of 1300 stock related tweets, the resulting accuracy is better, at about 55%. But still not a favorable result.

```
0.546923076923077
{'Positive': 317, 'Neutral': 846, 'Negative': 137}
```

Figure 6. Results from the RoBERTa model

Eventually, we choose the more accessible and easier to use TextBlob as the NLP model. It is also pretrained. We pre-process the text for better results, which removes “RT”, words that include “@” or “#”, “https” hyper links, and any punctuation. This text is then fed to Textblob for sentiment analysis and returns a BaseBlob object. BaseBlob.polarity then returns a float between -1.0 and 1.0, where -1.0 is very negative, 1.0 is very positive, and 0.0 is defined as neutral. These data are then sent and stored in the database, ready to be accessed.

3.3. Database and Machine Learning

As mentioned in the previous sections, the backbone of this project, the database, stores a stock’s prices and its sentiment analysis by date [13]. It has a data structure where the information about the stock is arranged by their date and is listed as subsets of that stock, as shown in Figure 7. The dataset can store multiple entries of stocks at the same time. It communicates with the web server frequently.



Figure 7. Screenshot of database

Before doing machine learning, we first request data from the database. A machine learning model is created for each individual stock. So we extract the data under that specific stock’s ticker name. The data we use as inputs are the dates and the 3 types of sentiments: positive, negative, and neutral. They are put into a list, which is then put into the 2D array called `input_data` (Figure 7). `input_data` contains all the data from previous days except for today’s, because today’s data doesn’t have a corresponding output. We will be using today’s data as the input for our prediction once the model is trained appropriately. The `output_data`, however, has only one category, which is the change in price of the stock in the next day. Essentially, it takes the price of the corresponding date and the next price value on the list to find the change in price from the day where the sentiment values are recorded to the next day. This combo of `input_data` and `output_data` is now our training data set for the linear regression machine learning model. We are using linear regression because it is the best performing model, proven in previous experiments (Section 4). After training using the method `model.fit()`, we can now predict the change in price between today and tomorrow. As shown in Figure 8, `model.predict()` produces the prediction, which is then turned into a percentage and bundled with ticker and name in the dictionary named `final_res`. `final_res` will be finally returned at the end of this method.

```

71 #
72 model = linear_model.LinearRegression()
73 model.fit(input_data, output_data)
74 res = model.predict([today_data])[0]
75 print(res)
76
77 # add stock info to join prediction
78 final_res = {
79     "prediction" : str(round(res[0] * 100, 2)) + "%",
80     "ticker" : ticker,
81     "name" : name
82 }

```

Figure 8. Screenshot of code 3

3.4. Web Server and Mobile App

A simple Flask server is set up and open for access. It facilitates the database and hosts the machine learning models. In Figure 9 is the setup of the web server. Its first method *stockUpdate* simply updates all the stocks that are already recorded in our database. It scrapes Twitter to find the new sentiment values, along with the price of the stock at the moment of update. The second method, *stockUpdateWithList* updates the stocks that are on a specific list. And if they were never recorded in the database, then they will be added to it. Method *stockInfo* takes in a parameter, the ticker symbol of a stock, in the form of a https hyperlink. It goes through the process of machine learning and predicting — a dictionary structured like *result* from Figure 9 is returned.

```

1 import random
2 from flask import Flask
3 import tweet
4 import data_processor
5 import json
6
7 app = Flask(__name__)
8
9
10 @app.route("/stockUpdate")
11 def stockUpdate():
12     data_processor.updateAllStocks()
13     return "Updated"
14
15 @app.route("/stockUpdateWithList")
16 def stockUpdateWithList():
17     data_processor.updateAllStocksWithList()
18     return "Updated with List"
19
20 @app.route("/stockInfo/<ticker>")
21 def stockInfo(ticker):
22     ticker = ticker.upper()
23     if ticker not in all_stocks:
24         result = {
25             "prediction" : "N/A",
26             "ticker" : ticker,
27             "name" : "Stock Data N/A"
28         }
29     else:
30         result = data_processor.getStockData(ticker)
31     return json.dumps(result)
32
33
34 stock_list_file = open("stock_list.txt")
35 all_stocks = []
36 for line in stock_list_file.readlines():
37     all_stocks.append(line.strip())
38
39 app.run(host = '0.0.0.0')

```

Figure 9. Screenshot of code 4

The web server allowed us to make a mobile application. It is targeted for people who don't know anything about coding or even investing. Thus, the app uses a very simplistic user interface.

This app is created with Thunkable. As shown in Figure 10, the panel on the left is our first screen, where the user types in a ticker symbol and presses the Search button, it navigates to the 2nd panel and sends a get request to our web server to get the information regarding this stock, including the predictions. After the web server responds, the information will be displayed on panel 2 in the order of *prediction*, *name*, and *lastly ticker*.

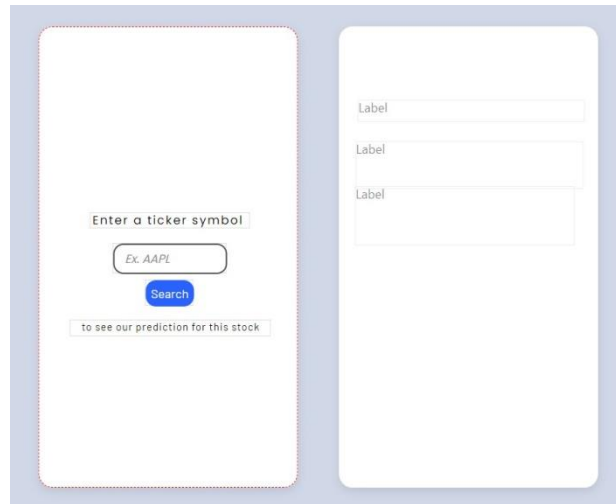


Figure 10. Home screen

Do note that Thunkable is a non coding app creation platform. They use logic statements to substitute for coding. Figure 11 shows the “code” for the first panel.



Figure 11. Screenshot of Thunkable

4. EXPERIMENT

4.1. Experiment 1

Certain aspects of machine learning can significantly improve its performance/accuracy if the right module is used. In the following experiments, we will look at the different factors that contribute to the accuracy of a machine learning module, and figure out which has the best performances.

All the following tests involve data that are artificially manipulated. Since we couldn't gather a big enough dataset of the real world twitter users' sentiments on certain stocks, we are using a dummy dataset to train the algorithms; and ultimately test each algorithm on their accuracy. The dataset consists of 2 categories: input data and output data. Input data includes Positive Sentiment, Negative Sentiment, Neutral Sentiment, Month, and Day. Positive Sentiment, Negative Sentiment, and Neutral Sentiment each takes a randomly generated integer between 0 and 50, while Month takes an integer between 1 and 12, and Day takes an integer between 1 and 30. The output data, however, is a single number that is the sum of Positive Sentiment, Negative Sentiment, and

Neutral Sentiment plus a randomly generated integer between -15 and 15. (should I explain why the output data doesn't take in Day and Month as its factors?)

We are using a relatively small batch of 100 generated samples of dummy data for each trial in the following experiments (add test results from 1000 data samples later)

The first experiment compares the accuracy between 3 different machine learning models: linear regression, polynomial regression, and Bayesian regression. As shown in Figure 12, polynomial regression (in green) slightly under-performed compared to linear regression and Bayesian regression. Do note that linear regression and Bayesian regression have very identical accuracy in each of the 20 trials, although they are not exactly the same.

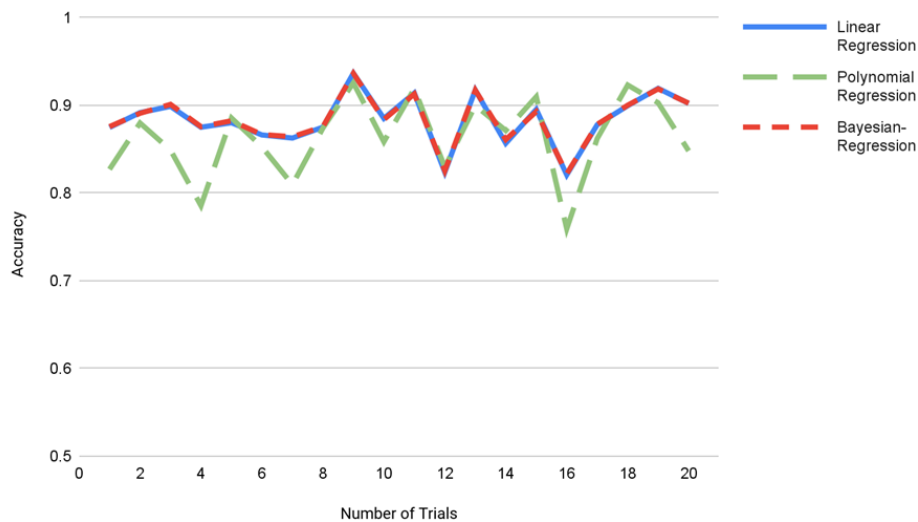


Figure 12. Graph of accuracy vs number of trials

However, the difference in accuracy between these 3 will turn to be neglectable when a bigger sample size is used.

4.2. Experiment 2

The second experiment focuses on polynomial regression. We want to find out which of its configurations produces the most favorable result. To give a short explanation, the *sklearn* polynomial regression takes in a Degree parameter and “generates a new feature matrix consisting of all polynomial combinations of the features with degree less than or equal to the specified degree.”

In Figure 13, we can see that *Degree 2* is performing significantly better than the other two, considering that the best accuracy we can get is 1; however it can go far down into the negatives. The farther it is into the negatives, the less accurate the model is. To better display all the values in one graph, we dropped out the result from Trial 15 of *Degree 4*, which came in at an astonishing -399.005058028454 accuracy rating. We can see that *Degree 3* and *Degree 4*'s accuracy not only underperform, but also fluctuate a lot throughout the trials, especially for *Degree 4*. Both of their results came out as unfavorable, staying below 0 for the most part. *Degree 2*'s performance is similar to the polynomial regression model in experiment 1, since that is also using degree 2, which is the default configuration for polynomial regression.

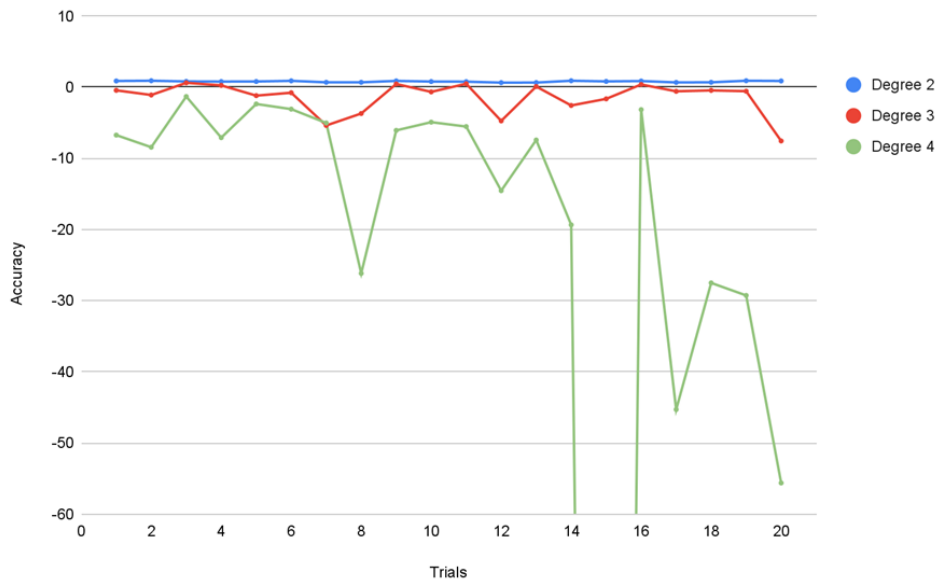


Figure 13. Graph of accuracy vs trials

4.3. Experiment 3

Figure 14 shows the result from experiment 3. In this experiment, we are trying to see what happens when we emit one of the five inputs. We are using a linear regression model for this experiment.

By emitting Month or Day, the accuracy stays virtually the same. Emit Month yields an average accuracy of 0.8614 from 20 trials, Emit Day yields 0.8595. Both are very close to the average accuracy of linear regression from experiment 1, which turns out to be 0.8836. This means that the output is likely not affected by neither Month or Day, which matches with our process of creating the dummy data. But it also proves that the linear model can function well even when one dimension (is it dimension? or) of the data is missing. But do note that Emit Month and Emit Day fluctuate with a wider range (at about 0.25) than that of the regular linear regression with the full dataset from experiment 1 (at a little more than 0.1). Still, these two results dusted the rest in their accuracy. Below them are the results from Emit Positive Sentiment, Emit Negative Sentiment, and Emit Neutral Sentiment, which are very inconsistent in their accuracy. Notably, Emit Neutral Sentiment displays a wider range of accuracy, although the three types of sentiment have the same weight when creating their matching outputs.

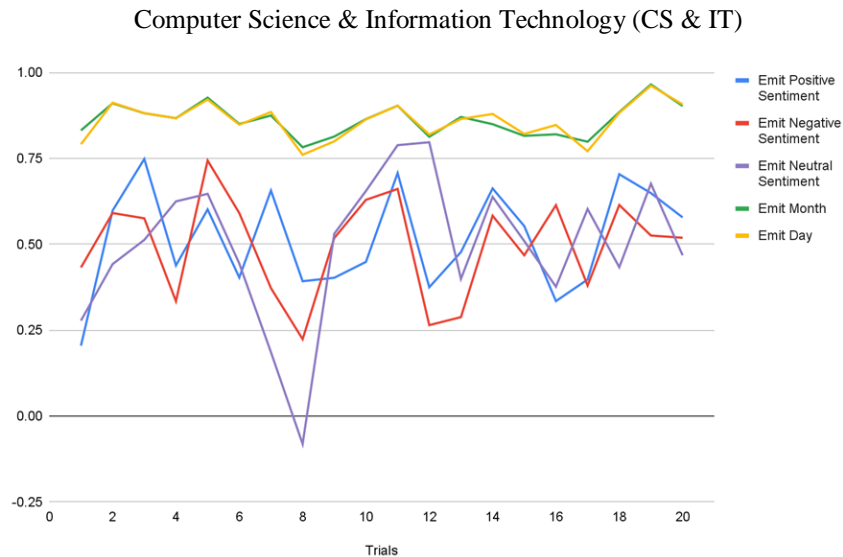


Figure 14. Graph of trials

5. RELATED WORK

Nguyen, T. H., & Shirai, K. [2] introduce a new feature called the TSLDA to the prediction model. The TSLDA model captures the topic and sentiment simultaneously. Their model was able to outperform a prediction model that only uses historical prices by 6.07% in accuracy. However, their work is different from ours in that they “chose SVM with the linear kernel as the prediction model”, which only predicts if the stock price goes up or down. In comparison, we used a linear regression model to predict the % change in the stock price.

Qiu, Y., Song, Z., & Chen, Z. [3] introduce a sentiment index that assigns different weights to different sources of sentiments. On top of that, their model takes in consideration stock market anomalies: the day-of-the-week effect and holiday effect. In comparison, our model takes in data as an input variable, which with a large amount of training data, the model can find out about these effects on its own. But do note that our paper doesn't include a system that assigns weights to the sentiments. Every tweet weighs the same.

Porshnev, A., Redkin, I., & Shevchenko, A. [4] analyzed Twitter user's sentiment in a detailed way: they “evaluate presence of eight basic emotions” in each tweet, which is then used to predict DJIA and S&P500 indicators [14]. While their work is focused on predictions on two of the most received stock market indicators, our application allows the public to search up any stock that is available in Yahoo! Finance. On top of that, our method simply analyzes the polarity of the tweets and divides them into 3 different categories: positive sentiment, negative sentiment, and neutral sentiment, instead of analyzing detailed emotions of each tweet like Porshnev and others did in their study.

6. CONCLUSIONS

The stock market is constantly evolving, to which machine learning might be its best suite for making predictions [15]. Among previous approaches that involve machine learning, social sentiment based stock predictions are definitely proven to have increased accuracy on stock price predictions. By experimenting with a dummy dataset, we found the best machine learning model to be linear regression, despite that Bayesian regression produced almost identical results. By

implementing the best performing machine learning model, and with the easy-to-use mobile application, more investors can now receive more accurate stock market predictions.

However, the dummy data that we used is artificially manipulated, which means that we assumed a relationship between the sentiments and stock prices. That is why the accuracy is as high as 0.8836. But due to the policy of Twitter not allowing us to obtain past tweets from any period of time, our actual dataset is too small to produce sufficient outcome. We had to work with the limitation and resulted in dummy data.

But in the future, once we gathered a big enough dataset. We can then perform Experiment 3 again. By cutting out a category of data and seeing how that affected the accuracy of prediction, we can tell how relevant that variable is to the price of the stock.

REFERENCES

- [1] Agrawal, J. G., Chourasia, V., & Mitra, A. (2013). State-of-the-art in stock prediction techniques. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 2(4), 1360-1366.
- [2] Nguyen, T. H., & Shirai, K. (2015, July). Topic modeling based sentiment analysis on social media for stock market prediction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 1354-1364).
- [3] Qiu, Y., Song, Z., & Chen, Z. (2022). Short-term stock trends prediction based on sentiment analysis and machine learning. *Soft Computing*, 26(5), 2209-2224.
- [4] Porshnev, A., Redkin, I., & Shevchenko, A. (2013, December). Machine learning in prediction of stock market indicators based on historical data and data from twitter sentiment analysis. In *2013 IEEE 13th International Conference on Data Mining Workshops* (pp. 440-444). IEEE.
- [5] Jordan, Michael I., and Tom M. Mitchell. "Machine learning: Trends, perspectives, and prospects." *Science* 349.6245 (2015): 255-260.
- [6] Hendler, Glenn. *Public sentiments: Structures of feeling in Nineteenth-Century American literature*. Univ of North Carolina Press, 2001.
- [7] Hoehle, Hartmut, and Viswanath Venkatesh. "Mobile application usability." *MIS quarterly* 39.2 (2015): 435-472.
- [8] Aggarwal, Rajesh K., and Guojun Wu. "Stock market manipulations." *The Journal of Business* 79.4 (2006): 1915-1953.
- [9] Gidofalvi, Gyoza, and Charles Elkan. "Using news articles to predict stock price movements." *Department of Computer Science and Engineering, University of California, San Diego* (2001): 17.
- [10] Gujjar, J. Praveen, and Prasanna Kumar HR. "Sentiment analysis: Textblob for decision making." *Int. J. Sci. Res. Eng. Trends* 7.2 (2021): 1097-1099.
- [11] Phan, Xuan-Hieu, Le-Minh Nguyen, and Susumu Horiguchi. "Learning to classify short and sparse text & web with hidden topics from large-scale data collections." *Proceedings of the 17th international conference on World Wide Web*. 2008.
- [12] Redman, Thomas C. "Measuring data accuracy: A framework and review." *Information quality*. Routledge, 2014. 33-48.
- [13] Dubitzky, Werner, et al. "The open international soccer database for machine learning." *Machine Learning* 108.1 (2019): 9-28.
- [14] Porshnev, Alexander, Ilya Redkin, and Alexey Shevchenko. "Machine learning in prediction of stock market indicators based on historical data and data from twitter sentiment analysis." *2013 IEEE 13th International Conference on Data Mining Workshops*. IEEE, 2013.
- [15] Kourou, Konstantina, et al. "Machine learning applications in cancer prognosis and prediction." *Computational and structural biotechnology journal* 13 (2015): 8-17.

TOWARDS DEVISING A FUND MANAGEMENT SYSTEM USING BLOCKCHAIN

Nibula Bente Rashid, Joyeeta Saha, Raonak Islam Prova,
Nowshin Tasfia, Md. Nazrul Huda Shanto and Jannatun Noor,

School of Data and Sciences, BRAC University, Dhaka, Bangladesh

ABSTRACT

State government operations comprise a large number of transactions for different processes that must be carried out across the state. This comprises new projects, maintenance and repairs, public employee compensation, and agricultural schemes. Low-level corruption, which is sometimes difficult to trace and hinders state growth, is a big challenge for the top administration. In order to eradicate corruption and bring transparency, technology can be used in an efficient way. An important task to exterminate corruption is to keep track of all the financial transactions of an undergoing project. This research uses blockchain technology to keep track of fund management systems and assure the transparency of any financial statement. This paper proposes to use a gateway where all transaction records are updated in the system and visible to all stakeholders. We find research gaps in the literature and focus on including government funds and local currency usage. The proposed model's motive is to generate a funding model that attains two sub-goals: designing a fund management methodology in which authorized individuals can receive and withdraw allocated funds in crypto currency, and evaluating a smart contract to incorporate the money and identify transparency and tracking. The proposed model executes every feature of our system in just 8.3786ms on average.

KEYWORDS

Blockchain, Ethereum, Smart contract, Government Funding.

1. INTRODUCTION

World economic system is driven on the basis of faithfulness and loyalty. People choose reliable and profitable institutions and media for investment. However, in developing countries like Bangladesh, due to corruption people have lost their faith in the country's financial system. It obstructs the development of the country and creates economic contortion in different public sectors. According to researchers, 1% corruption reduces 0.72% growth rate and 2% productivity of a country [1].

In order to get rid of corruption and bring transparency, we can use blockchain. An important task to exterminate corruption is to keep track of all the financial transactions of an undergoing project [25][26]. Blockchain provides transparency, security, decentralization, non-corruptibility, immutability, consistency, and speed. These characteristics are essential for a trustworthy fund management system.

In developing countries, corruption mainly takes place in the form of bribery and money embezzlement. Whenever the government or any other organization undertakes a complex project, they create funds. Corrupted people target these funds to achieve their self-gain. Often, it is seen that there are no proper records of how fund was utilized. Moreover, Bangladesh is one of the countries with the highest rates of informal payments in regard to public services (GCR 2015-2016). When getting operational licenses such as an electricity connection, almost 60% of

companies intend to make informal payments (ES 2017)[31]. Moreover, Bangladesh has made great progress in a variety of social sectors. On the other hand, corruption, nepotism, malinvestment, and misdirected funds have delayed economic growth and stopped the country from progressing. According to a study conducted by Stockholm University, Bangladesh diverts public resources to unproductive sectors, obstructs the government's ability to apply good policies, and reduces public trust in the government [32]. In Bangladesh, in a report of TIB, it is shown that there is more than 61% misuse of total allocated funds for the implementation of a forest project [2]. In another report, it is shown that about 14.36% to 76.92% corruption and misinformation occur in the climate projects [3].

According to a study conducted by the Stockholm University, corruption in Bangladesh diverts public resources to unproductive sectors, obstructs the government's ability to apply good policies, and reduces public trust in the government [20]. According to an NGO survey on everyday corruption, 66 percent of the population paid bribes to authorities in order to get basic government welfare services [21]. However, there are no existing fund management systems that include government projects. The World Bank has pulled out of a project to build Bangladesh's largest bridge, citing corruption concerns [19]. As a result, Bangladesh's image at the international level is now questionable. If the problem is not solved on an urgent basis, Bangladesh will soon be deprived of financial help from foreign organizations [4].

In this paper, the proposed method is to use blockchain technology for tracking of fund management systems for government, private, non-profit organizations, and also on individual levels. Previous papers on blockchain fund management do not include government and governmental funds together. Moreover, The proposed system uses Ethereum which is a smart contract and digital certification platform that offers Ether crypto currency. Here, payments are cryptographically validated and executed by a network of computers with equality. Furthermore, the proposed method includes the use of local currency to convert to crypto currency in the management system for ease of use. To our knowledge, no previous papers proposed to convert local currency to digital currency.

2. RELATED WORK

A fund tracking management system is needed for the authority to decrease the rate of corruption. However, corruption exists in many sectors.

2.1. Corruption in Bangladesh

In Bangladesh, corruption mainly takes place in the form of bribery and embezzlement. High officials working in a government or non-government organization are mainly responsible for corruption. Bangladesh receives about 63% of the foreign aid as loan and the rest of the aid (37%) is received as grants. In the economic year 2010-11, Bangladesh received \$1721.771 millions in terms of commitment and disbursement. OPEC, ADB, and IDA are the leading aid donor organizations [33]. Though Bangladesh has achieved substantial steps to improve assistance efficiency, doubts persist about who are the main beneficiaries of US \$1.5 billion foreign aid that the country gets each year. "Whether foreign assistance helps the nation or not is a tricky subject," said Piash Karim, a sociology and economics professor at BRAC University. "People get a relatively little portion of the entire sum, while a crooked clique of NGOs and government leaders profit. There is no way to provide proof, but there have been claims of wrongdoing in foreign-aid projects," he continued. According to Muhammad from the Jahangirnagar University, fighting corruption is the only solution to the issue. "We should assess the whole aiding process and determine what value we really get since we have been receiving foreign help for quite some time," he added. "The whole assistance system's approach is flawed.

It's time to learn how to use our own resources". If the problem is not solved on an urgent basis, our country will soon be deprived of financial help from foreign organizations [25].

Moreover, in order to transfer funds from donor to receiver, there are several payment services that offer money transferring in different currencies. Bangladesh Electronic Funds Transfer Network (BEFTN) is the first paperless digital interbank money transfer system of Bangladesh, launched in February 2011. As a lean-over checkbook clearing system, it supports both debit and credit transactions. Payroll, domestic and international remittances, welfare payments, bill payments, business dividends, security payments, corporate payments, federal tax payments, and individual payments, all types of credit transfers can be handled through this network. Similarly, it accepts debit transactions such as payments like utility bill, insurance premium, association, EMI, and so on [5].

2.2. Fund Management with Blockchain

Since funding procedure includes monetary values, it needs high security and user data privacy. Meanwhile, one of the primary components of blockchain is secure identification and anonymity, which makes blockchain technology the perfect match for funding mechanisms. The blockchain ecosystem comprises ICOs, wallets, and exchanges, which are essential parts of crypto currency transit and administration. On top of these fundamental qualities, blockchain offers infrastructure for the development of D Apps that give a user interface to its clients, in this instance contributors and recipients. Furthermore, other blockchain ecosystem elements, such as distributed ledger and distributed storage, consider making blockchain appropriate for funds collection processes by giving organizations, beneficiaries, donors, and legal authorities more control over the information and promoting data transparency [7]. Hence, blockchain may be used to establish a safe money trail [6].

Formal initiatives to make fundraising systems visible, safe, and traceable are very few, and the majority of platforms are unproductive. There have been several initiatives to model the fund gathering process using blockchain technology. A scholarly study [6], offers its own virtual currency named Charity Coin (CC) and covers the operation of charity fundraising with crypto currency. However, they focus on nonprofit organizations. There is no mention of how their system can be utilized to eradicate corruption in governmental large projects funding. Another research paper [8], has created a Hyperledger-based donation system for NGOs. Charity collection deals with dollars but anything in Hyperledger use, restricts their approach. Their system works with predefined users. However, their work is for charitable organizations. Since there is no specific donor for all the time, it changes over time. As a result, it poses a barrier to their system's usability.

In another paper [9], the authors discuss the basic mechanism to track charitable donations focusing on blockchain technology. The research work [10], proposed a prototype of a blockchain application for Government fund tracking that helps to track transactions using Hyperledger Composer. However, they do not mention which consensus algorithm they use for system reliability. As they used a permissioned blockchain, if the system is managed only by the government officials and if most of the officials come out dishonest then the system can easily be fooled. In a research paper [11], they propose a transparent tendering system using blockchain where the citizens are able to participate in tendering directly and track all fund transactions by using the combination Hyperledger Fabric and Hyperledger Composer. Because of using private networks, a restriction comes forward for the general citizens to visualize the tendering and funding procedures which creates doubt about the system's reliability. Lastly, In paper [9] The author D.Fiergbor describes the blockchain technology of the fund management system in Ghana.

Here, the author focuses on transparency, accuracy, and accountability among the fund managers. The main reason for that proposal is the security issues of data storage.

3. METHODOLOGY

In this research experiment, using blockchain technology, the proposed model is to generate a funding model that attains two goals. First, designing a fund management methodology in which authorized individuals can receive and withdraw allocated funds in cryptocurrency. Secondly, evaluating a smart contract to incorporate the money and identify transparency and traceability.

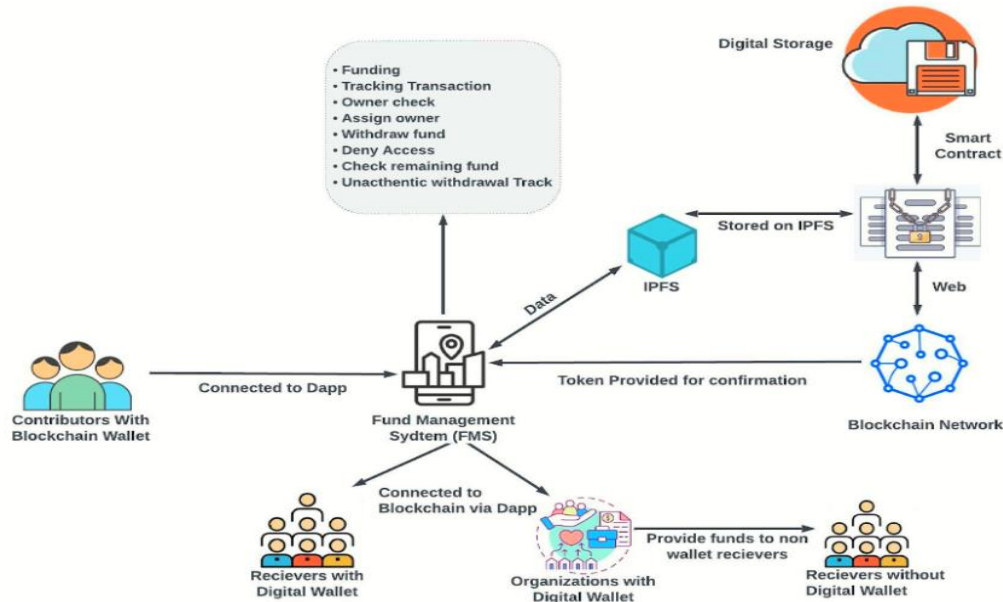


Fig. 1. Framework Architecture

3.1. Proposed Framework Architecture

The fund management system can be accessed by the system's three major users: contributors, organizations, and receivers via mobile and web-based applications using blockchain technology. In this system, we use layer 2 scaling mechanism of Ethereum, polygon. As polygon uses different side chains alongside the main chain, it boosts the transactional speed and lowers gas expenses while maintaining the system's decentralization nature as well as security. The proposed framework's high-level structure is shown below in Fig.1 In the proposed model, the person who gives funds in the system is the deployer and he/she is the owner of the transaction. The deployer has full accessibility to track and control the whole transaction system.

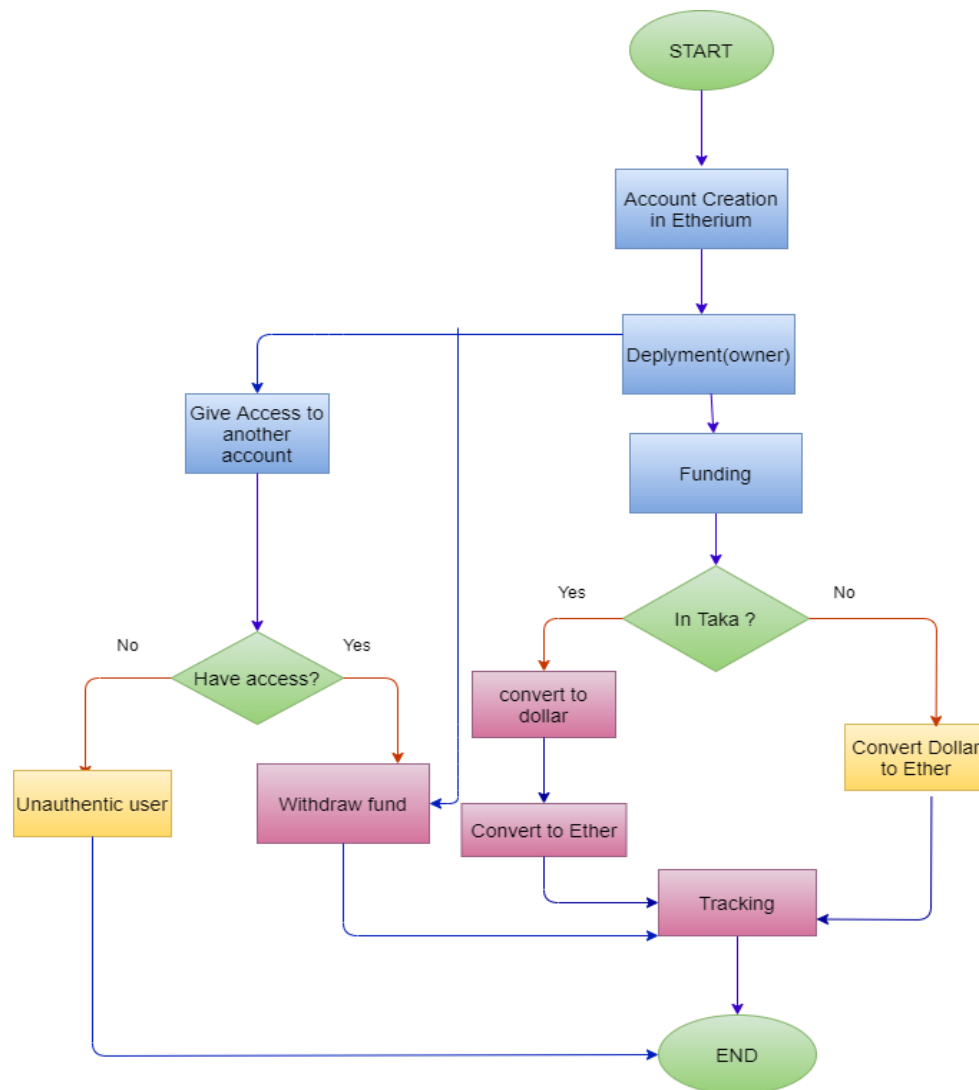


Fig. 2. Sequential Activity Model

The funder can provide money in local currency through this system. Later, the currency gets converted into crypto currency via the system. Here, only the deployer can withdraw money and also decide whom he or she wants to give access to withdraw money, along with keeping the records of the whole transaction process. The owner has the ability to delete the assigned owner if he or she wants to. If the owner sees any suspicious behavior, then the deployer can delete the suspicious user from the system. Users can check the leftover balance of the funder's account. Users can check and verify the actual owner of the current funding system by the feature. Organizations can also provide documentation by returning images of receipts to ensure that their money is properly distributed among targeted individuals.

3.2. Sequential Activity Model of Proposed Model

Fig. 2 shows the system activity model of the system. Here, deployers and authorized users assigned by the owner must register to access the system. The deployer provides access to the user. After gaining access, an authorized individual enters the system using their private key and withdraws funds. Unauthorized users get denied to access the system and get identified as inauthentic users. All other aspects are handled by system deployers, and only authorized people

can withdraw funds from the system. After funding, the currency is verified to see if it is in taka(Local Currency) or in dollar using the model. If the currency is in taka, it gets converted to dollars and subsequently to crypto currency. Else, the dollars are turned into crypto currency directly. The whole transaction is recorded and tracked by this blockchain-based system to assure authenticity of the management.

3.3. The Layer Structure Of Proposed Model

The presented framework is built on a layered architectural form, as seen in Fig. 3. Interface layer, business logic layer, application layer, transaction layer, trust layer, blockchain layer, security and administrative layer, and infrastructure layer are the eight layers we've created here.

Interface Layer: Internet-based websites and DApps of the fund management system are encapsulated in the interface layer. The purpose of this layer is to offer an interface for funders, wallet beneficiaries, and organizations. This layer is used by these individuals to initiate the donation process.

Business logic layer: Smart contracts make up the business logic layer, which deals with terms, rules, and interaction requirements. As a result, this layer might be thought of as an operational database of smart contracts, equipped with all interaction, contract activation, and execution rules.

Application layer: Online records, donations records, identity verification, and transaction information are all part of the application layer. The application layer combines the interface layers with the business logic as in kind of a smart contract. The application layer combines the interface layers with the business logic as in kind of a smart contract.

Trust Layer: The trust layer contains the smart contract's security analysis, formal identity verification, and arbitration methods like Proof-of-Work. The trust layer also interacts with transaction consensus methods and recently introduced block verification, while the blockchain layer stores the results of execution.

The blockchain layer: This layer stores data about blocks' and nodes; it also holds the distributed ledger's basic information and hashes of each transaction executed by investors, organizations, and recipients with their secret and public keys addresses.

Security and administration layer: The system's security and maintenance is assured by this layer. Several security attacks aim at blockchain, with the 51 percent attack being the most common. This layer is interconnected to the system and works in tandem with it, containing several security algorithms and protocols as well as administrative responsibilities to ensure the system's integrity.

Transaction layer: The transaction layer is in charge of transactions, conducted by smart contracts or users of the fund management system.

Infrastructure Layer: It is made up of a peer-to-peer network that verifies, forwards, and distributes the Ethereum blockchain transactions. It also covers interaction, authentication, and

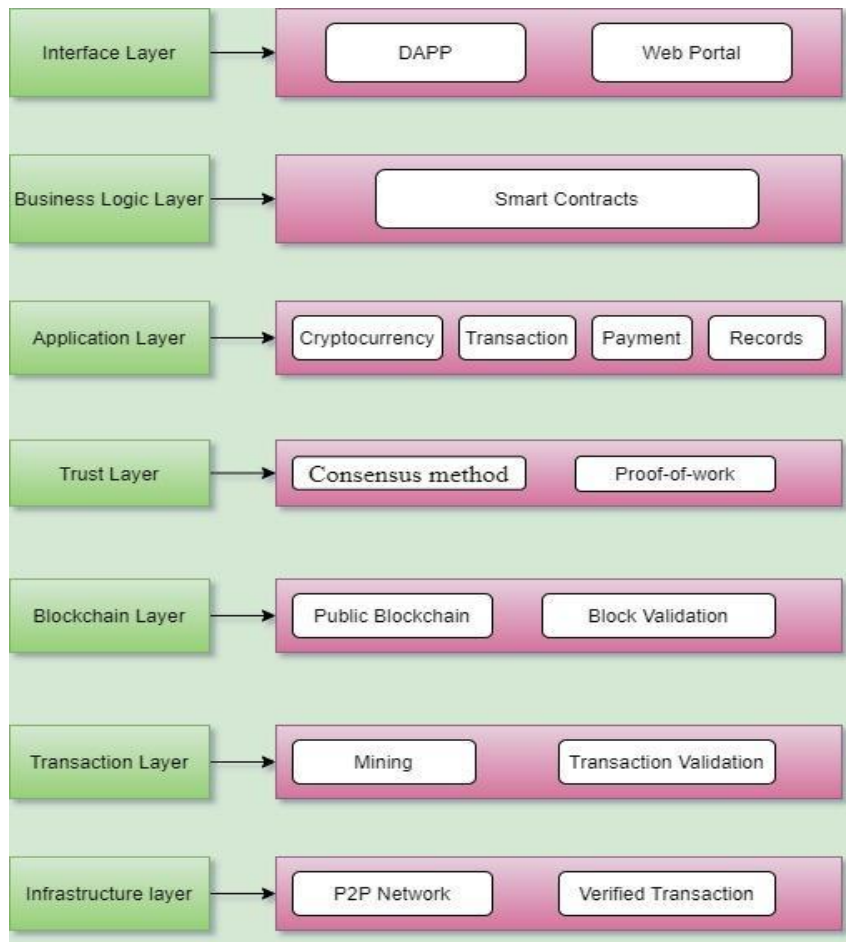


Fig. 3. Layered Structure

distributed networking. When one transaction is completed, it is transmitted to all network nodes, and each node verifies the transaction using established parameters. The validated transaction is then saved in the database.

4. EXPERIMENTAL EVALUATION

Firstly, we discuss the experimental setup and measures of our study. Later, we share some of the results of our proposed system including cost analysis. Lastly, a theoretical comparison of our study and some related studies is shown.

4.1. Experimental Setup

For the setup, a metamask extension is integrated into the browser. Metamask is a secure way to connect to blockchain-based applications [13]. We create an account and have to set up polygon network in our metamask. We borrow some matic crypto for our testing purpose. Also, remix IDE is used.

```

currency conversion
Receiver's address

2. TK
Enter your choice: 1
Enter your dolar amount: 90
Dolar to eth: 0.025106073159097186
account indicator: 0xD79B57a9d4221940008BD49f6947DBd261265721 0.025106073159097186 25106073159097186
Transaction sent: 0x3456684ec8fc03f34bbba81a330b01c30cd401ab987e65279f543954doe188c6
Gas price: 2.241460064 gwei Gas limit: 120209 Nonce: 6
Waiting for confirmation... -
FundMe.fund confirmed Block: 10687109 Gas used: 109281 (90.91%)

Funding working well 0xD79B57a9d4221940008BD49f6947DBd261265721 0.025106073159097186
Currency Converter Response Time 11.936228513717651

Response time
Sender's address

```

Fig. 4. Fund tracking process

[14]. It is used for smart contract development by using solidity language[15]. Remix IDE is used for testing, debugging, and deploying smart contracts. To execute smart contracts, brownie framework is used, which is a python based development and testing framework [16]. Moreover, to call the solidity language or to execute the whole system we use a brownie framework. In this whole system, we use proof of work to process transactions from peer to peer without any third party[17]. It is a decentralized consensus mechanism. On the other hand, before a transaction, the system goes through the currency conversion. Which means, if the currency is in taka, the system converts it to dollars, and from dollars, it converts to ether. We use oracle [18] for converting currency.

4.2. Experimental Findings

In the funding process, the system can track the sender's and receiver's funding information and can also track the withdrawal information as well. In Fig. 4, we see the tracking of transactions after the funding process. Fig. 5 shows the funding of account 2 after the transaction along with the Funding process. Also, Fig. 6 shows the withdrawal of account 1 through which we can check the system's withdrawal feature process. Now, in the proposed model, we calculate response time on the basis of our features. Hence, the execution process in every account remains the same. So, here we differentiate the response time to execute every feature in our system.

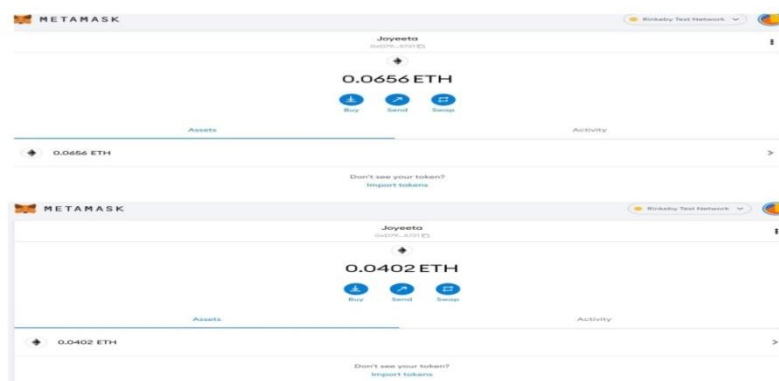


Fig 5. Funding result of account 2

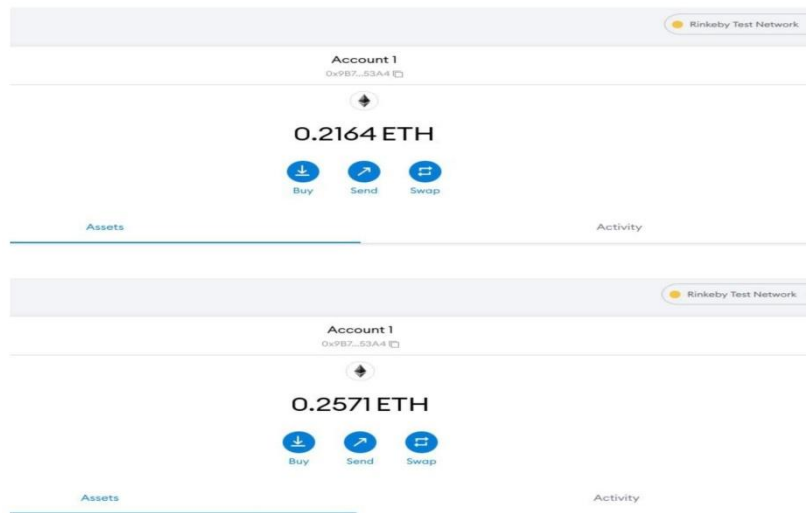


Fig 6. Withdrawal process result of account 1.

Table 1. Response time for each feature of the proposed model.

Features	Response time
Owner check	0.3063 ms
Owner's balance check	0.2831 ms
Owner's withdrawal process	6.212270 ms
Assign new owner	39.51796 ms
Funding	0.046349 ms
Currency convert in Dollar	11.9362285 ms
Currency Convert in Taka	11.9585 ms
Get Contract Balance	0.34818 ms

Table 1 shows the response time for each feature of our proposed model. In our proposed model, it takes 0.304 ms to 40 ms for executing every feature. Hence, on average, to execute the whole system, per account it takes 8.3786 ms including the currency converter in dollars. However, on the basis of currency converter and withdrawal by the assigned owner, the average time may vary. Since the assigned owner may have to go through some process which may cause more time than the withdrawal by the owner itself. The reason for the short response time is the polygon network [12]. We use layer 2 Ethereum scaling of polygon network.

4.3. Cost Analysis and Discussion

Depending on the aforementioned scenarios, we estimate the cost of our proposed model:

- Procedure
- A Blockchain App's complexities
- Development Tools

Costs of deployment and third-party services:

Public Blockchain: \$0.01 per transaction + \$750 for third parties.

Costs of maintenance: Approximately 10% to 15% of the total project budget.

The following are some of the third-party tools which Blockchain Applications may require:

Amazon Web Services: Computation, Memory, Delivery. (Depends on the user's number, \$100 - \$1000).

Amazon SNS, Twilio: Provide notifications inside the interface. (\$10 - \$50).

Mixpanel/Flurry: Analysis of Data, channel, and monitoring. (\$0 - \$150).

The apps can be shifted to multiple platforms depending on their stability, adaptability, and confidentiality since blockchain technology is indeed new to the industry and new systems are arriving in the markets day after day. Hence, further additional costs for development are needed for this system. In our proposed model, we have made a medium complexity blockchain app which is dApps, built on the blockchain platforms like Ethereum. That's why we are able to work with multiple features at a minimal cost.

4.4. Theoretical Comparison

As shown in Table 2, related papers do not place government funds as a stakeholder. Also, private and permission less blockchain usage makes our study different than state-of-the-Arts.

Table 2. Theoretical comparison with related works

Paper Title	Research Gap	Our Proposed Remedy
A framework to make charity collection transparent and auditable using blockchain technology[6]	No mention of tracking the government funds.	We focus on both governmental and non-governmental organizations
E-Governance, A Tendering Framework Using Blockchain[11]	Because of using private network, a restriction comes forward for the general citizens to visualize the tendering and funding procedures which creates confusion on the system's reliability	We use public blockchain in our system.
Blockchain for government fund tracking using Hyperledger[10]	As used a permissioned blockchain, if the system is managed by only the government officials and most of the officials comes out as dishonest, then the system can become vulnerable	We use permissionless blockchain in our system
A blockchain based tracking system for university donation[22].	No implementation on how the study materials are going to be tracked	We use Proof of work in our system
Blockchain-Based One-Off Address System to Guarantee Transparency and Privacy for a Sustainable Donation	No proper information about how the donors can trust the receivers if their provided donation is properly being	We use decentralized consensus mechanism

Environment[23]	utilized or not	
Platform for Tracking Donations of Charitable Foundations based on Blockchain Technology[24]	Lack of practical implementation, only talked about the tools but not any system prototype.	Implementation Shown

4.5. Potential interdisciplinary applications of this study

Blockchain is not only limited to crypto currency, many other interdisciplinary applications can benefit from Blockchain as well. IoT is a field that is used for solving many social issues. Study [28] works for the safety of elderly populations by monitoring indoor navigational Hazards with the help of IoT. As IoT is vulnerable to security attacks, blockchain can provide security to IoT applications which will make them more reliable. Study [29] discusses how IoT can benefit from blockchain integration.

Study[27], discusses a major issue faced by the elderly populations which is loneliness. To mitigate loneliness government can build infrastructure to create interactive environments for elderly populations to improve their social interaction. In that case, the fund can be tracked by

blockchain so that there is no corruption and hence, infrastructure to mitigate loneliness of elderly populations will be ensured. Also, creative industries like music industry can leverage blockchain technology by creating unchangeable database for music copyright information and conducting frictionless royalty payments [30].

5. LIMITATIONS AND FUTURE WORKS

A limitation of our study is that we only worked on currency converters of Taka and Dollar. We do not provide options for other currencies. Another limitation is that we use Ethereum which is very costly to handle. As there is a chance that Ethereum will become more expensive to use, that can be a problem since some people may not effort this. On the other hand, in our system, we have a limitation of funding amount. The least amount of fund is 50 dollars. Furthermore, we are now using the proof-of-work mechanism, however, there are other mechanisms that can reduce response time.

In the future, we will try to include different currencies as well. Hence, it would be easier to transfer money or take donations from other countries. Also, in the future, we will try to exclude fund limitations and mention specific withdrawal amount that will give users a more friendly environment. In the future, we will look forward to find a more suitable public blockchain network or try to make one of our own to reduce cost. We can use stable coins instead of Ethereum. Lastly, we will try to shift our mechanism from proof-of-work to proof-of-stake to reduce the response time.

6. CONCLUSION

To conclude, in this paper, a blockchain-based fund management system has been introduced. The system can be used in a variety of business settings. The system has taken advantage of blockchain properties like irreversibility and security to create a trustworthy-decentralized system. It outperforms the existing standard methods by implementing a more reliable tracking

system which would put an end to any wrongdoings or anomalies. And the smart contract authorities are in charge of making sure that practically all of the system's transactions can be tracked. We use Ethereum as it is a public network which means all the records can be visible and accessible by every peer. Here, by using blockchain, we make a structure for fund management where all the transactional data will be recorded. This data is undeletable and unchangeable. Hence, there is no way to corrupt this structure. With this structure, we can keep records of funds of different sectors so that it would be more trustable to rely on this and will help to reduce corruption as well. Hence, in the government sector or in any organization, this system can be utilized whenever it comes to any trustable fund issues.

7. CONCLUSIONS

This paper addresses the lack of strength of OLLVM obfuscation in control flow protection and the gap in identifier obfuscation by proposing two broad categories of enhancements. In control flow obfuscation, first, adding nested switches at the control flow level and adding the switch structure again in the flattened code, thus increasing the complexity of the code while resisting existing scripting attacks; second, proposing an in-degree treatment for bogus blocks to increase the confusion of bogus blocks further. Further, at the level of identifier obfuscation, four algorithms are proposed and bridge the gap of OLLVM in identifier obfuscation. By comparing with OLLVM, this paper can significantly improve the original control flow complexity in obfuscation effect; replace 65.2% of custom identifiers while guaranteeing program functionality. Furthermore, the time overhead from obfuscation is almost negligible. The space overhead is at 1.5 times.

In future work, we will pay attention to generating more secure opaque predicates and are not limited to the number-theoretic model. Meanwhile, the practical effectiveness of existing obfuscation algorithms in large projects remains tested. Therefore, we will focus on how to provide more accessible use of the obfuscation framework model in large projects.

ACKNOWLEDGEMENTS

*Authors have equal contributions to this paper.

REFERENCES

- [1] G. Locatelli, G. Mariani, T. Sainati, and M. Greco, "Corruption in public projects and megaprojects: There is an elephant in the room!," *International Journal of Project Management*, vol. 35, no. 3, pp. 252–268, 2017.
- [2] T. R. . December and T. Report, "61% fund embezzlement in forest projects: Tib." <https://www.tbsnews.net/bangladesh/corruption/61-fund-embezzlement-forest-projects-tib-178726>, Dec 2020.
- [3] "Tibstudy: 14% to 7% corruption found in climate change projects." <https://archive.dhakatribune.com/bangladesh/corruption/2020/12/24/tib-study-14-to-76-corruption-found-in-climate-change-projects>, Dec 2020.
- [4] Bangkok, "What's happening with aid to bangladesh?" <https://www.thenewhumanitarian.org/report/96902/analysis-what%E2%80%99s-happening-aid-bangladesh>, Sep 2017.
- [5] "Payment and settlement systems." <https://www.bb.org.bd/en/index.php/financialactivity/paysystem..>
- [6] M. S. Farooq, M. Khan, and A. Abid, "A framework to make charity collection transparent and auditable using blockchain technology," *Computers Electrical Engineering*, vol. 83, p. 106588, 2020.
- [7] G. Wood, "Ethereum: a secure decentralized generalized transaction ledger", *Ethereum Proj Yellow Pap* 2014;151:1–32.
- [8] A. Mehra, S. Lokam, A. Jain, M. Sivathanu, S. Singanamalla, and J. O'Neill, "Vishrambh: Trusted philanthropy with end-to-end transparency," in *HCI for Blockchain: a CHI 2018 Workshop on*

- Studying, Critiquing, Designing and Envisioning Distributed Ledger Technologies, Montreal, QC, Canada, 2018
- [9] D. D. Fiergbor, "Blockchain technology in fund management," in *International Conference on Application of Computing and Communication Technologies*, pp. 310–319, Springer, 2018
- [10] A. Mohite and A. Acharya, "Blockchain for government fund tracking using hyperledger," in *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, pp. 231–234, IEEE, 2018
- [11] Y. Goswami, A. Agrawal, and A. Bhatia, "E-governance: A tendering framework using blockchain with active participation of citizens," in *2020 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, pp. 1–4, IEEE, 2020
- [12] S. Neiwai, "Polygon (MATIC): The Swiss Army Knife of Ethereum Scaling", <https://www.gemini.com/cryptopedia/polygon-crypto-matic-network-dapps-erc20-token>, August 2021
- [13] "The crypto wallet for defi, Web3 Dapps and nfts," MetaMask. [Online]. Available: <https://metamask.io/>. [Accessed: 28-May-2022].
- [14] ReadTheDocs, "Creating and Deploying a Contract," [Online]. Available: <https://remix-ide.readthedocs.io/en/latest/createdeploy.html>
- [15] J. Frankenfield, "Smart contracts: What you need to know," Investopedia, 24-Mar-2022. [Online]. Available: <https://www.investopedia.com/terms/s/smart-contracts.asp>. [Accessed: 28-May-2022]
- [16] "How to deploy a smart contract with brownie," QuickNode. [Online]. Available <https://www.quicknode.com/guides/web3-sdks/how-to-deploy-a-smart-contract-with-brownie>. [Accessed: 28-May-2022].
- [17] J. Frankenfield, "Proof of work (POW)," Investopedia, 12-May-2022. [Online]. Available: <https://www.investopedia.com/terms/p/proof-work.asp>. [Accessed: 28-May-2022].
- [18] "Oracle Supply Chain amp; Manufacturing 22A – get started," Oracle Help Center, 19- Apr-2022. [Online]. Available: <https://docs.oracle.com/en/cloud/saas/supply-chain-management/22a/index.html>. [Accessed: 28-May-2022].
- [19] "World Bank cancels Bangladesh bridge loan over corruption" BBC News, 30-June-2012. [Online]. Available: <https://www.bbc.com/news/world-south-asia-18655846>. [Accessed: 19-June-2022].
- [20] "Corruption in Bangladesh Composition" A. Islam, 6-May-2020. [Online]. Available : <https://lekhapora.org/corruption-in-bangladesh-composition/>. [Accessed: 19-June-2022].
- [21] B. Report, "Bangladesh ranks 13th most corrupt country in the world." <https://www.businessinsiderbd.com/national/news/16468/bangladesh-ranks-13th-most-corrupt-country-in-the-world>, Jan 2022.
- [22] E. A. FERWANA, A BLOCKCHAIN-BASED TRACKING SYSTEM FOR UNIVERSITY DONATION. PhD thesis, 2021.
- [23] J. Lee, A. Seo, Y. Kim, and J. Jeong, "Blockchain-based one-off address system to guarantee transparency and privacy for a sustainable donation environment," *Sustainability*, vol. 10, no. 12, p. 4422, 2018.
- [24] H. Saleh, S. Avdoshin, and A. Dzhonov, "Platform for tracking donations of charitable foundations based on blockchain technology," in *2019 Actual Problems of Systems and Software Engineering (APSSE)*, pp. 182–187, IEEE, 2019.
- [25] Benjamin A. Olken, "Monitoring Corruption: Evidence from a Field Experiment in Indonesia," *Journal of Political Economy* 115, no. 2 (April 2007): 200-249. doi: 10.1086/517935
- [26] ONE Campaign, 2022. Public monitoring of government projects reduced corruption by 20%. Retrieved July 15, 2022, from <https://www.one.org/international/follow-the-money/public-monitoring-of-government-projects-reduced-corruption-by-20/>
- [27] Thakur, N., Han, C.Y.: A framework for facilitating human-human interactions to mitigate loneliness in elderly. In: Ahram, T., Taiar, R., Langlois, K., Choplin, A. (eds.) *IHIET 2020. AISC*, vol. 1253, pp. 322–327. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-55307-4_49
- [28] Thakur, N., Han, C.Y. (2020). A Framework for Monitoring Indoor Navigational Hazards and Safety of Elderly. In: Stephanidis, C., Antona, M., Gao, Q., Zhou, J. (eds) *HCI International 2020 – Late Breaking Papers: Universal Access and Inclusive Design. HCII 2020. Lecture Notes in Computer Science()*, vol 12426. Springer, Cham. https://doi.org/10.1007/978-3-030-60149-2_56
- [29] Abdelmaboud A, Ahmed AIA, Abaker M, Eisa TAE, Albasheer H, Ghorashi SA, Karim FK. Blockchain for IoT Applications: Taxonomy, Platforms, Recent Advances, Challenges and Future Research Directions. *Electronics*. 2022; 11(4):630. <https://doi.org/10.3390/electronics11040630>

- [30] Arcos, L. C. (2018), "The blockchain technology on the music industry", Brazilian Journal of Operations & Production Management, Vol. 15, No. 3, pp. 439-443, doi:10.14488/BJOPM.2018.v15.n3.a11, available from: <https://bjopm.emnuvens.com.br/bjopm/article/view/449> (accessed on 13/7/2022).
- [31] <https://www.ganintegrity.com/portal/country-profiles/bangladesh/>, Nov 2020.
- [32] A. A. I. a. Arif, "Contact." <https://lekhapora.org/corruption-in-bangladesh-composition/>, May 2020.
- [33] M. H. Ahamad, "Foreign grants and loans in bangladesh," 2018.
- [34] Bangkok, "What's happening with aid to bangladesh?." <https://www.thenewhumanitarian.org/report/96902/analysis-what%E2%80%99s-happening-aid-bangladesh>, Sep 2017.

AUTHORS

Nibula Bente Rashid: recently graduated in computer science and engineering from Brac University, Dhaka, Bangladesh. Her research areas include cloud computing, blockchain, and cybersecurity

Joyeeta Saha has recently graduated from Brac University. She was in Computer Science and Engineering Department. Her research interest is related to Blockchain, Cloud computing, and Deep learning. Now she is working as a project employee at Quantanite.

Raonak Islam Prova is currently working as a trainee automation developer in Summit14 Communications Limited. She received her B.Sc degree in Computer Science and Engineering from BRAC University. Her fields of interest in research includes cloud computing, blockchain, automation, networking and cybersecurity.

Nowshin Tasfia has recently graduated from Brac University. She was in Computer Science and Engineering Department. Her research interest is related to Blockchain, Cloud computing, and Deep learning

Md. Nazrul Huda Shanto is working as a Research Assistant at BRAC University. He is currently pursuing a bachelor's degree in Computer Science from the School of Data and Sciences, Brac University, Dhaka, Bangladesh. His area of research interests includes HCI4D, Cloud Computing, Cybersecurity, and Image Processing

Jannatun Noor has been working as a Lecturer at the Department of CSE, BRAC University since September 2018. Besides, she is working as a Graduate Research Assistant under the supervision of Prof Dr. A. B. M. Alim Al Islam in the Department of CSE at Bangladesh University of Engineering and Technology (BUET). Prior to BRACU, she worked as a TEAM LEAD of the IPV-Cloud team in IPvision Canada Inc. She received her B.Sc. and M.Sc. degrees in CSE from BUET. Her research work covers Cloud Computing, Wireless Networking, Data Mining, Big Data Analysis, Internet of Things, etc.

F-LOW: A PROMISING COUNTERMEASURE AGAINST DDoS ATTACKS BASED ON SPLIT SKETCH AND PCA

Fei Wang, Zhenxing Li* and Xiaofeng Wang

School of Computer,
National University of Defense Technology, Changsha, China

ABSTRACT

Distributed Denial of Service (DDoS) is Achilles' heel of cloud security. This paper thus focuses on detection of such attack, and more importantly, victim identification to promote attack reaction. We present a collaborative system, called F-LOW. Profiting from bitwise-based hash function, split sketch, and lightweight IP reconstruction, F-LOW can defeat shortcomings of principle component analysis (PCA) and regular sketch. Outperforming previous work, our system fits all Four-LOW properties, low profile, low dimensional, low overhead and low transmission, of a promising DDoS countermeasure. Through simulation and theoretical analysis, we demonstrate such properties and remarkable efficacy of our approach in DDoS mitigation.

KEYWORDS

DDoS detection, victim identification, principle component analysis, split sketch, bitwise-based hash.

1. INTRODUCTION

As cloud computing becomes ubiquitous on the Internet, it opens the door to many serious attacks, particularly Distributed Denial of Service (DDoS) attack. As one of the most serious threats to cloud security, DDoS attack affects cloud and datacenter service even worse than to regular Internet service. Regarding cloud and datacenter as an inherent multi-tenant infrastructure, DDoS attack against a single customer is against all customers in the infrastructure [1]. In recent years, DDoS attacks evolve both in the quantity and the destructive power of a single attack. The peak bandwidth of the largest DDoS attack in 2021 exceeds 3Tbps [2]. DDoS attacks in such scale can easily breakdown arbitrary online services and incur huge financial losses. Cloud providers need to do more to ensure the availability of their cloud services.

Attention is thereby devoted to countermeasures against DDoS attacks towards online services such as cloud computing. This paper is concerned with detection, and more importantly, victim identification of network-wide DDoS attacks, so that outcomes can contribute to quick reaction to such devastating attacks. The challenging is, the two countermeasures suffer from the increasing link bandwidth of current Internet as well as inconspicuous sources of DDoS attacks. From this perspective, a promising DDoS countermeasure system must have the following Four-LOW properties. (1) low profile: the system should have the capability of detecting low-profile network anomaly, so that DDoS attacks can be detected as early as possible; (2) low dimensional: in order to identify victim IPs of DDoS attacks, dimensional reduction mechanism is necessary for processing high-dimensional data, such as per-IP flow statistic; (3) low overhead: expensive computational cost and memory consumption should be avoided in the system; (4) low

transmission: if the system is distributed, collaborative nodes transmit as little data as possible, not exacerbating network congestion caused by an ongoing DDoS attack.

Previous works make progress in DDoS detection or identification, but they do not fit all Four-LOW properties described above. Many of them are insensitive to low-profile DDoS attacks and only can detect anomalies when attack traffic is aggregated conspicuously near the victims [3][4][5]. Lakhina et al. [6] apply PCA on origin-destination (OD) flows, the aggregations of traffic from a source router to a destination router. Taking advantage of PCA, this method can figure out malicious OD flows that includes low-profile attack traffic. However, this method cannot result in particular victim IPs due to coarse-grained aggregation. Applying PCA on fine-grained per-IP flows may help to distinguish victim IPs, but the data dimensionality jumps sharply to a high degree that PCA cannot afford to. To tackle this problem, Li et al. [7] use sketch to randomly aggregate IP flows. Although this method keeps input data of PCA in low dimensional space, it therewith poses another high-overhead problem. That is to infer key IPs reversely from a number of particular buckets. This process always involves great space in memory for storing mapping tables, as well as high computational cost for calculate intersections of multiple large sets [8][9][10]. Another problem that obstructs practical application of [6] and [7] is too much data sharing among collaborative nodes. Original network measurements, such as OD flows statistic [6] and sketches [7] are transmitted over the network to a centralized device, challenging sparing bandwidth capability of DDoS-compromised links. Recently, software-defined DDoS detection methods are proposed to confront DDoS attacks [15,16] for cloud and datacenters.

To conquer above challenges, this paper proposes a collaborative DDoS detection and victim identification system, called F-LOW. Our system can satisfy all Four-LOW properties, thus being a promising countermeasure against network-wide DDoS attacks. Our main contributions are summarized as follows.

- We pioneerly propose split sketch which divides a sketch structure into pieces and deploy them on distributed network devices. In this way, network measurements all over the network seems to be in a sketch and can be treated relevantly, even without being sent to a centralized device.
- We formally define bitwise-based hash function and adopt it into split sketch for traffic aggregation. The aggregation mechanism not just significantly reduces data dimensionality in local detection, but also makes victim identification easy-to-implement.
- We extend existing PCA-based anomaly detection algorithm to detect multiple anomalous flows, suiting the needs of practical situations. Heuristic rules are presented to improve computational cost.
- We present a lightweight IP reconstruction algorithm to identify victim IPs of DDoS attacks. It can accurately infer victim IPs or partial victim IP with very low computational cost. Both outcomes help to effectively filter attack traffic during reaction.
- We demonstrate Four-LOW properties of our approach through simulation and theoretical analysis. When compared with the most relevant work, our approach has much lower overhead and greater scalability. We experimentally show that outcomes of our system, even being partial victim IPs, can help to filter DDoS traffic in reaction. The false positive rate is less than 3%.

The rest of this paper is organized as follows: Section 2 presents overview of the F-LOW system and Section 3 anatomizes the design of F-LOW. We evaluate our F-LOW system in Section 4 and draw a conclusion in Section 5.

2. F-LOW OVERVIEW

F-LOW is a distributed framework for DDoS detection and victim identification. It consists of global detectors (GD) and local detectors (LD). LDs are widely deployed all over the network. For simplicity, we consider a classic distributed architecture as shown in Figure 1, which is adopted in many literatures [6][7]. In the F-LOW system, GDs communicate with each other and LDs using reliable channels.

In general, LDs detect traffic anomaly based on measurements of local traffic, while the GD confirms whether reported anomalies arise from network-wide DDoS attacks, and infers victim IPs (or partial victim IPs) based on local detection results. Particularly, an LD performs

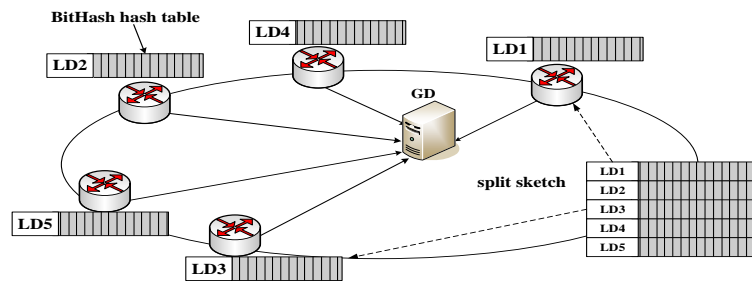


Figure 1. The architecture of FLOW system

a piece of a sketch, that is, a hash table of a special designed bitwise hash function; in this way the LD aggregates packets into bitwise flows according to certain bits in their destination IPs (dstIP). The aggregation significantly reduces the scale of time-series measurement data of network traffic. LDs thus can detect when network anomalies occur and identify flows that involves these anomalies, using a PCA-based anomaly detection method. The method is like [6] and extended for identifying more than one anomalous flow. LDs send very little data to the GD including IDs of Anomalous Flows (AF). In view of the whole F-LOW system, all hash tables performed by LDs constitute a virtual sketch, called split sketch. The GD integrates flow IDs into split sketch and determines anomalies pervading all over the network as DDoS attacks. Taking advantage of bitwise hash functions, the GD can reconstruct (partial) victim IPs of DDoS attacks reversely with much lower computational cost and smaller space requirement than previous work [8][9][10].

There are three principles behind the F-LOW system, making it scalable and practical in network-wide DDoS detection and identification

Principle 1: Flow aggregation based on bitwise information. Aggregating packets according to certain bits in their destination IPs, on one hand, reduces data dimensionality to yield PCA-based method. On the other hand, since flow ID reifies part of victim IP, it benefits determination of victim IP.

Principle 2: Combination of distributed traffic monitoring and minimum data transmission. Considering the nature of DDoS attacks, F-LOW leverages pieces of a sketch on different LDs to gather network-wide measurements. Meanwhile, without sending original measurements, F-LOW greatly reduces the amount of transmitted data between nodes, respecting limited link bandwidth

Principle 3: Lightweight reconstruction of victim IP based on popularity. As a profit from bitwise hash functions, F-LOW provides a lightweight way to infer victim IPs of DDoS attacks,

having no use of storing mapping tables and calculating intersection of large sets of IPs like reverse sketch [10]. Determining a bit in victim IPs relies on the popularity of the bit being reified as “0” or “1”, in accordance with the multi-source feature of DDoS attacks

3. ANATOMY OF F-LOW

Our F-LOW system achieves desirable Four-LOW properties through four essential components: bitwise-based flow aggregation, split sketch structure, PCA-based local detection method, and novel IP reconstruction algorithm. In this section, we present details of these components

3.1. Bitwise-based Flow Aggregation

In F-LOW, LDs aggregate packets into flows according to certain bits in their dstIPs rather than whole dstIPs, thus obtaining much fewer flows. The principle of bitwise-based flow aggregation is, specifying k bit positions in an L -length IP address (e.g. $L=32$ in IPv4), packets whose dstIPs have identical bits in the same positions are assigned to the same flow.

Let us define a function $F(ip,mask)$, where ip is an IP address and $mask$ indicates which k bits are specified. $mask$ is viewed as an L -length bit string in which specified bits are set to “1” while others are set to “0”. $F(ip,mask)$ combines selected bits in ip orderly and results in a k -length string. Let ip_1 and ip_2 be two different IPs. Then we assign packets that are destined for ip_1 and ip_2 into the same bitwise flow if and only if

$$F(ip_1,mask) = F(ip_2,mask)$$

To identify flows, $F(ip,mask)$ is viewed as flow ID. Figure 2 illustrates an example of bitwise-based flow aggregation. $IP_1 \sim IP_4$ are different dstIPs of packets. Specifying five bit positions in an IP address, packets destined for $IP_1 \sim IP_4$ thus are divided into three flows, identified by “010011”, “011010” and “110110”, respectively. Given a fixed k , there are at most 2^k bitwise flows. LDs thus can achieve an affordable space overhead by varying k

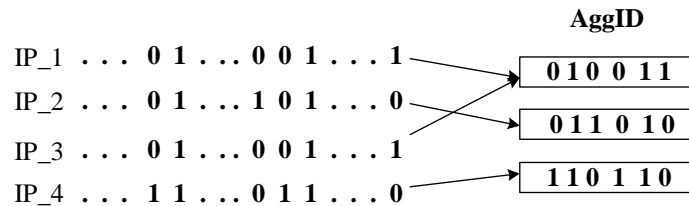


Figure 1. An example of bitwise-based flow aggregation

Considering functionality of bitwise-based flow aggregation, we then regard $F(ip,mask)$ as a particular hash function, called bitwise hash function. Accordingly, $mask$ is the seed. In the following, we explain how to use such hash function to generate a split sketch and show the remarkable benefits obtained from our novel design.

3.2. Split Sketch

In this section, we present the deployment of split sketch for network-wide traffic measurement. The split sketch is a virtual sketch structure that is divided into pieces and deployed on multiple LDs in F-LOW. In split sketch, bitwise hash functions are adopted instead of regular universal hash functions.

Imagine a sketch that consists of hash tables of N bitwise hash functions. Each hash table is a piece of the sketch puzzle and each LD only performs one piece. An LD generates its own bitwise hash function independently, by randomly choosing k integers from 0 to $L-1$. These integers correspond to selected bit positions in an IP address. Then the LD allocates memory space to keep a hash table of size 2^k . Since our F-LOW system uses traffic volume as a feature to detect DDoS attacks, each bucket in the hash table is a counter for counting the number of packets that are mapped into the bucket. Although many other features can be adopted in our system, we choose the simplest but promising one, so as to simplify traffic measurement as well as reduce computational cost. When a packet arrives, the LD updates the bucket with index $F(\text{dstIP}, \text{mask})$ by adding one. Since each LD independently determines bitwise hash function, their choices may clash. However, the probability that two LDs select the same hash function is $(1/L)^k$. It is extremely small when $k=10$ and $L=32$. We thus say hash functions used in a split sketch are different.

Split sketch does not actually gather traffic measurements all over the network into a centralized device. In fact, each row of split sketch presents statistic of different traces. We thus let LDs detect network anomaly based on their own pieces of split sketch. On the other hand, local detection results (anomalous flow IDs) are integrated into a whole sketch and processed at GDs to confirm DDoS attacks and infer their victim IPs

3.3. Extended PCA-Based Local Detection

At LDs, we adopt PCA-based local detection method to detect potential DDoS attacks and identify abnormal traffic aggregates, which is similar with subspace method in [6]. Unfortunately, subspace method in [6] only obtains the first AF that causes greatest anomalous change among all candidate flows. In practice, it is possible that more than one anomaly occurs simultaneously, legitimate or malicious. Furthermore, to reduce false negative, it is essential to capture as many AFs as possible. We thus extend identifying part in [6] to find a set of primary AFs.

Through a piece of split sketch performed by an LD, packets are naturally aggregated into bitwise flows. The LD obtains a time-series measurement of local traffic, forming an $n \times m$ matrix D , where n is the number of time intervals and m the size of hash table, $m = 2^k$. Let y represent a row of D , which is viewed as traffic fingerprint at a particular time. At first, we assume only one AF and identify first AF using subspace method (see details in [6]).

To find residual important AFs, we eliminate the change caused by previous found AFs. Let Ω be the set of previous found AFs. Each flow in Ω corresponds to a four-tuple $\langle sn, F_i, \tilde{y}_i^*, \tilde{\Delta}_i \rangle$, where sn represents the order of flow F_i being found. For an anomalous $y = y_i^* + \Delta_i$, y_i^* represents the sample vector for normal traffic conditions and Δ_i represents the amount of change due to flow F_i . Then \tilde{y}_i^* and $\tilde{\Delta}_i$ are best estimates of y_i^* and Δ_i , which are calculated in the procedure to find F_i (see details in [6]).

Then we can construct a new traffic fingerprint y' , getting rid of effects of AFs that have been verified to be anomalous.

$$y' = y - \sum_{F_i \in \Omega} \tilde{\Delta}_i$$

In terms of y' , the next AF can be found using the same method as [6] does. The procedure is repeated until the change caused by any residual flow does not exceed a threshold.

$$\|y_n - \tilde{\Delta}_i\| < \rho$$

In view of traffic increase that accompanies nearly all flooding-based DDoS attacks, we improve above local detection method in two greedy ways: 1) Further compress D to D' . The latter only consists of columns having increased amounts along the time axis in D . Accordingly, P is generated based on D' . 2) Measure the amount of packets that arrive in each time interval, and perform anomaly detection upon y , only if the increasing scale of packet amount in the corresponding interval exceeds certain threshold.

Local detection results in a set of AFs (or a set of flow IDs). According to the design of bitwise-based hash function, a flow ID reifies some bits in an IP address and mask indicates which bits they are. Namely, if an AF includes DDoS traffic, segments of victim IP of the DDoS attack is fixed by the flow ID. In the next subsection, we explain how the GD infers victim IPs based on such incomplete IP segments.

3.4. Alert-burst-based Global Detection

When a DDoS attack is launched, the malicious attack traffic accessing the attack target is relatively concentrated. From the point of view of distributed attack detection, the traffic abnormal alarm generated by each local detection device should also have time correlation. Therefore, when the global detection device finally determines the DDoS attack, it should consider the time when the abnormality occurs. If many abnormal traffic alarms occur in a short time, it can be considered that a large-scale DDoS attack has occurred in the network. As showed in Figure 3, a DDoS attack can be confirmed when LD0, LD1 and LD2 all reports local anomaly alerts nearly in the same time.

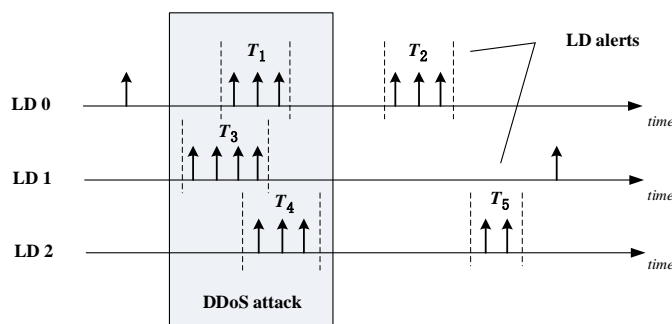


Figure 3. An outbreak period of local anomaly alerts

Assuming the number of abnormal alerts generated in an observation window T is N_{alert} , the abnormal alert density (DAA) is defined as the number of abnormal alarms generated in a unit time,

$$Dens(T) = \frac{N_{alert}}{T}$$

If the observation window t , the number of abnormal alarms generated by the local detection equipment LD is N_{alert} , and the abnormal alert density is $Dens(t)$. Expand the observation window along any direction of the time axis to obtain a new observation window t' , let the abnormal alert density in t' be $Dens(t')$. If $Dens(t') < Dens(t)$ is always true, then the observation time t is called the abnormal burst period (BPA).

The determination process of abnormal burst period starts from the two abnormal alarms with the closest time interval. The initial observation window is set as τ and then slide the observation window to cover the above two abnormal alarms. Stretch the observation window along the two

directions of the time axis then the first abnormal burst period of the local detection equipment can be obtained. As the network environment is complex and changeable, security threats occur all the time, each local detection device may have multiple abnormal outburst periods. The above operation is repeated until the initial observation window is greater than the threshold τ_0 . Threshold τ_0 is the maximum relevant abnormal time interval, and 10 times the sampling interval is fine.

Based on the abnormal burst period of each local detection device, the global detection device finally determines DDoS attack in the network. DDoS attack sources are distributed all over the network, and the attack traffic triggers an abnormal alarm at multiple local detection devices. Therefore, a concentrated abnormal burst period of multiple local detection devices in a short time means a large-scale DDoS attack. By comparing the abnormal outbreak period of each local detection equipment, if the number of equipment whose abnormal burst period overlap in time is greater than a threshold n , then global detection confirm a DDoS attack in the network. As shown in Figure 3, three local detection devices detect a large number of network traffic abnormalities at almost the same time (abnormal outburst periods T1, T3, T4), so it can be judged that there are large-scale DDoS attacks in the network. Local detection devices LD0 and LD2 also have anomaly alert burst in T2 and T5 respectively. However, since no global traffic anomaly are found by other detection devices in the network, these alerts do not conform to the distributed characteristics of DDoS attacks.

3.5. Victim IP Reconstruction

In order to confirm ongoing DDoS attacks and figure out victim IPs, GDs reconstruct anomalous dstIPs based on detection results obtained from LDs. In a DDoS attack, tens of thousands of hosts all over the network generate attack traffic simultaneously. The intuition thus is that a dstIP being widely regarded as anomalous indicates a network-wide DDoS attacks towards the dstIP. Following this lead, we show how GDs reconstruct victim IPs in details. Benefiting from bitwise hash function used in split sketch, our IP reconstruction algorithm has very low memory requirement and computational cost.

Let us begin with formatting outcomes of LDs. Through local detection, LDs obtain a number of AFs, which are sent to the GD in the form of (timestamp, flowID, mask). flowID is the ID of AF. timestamp represents the time when the anomaly occurs. mask implies the particular bitwise-based hash function adopted by an LD. If an LD report more than one AF to the GD, mask is only sent once. Combining flowID and mask, we can extract an L-length vector, in which bits specified by mask are set to "0" or "1", according to flowID. We call the vector Discrete Segment of IP (DSIP). Figure 4 shows a simple example of DSIP which gives a partial segment of the destination IP address. L is the length of the IP address. If the IPv4 address is used in the network, $L=32$.

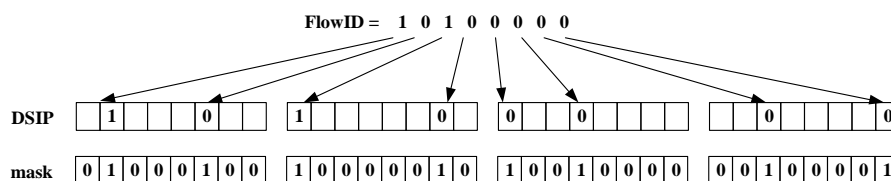


Figure 4. An example of DSIP

Regarding IP reconstruction as a vote for bits in victim IPs, then DSIP is the ballot. The GD first creates an empty vector B of length L to imitate an IP address (victim IP), and then deduces the

IP bit-by-bit on the basis of amounts of DSIPs which set the bit to “0” and “1”. IP reconstruction consists of three steps: eliminate noise DSIPs, decide bit statuses and reduce uncertainty.

Ideally, LDs perform accurate detection and all DSIPs generated are segments of victim IPs. However, due to inescapable false positive, even very small, there might be wrong DSIPs being segments of normal IPs. Thus we filter such noise DSIPs. A DSIP is considered as noise if it sets some bits just in contrast to what most other DSIPs set. For the i -th bit in B , we count the numbers $q_{0,i}$ and $q_{1,i}$ of DSIPs that reify the bit as “0” and “1” respectively. When comparing $q_{0,i}$ and $q_{1,i}$, we simply decide the one less than half of the other is smaller, like $q_{0,i}$ when $q_{0,i} < q_{1,i}/2$. If a DSIP reifies at least one such bit, we consider it as a noise DSIP. After obtained a more reliable set Γ' by eliminating noise DSIPs, we decide each bit in B with one of the following statuses, zero, one, amphibious, and unknown. The details of determining algorithm is given in our previous work[14]. To reduce the uncertainty in resultant victim IPs, we compare all DSIPs in Γ' with B and reify unknown and amphibious bits as many as possible. If most reified bits in a DSIP match with B , we believe the DSIP is truly part of a victim IP. Other reified bits are unmatched only because no enough LDs select these bits in their hash functions. We thus can adjust statuses in B according to unmatched bits in the DSIP, reducing unknown bits. For amphibious bits, such mostly matched DSIPs help to decompose B into two or more vectors. Each vector corresponds to a distinct victim IP.

As a result, the GD obtains a number of victim IPs (or partial victim IPs), which also proves ongoing DDoS attacks. If there are enough LDs participating in the F-LOW system, the GD can identify whole victim IPs accurately. Even in lacking adequate LDs, partial victim IPs are still efficient as filtering rules in DDoS reaction. Our IP reconstruction algorithm infers victim IPs without storing mapping tables and calculating intersection of large sets, thus having low computational cost and memory requirement. We evaluate our algorithm and compare its overhead with previous work in the next section

4. EVALUATION

In this section, we evaluate the performance of F-LOW system through simulation and theoretical analysis. The low-profile property of F-LOW is achieved through PCA-based anomaly detection method. Since the capability of the method has been thoroughly verified in many previous works [6] [7] [13], we lay particular emphasis on global detection of our approach. The focal point of evaluation is accuracy of IP reconstruction. We also measure the overhead of LDs and GDs, in order to demonstrate scalability of proposed approach

4.1. Coverage Rate

We first measure an important metric, coverage rate of selected bits by LDs in an IP address, which is relevant to performance of IP reconstruction. IP reconstruction relies on flow IDs reported by LDs. Namely, GD deduces a victim IP based on LDs' reification on every bit of IP. Therefore, the more bits are selected by LDs, more precise the reconstructed victim IP is.

We define coverage rate as the proportion of bits in an IP address that are selected by LDs at least once. Each LD chooses k bits in an IP address to generate its own bitwise hash function. The probability of a bit being selected by an LD is k/L . Suppose N LDs exists in the network. Then, for a bit, it is not selected by any LDs with the probability $q = (1 - k/L)^N$

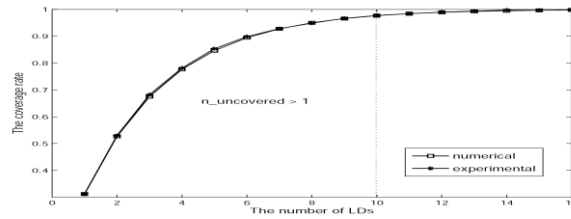


Figure 5. The numerical and experimental coverage rates with respect to the number of LDs

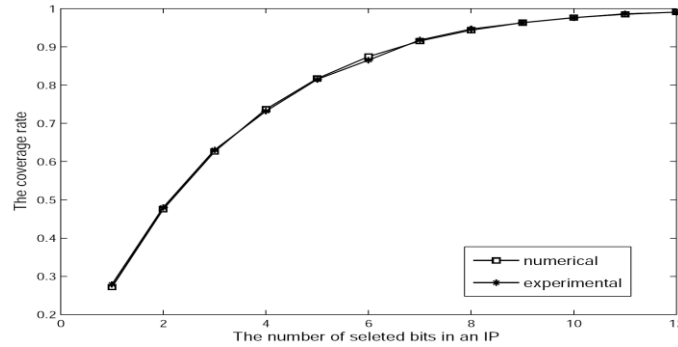


Figure 6. The numerical and experimental coverage rates with respect to the number of selected bits in an IP

Accordingly, the theoretical expected coverage rate is $1-q$. We also calculate the coverage rate in the simulation. As we can see in Figure 4 and 5, the numerical and experimental coverage rates perfectly match with each other. When $k=10$, $L=32$, and $N=10$, the coverage rate is more than 97%. Uncovered bits in an IP address are no more than one. Namely, only 10 LDs participating in the F-LOW system can cover almost whole IP address, thus potentiating precise IP reconstruction.

4.2. Accuracy of DDoS Detection

We qualitatively compare our F-LOW system with three existing approaches with respect to system overhead: sketch subspace (SS) [7], two-level sketch (TS) [8], and reverse sketch (RS) [10], which are closely related to our approach. These three methods all take the identification of the IP address (DDoS attack victim) causing the network anomaly as the final target of detection, and all reduce the statistical data dimension with the help of the summary data structure (sketch).

Figure 7 shows the experimental statistical results. Where FP represents false positive rate and FN represents false negative rate. Since RS is only a method of inversely deriving the IP address corresponding to the abnormal sketch entry, and there is no specific anomaly detection process, RS is used to replace bithash based sketch in F-LOW system to compare the accuracy of RS and F-LOW in victim identification. In addition, the SS method is a single point detection algorithm, 10 points are selected in the experimental topology to deploy SS. As shown in Figure 7, F-LOW performs better than other methods in all aspects. The victim identification FN of F-LOW system is slightly higher than that of RS method because only the number of complete IP addresses is calculated when the experimental data are counted.

Since our approach may result in partial victim IPs, it is difficult to evaluate its accuracy in DDoS detection and victim identification. We thus incarnate the performance in an indirect but persuasive way. That is to evaluate the efficacy of outcomes of our F-LOW system in filtering traffic towards victims (distinguishing attack packets from normal packets towards victims is

beyond the scope of this paper). We show the simulation result in Figure 7. In our simulation, most of partial victim IPs only have one indeterminate bit. Using them as rules, we can still successfully filter attack packets of DDoS attacks, with a false positive rate of less than 3%. All traffic toward victims is captured as their IPs matches with partial IPs.

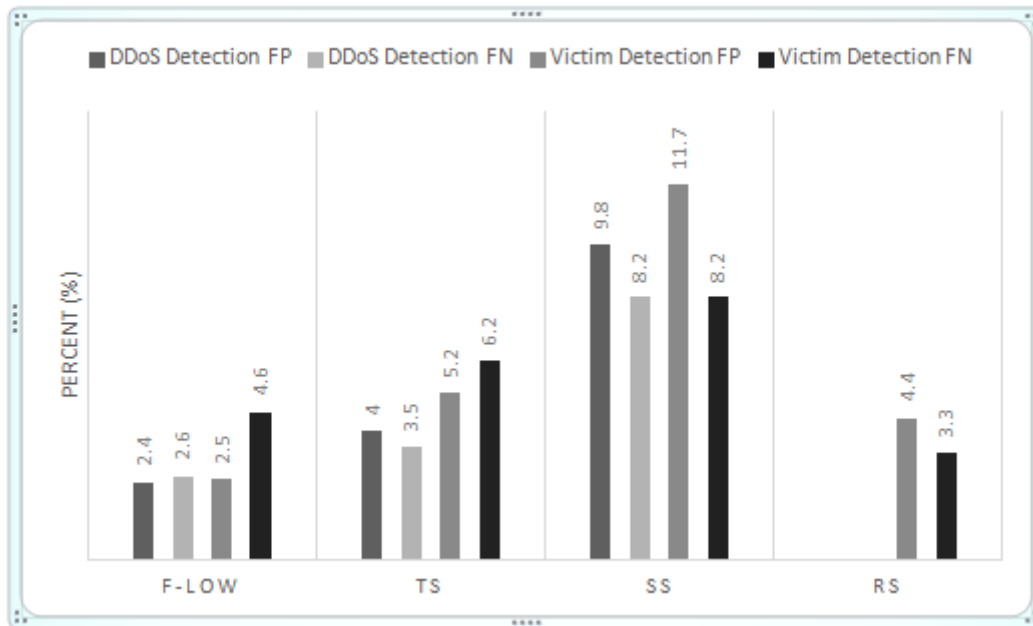


Figure 7. Accuracy of DDoS Detection

4.3. Overhead Comparison

Finally, we qualitatively compare our F-LOW system with sketch subspace (SS) [7], two-level sketch (TS) [8], and reverse sketch (RS) [10], which are closely related to our approach. The goal is to demonstrate the advantages of our approach in computational complexity, space requirement and Communication overhead.

When comparing computational complexity, we consider two aspects: the cost to update corresponding structures, which can be approximately reflected through the Number of calculating Hash Value per Packet (NHVP); the cost to identify victim IPs after detecting anomaly (COIV). We leave the overhead of PCA for this moment, as three of four approaches adopt PCA to detect network anomaly. Since all compared approach use sketch, we thus separate space requirement for sketch structure (sketch size) with other storage in the comparison. Communication overhead, represented by the amount of transmitted data between nodes, exists only in collaborative systems. Thus, reverse sketch has no communication overhead.

In F-LOW, each collaborative node maintains only one hash table of size M . One of the major differences between our approach with the other three is the way to map packets in a sketch. Our approach uses special designed bitwise-based hash while others use random hash in sketches, which induces completely different method to identify victims. Most COIV of our approach arises from counting DSIPs and comparing every DSIP with the vector B , so the cost increases with the number n_{dsip} of received DSIPs proportionally.

Table 1. The overhead comparison.

Method	Comput. Complexity		Space Requirement		Commun.
	<i>NHVP</i>	<i>COIV</i>	sketch size	other storage	Overhead
F-LOW	1	$O(Ln_{dsip})$	M	0	$O(n_{dsip})$
SS [7]	$4H$	$O(n_{flow})$	$4HM$	$O(n_{flow})$	$O(HM)$
TS [8]	$H + H_{bc} + H_{bf}$	$O(1)$	$HM + H_{bc}M_{bc}$	$H_{bc}L_{bf}$	$O(n_{dip})$
RS [10]	H	$qHM(\frac{n}{M})^{\frac{1}{q}}$	HM	$H \log mn^{\frac{1}{q}}$	0

We summarize the comparison result in Table 1, in which n_{flow} represents the number of distinct flows that SS needs to store states, and n_{dip} represents the number of anomalous destination IPs. H is the number of hash functions in a sketch. L is the length of an IP address. The parameter $q = 4$ in [10]. L_{bf}, H_{bf}, H_{bc} and M_{bc} are defined in [8]. Considering the number of DDoS attacks or victim IPs, the number n_{dsip} is much smaller than M and n_{flow} . In this context, our F-LOW system outperforms SS in all three aspects. Relatively, TS has lowest cost $O(1)$ during victim identification at the expense of space. Even though approaches TS and RS are slightly better than F-LOW in some aspects, they all require much more memory to store statistic of network traffic.

5. CONCLUSION

In this paper, we study countermeasures against DDoS attacks and present F-LOW, a collaborative system that can detect DDoS attacks and identify victim IPs of these attacks. Innovative bitwise-based hash function and split sketch are designed to digest network traffic. Meanwhile, an efficient IP reconstruction is proposed to reversely calculate anomaly IPs. Benefitting from those innovations, F-LOW system can accurately detect network anomaly caused by DDoS attacks and infer victim IPs, thus contributing to quick and efficient DDoS reaction. In summary, F-LOW has Four LOW-characteristics: low profile, low dimensional, low overhead, and low transmission, while previous DDoS countermeasures only fit parts of them. The outcome of F-LOW can help to filter attack traffic with very low false positive.

This paper proposes a promising approach against DDoS attacks. Since the F-LOW system is a collaborative system which involves plenty of cooperative detection nodes in the network, it is very difficult to apply and deploy. In the future, a flexible cooperation mechanism should be put on the agenda, to provide a practicable collaboration platform for F-LOW system

REFERENCES

- [1] Biswas, R., Kim, S., & Wu, J., (2021) "Sampling rate distribution for flow monitoring and DDoS detection in datacenter", IEEE Transactions on Information Forensics and Security, Vol. 16, pp2524-2534.
- [2] Toh, A., (2022) "Azure DDoS Protection—2021 Q3 and Q4 DDoS attack trends".
- [3] Wang, F., Wang, H., Wang, X., & Su, J., (2012) "A new multistage approach to detect subtle DDoS attacks", Mathematical and Computer Modelling, Vol. 55, No. 1-2, pp198-213.
- [4] Wang, H., Zhang, D., & Shin, K. G., (2002, June) "Detecting SYN flooding attacks" Proceedings, Twenty-first annual joint conference of the IEEE computer and communications societies, Vol. 3, pp1530-1539.
- [5] Chen, Y., Hwang, K., & Ku, W. S., (2007) "Collaborative detection of DDoS attacks over multiple network domains", IEEE Transactions on Parallel and Distributed Systems, Vol. 18, No. 12, pp1649-1662.
- [6] Lakhina, A., Crovella, M., & Diot, C., (2004) "Diagnosing network-wide traffic anomalies", ACM SIGCOMM computer communication review, Vol. 34, No. 4, pp219-230.

- [7] Li, X., Bian, F., Crovella, M., Diot, C., Govindan, R., Iannaccone, G., & Lakhina, A., (2006, October) "Detection and identification of network anomalies using sketch subspaces", In Proceedings of the 6th ACM SIGCOMM conference on Internet measurement, pp147-152.
- [8] Liu, H., Sun, Y., & Kim, M. S., (2011) "A Scalable DDoS Detection Framework with Victim Pinpoint Capability, Journal of Communications, Vol. 6, No. 9, pp660-670.
- [9] Salem, O., Vaton, S., & Gravey, A., (2010) "A scalable, efficient and informative approach for anomaly-based intrusion detection systems: theory and practice", International Journal of Network Management, Vol. 20, No. 5, pp271-293.
- [10] Schweller, R., Gupta, A., Parsons, E., & Chen, Y., (2004, October) "Reversible sketches for efficient and accurate change detection over network data streams", In Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, pp207-212.
- [11] Cormode, G., & Muthukrishnan, S., (2005) "An improved data stream summary: the count-min sketch and its applications", Journal of Algorithms, Vol. 55, No. 1, pp58-75.
- [12] Smith, L. I., (2002) "A tutorial on principal components analysis", Cornell University.
- [13] Huang, L., Nguyen, X., Garofalakis, M., Jordan, M., Joseph, A., & Taft, N., (2006) "In-network PCA and anomaly detection", Advances in neural information processing systems, Vol. 19.
- [14] Wang, F., Wang, X., Hu, X., & Su, J., (2012, October) "Bitwise sketch for lightweight reverse IP reconstruction in network anomaly detection", In 2012 IEEE 9th International Conference on Mobile Ad-Hoc and Sensor Systems (MASS 2012), pp1-4.
- [15] Cui, Y., Qian, Q., Guo, C., Shen, G., Tian, Y., Xing, H., & Yan, L., (2021) "Towards DDoS detection mechanisms in software-defined networking", Journal of Network and Computer Applications, Vol. 190, pp103156.
- [16] Wytrębowicz, J., (2018, October) "Software-defined anti-DDoS: Is it the next step?", In Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2018, Vol. 10808, pp652-663.

AUTHORS

Fei Wang received the PhD. degree from the National University of Defense Technology, China. She is now an associated professor in School of Computer, National University of Defense Technology. Her current research interests are next generation networks, network security and network monitoring and forensics.



Zhenxing Li received the master's degree from Guilin University Of Electronic Technology, China. He is now an R&D Engineer in School of Computer, National University of Defense Technology. His Current research interests are Network Traffic Analysis and Network Security



Xiaofeng Wang is an assistant professor in School of Computer, National University of Defense Technology (NUDT), China. He completed his PhD at NUDT in 2009. His current research interests are in trust and security of networking systems, distributed and intelligent data processing. He has published several papers in renowned journals and conferences like IEEE/ACM CCGrid, AINA, IEEE Transactions on Services Computing and Elsevier FGCS etc.



AN AUTOMATIC SHEET MUSIC GENERATING ALGORITHM BASED ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Ruize Yu¹ and Yu Sun²

¹Santa Margarita Catholic High School,
22062 Antonio Pkwy, Rancho Santa Margarita, CA 92688

²California State Polytechnic University, Pomona,
CA, 91768, Irvine, CA 92620

ABSTRACT

Due to the ever growing popularity of music as a part of everyday life, and with the continuous advances in AI technology, it is now possible for computers to listen to and recognize music [1]. However, there still exist limitations on machines' ability to recognize audio. This paper proposes an application to simplify the process of music transcription and reduce its runtime [2]. This application was tested in a different range of settings and evaluated. The results show what can be further improved on this application.

KEYWORDS

MIDI, Pitch Recognition, Magenta.

1. INTRODUCTION

The human ear is able to detect a range of frequencies, allowing it to “register” different sounds upon hearing it. Due to ever-advancing technology, it is no surprise that researchers will eventually create pitch detection algorithms that capture sounds and send the information to the algorithm [4]. Ever since pitch detection algorithms have been introduced, people have found creative ways to utilize such algorithms to solve various problems that may present themselves [5]. One of such problems is recognizing the various pitches of a music file, and writing them in a sheet music format. This has always been done with human hearing, thus increasing the room for error if the said person does not have a perfect pitch. Thankfully, researchers have looked into this problem and have created pitch detection algorithms described earlier to deal with problems like such.

Similar programs have already been created to help the user convert a raw music file into MIDI, many of which exist as online conversion tools [6]. An implementation of such programs is Ableton, which comes with many additional settings on top of converting music to MIDI. However, if a user wishes to write the said file into sheet music, they have to go through various processes until they can reach the final result. Another problem that arises with the process is that the MIDI file might take extra space, making it very inefficient if the user only wants to convert the raw audio into a piece of sheet music. Most of the existing programs lack a comprehensive workflow that allows the user to complete the entire process without interruptions.

In this paper, the method used is very similar to other music-to-MIDI programs. However, our method includes converting the MIDI file to sheet music and writing it as an image for the user. The goal of this method is to attempt to create an algorithm that writes an audio file into sheet music and reduces the number of steps needed to do so. Some of the features included in this algorithm are connected together to provide a satisfactory result. First, the audio is converted from a .wav file into a .midi file, the MIDI file is manipulated by Mido [7]. Magenta and TensorFlow AI provides the conversion needed from raw audio to MIDI. Second, the MIDI file is converted into sheet music, this step requires the use of Music21, which provides all the components needed for a piece of sheet music, such as the tempo, time signature, measures, and such. The third feature of this algorithm is the storage of the converted sheet music, which uses Google Firebase to store the said sheet music, which is an image at the current stage. This image is then sent to the website, which is created using both ngrok and flask.

To demonstrate how the above techniques reduce steps taken to convert audio to sheet music, and how the program is accurate, there will be two different approaches taken. The first approach tests each individual process using google colab [8]. The program is separated into three distinct steps. The first step transcribes a raw audio file into a MIDI file, the second step converts the MIDI file into sheet music, and the last step opens up a webpage that displays the converted sheet music. As shown in a later experiment, the three steps require building up off of each stage to provide the final result. The second approach is through repl.it, and tests the program from end-to-end. This is done to show how the program functions without the need to go through individual steps.

The rest of the paper is organized as follows: Section 2 provides the details on the challenges faced during the experiment and designing stages; Section 3 focuses on the details of the solutions corresponding to the challenges in Section 2; Section 4 presents the details about the experiments done, following by presenting the related work in Section 5; Section 6 will show the concluding remarks, as well as the future work of this project.

2. CHALLENGES

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

2.1. Identifying the Pitches

The first challenge that presented itself is identifying the different pitches within a raw audio file. Since the computer reads in pitches and processes it in a different way than the human ear, it needs to first know which pitch matches which music note. Additionally, it needs to be able to recognize the length of each pitch and record it inside the MIDI file. The second part of this challenge is writing into a MIDI file, which requires an algorithm and tuning [9]. The program cannot proceed without an accurate MIDI file, which will then be converted into a music sheet.

2.2. Recognizable Sheet Music

Another problem that came up was ensuring that the MIDI conversion writes an accurate and readable sheet music for the reader. Because music sheets have different clefs and such, the program needs to convert notes to their respective clef and will need to transcribe the notes depending on the clef. Another problem within this situation is writing in a different tempo and music notes, such as eighth notes, or the common quarter notes. The program needs to differentiate between when to use different notes and when to just change the tempo to make the sheet easier to read.

2.3. Creating the Website

After finishing the algorithm for the conversion, a website needs to be made. While a website is not very hard to create, the problem resides in connecting the algorithm, which uses google firebase, with the website itself.

3. SOLUTION

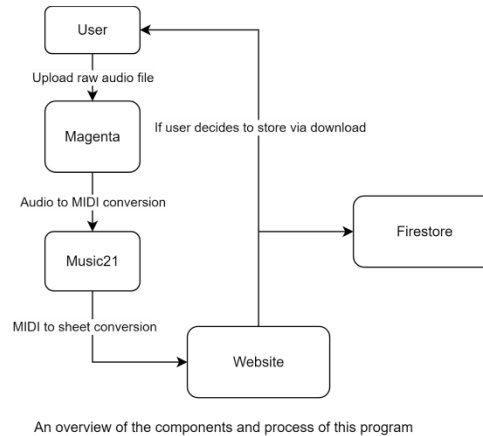


Figure 1. An overview of the components and process of this program

First, this process requires the user to upload a raw audio file in the form of .wav. This audio file is received by Magenta, a machine learning library focused and trained on music provided by Google, as an input to convert to MIDI [3]. Magenta will then use its Onset and Frames feature to convert the file to MIDI and will return it as an output. Then, the output will be received by Music21, a python-based toolkit for musicology. Music21 will primarily be used to write the MIDI file to sheet music in this process [15]. After Music21 finishes running, the output will be returned as a png image of the sheet music. This file will then be returned to the website, hosted by flask and ngrok, which will display the finished image to the user. The file will also be uploaded to Firestore, a cloud storage location.

Show File-select Fields

Show a file-select field which allows only one file to be chosen:

Select a file: Alesis-Fusio...-ass-C2.wav

Figure 2. Select file screen

The website first instructs the user to upload a file and press submit. Using python, the music to sheet conversion is done with the help of Magenta, specifically, the use of its onset and frames function.

```

import os
#return the midi file path
def generateMidi(filename:str):
    #generate Midi file and saved as song.wav.midi
    command_line = f'onsets_frames_transcription_transcribe \
--model_dir="models/maestro/train" \
--load_audio_with_librosa=True \
--transcribed_file_suffix="" \
{filename}'
    print(filename)
    os.popen(command_line).read()
    path = f'{filename}.midi'
    return path

```

Figure 3. The Screenshot of code 1

The `command_line` provides the necessary program configuration, with the audio file as the target. This method returns the transcribed audio file as a MIDI file, which is then used to convert to sheet music. To convert MIDI to sheet music, the program uses Music21 to parse the MIDI file (shown below as `midi_path`) to parse it into a music21 object.

```

def midi2sheet_music(midi_path:str, song_path):
    #print("hi")
    filename = pathlib.Path(song_path).stem
    create_directory(filename)
    print(filename)
    print(f"#####{midi_path}")
    parsed = music21.converter.parse(midi_path)
    print (type(parsed))
    #parsed.show()
    #parsed.write('musicxml.png', f'{filename}/{filename}.png')
    return parsed

```

Figure 4. Screenshot of code 2

The annotated object is then used as a parameter for the `generate_sheet_music` function, which will use it to return a png file of the sheet music.

```

def generate_sheet_music(song_path, folder_path = '/content/drive/MyDrive/Music_Sheet_Thing/'):
    midi_path = generateMidi(filename = song_path)
    print(f"*****{midi_path}*****")
    sheet_music_dir = midi2sheet_music(midi_path = midi_path, song_path= song_path)
    print(f'SHEET MUSIC DIR: {sheet_music_dir}')
    print(f'{folder_path}{midi_path}.png')
    #find why midi_path adds a s
    sheet_music_dir.write("musicxml.png", f'{folder_path}{midi_path}.png')
    url = firebase_storage.upload_file(local_path=f'{folder_path}{midi_path}.png', file_name= f'music/temp')
    return url

```

Figure 5. Screenshot of code 3

The program uses music21's `.write` function to write the object into a png file, which is then returned after being stored in the firestore.

Final image

<https://storage.googleapis.com/music-to-sheet.appspot.com/music/Alesis-Fusion-Acoustic-Bass-C2.wav.midi-1>

♩ = 120



Figure 6. Final image

The png file is then displayed on the website that originally prompted the user to input a .wav file.

4. EXPERIMENT

4.1. Experiment 1: Individual Steps via Google Colab

In this experiment, we tested and timed each function of the program bit-by-bit. This is to find where the algorithm takes the most time to execute, and where the bottleneck occurs.

To proceed with the experiment, each code block is run individually using Google Colab, and their runtime and memory usage are all recorded with %%timeit, %%prun, and %%memit.

4.2. Experiment 2: Complete Process Using the Flask Website

Have multiple testing data be uploaded to the website and be converted to sheet music check for errors, inaccuracies, and see what can be improved. The result will be collected and will provide feedback on the accuracy of the algorithm. This experiment will provide us with data on how to improve the algorithm, and where the errors are occurring.

To test the accuracy of the converter, three different wav files of C major notes were created using MuseScore. The three different data were uploaded to the website, where it is then converted back to sheet music. The first wav file is a C Major scale played on a piano in common time. The second file is a scale of C Major triads, played on a piano in common time and in ascending scale. The last file is a variation of the first file but split up with a quarter rest between the notes.

After creating the sample data and running them through the algorithm, it appears that the general shape and notes of the audio file are preserved. However, there are cases where the converter showed inaccuracies. For example, in the test case where chords are uploaded to the website (Figure 2), the only inaccuracy shown is in the first chord, where the algorithm missed concert C4 (Figure 2b). Another example is Figure 1b, where the C Major ascending & descending scales showed that some notes are carrying on the next beat. When separated (Figure 3), the algorithm works fine again.



Figure 7. Original Version (Figure 1a)



Figure 8. Magenta Version (Figure 1b)

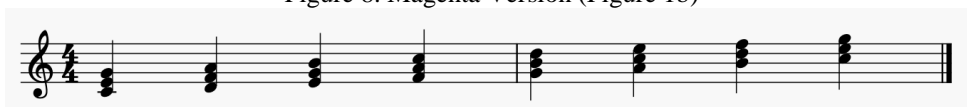


Figure 9. Original Version (Figure 2a)

reaching a F1 score of 0.95 compared to Magenta's F1 score of 0.9. Their AMT design proved to be better than Magenta, achieved through their neural network design and training of data. This AMT tool will have an edge over other music transcription algorithms due to it using a different deep learning API trained by themselves.

Mingheng Liang, in his research Music Score Recognition and Composition Application Based on Deep Learning, presents a deep-learning based music score recognition model [12]. The said model employs a deep network, accepts the entire score image as input, and outputs the note's time value and pitch directly. The work in question is almost the opposite of the algorithm described in our work; instead of accepting and reading an audio file as an input, Liang's work processes a musical score image and transcribes it into notes to be played.

In the work Magenta Studio: Augmenting Creativity with Deep Learning in Ableton Live by researchers from Google and the University of California, Berkeley, a design of a music generation application is provided, as well as its implementation as a plugin for the audio workstation Ableton Live [13]. The authors provided a flowchart of their application, which shows that the API references Magenta; the same library that the algorithm in this paper uses. However, the major difference between the two works is that Magenta Studio features multiple different plug-ins in addition to music generation, while this algorithm only focuses on the conversion of audio files to sheet music.

6. CONCLUSION

In this paper, we proposed a method that combines the steps of converting audio to MIDI, converting MIDI to sheet music, and uploading the sheet music to a website, into one algorithm. To test this algorithm's efficiency and accuracy, two different experiments were conducted, one tests the runtime and memory usage, while the other one tests the accuracy of the application. In the first experiment, each stage of the music transcription algorithm was timed during its runtime and recorded. In the second experiment, three different piano audio files were prepared using MuseScore 3, and the accuracy of this algorithm is tested by running it on all three audio files and comparing the result with the original music sheet [14]. The first experiment showed that the average run time is prolonged due to an unnecessary line during testing, which was eventually removed to reduce the run time. The second experiment showed Magenta's limitations on music transcription.

Currently, this system's accuracy is limited by using Google's Project Magenta [10]. Although accurate, it only achieves an F1 score of 0.9, which shows that it is still prone to external noises such as lingering notes and audio. This will only worsen when this application is used to transcribe any complex composition. During testing, it is also shown that there are no current ways to make adjustments midway, as the transcription is done from end-to-end.

These limitations will be solved if the future implementation of this program allows for access to the algorithm in between steps to adjust for error, as well as possibly training a deep learning network similar to Gossman's AMT system in work 1.

REFERENCES

- [1] Extance, Andy. "How AI technology can tame the scientific literature." *Nature* 561.7722 (2018): 273-275.
- [2] Benetos, Emmanouil, et al. "Automatic music transcription: An overview." *IEEE Signal Processing Magazine* 36.1 (2018): 20-30.

- [3] Benetos, Emmanouil, et al. "Automatic music transcription: challenges and future directions." *Journal of Intelligent Information Systems* 41.3 (2013): 407-434.
- [4] Rabiner, Lawrence, et al. "A comparative performance study of several pitch detection algorithms." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 24.5 (1976): 399-418.
- [5] Luengo, Iker, et al. "Evaluation of pitch detection algorithms under real conditions." *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*. Vol. 4. IEEE, 2007.
- [6] Loy, Gareth. "Musicians make a standard: the MIDI phenomenon." *Computer Music Journal* 9.4 (1985): 8-26.
- [7] Snyder, David, Guoguo Chen, and Daniel Povey. "Musan: A music, speech, and noise corpus." *arXiv preprint arXiv:1510.08484* (2015).
- [8] Alves, Francisco Regis Vieira, and Renata Passos Machado Vieira. "The Newton fractal's Leonardo sequence study with the Google Colab." *International Electronic Journal of Mathematics Education* 15.2 (2019): em0575.
- [9] Yamamoto, Kotaro, and Munetoshi Iwakiri. "A standard MIDI file steganography based on fluctuation of duration." *2009 International conference on availability, reliability and security*. IEEE, 2009.
- [10] Roberts, Adam, et al. "Magenta studio: Augmenting creativity with deep learning in ableton live." (2019).
- [11] Grossman, Aitan, and Josh Grossman. "Automatic Music Transcription: Generating MIDI From Audio." (2020).
- [12] Byrski, Aleksander, and Marek Kisiel-Dorohinicki. *Evolutionary Multi-Agent Systems: from inspirations to applications*. Vol. 680. Springer, 2016.
- [13] Roberts, Adam, et al. "Magenta studio: Augmenting creativity with deep learning in ableton live." (2019).
- [14] Todea, Diana. "The Use of the MuseScore Software in Musical E-Learning." *Virtual Learn* (2015): 88.
- [15] Cuthbert, Michael Scott, and Christopher Ariza. "music21: A toolkit for computer-aided musicology and symbolic music data." (2010).

LOW-CARBON INNOVATION DECISION CONSIDERING QUALITY DIFFERENCES AND GOVERNMENT SUBSIDIES UNDER THE THREE-PARTY TRADING PLATFORM

Xu Wang^{1,*}, Longzhen Zhou¹, Zusheng Zhang¹,
Yingbo Wu² and Longxiao Li³

¹College of Mechanical and Vehicle Engineering,
Chongqing University, Chongqing, 400030, China

²School of Big Data & software engineering,
Chongqing University, Chongqing, 401331, China

³School of Business Administration,
Chongqing University of Science and Technology, Chongqing, 401331, China

ABSTRACT

In the context of low-carbon innovation, reasonable subsidy, innovation, and pricing strategies are important to achieve resource decarbonization and supply-demand matching, while the quality differentiation of resources has a significant impact on the strategy formulation. In this paper, we study low-carbon innovation and government subsidy in different innovation scenarios with two providers offering differentiated manufacturing resources on a resource trading platform, integrating two variables of resource quality difference and demand-side low-carbon preference. Using utility theory and the Stackelberg game, a decision model of low carbon innovation and government subsidy is constructed, and the equilibrium solution is obtained with inverse induction. Then, the low-carbon innovation and subsidy strategies under different innovation scenarios are compared and the effects of relative coefficients of quality and innovation cost coefficients on the strategies are analyzed. The findings show that when the difference in resource quality is small, the level of green innovation is higher in the low carbon innovation scenario with high-quality resources compared to the low carbon innovation scenario with low-quality resources, and the rate of government subsidy for innovation investment is also higher. In case of the large difference in resource quality, the relative magnitudes of green innovation level and government subsidy rate for innovation inputs in different scenarios are related to innovation cost coefficients.

KEYWORDS

Low-carbon innovation, Quality differentiation of resources, Government subsidies, Manufacturing resource trading platform.

1. INTRODUCTION

With the reduction of renewable resources, ecological destruction, and increasing pressure on environmental protection, especially the convening of the United Nations Climate Change Conference, more and more countries begin to pay attention to environmental protection[1].Some governments subsidize enterprises and consumers who produce or use green and energy-saving products to improve the environment and support the development of the environmental protection industry. In recent years, subsidies for green manufacturing enterprises have been

David C. Wyld et al. (Eds): AI, AIMLNET, BIOS, BINLP, CSTY, MaVaS, SIGI - 2022

pp. 259-278, 2022. CS & IT - CSCP 2022

DOI: 10.5121/csit.2022.121823

introduced in Beijing, Shenzhen, Guangzhou, and other places in China, including subsidies for low-carbon products, support for low-carbon projects, and one-time funding incentives.

In this context, some scholars began to study the impact of government intervention on the green supply chain. Sheu studied the impact of government financial intervention on the competitive green supply chain. The study shows that the government should adopt green tax and subsidy policies to ensure that the profit of green product production is non-negative[2]. Madani first discussed the competition strategy between the green supply chain and the non-green supply chain under the leadership of the government, considering the government subsidies for green products, and taxing non-green products. The study shows that the impact of the government to improve the subsidy rate is far greater than the tax rate, which will also lead to increased profits for the government and the supply chain and product sustainability[3]. Zhu Qinghua studied the green supply chain management problem based on government subsidies and considered the three-stage game model of product green degree and government subsidies. The research shows that when the production cost coefficient of green products is high and the consumer's environmental awareness is low, the government should appropriately reduce the lower limit of subsidies and reduce the green input of products. When the production cost coefficient of green products is low and consumers' environmental awareness is high, the opposite is true[4].

In response, a growing number of firms have devoted attention to green technologies investment in R&D and production processes to curb carbon emissions[5, 6]. For example, Apple invested heavily in green technologies such as renewable energy in 2017. With these investments, the company cut emissions by nearly 2m tonnes from last year. However, green investment is not free; Companies have to bear a lot of investment costs, which may reduce the benefits of green investment and even become an important obstacle to the adoption of green technology[7, 8]. Therefore, whether to invest in green technology has become an important issue for enterprises. In addition to the price and greenness of the product, the quality of the product is also a powerful means of attracting customers and even determines the position of an enterprise in the market. When enterprises with different product quality choose low-carbon innovation strategies, they will be different due to different market positions or consumer preferences. How to improve their product's greenness to improve their profitability in the context of product quality that cannot be improved simply has become a key issue for enterprises in the competitive environment, especially in the transformation to green production. However, so far, most studies haven't studied the quality differentiation in low-carbon innovation decisions. Quality is defined as the environmental quality that reflects the green degree of products[8] or considers the relationship between quality improvement investment and low-carbon innovation investment[9].

The third-party platforms can reduce the technological input of manufacturing enterprises and integrate resources and capabilities among manufacturing enterprises, which is of great significance to promoting the development of the platform economy and manufacturing industry. With the continuous integration of information technology and the manufacturing industry, the third-party platform of the manufacturing industry has developed rapidly, which has attracted the attention of many scholars. For example, studies have shown that the main motivation for enterprises to join the platform is that the services provided by the platform can help enterprises improve efficiency and reduce transaction costs[10]. Yoo compares the profits obtained by enterprises joining the third-party platform with those that have self-built platforms respectively and comprehensively analyzes the influence factors of enterprises joining the third-party platform. They found that companies' IT capabilities, costs, and purchasing needs were the main drivers and revealed that SMEs were better suited to join third-party platforms[11]. Many scholars have studied the third-party platform from the basic functions, architecture, operation model, and implementation technology of the platform[12-14]. In addition, there are few studies

on service investment and pricing strategies on third-party platforms, especially on resource quality or greenness improvement on third-party platforms.

Based on the above research, considering the impact of third-party platform transactions and resource quality differences on resource providers' low-carbon innovation strategy and government subsidies strategy, this paper studies the problem of low-carbon innovation of third-party platform resource providers under government subsidies. This paper considers a third-party trading platform composed of low-quality resource providers(), high-quality resource providers(), and resource demanders(). The government will issue a certain subsidy strategy for low-carbon innovation projects, and resource providers are likely to invest in low-carbon innovation of resources. Through the research of this paper, some different results are obtained from previous studies. Firstly, resource quality differences and low-carbon innovation cost coefficients will affect the low-carbon innovation decisions of quality resource providers, and low-quality resource providers should also consider platform transaction rates. Different from previous studies, this paper finds that high-quality resource providers have lower innovation willingness than low-quality resource providers in terms of market share and profit. Then, government subsidies are not always biased towards high-quality resource providers, and low-quality resource providers are given higher rates of subsidies when the quality differences between the two sides are large and innovation cost coefficients are low. Finally, the platform development is not necessarily committed to improving the overall level of resources, in different cases, different resource providers can have different incentive strategies.

2. PROBLEM DESCRIPTION AND SYMBOLIC DESCRIPTION

2.1. Problem Description

On the third-party manufacturing resource trading platform, a type of manufacturing resource is provided by two competing manufacturing resource providers $RSP_i(i=1,2)$, and the two providers provide two differentiated resources $R_1 \setminus R_2$, which R_2 has a more reliable quality level and R_1 is inferior R_2 in this respect. RSD can choose any of the resources for direct transactions, while RSP_i paying a certain commission fee to the resource-sharing platform. As two competing suppliers, they will invest in low-carbon innovation projects to improve their product green degree, thereby affecting the choice of resource demanders to maximize their profits. At the same time, the regulatory authorities of the government also have the responsibility to enhance social-environmental benefits and to give resource providers certain project subsidies. Therefore, under the influence of different government subsidy strategies on the low-carbon innovation level of resource providers, how do subsidies affect their low-carbon innovation decisions and pricing for suppliers with different quality resources?

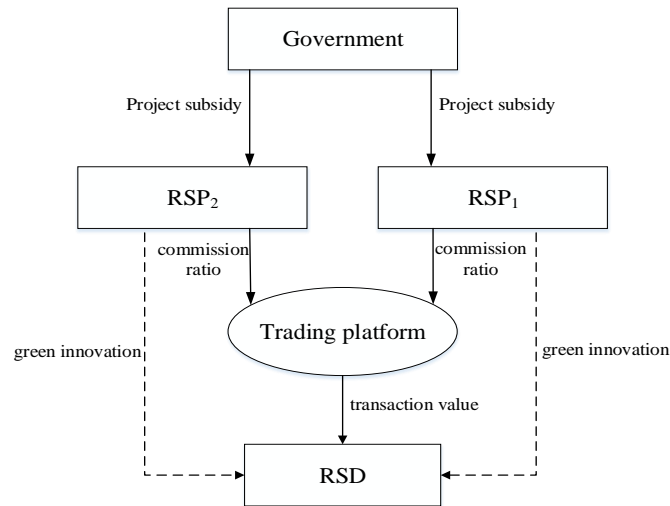


Figure 1. Operation Mode of Tripartite Trading Platform

2.2. Symbolic Description

The relevant symbols in this paper are as follows shows:

Table 1. Symbolic Description

	Symbolic	Description
Parameter	v	the initial utility of resource demanders for high-quality manufacturing resources ($0 < v < 1$)
	θ	relative coefficients of resources quality ($0 < \theta < 1$)
	β	transaction rates on trading platforms
	k	low-carbon innovation cost coefficients
	$u_i (i = 1, 2)$	utility of resource demanders
	$Q_i (i = 1, 2)$	market share of resource providers
Decision variable	$P_i (i = 1, 2)$	transaction prices of resource
	$m_i (i = 1, 2)$	low-carbon innovation effort coefficient of resource providers
	η	government subsidy rate
Other symbols	RSD	resource demanders
	RSP_i	resource providers
	R_i	resources
	$\pi_i (i = 1, 2, g)$	profit (utility) of different themes

3. MODEL CONSTRUCTION AND SOLUTION

Hypothesis 1: In some literature, the linear combination of price and non-price variables is generally used to represent the demand function[15-17], and some empirical studies also found that the level of green degree of enterprise products can affect consumers' purchasing behavior. This paper refers to Gao Juhong's research[18], the relationship between the nature of

manufacturing resources and perceived utility is expressed by the following utility function: $u = V - e_1 P + e_2 m$ (wherein, the initial utility is expressed as $v \sim U(0,1)$, a is the price elasticity coefficient, b is the environmental sensitivity coefficient of demand, and the impact of the green degree of resources on demand is bm). To facilitate the analysis results, and without losing generality, set $e_1 = e_2 = 1$.

Hypothesis 2: With the continuous improvement of the green degree of manufacturing resources, it is more and more difficult to improve the green degree, so the cost function of low-carbon innovation should have the characteristics of increasing marginal cost. At the same time, combined with reality, there is a certain fixed investment in green innovation that can't be recovered in time. This paper sets this fixed cost as a . Therefore, the cost function of green innovation of manufacturing resources can be expressed as $f(m) = 1/2km^2 + a$, which k is the low-carbon innovation cost coefficient.

Hypothesis 3: To simplify the calculation without losing generality, the cost of providing unit manufacturing resources is assumed to be zero.

Hypothesis 4: Referring to previous research results[19], social welfare is composed of total industry profits and environmental benefits. So the industry social welfare function considered by the government can be set to:

$$\pi_g = \pi_1 + \pi_2 + \pi_p + m_l Q_1 + m_h Q_2 \quad (1)$$

Hypothesis 5: The resource sharing platform introduces manufacturing resources with an audit mechanism, the quality difference between the two RSP_i would not be too large, assuming the quality difference $0.6 < \theta < 0.99$.

Hypothesis 6: The above information is public knowledge.

3.1. Basic model

This paper constructs a government- resource trading platform -manufacturing resource providers and demands decision-making model composed of two manufacturing resource platform suppliers and demanders, which meets the manufacturing resource demand at the transaction price P_i and pays a certain tariff as βP_i to the platform. At this time, the government has not issued the low-carbon innovation subsidy policy and RSP_i has decided not to carry out low-carbon innovation, only through price decision-making to maximize its profits. According to the above assumptions, the initial utility RSD is:

$$u_1 = \theta v - P_1 \quad (2)$$

$$u_2 = v - P_2 \quad (3)$$

When the conditions $\theta v - P_1 > 0$ $\theta v - P_1 < v - P_2$ are satisfied, RSD select R_1 to trade. So after simplification, we get when $P_1 < \theta P_2$ is satisfied, the high-quality resources and low-quality resources on the platform exist in the market and compete with each other, and the demand function R_1 is:

$$Q_1^0 = \int_{\frac{P_1}{\theta}}^{\frac{P_2-P_1}{1-\theta}} dv = \frac{P_2-P_1}{1-\theta} - \frac{P_1}{\theta} \quad (4)$$

When $P_1 > \theta P_2$ we get high-quality resources completely replace low-quality resources RSP_1 and exit the competition on the platform, high-quality resources form a monopoly RSD and have only one choice on this platform. This paper does not consider this situation. Similarly, when RSD chosen R_2 , at this point satisfies conditions $v - P_2 > 0$ $\theta v - P_1 > v - P_2$, the demand function R_2 is:

$$Q_2^0 = \int_{\frac{P_2-P_1}{1-\theta}}^1 dv = 1 - \frac{P_2-P_1}{1-\theta} \quad (5)$$

At this time, the profit functions of RSP_i the resource trading platform are:

$$\pi_1 = (1-\beta)P_1Q_1^0, \quad \pi_2 = (1-\beta)P_2Q_2^0, \quad \pi_p = \beta(P_1Q_1^0 + P_2Q_2^0)$$

The RSP_i trader will determine the transaction price of manufacturing resources according to the known market conditions, to maximize profits. The decision model is:

$$\max_{\{P_1, P_2\}} (\pi_1^0, \pi_p^0), \text{ s.t. : } P_1 > 0, \quad P_2 > 0, \quad P_1 < \theta P_2$$

Theorem 1: When RSP_i does not carry out low-carbon innovation, there is a balanced price strategy P_1^{0*} P_2^{0*} , and the optimal profits of RSP_i and platform profit are respectively:

$$P^{0*} : \begin{cases} P_1^{0*} = \frac{(\theta-1)\theta}{\theta-4} \\ P_2^{0*} = \frac{2(\theta-1)}{\theta-4} \end{cases} \quad (6)$$

The optimal profit RSP_i and platform profit obtained from Equation (6) are:

$$\pi^{0*} : \begin{cases} \pi_1^{0*} = \frac{\theta(-1+\theta)(-1+\beta)}{(\theta-4)^2} \\ \pi_2^{0*} = \frac{4(-1+\theta)(-1+\beta)}{(\theta-4)^2} \end{cases} \quad (7)$$

$$\pi_p^0 = -\frac{(\theta+4)(-1+\theta)\beta}{(\theta-4)^2} \quad (8)$$

Theorem 1 shows that in the absence of government intervention, P_1^{0*} is always less than P_2^{0*} , and the equilibrium price strategy is only affected by the relative coefficients of quality.

Lemma 1: In the basic model, the equilibrium price strategy of RSP_i decreases with the increase θ . For the three profit functions, they all decrease with the increase θ . Lemma 1 is in line with traditional cognition. When the resource trading platform doesn't conduct such behavior as low-carbon innovation to enhance the utility of demanders at the other end, its income will not

increase with the homogenization of manufacturing resources, that is, such trading intermediaries as the platform need to provide sufficient choices of demanders to maximize their interests. For RSP_1 , when the quality difference between the two RSP_i is large, the price can rise by improving their resource quality. However, when the relative coefficients of quality are small, the two RSP_i are at the same consumption level. To maximize their profit, the price war can only be used to occupy the market, RSP_2 . But when both sides take such a price war, their profit margins will all decline.

3.2. Model for Low-Carbon Innovation of Resource Provider

When implementing the government subsidy mechanism, the decision-making problem of the government-enterprise low-carbon innovation-decision problem can be understood as the Stackelberg game of the government as a leader and the enterprise as a follower. At the same time, considering that low-carbon innovation investment is a strategic decision, so the decision-making order is shown in figure2, and the inverse order solution method is used to obtain the equilibrium results. Considering the different innovation ability of RSP_i in improving the green degree of resources, two scenarios are set: (1) RSP_1 conducts low-carbon innovation; (2) RSP_2 conducts low-carbon innovation.

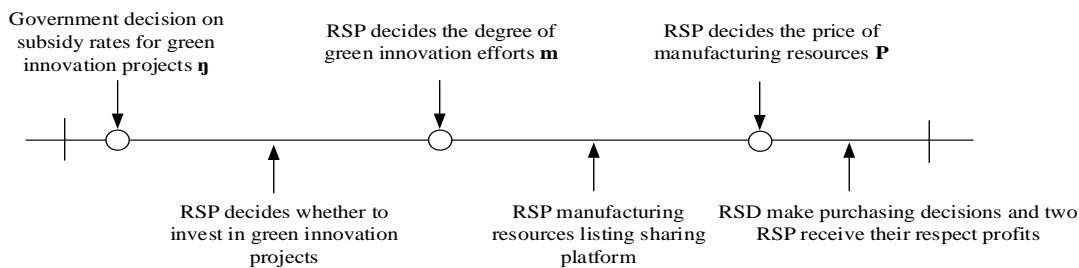


Figure 2. Game timeline

3.2.1. RSP_1 conducts low-carbon innovation

After the government released the subsidy policy for low-carbon innovation projects, in this scenario RSP_1 becomes an innovative enterprise for project investment, its profit function is $\pi_1^i = (1-\beta)P_iQ_1^i - (\frac{km_1^2}{2} + a) \cdot (1-\eta)$, and the profit functions of RSP_2 and platform unchanged. Similarly, the demand functions R_i are: $Q_1^i = \frac{P_2 + m_1 - P_1}{1-\theta} - \frac{P_1 - m_1}{\theta}$ and $Q_2^i = 1 - \frac{P_2 + m_1 - P_1}{1-\theta}$, the industry social welfare function considered by the government is $\pi_g^i = \pi_s^i + \pi_r^i + \pi_p^i + m_1^i Q_1^i$. At this time the two enterprises play Bertrand games, the decision model is:

First stage: $\max_{(\eta)} (\pi_g^1), s.t.: 0 < \eta < 1$

The second stage: $\max_{\{P_1, P_2, m_1\}} (\pi_1^1, \pi_2^1), s.t.: P_1 > 0, P_2 > 0, m_1 > 0, P_1 < \theta P_2 + m_1$

Theorems 2: When RSP_1 conducting green innovation, the optimal subsidy rate of the government, the low-carbon innovation effort coefficient RSP_1 , the pricing policies of RSP_i P_1^{1*} and P_2^{1*} , their optimal profits and platform profits are respectively:

$$\eta^{*} = \frac{k(2\beta + 1)\theta^3 + (-6\beta - 1)k\theta^2 + (4\beta k + 2\beta - 2)\theta - 4\beta + 4}{k\theta(3\theta^2 - 7\theta + 4)} \tag{9}$$

$$m_1^{*} = \frac{-3\theta^3 + 7\theta^2 - 4\theta}{24 + k\theta^4 - 9k\theta^3 + (24k + 4)\theta^2 + (-16k - 18)\theta} \tag{10}$$

$$P^{*} : \begin{cases} P_1^{*} = \frac{\theta(-1 + \theta)(\theta - 4)(k\theta^2 - k\theta + 1)}{24 + k\theta^4 - 9k\theta^3 + (24k + 4)\theta^2 + (-16k - 18)\theta} \\ P_2^{*} = \frac{(-1 + \theta)(2k\theta^3 - 10k\theta^2 + 8k\theta + 5\theta - 12)}{24 + k\theta^4 - 9k\theta^3 + (24k + 4)\theta^2 + (-16k - 18)\theta} \end{cases} \tag{11}$$

The optimal profit and platform profit obtained from Equations (9)-(11) are :

$$\pi^{*} : \begin{cases} \pi_1^{*} = \frac{2(k\theta^2 - k\theta + 1)(k(a + \frac{3}{2})\theta^5 + (-11a - 5)k\theta^4 + ((42a + \frac{11}{2})k + 4a)\theta^3 + ((-64a - 2)k - 26a)\theta^2 + (32ak + 60a)\theta - 48a)(-1 + \beta)}{3k\theta(\theta - \frac{4}{3})(-1 + \theta)(24 + k\theta^4 - 9k\theta^3 + (24k + 4)\theta^2 + (-16k - 18)\theta)} \\ \pi_h^{*} = \frac{4(-1 + \beta)(-6 + k\theta^3 - 5k\theta^2 + (4k + \frac{5}{2})\theta)^2(-1 + \theta)}{(24 + k\theta^4 - 9k\theta^3 + (24k + 4)\theta^2 + (-16k - 18)\theta)^2} \end{cases} \tag{12}$$

$$\pi_p^{*} = -\frac{\beta(k^2\theta^7 - 6k^2\theta^6 + (-7k^2 + 2k)\theta^5 + (92k^2 + 2k)\theta^4 + (-144k^2 - 100k + 1)\theta^3 + (64k^2 + 288k + 17)\theta^2 + (-192k - 104)\theta + 144)(-1 + \theta)}{(24 + k\theta^4 - 9k\theta^3 + (24k + 4)\theta^2 + (-16k - 18)\theta)^2} \tag{13}$$

Lemma 2: In this scenario, all k values satisfying the constraint are $(\frac{-5\theta + 12}{2\theta^3 - 10\theta^2 + 8\theta}, \frac{-2(\theta - 2)(\beta - 1)}{\theta(-1 + \theta)(2\beta\theta - 4\beta + \theta)})$.

$\frac{\partial m_1^{*}}{\partial k} < 0$, and for $\frac{\partial m_1^{*}}{\partial \theta}$: (1) When $\theta \in (0.6, \frac{(1052 + 12\sqrt{5865})^{\frac{1}{3}}}{6} - \frac{32}{3(1052 + 12\sqrt{5865})^{\frac{1}{3}}} + \frac{11}{3})$,

for $\frac{-5\theta + 12}{2\theta^3 - 10\theta^2 + 8\theta} < k < \frac{12\theta^4 - 108\theta^3 + 326\theta^2 - 336\theta + 96}{3(\theta + \frac{4}{3})(\theta - 1)^2(\theta - 4)\theta^2}$ we get $\frac{\partial m_1^{*}}{\partial \theta} > 0$.and for

$\frac{12\theta^4 - 108\theta^3 + 326\theta^2 - 336\theta + 96}{3(\theta + \frac{4}{3})(\theta - 1)^2(\theta - 4)\theta^2} < k < \frac{-2(\theta - 2)(\beta - 1)}{\theta(\theta - 1)(2\beta\theta - 4\beta + \theta)}$ we get $\frac{\partial m_1^{*}}{\partial \theta} < 0$;(2) When

$\theta \in (-\frac{(1052 + 12\sqrt{5865})^{\frac{1}{3}}}{6} - \frac{32}{3(1052 + 12\sqrt{5865})^{\frac{1}{3}}} + \frac{11}{3}, 0.8351771205)$ and $\beta < -\frac{\theta(\theta^3 - 11\theta^2 + 35\theta - 20)}{5\theta^4 - 43\theta^3 + 128\theta^2 - 148\theta + 48}$, $\frac{\partial m_1^{*}}{\partial \theta} > 0$.But when

$\theta \in (-\frac{(1052 + 12\sqrt{5865})^{\frac{1}{3}}}{6} - \frac{32}{3(1052 + 12\sqrt{5865})^{\frac{1}{3}}} + \frac{11}{3}, 0.8351771205)$ and $\beta > -\frac{\theta(\theta^3 - 11\theta^2 + 35\theta - 20)}{5\theta^4 - 43\theta^3 + 128\theta^2 - 148\theta + 48}$, for

$\frac{-5\theta + 12}{2\theta^3 - 10\theta^2 + 8\theta} < k < \frac{12\theta^4 - 108\theta^3 + 326\theta^2 - 336\theta + 96}{3(\theta + \frac{4}{3})(\theta - 1)^2(\theta - 4)\theta^2}$ we get $\frac{\partial m_1^{*}}{\partial \theta} > 0$, and for $\frac{12\theta^4 - 108\theta^3 + 326\theta^2 - 336\theta + 96}{3(\theta + \frac{4}{3})(\theta - 1)^2(\theta - 4)\theta^2} < k < \frac{-2(\theta - 2)(\beta - 1)}{\theta(\theta - 1)(2\beta\theta - 4\beta + \theta)}$

we get $\frac{\partial m_1^{*}}{\partial \theta} < 0$;(3) When $\theta \in (0.8351771205, 1)$ we get $\frac{\partial m_1^{*}}{\partial \theta} < 0$. For the sensitivity of government subsidy

rate: $\frac{\partial \eta^{*}}{\partial k} < 0$, $\frac{\partial \eta^{*}}{\partial \beta} > 0$, and for $\frac{\partial \eta^{*}}{\partial \theta} < 0$: (1) When $\theta \in (0.6, \frac{7}{4} - \frac{\sqrt{17}}{4})$ we get $\frac{\partial \eta^{*}}{\partial \theta} > 0$;(2) When

$\theta \in (\frac{7}{4} - \frac{\sqrt{17}}{4}, 0.830884198515145)$ we get $\frac{\partial \eta^{*}}{\partial \theta} > 0$;(3) When $\theta \in (0.830884198515145, 1)$ and $k < \frac{6\theta^3 - 25\theta^2 + 28\theta - 8}{2\theta^2(\theta - 1)^2}$, we get

$\frac{\partial \eta^{*}}{\partial \theta} > 0$. But when $\theta \in (0.830884198515145, 1)$ and $k > \frac{6\theta^3 - 25\theta^2 + 28\theta - 8}{2\theta^2(\theta - 1)^2}$, we get $\frac{\partial \eta^{*}}{\partial \theta} < 0$. For equilibrium price:

when $k \in (\frac{-5\theta + 12}{2\theta^3 - 10\theta^2 + 8\theta}, \frac{-\theta^2 + 9\theta - 12}{\theta(\theta^3 - 7\theta^2 + 14\theta - 8)})$ we get $P_1^{*} > P_2^{*}$, but when $k \in (\frac{-\theta^2 + 9\theta - 12}{\theta(\theta^3 - 7\theta^2 + 14\theta - 8)}, \frac{-2(\theta - 2)(\beta - 1)}{\theta(\theta - 1)(2\beta\theta - 4\beta + \theta)})$ we get

$P_1^{*} < P_2^{*}$. And $\frac{\partial P_1^{*}}{\partial k} < 0$, $\frac{\partial P_2^{*}}{\partial k} > 0$, $\frac{\partial P_1^{*}}{\partial \theta} < 0$.

Lemma 2 shows that with the increase of innovation cost coefficients, the government’s subsidy rate would decrease, and RSP_1 would reduce the investment in low-carbon innovation. For the resource pricing that directly affects the RSP_1 ’s profit, the increase in innovation cost will make the pricing lower. At this time, it will not choose to increase the resource pricing to recover the

investment because of the increase of the previous investment and P_2^{1*} show the correlation of differentiated green innovation cost. Because there is no threat caused by low-carbon improvement RSP_2 and RSP_2 will seek greater profits by increasing the pricing, which causes the result that $P_1^{1*} < P_2^{1*}$. When the innovation cost is low, low-carbon behavior implementers, RSP_1 tend to take greater risks to improve pricing to obtain higher benefits and RSP_2 will reduce pricing because of the threat posed, so $P_1^{1*} > P_2^{1*}$.

Compared with the impact of low-carbon innovation cost on various strategies, the influencing mechanism of resource quality difference between the two sides is more complex: (1) When the θ is large, the η^{1*} is positively correlated with θ . In this situation, when the k is small, under the influence of low innovation cost and high government subsidy rate, RSP_1 is willing to carry out low-carbon innovation on its own high-quality resources to enhance its market position, so m_1^{1*} is positively correlated with θ . If the k is large, although low-carbon innovation is beneficial to the rise of market position, the reduction of θ and the high cost of innovation are also easy to make RSP_1 be content with the status quo, and then make m_1^{1*} negatively correlated with θ ; (2) When the θ is small, the η^{1*} is positively correlated with θ . If the β is small, with the narrowing of the quality difference and the improvement of η^{1*} , although the θ is not big, but RSP_1 also intends to stabilize their market position and carry out innovation investment, m_1^{1*} is positively correlated with θ . If the β is large and the k is small, the increase in the β will increase η^{1*} , and m_1^{1*} is also positively correlated with θ . However, when the β and k are large, although the η^{1*} increases appropriately with the increase of θ and β , the high innovation cost hinders the low-carbon innovation behavior of RSP_1 , so m_1^{1*} is negatively correlated with θ ; (3) When the resource quality of both sides is very close and the k is small, the η^{1*} increases with the increase of θ , while m_1^{1*} decreases with the increase of θ . When the k is large, the η^{1*} decreases with the increase of θ , which is different from the situation that the quality difference is obvious before or the k is small. The larger θ let the government is more convinced that at this time RSP_1 has enough capital to carry out low-carbon innovation, and the increase of subsidy cost reduces the support, and m_1^{1*} always decreases with the increase of θ , so in this situation, regardless of the level of k , RSP_1 thinks that its resources have been able to meet the requirements of market competition and reduce the investment in low-carbon innovation. In this case, market resources are in the most difficult stage of low-carbon innovation.

3.2.2. RSP_2 conducts low-carbon innovation

In this scenario, RSP_2 as an innovative enterprise to conduct low-carbon innovation, its profit function is: $\pi_h^2 = (1 - \beta)P_2Q_2^2 - (\frac{km_2^2}{2} + a) \cdot (1 - \eta)$, at this time the profit function of RSP_1 and the platform is unchanged. Similarly, the demand functions of R_1 are respectively: $Q_1^2 = \frac{P_2 - m_2 - P_1}{1 - \theta} - \frac{P_1}{\theta}$ and $Q_2^2 = 1 - \frac{P_2 - m_2 - P_1}{1 - \theta}$. The industry social welfare function considered by the government is: $\pi_g^2 = \pi_h^2 + \pi_r^2 + \pi_p^2 + m_2^2Q_2^2$. At this time, the two enterprises conduct the Bertrand game, the decision model is:

First stage: $\max_{\{\eta\}}(\pi_g^2), s.t.: 0 < \eta < 1$

The second stage: $\max_{\{P_1, P_2, m_2\}} (\pi_r^2, \pi_h^2), s.t.: P_1 > 0, P_2 > 0, m_2 > 0, P_1 < \theta P_2 - \theta m_2$

Theorems 3: When RSP_2 conducting green innovation, the optimal subsidy rate of the government, the low-carbon innovation effort coefficient of RSP_2 , the pricing policies of RSP_i : P_1^{1*} and P_2^{1*} , their optimal profits and platform profits are respectively:

$$\eta^{2*} = \frac{(-1 + \theta)(\theta - 4)(\beta + 1)k - 2\beta + 2}{(2\theta^2 - 10\theta + 8)k} \tag{14}$$

$$m_2^{2*} = -\frac{8(-1 + \theta)(\theta - 2)}{k\theta^3 + (-9k + 4)\theta^2 + (24k - 18)\theta - 16k + 24} \tag{15}$$

$$P^{2*} : \begin{cases} P_1^{2*} = \frac{(-1 + \theta)(k\theta^2 + (-5k + 4)\theta + 4k - 10)\theta}{k\theta^3 + (-9k + 4)\theta^2 + (24k - 18)\theta - 16k + 24} \\ P_2^{2*} = \frac{2(-1 + \theta)(k\theta^2 - 5k\theta + 4k - 2)}{k\theta^3 + (-9k + 4)\theta^2 + (24k - 18)\theta - 16k + 24} \end{cases} \tag{16}$$

The optimal profit and platform profit obtained from Equations (14)-(16) are:

$$\pi^{2*} : \begin{cases} \pi_r^{2*} = \frac{(-1 + \beta)(k\theta^2 + (-5k + 4)\theta + 4k - 10)^2(-1 + \theta)\theta}{(k\theta^3 + (-9k + 4)\theta^2 + (24k - 18)\theta - 16k + 24)^2} \\ \pi_h^{2*} = \frac{(-1 + \beta)((a\theta^2 + (-8a + 8)\theta + 16a - 8)(-1 + \theta)k + 4a(\theta^2 - \frac{9}{2}\theta + 6))(-2 + k(\theta^2 - 5\theta + 4))}{2((\theta - 4)^2(-1 + \theta)k + 4\theta^2 - 18\theta + 24)(-1 + \theta)k(\theta - 4)} \end{cases} \tag{17}$$

$$\pi_p^{2*} = -\frac{(-1 + \theta)\beta(k^2\theta^5 + (-6k^2 + 8k)\theta^4 + (-7k^2 - 60k + 16)\theta^3 + (92k^2 + 116k - 80)\theta^2 + (-144k^2 + 100)\theta + 64(-\frac{1}{2} + k)^2)}{(k\theta^3 + (-9k + 4)\theta^2 + (24k - 18)\theta - 16k + 24)^2} \tag{18}$$

Lemma 3: When considering RSP_2 conducting low-carbon innovation, $P_1 < \theta P_2 - \theta m_2$ would be satisfied if $k \in (\frac{-4\theta + 10}{\theta^2 - 5\theta + 4}, +\infty)$. $\frac{\partial m_2^{2*}}{\partial k} < 0, \frac{\partial m_2^{2*}}{\partial \theta} < 0, \frac{\partial \eta^{2*}}{\partial k} < 0, \frac{\partial \eta^{2*}}{\partial \theta} > 0, \frac{\partial \eta^{2*}}{\partial \beta} > 0$. For the equilibrium price: $P_1^{2*} < P_2^{2*}$, $\frac{\partial P_1^{2*}}{\partial k} > 0, \frac{\partial P_1^{2*}}{\partial \theta} < 0, \frac{\partial P_2^{2*}}{\partial k} < 0, \frac{\partial P_2^{2*}}{\partial \theta} < 0$.

Lemma 3 shows that when the k increases, the government also decreases. The investment in low-carbon innovation projects should be reduced to avoid risks. At the same time, the P_2^{2*} will also be reduced to improve market share to recover the investment, but the P_1^{2*} will be increased to seek profits, but $P_1^{2*} < P_2^{2*}$ holds.

When the θ increases, RSP_2 will reduce investment in a low-carbon innovation project, while reducing resource pricing to ensure market priority, but at this time the market competition is more intense, and transaction prices will be reduced. For the government, when the subsidy object is RSP_2 , the η^{2*} will increase with the reduction of quality difference, that is, when the market lacks differentiated products, government incline to the enterprise caused more benefit social welfare.

4. COMPARATIVE ANALYSIS

The above analysis of resource sharing platforms with different quality levels of resource providers-government emission reduction model, the following from the social green innovation efficiency, government subsidies efficiency, and market share to be compared to get some management inspiration. By comparing the range of low-carbon innovation costs in the two

cases, it can be obtained that when $k \in (\frac{-4\theta + 10}{\theta^2 - 5\theta + 4}, \frac{-2(\theta - 2)(\beta - 1)}{\theta(\theta - 1)(2\beta\theta - 4\beta + \theta)})$ the two scenarios can be compared at the same time.

4.1. Comparison of Environmental Benefits

Lemma 3: For low-carbon innovation efforts coefficient difference $m_1^{1*} - m_2^{2*}$, when $\theta \in (0.6, -\frac{(51+12\sqrt{13})^{\frac{1}{3}}}{4} - \frac{9}{4(51+12\sqrt{13})^{\frac{1}{3}}} + \frac{9}{4})$ and $k \in (\frac{-4\theta + 10}{\theta^2 - 5\theta + 4}, \frac{12\theta^4 - 102\theta^3 + 352\theta^2 - 576\theta + 384}{5(\theta - 4)^2(-1 + \theta)(\theta - \frac{12}{5})\theta})$ we can get $m_1^{1*} > m_2^{2*}$, when then $k \in (\frac{-4\theta + 10}{\theta^2 - 5\theta + 4}, \frac{12\theta^4 - 102\theta^3 + 352\theta^2 - 576\theta + 384}{5(\theta - 4)^2(-1 + \theta)(\theta - \frac{12}{5})\theta})$ we get $m_1^{1*} < m_2^{2*}$; When $\theta \in (-\frac{(51+12\sqrt{13})^{\frac{1}{3}}}{4} - \frac{9}{4(51+12\sqrt{13})^{\frac{1}{3}}} + \frac{9}{4}, 1)$ we get $m_1^{1*} < m_2^{2*}$.

Lemma 4 shows that even if RSP_i carrying out low-carbon innovation on the same resource trading platform, the degree of low-carbon innovation efforts is also very different: in most cases, RSP_2 can pay more cost to improve the green degree of its resources. Only when the quality difference between the two sides is large enough and the cost of green innovation is small, the investment cost of RSP_1 is higher than RSP_2 . In this case, RSP_1 is at a disadvantage and the marginal cost of low-carbon innovation is low. At this time, RSP_1 has a greater determination to change than RSP_2 .

4.2. Comparison of Government Subsidy Rate

Lemma 5: Comparing the η^{i*} in different scenes, when $\theta < \frac{7}{2} - \frac{\sqrt{33}}{2}$ and $k \in (\frac{-4\theta + 10}{\theta^2 - 5\theta + 4}, \frac{-10\theta^2 + 32\theta - 32}{\theta(\theta - 1)(\theta - 4)^2})$ we get $\eta^{1*} > \eta^{2*}$, when $k \in (\frac{-10\theta^2 + 32\theta - 32}{\theta(\theta - 1)(\theta - 4)^2}, \frac{-2(\theta - 2)(\beta - 1)}{\theta(\theta - 1)(2\beta\theta - 4\beta + \theta)})$ we get $\eta^{1*} < \eta^{2*}$; When $\theta > \frac{7}{2} - \frac{\sqrt{33}}{2}$ we get $\eta^{1*} < \eta^{2*}$. Sensitivity analysis of government subsidy rate difference: $\frac{\partial(\eta^{1*} - \eta^{2*})}{\partial k} < 0$, and when $\theta < \frac{7}{2} - \frac{\sqrt{33}}{2}$ and $k \in (\frac{-4\theta + 10}{\theta^2 - 5\theta + 4}, \frac{-10\theta^2 + 32\theta - 32}{\theta(\theta - 1)(\theta - 4)^2})$ we get $\frac{\partial(\eta^{1*} - \eta^{2*})}{\partial \beta} < 0$, but when $k \in (\frac{-10\theta^2 + 32\theta - 32}{\theta(\theta - 1)(\theta - 4)^2}, \frac{-2(\theta - 2)(\beta - 1)}{\theta(\theta - 1)(2\beta\theta - 4\beta + \theta)})$ we get $\frac{\partial(\eta^{1*} - \eta^{2*})}{\partial \beta} > 0$; When $\theta > \frac{7}{2} - \frac{\sqrt{33}}{2}$ we get $\frac{\partial(\eta^{1*} - \eta^{2*})}{\partial \beta} > 0$.

Lemma 5 shows that for the government when the θ is large and the k is low, $\eta^{1*} > \eta^{2*}$, RSP_1 with greater innovation risk will receive greater support to increase its willingness to innovate; when k is high, $\eta^{1*} < \eta^{2*}$, the government would measure the efficiency of subsidies for improving social and environmental benefits, that is, choosing RSP_2 with more market influence to subsidize, as well as when the θ is large. When k is small, the difference η^{i*} decreases with k 's increase, while when k is large, the difference will increase with k increase. That is to say when encouraging the innovation willingness caused by the quality gap between R_i , the government does not ignore the cost, but fully considers the decisions made by the subsidies. For the rate of transaction cost, When RSP_i are on the platform with a higher transaction rate, if the quality difference is large but the k is small, the difference of the η^{i*} for different RSP_i will be smaller, but in this case, when the k is large, the difference of η^{i*} will become larger, and the government will become another factor to promote the differentiation of RSP_i 's income. When the quality difference is not large, higher transaction rate will result in smaller subsidy rate

differences, that is, when trading platforms tend to commercialize (higher transaction cost), η^{i*} will gradually increase market intervention.

4.3. Comparison of Market Share

Lemma 6: (1) Comparison of pricing strategies: The comparison of pricing strategies for RSP_i in different scenarios, we get $P_1^{2*} < P_1^{0*} < P_1^{1*}$ and $P_2^{1*} < P_2^{0*} < P_2^{2*}$. (2) Comparison of market share: For

the comparison of market share of RSP_1 in different scenarios: $Q_1^1 > Q_1^0 > Q_1^2$ and $\frac{\partial Q_1^1}{\partial k} < 0$, when

$$\theta > \frac{(63 + 8\sqrt{62})^{\frac{1}{3}}}{6} - \frac{1}{6(63 + 8\sqrt{62})^{\frac{1}{3}}} + \frac{3}{2} \text{ we get } \frac{\partial Q_1^1}{\partial \theta} > 0, \text{ when } \theta < \frac{(63 + 8\sqrt{62})^{\frac{1}{3}}}{6} - \frac{1}{6(63 + 8\sqrt{62})^{\frac{1}{3}}} + \frac{3}{2} \text{ and}$$

$$\frac{-5\theta + 12}{2\theta^3 - 10\theta^2 + 8\theta} < k < \frac{-\theta^4 + 2\theta^3 - 14\theta^2 + \sqrt{-15\theta^8 + 284\theta^7 - 1968\theta^6 + 6632\theta^5 - 11260\theta^4 + 8256\theta^3 + 128\theta^2 - 3072\theta + 1024} + 48\theta - 32}{2\theta^2(\theta^4 - 10\theta^3 + 33\theta^2 - 40\theta + 16)} \text{ we get}$$

$$\frac{\partial Q_1^1}{\partial \theta} > 0, \text{ so when } \frac{\theta^4 - 2\theta^3 + 14\theta^2 - 48\theta + 32 + \sqrt{-15\theta^8 + 284\theta^7 - 1968\theta^6 + 6632\theta^5 - 11260\theta^4 + 8256\theta^3 + 128\theta^2 - 3072\theta + 1024}}{2\theta^2(\theta^4 - 10\theta^3 + 33\theta^2 - 40\theta + 16)} < k < \frac{-2(\theta - 2)(\beta - 1)}{\theta(\theta - 1)(2\beta\theta - 4\beta + \theta)}.$$

But when $\frac{-\theta^4 + 2\theta^3 - 14\theta^2 + \sqrt{-15\theta^8 + 284\theta^7 - 1968\theta^6 + 6632\theta^5 - 11260\theta^4 + 8256\theta^3 + 128\theta^2 - 3072\theta + 1024} + 48\theta - 32}{2\theta^2(\theta^4 - 10\theta^3 + 33\theta^2 - 40\theta + 16)} < k < \frac{\theta^4 - 2\theta^3 + 14\theta^2 - 48\theta + 32 + \sqrt{-15\theta^8 + 284\theta^7 - 1968\theta^6 + 6632\theta^5 - 11260\theta^4 + 8256\theta^3 + 128\theta^2 - 3072\theta + 1024}}{2\theta^2(\theta^4 - 10\theta^3 + 33\theta^2 - 40\theta + 16)}$

We get $\frac{\partial Q_1^1}{\partial \theta} < 0$, $\frac{\partial Q_1^2}{\partial k} > 0$, $\frac{\partial Q_1^2}{\partial \theta} < 0$; For the comparison of market share of RSP_2 in different scenarios:

When $\frac{-5\theta + 12}{2\theta^3 - 10\theta^2 + 8\theta} < k < \frac{-\theta^2 + 14\theta - 24}{\theta(\theta - 1)(\theta - 4)^2}$ we get $Q_2^1 < Q_2^0 < Q_2^2$, when $\frac{-\theta^2 + 14\theta - 24}{\theta(\theta - 1)(\theta - 4)^2} < k < \frac{-2(\theta - 2)(\beta - 1)}{\theta(\theta - 1)(2\beta\theta - 4\beta + \theta)}$ we get

$Q_2^0 < Q_2^1 < Q_2^2$, and $\frac{\partial Q_2^1}{\partial k} > 0$, $\frac{\partial Q_2^2}{\partial k} < 0$, $\frac{\partial Q_2^2}{\partial \theta} > 0$. (3) Comparison of platform trading volume: The sum of

market share of RSP_1 and RSP_2 can be regarded as the total trading volume of the platform, when

$\theta < \frac{7}{4} - \frac{\sqrt{17}}{4}$ and $\frac{-4\theta^2 + 18\theta - 24}{4\theta^4 - 31\theta^3 + 75\theta^2 - 64\theta + 16} < k < \frac{-2(\theta - 2)(\beta - 1)}{\theta(\theta - 1)(2\beta\theta - 4\beta + \theta)}$ we get $Q_1^0 + Q_2^0 < Q_1^1 + Q_2^1 < Q_1^2 + Q_2^2$, but when

$\frac{-4\theta + 10}{\theta^2 - 5\theta + 4} < k < \frac{-4\theta^2 + 18\theta - 24}{4\theta^4 - 31\theta^3 + 75\theta^2 - 64\theta + 16}$ we get $Q_1^0 + Q_2^0 < Q_1^2 + Q_2^2 < Q_1^1 + Q_2^1$, $Q_1^0 + Q_2^0 < Q_1^1 + Q_2^1 < Q_1^2 + Q_2^2$. And

in scenes 2 and 3, we can get $\frac{\partial(Q_1^1 + Q_2^1)}{\partial k} < 0$, $\frac{\partial(Q_1^1 + Q_2^1)}{\partial \theta} > 0$. (4) Comparison of sensitivity: ① Comparing

the sensitivity of resource pricing to k and θ in two innovation scenarios, we can get $\frac{\partial P_1^{1*}}{\partial k} < \frac{\partial P_2^{1*}}{\partial k}$,

$\frac{\partial P_1^{2*}}{\partial k} < \frac{\partial P_2^{2*}}{\partial k}$, $\frac{\partial P_1^{2*}}{\partial k} > \frac{\partial P_2^{2*}}{\partial k}$, $\frac{\partial P_1^{1*}}{\partial \theta} > \frac{\partial P_2^{1*}}{\partial \theta}$, $\frac{\partial P_1^{2*}}{\partial \theta} > \frac{\partial P_2^{2*}}{\partial \theta}$. ② Comparing the sensitivity of RSP_i 's market share to k and

θ in two innovation scenarios, we can get $\frac{\partial Q_1^1}{\partial k} < \frac{\partial Q_2^1}{\partial k}$, $\frac{\partial Q_1^{1*}}{\partial k} < \frac{\partial Q_2^{1*}}{\partial k}$, $\frac{\partial Q_1^2}{\partial k} > \frac{\partial Q_2^2}{\partial k}$, $\frac{\partial Q_1^{2*}}{\partial k} < \frac{\partial Q_2^{2*}}{\partial k}$, $\frac{\partial Q_1^1}{\partial \theta} < \frac{\partial Q_2^1}{\partial \theta}$.

Lemma 6 shows that when RSP_i carrying out green innovation, they always tend to higher pricing to recover the cost of investment, while the other RSP_i will take lower pricing to seek greater market share. At the same time, the sensitivity of the price strategy k is always lower than that of the other scenario, and in the same scenario, the RSP_i 's price sensitivity k is always lower than that of the other the other RSP_i 's, which indicates that the first one to invest would reduce the impact of low-carbon innovation cost on resource pricing.

When RSP_1 carrying out low-carbon innovation, its pricing sensitivity θ is higher than that of its counterpart, while RSP_2 also having the same effect on low-carbon innovation cost when conducting low-carbon innovation, but its sensitivity is less than RSP_1 when the low-carbon innovation cost is high. These show that when RSP_1 considering low-carbon innovation, it is more likely to carry out low-carbon innovation activities based on quality differences, and its pricing will be more sensitive with θ ; when RSP_1 carries out green innovation, the larger k will make RSP_1 pay attention to the determination of RSP_2 and pay more attention to the impact of θ on the reasonable market price.

For the market share of RSP_1 , the high pricing in the low-carbon innovation scene can also make it occupy the largest market share in three scenarios, while its market share in the basic model is slightly smaller and is the smallest in scene 3. But when its market share is the largest, if the k increases, its market share will gradually decline due to the increase in pricing; When the quality difference between R_i is large enough and the k is in a suitable area. The market share of RSP_1 will decrease with the narrowing of quality difference. In other cases, with the narrowing of quality difference, the market share of RSP_1 after low-carbon innovation tends to increase. When the low-carbon innovation cost is low, the market share of RSP_2 is the largest in scene 3, and it is the smallest under the basic model. When the k is high, the market share under the basic model is greater than that of scene 2. As above analysis of RSP_1 , the changing trend of Q_2^i on k is only related to whether to conduct low-carbon innovation. However, the relationship between Q_1^2 and Q_2^2 with respect to θ is opposite. When RSP_2 carrying out low-carbon innovation, its original quality advantage is greater, and its market share advantage is smaller after low-carbon innovation.

For the total trading volume of the platform, when the quality difference is large and low-carbon innovation cost is small, the total market volume in scene 3 is the largest, followed by that of scene 2, and the smallest is in the basic model; when the quality difference is large but low-carbon innovation cost innovation is high, the total market volume of scene 2 is the largest, and the basic model is still the smallest. When the quality difference is not large, the total market volume of scene 3 is the largest. It can be seen that although the previous conclusion tells us that $m_2^{2*} > m_1^*$ is always true, but m_2^{2*} will be relatively smaller when the quality difference is large and the low-carbon innovation cost is high. At the same time, due to a slightly higher P_2^{2*} , all these factors will make the market share of RSP_2 , which is relatively larger, become smaller than that before, and the total market will become smaller. The platform trading volume is a reduction function of k and an increasing function of θ . However, for the trading platform, the total amount of transactions is directly related to the interests of the platform. Therefore, from the perspective of platform development, it can consider raising the upper limit threshold or providing appropriate low-carbon innovation subsidies to RSP_i for expanding the trading market.

5. NUMERICAL ANALYSIS

In order to analyze the equilibrium price strategy and profit of RSP_i under different scenarios more intuitively, the government's project subsidy rate and low-carbon innovation effort coefficient, and the influence of different factors on them, this section verifies the above model and conclusion through case analysis. The common parameters in the above model are set to: $\beta = 0.01, a = 0.1$. Thus we can get the relationship between the price strategy of RSP_i and the difference of k and θ in scene 2 and 3, as figure 3-4 shows:

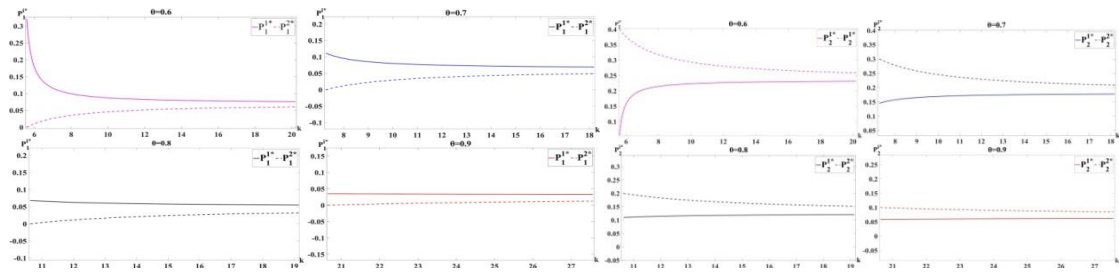


Figure 3. The price strategy of RSP_i in different scenes

It can be concluded from Figures 3-4 that the P_i^{i*} decreases with the increase of k , but tends to be stable. When RSP_i does not carry out low-carbon innovation, its pricing increases to a stable value with the increase of k . And under different quality differences, the decline/rise is not the same, it can be seen that the decline/rise trend will be significantly slowed down when the quality difference is small, that is, the market pricing of both sides will not change greatly with low carbon innovation investment. The smaller the quality difference is, the lower the price strategy for green innovation will be, and the P_i^{i*} will also be reduced. Comparing the price strategies of the two sides in different scenarios, it is found that P_x^{y*} is always less than P_x^{x*} , which also verifies the previous conclusion.

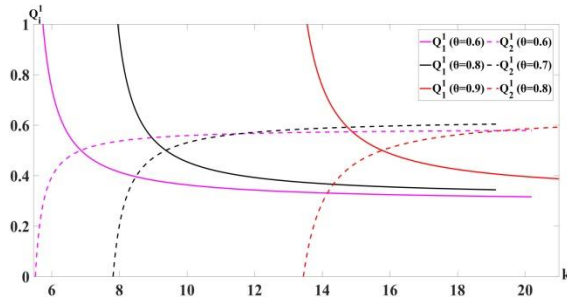


Figure 4. RSP_i 's market share in scene 2

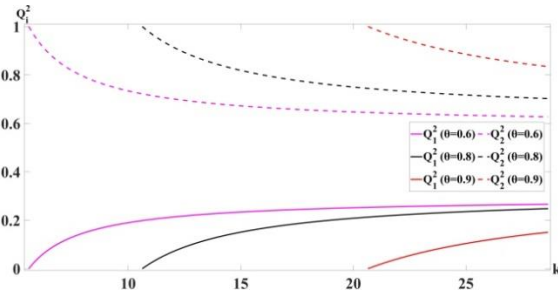


Figure 5. RSP_i 's market share in scene 3

As Figure 5-6 shows, under different θ values, k has different value range to make Q_i^i greater than zero, that is, our research is based on the competition between RSP_i , so we compare in the k value range under different scenes. It can be found that: compared with the market share gap between RSP_i under scene 2 and scene 3, the gap between the two sides under scene 2 is significantly smaller than that of scene 3, and when k is small in scene 2, the market share of RSP_1 is even higher than RSP_2 , and this critical value increases with the increase of θ between R_i . When one party carries out green innovation, its market share will decrease with the increase of k , while the market share of the other party will increase with the increase of k , but eventually will tend to be flat. It is worth noting that the Q_1^* convergence value in Scene 2 is significantly greater than that in scene 3, while the Q_2^* convergence value when θ is small has no significant difference in the two scenes. When the θ is large, the convergence value in scene 3 is greater than that in scene 2. Therefore, no matter who carries out low-carbon innovation, it is beneficial for the expansion of its market. For RSP_1 in scene 2, its market share increases with the increase of θ , while for RSP_2 , the correlation between market share and θ depends on the size of

k . But in scene 3, Q_1^* decreases with the increase of θ , and Q_2^* is monotonically positively correlated with θ .

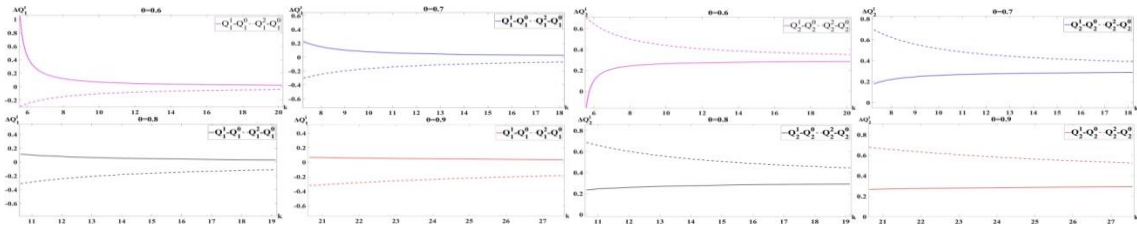


Figure 6. Comparison of RSP_i 's market share differences (scene2-1 and scene2-3)

For the market share difference of RSP_1 between scene 2\3 with scene 1, as shown in Figure 7-8, the market share of RSP_1 after green innovation is significantly higher than that of scenes 1 and 3. The market share difference between scenes 2-1, 3-1 and 2-3 decreases with the increase of k . Except for the situation that the θ is small and k is low, the market expansion of RSP_1 after low-carbon innovation is stable at a low level (the change relationship with k and θ is not obvious), but the maximum expectation of RSP_1 's market expansion is slightly higher when θ is small, but the minimum expectation is low.

For the market share of RSP_2 , except for the case that the θ is small and k is low in scene 2, as long as RSP_i carries out green innovation, its market share will increase, but the increase in market share of itself is significantly greater than that of the other party when it carries out low-carbon innovation, and the difference decreases with the increase of k , but the decrease slows down with the increase of θ . This difference also increases with the increase of θ , indicating that RSP_2 's low-carbon innovation brings more benefits to its market expansion when the θ is large.

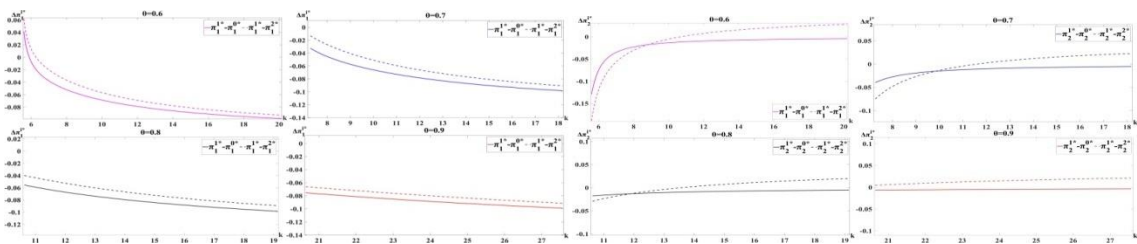


Figure 7. Comparison of RSP_i 's profit differences (scene2-1 and scene2-3)

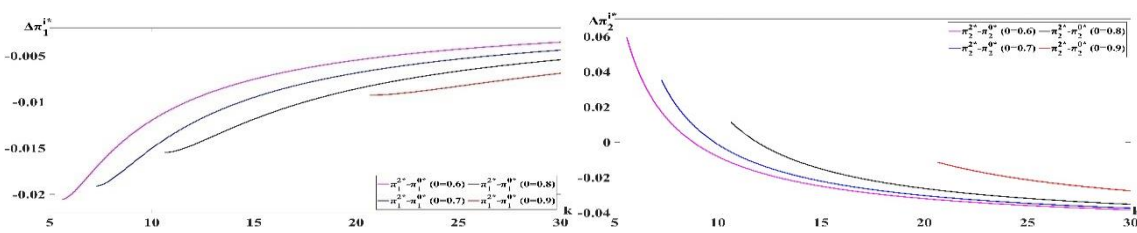


Figure 8. Comparison of RSP_i 's profit differences (scene3-1)

As shown in Figure 9-11, when RSP_1 conducting low-carbon innovation, the profit will increase relative to scenes 1 and 3 only when the θ and k are small, and the profit under scene 3 is always less than scene 1, that is, $\pi_1^{1*} < \pi_1^{2*} < \pi_1^{0*}$ is established in general and $\pi_1^{2*} < \pi_1^{0*} < \pi_1^{1*}$ is only established in special cases. The profit difference between scenes 2-1 and scenes 2-3 generally increases with the increase of k , and tends to be flat with the increase of θ . The expected value of RSP_1 profit loss (scene 2-1) increases with the increase of θ , while the profit difference of RSP_1 between scenes 3-1 decreases with the increase of k , and the profit difference is smaller with the decrease of the θ .

For the RSP_2 's profit in scene 2, compared with scene 1, its profit decreases, we can get $\pi_2^{0*} > \pi_2^{1*}$, and this difference decreases monotonically with k ; The comparison with profits under scene 3 depends on the size of θ and k , when θ is relatively small ($\theta < 0.9$), there are two boundary points 1 and 2 of k , $\pi_2^{1*} < \pi_2^{0*} < \pi_2^{2*}$ when $k < k_1$, $\pi_2^{1*} < \pi_2^{2*} < \pi_2^{0*}$ when $k_1 < k < k_2$, $\pi_2^{2*} < \pi_2^{1*} < \pi_2^{0*}$ when $k > k_2$; When θ is large ($\theta \geq 0.9$) we get $\pi_2^{2*} < \pi_2^{1*} < \pi_2^{0*}$. The profit function of RSP_2 is always a decreasing function of k , and when k is large, the yield profit of RSP_2 in scene 3 is an increasing function of θ .

Combined with the above analysis, the innovation willingness of RSP_i for low-carbon innovation (equal to the estimated profit minus the basic profit) will show great differences with the difference in the θ and the k . In most cases, the innovation willingness of RSP_1 is generally low, which is far lower than that of RSP_2 for low-carbon innovation, and decreases with the increase of θ and the k . If RSP_1 starting from the perspective of their own short-term interests, because the reduction of profits may not lead to green innovation, but the market share in scene 1 or scene 3 will be lower than that in scene 2, and the pricing of RSP_1 in scene 2 is at the maximum value in three cases. The profit and loss of short-term profits are only due to the input of early innovation costs, so in the long run, RSP_1 should take the initiative to seek opportunities for the improvement of product greenness, so that the market share it occupies will increase significantly, and there is a chance to exceed the market share of RSP_2 . For RSP_2 , the smaller the quality difference is, the lower the innovation willingness is. That is to say, when the quality difference between the two sides is large, although the market expansion degree is small at this time, the pricing is slightly higher, so the profit gain of RSP_2 is larger at this time.

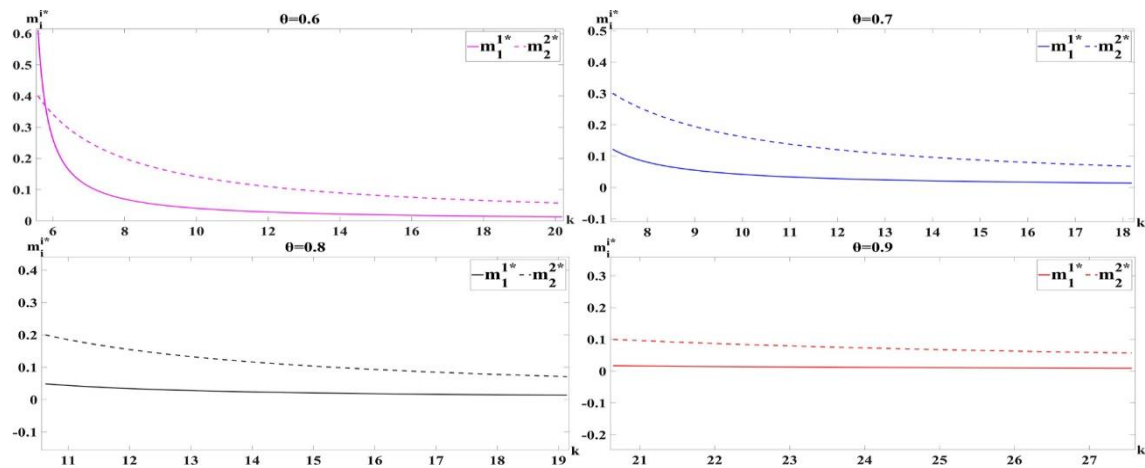


Figure 9. Comparison of RSP_i 's low-carbon innovation effort coefficient

As shown in Figure 12, only when θ and k are small, $m_1^{1*} > m_2^{2*}$, and m_i^{i*} is negatively correlated with k . When RSP_1 carrying out low-carbon innovation, the impact of k on its effort decreases with the increase of θ , and m_1^{1*} tends to zero in most cases. When RSP_2 carries out low-carbon innovation, its effort decreases with the increase of θ , and the influence of k on its effort decreases with the increase of θ . It is worth noting that when m_i^{i*} reaches the maximum value, the corresponding θ and k values are consistent with the highest innovation intention, that is, innovation intention is positively correlated with innovation effort.

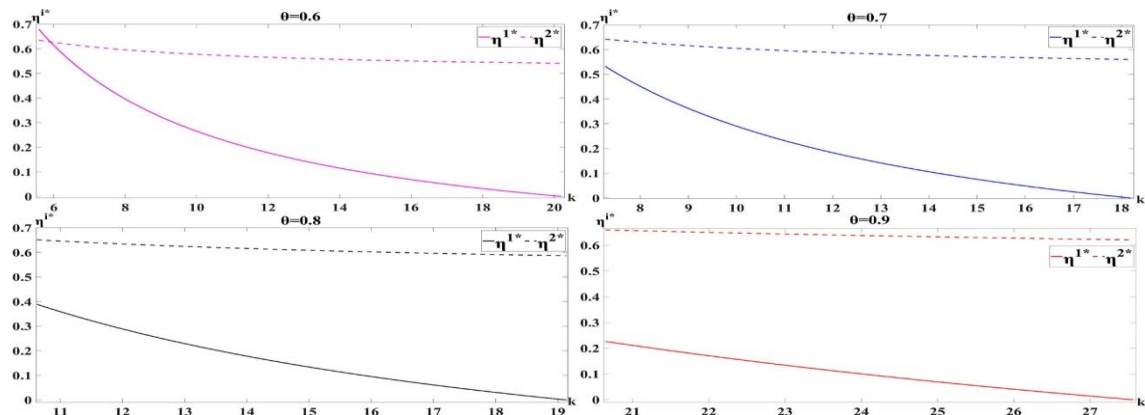


Figure 10. Comparison of government subsidy rate

As shown in Figure 13, only if the θ and k are small, $\eta^{1*} > \eta^{2*}$ can be obtained. But in other situations, we only can get $\eta^{1*} < \eta^{2*}$. And when k is larger, η^{i*} is smaller; With the increase of θ , the value range of η^{1*} will gradually become smaller, and the decrease rate of k will gradually slow down, while η^{2*} is maintained in a similar interval and slightly increased, and the influence of k on η^{2*} is gradually reduced.

6. CONCLUSION AND FORESIGHT

This paper studies the resource providers' low-carbon innovation strategy and government subsidy strategy under the background of collaborative sharing of manufacturing resources. Considering the impact of resource quality differences and low-carbon preferences of demanders on resource providers' low-carbon innovation investment, operation strategy, and government subsidy rate, under the three scenes of neither low-carbon innovation, low-quality resource providers' low-carbon innovation, and high-quality resource providers, the utility theory and Stackelberg game method are used to construct resource providers' decision model with profit maximization as the goal and the decision model with social welfare maximization as the goal. The optimal government subsidy rate, the optimal low-carbon innovation effort coefficient, and the optimal price strategy in different scenes are obtained by using the reverse induction method. Then it compares the equilibrium schemes of resource providers and government in different scenarios, and the influence of quality relative coefficient and innovation cost coefficient on the equilibrium scheme in different innovation scenes. Finally, in order to analyze the equilibrium results more intuitively, this paper makes further numerical analysis. In this process, the following research conclusions are obtained:

- (1) The degree of low-carbon innovation efforts is only related to low-carbon innovation cost and quality differences between RSP_i , and when the cost of low-carbon innovation is high, both RSP_i should invest in low cost of innovation. ①For low-carbon innovation in scene 2: When RSP_i deciding on low-carbon innovation, it should consider: **a.** When θ is small, it should avoid the impact of the high cost of low-carbon innovation and reduce investment appropriately, and instead choose lower resource pricing to compensate for market weaknesses; **b.** When θ is large, the investment in low-carbon innovation should be comprehensively considered by combining the platform transaction rate and innovation cost. In most cases, the investment should increase with the narrowing of the quality difference, but when the innovation cost and transaction rate are large, the investment in low-carbon innovation should be reduced to avoid risks with the narrowing of the quality difference; **c.** When the quality of resources between the two sides is very close, innovation input should be reduced and resource pricing should be reduced to gradually encroach on the market as quality differences narrow. ② For low-carbon innovation in scene 3: As the quality difference between the two sides shrinks, innovation input should be reduced.
- (2) ①From the perspective of resource pricing: When RSP_i conducting low-carbon innovation, they should pay more attention to θ developing higher pricing to recover the cost of investment, but at this time they should also pay more attention to low-carbon innovation costs to adopt lower pricing to gain greater market share. ②From the perspective of market share: **a.** RSP_1 should consider k when to determine the amount of investment in innovation costs, after which only when is k too large, its market share will be lower than RSP_2 , and innovation initiatives, in this case, can also significantly narrow the gap with RSP_2 's market share. And in most cases (except that there is almost no difference in quality and the innovation cost is in the middle value, the greater the quality disadvantage is, the greater the benefit of low-carbon innovation is), the larger the θ is, and the larger the market share that this measure can occupy is. **b.** RSP_2 should be based on k to decide whether to carry out low-carbon innovation. When k is low, it is bound to carry out low-carbon innovation, and the greater its original quality advantage, the smaller its market share advantage after low-carbon innovation; When k is high, it is unnecessary to carry out low-carbon innovation. ③ From the perspective of RSP_i profits: **a.** When RSP_1 conducting low-carbon innovation, it is generally necessary to bear the early profit loss to obtain market expansion and reputation improvement. **b.** When low-carbon innovation is carried out by RSP_2 , the θ and k should be considered to measure the profit and loss. When θ is relatively small ($\theta < 0.9$), there are two boundary points for k . When k is low, the profit increases after innovation and the profit loss is the largest in the opposite situation. When the innovation cost is in the middle value, the profit is lost after innovation, but the loss is larger in the opposite situation. When the innovation cost is large, the profit is lost after innovation, and the profit loss is small in the opposite situation. Deciding to conduct low-carbon innovation is always the most defective decision of RSP_2 when there is no difference in resource quality between RSP_i . And the lost profit in scene 3 increases with the decrease of θ .
- (3) For the government, the formulation of its subsidy rate is related to low-carbon innovation costs, resource quality differences, and platform transaction rates. ①Subsidy for low-carbon resource innovation on the same trading platform: When k is low and the quality of platform resources varies greatly, the government should support RSP_1 with higher risk to conduct low-carbon innovation with a higher subsidy rate; When k is high or θ is large, the government should fully measure the efficiency of subsidies for improving social and

environmental benefits, and make greater subsidies for RSP_2 with more market influence. However, the subsidy rate difference between them decreases with the increase of k . ② Low-carbon innovation subsidies for resources on platforms with different transaction rates: For the platform government with a high transaction rate, it should give a higher subsidy rate. If the quality of platform resources varies greatly and k is low, the government should set a smaller subsidy rate difference to encourage innovation, while the government should set a larger subsidy rate difference when k is high. And the government should set a smaller subsidy rate difference based on higher commission rates if the quality of platform resources varies a little.

- (4) For the total market volume of trading on the platform, it has an extremely important impact on the development of the platform: When θ and k are small, the platform should introduce corresponding policies to encourage high-quality resource providers to carry out low-carbon innovation to promote the development of the platform, and when θ is small but k is high, the platform should introduce corresponding policies to encourage low-quality resource providers to carry out low-carbon innovation to promote the development of the platform; When θ is large, the platform should introduce corresponding policies to encourage high-quality resource providers to carry out low-carbon innovation to promote the development of the platform. In these low-carbon innovation scenes, the platform should also consider raising the quality threshold of resources online or publishing appropriate preferential subsidies for low-carbon innovation in order to achieve a better market expansion effect.

This study considers the impact of manufacturing resource quality on low-carbon innovation decision-making and effect, and further considers the after-sales service attributes of manufacturing resources. Low-carbon innovation under platform incentives is also a feasible way to improve the green degree of resources under the background of resource sharing. With the increasing awareness of environmental protection among consumers/manufacturers/markets, subsidy targets have become feasible research points affecting the green degree of resources, which have a crucial impact on the development and promotion of low-carbon products. Subsequent research can be carried out in combination with the relevant attributes of resources and different mechanisms of subsidies.

ACKNOWLEDGEMENTS

This paper is supported by the National Key Research and Development Program of China (2019YFB1706101) and the Humanities and Social Sciences Research Planning Project of Chongqing Education Commission in 2022 (No.22SKGH433). The authors would also like to express appreciation to the anonymous reviewers and editors for their very helpful comments that improved the paper.

REFERENCES

- [1] Chinaenvironment.com. "United Nation Conference on Environment and Development (1992)." <http://www.chinaenvironment.com/view/ViewNews.aspx?k=1992060-3154607437> (accessed).
- [2] J.-B. Sheu and Y. J. Chen, "Impact of government financial intervention on competition among green supply chains," *International Journal of Production Economics*, vol. 138, no. 1, pp. 201-213, 2012, DOI: 10.1016/j.ijpe.2012.03.024.
- [3] S. R. Madani and M. Rasti-Barzoki, "Sustainable supply chain management with pricing, greening and governmental tariffs determining strategies: A game-theoretic approach," *Computers & Industrial Engineering*, vol. 105, pp. 287-298, 2017, doi:10.1016/j.cie.2017.01.017.
- [4] Chen Z S. A Bargaining-Coordination Mechanism for Green Supply Chain of Wood-based Panel under the Government's Incentive Policy[J]. *Chinese Journal of Management Science*, 2016, 24(02): 115-124

- [5] H. K. Chan, H. He, and W. Y. C. Wang, "Green marketing and its impact on supply chain management in industrial markets," *Industrial Marketing Management*, vol. 41, no. 4, pp. 557-562, 2012, doi: 10.1016/j.indmarman.2012.04.002.
- [6] M. Frostenson and F. Prekert, "Sustainable supply chain management when focal firms are complex: a network perspective," *Journal of Cleaner Production*, vol. 107, pp. 85-94, 2015, DOI: 10.1016/j.jclepro.2014.05.034.
- [7] D. Krass, T. Nedorezov, and A. Ovchinnikov, "Environmental Taxes and the Choice of Green Technology," *Production and Operations Management*, pp. n/a-n/a, 2013, doi: 10.1111/poms.12023.
- [8] A. C. da Silva, M. P. Méxas, and O. L. G. Quelhas, "Restrictive factors in implementation of clean technologies in red ceramic industries," *Journal of Cleaner Production*, vol. 168, pp. 441-451, 2017, DOI: 10.1016/j.jclepro.2017.09.086.
- [9] X. Zhang, T. Chen, and C. Shen, "Green investment choice in a duopoly market with quality competition," *Journal of Cleaner Production*, vol. 276, 2020, doi: 10.1016/j.jclepro.2020.124032.
- [10] Y. Yao and K. X. Zhu, "Research Note—Do Electronic Linkages Reduce the Bullwhip Effect? An Empirical Analysis of the U.S. Manufacturing Supply Chains," *Information Systems Research*, vol. 23, no. 3-part-2, pp. 1042-1055, 2012, doi: 10.1287/isre.1110.0394.
- [11] "A Model of Neutral B2B Intermediaries," *Journal of Management Information Systems*, vol. 19, no. 3, pp. 43-68, 2014, doi: 10.1080/07421222.2002.11045734.
- [12] W. Thitimajshima, V. Esichaikul, and D. Krairit, "A framework to identify factors affecting the performance of third-party B2B e-marketplaces: A seller's perspective," *Electronic Markets*, vol. 28, no. 2, pp. 129-147, 2017, doi: 10.1007/s12525-017-0256-3.
- [13] A. W. Siddiqui and S. A. Raza, "Electronic supply chains: Status & perspective," *Computers & Industrial Engineering*, vol. 88, pp. 536-556, 2015, doi: 10.1016/j.cie.2015.08.012.
- [14] O. Sohaib, M. Naderpour, W. Hussain, and L. Martinez, "Cloud computing model selection for e-commerce enterprises using a new 2-tuple fuzzy linguistic decision-making method," *Computers & Industrial Engineering*, vol. 132, pp. 47-58, 2019, doi: 10.1016/j.cie.2019.04.020.
- [15] D. Ghosh and J. Shah, "A comparative analysis of greening policies across supply chain structures," *International Journal of Production Economics*, vol. 135, no. 2, pp. 568-583, 2012, doi: 10.1016/j.ijpe.2011.05.027.
- [16] Z. Liu, T. D. Anderson, and J. M. Cruz, "Consumer environmental awareness and competition in two-stage supply chains," *European Journal of Operational Research*, vol. 218, no. 3, pp. 602-613, 2012/05/01/ 2012, doi: <https://doi.org/10.1016/j.ejor.2011.11.027>.
- [17] R. C. Savaskan, S. Bhattacharya, and L. Wassenhove, "Closed-Loop Supply Chain Models with Product Remanufacturing," *Management Science*, vol. 50, no. 2, pp. 239-252, 2004.
- [18] J. Gao, H. Han, L. Hou, and H. Wang, "Pricing and effort decisions in a closed-loop supply chain under different channel power structures," *Journal of Cleaner Production*, vol. 112, pp. 2043-2057, 2016, doi: 10.1016/j.jclepro.2015.01.066.
- [19] A. Atasu and G. C. Souza, "How Does Product Recovery Affect Quality Choice?," *Production and Operations Management*, vol. 22, no. 4, pp. 991-1010, 2013, doi: 10.1111/j.1937-5956.2011.01290.x.

IMAGE SEGMENTATION IN SHAPE SYNTHESIS, SHAPE OPTIMIZATION, AND REVERSE ENGINEERING

Milan Ćurković¹, Andrijana Ćurković²,
Damir Vučina¹ and Domagoj Samardžić¹

¹Faculty of Electrical Engineering, Mechanical Engineering and Naval
Architecture, University of Split, Croatia

²Faculty of science Split, University of Split, Croatia

ABSTRACT

Image segmentation and segmentation of geometry are one of the basic requirements for reverse engineering, shape synthesis, and shape optimization. In terms of shape optimization and shape synthesis where the original geometry should be faithfully replaced with some mathematical parametric model (NURBS, hierarchical NURBS, T-Spline, ...) segmentation of geometry may be done directly on 3D geometry and its corresponding parametric values in the 2D parametric domain. In our approach, we are focused on segmentation of 2D parametric domain as an image instead of 3D geometry. The reason for this lies in our dynamic hierarchical parametric model, which controls the results of various operators from image processing applied to the parametric domain.

KEYWORDS

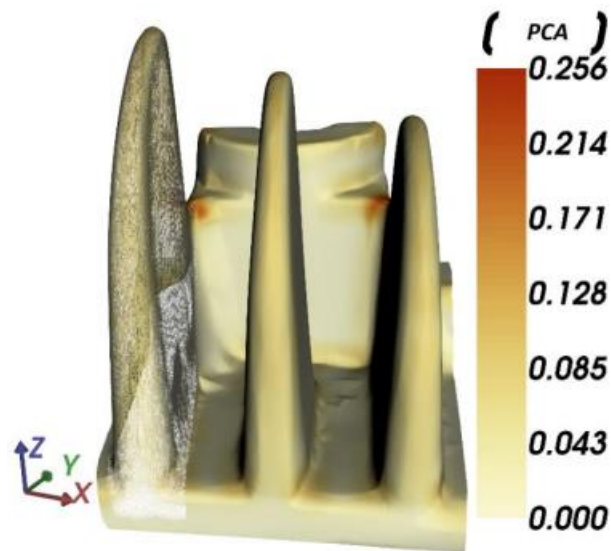
Image Segmentation, Shape Optimization, Shape Synthesis, Reverse Engineering.

1. INTRODUCTION

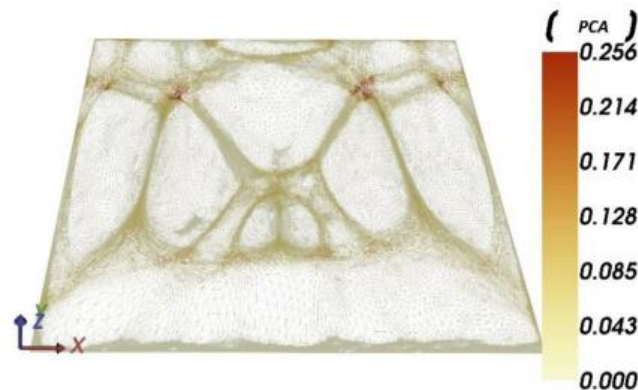
Segmentation of geometry is one of the basic requirements for reverse engineering, shape optimization, and shape synthesis. There are successful algorithms for segmenting 3D geometry (3D point cloud / triangulated surface) [1,2,3], and successful algorithms for segmenting 2D images [4,5,6,7,8]. Algorithms for both dimensions, 2D and 3D, are based on the same ideas/approaches. The key difference is in simply topology in the case of 2D image (matrix representation). Those algorithms use information about geometric features: edges, peaks, gaps, and for that use PCA (principal component analysis), and Gaussian maps, etc. To determine the boundaries between regions, mainly the principle of 3D water shadow and its variations are used. In terms of shape optimization and shape synthesis, segmentation of geometry may be done directly on 3D geometry and on its corresponding parametric values in the projected 2D parametric domain. In our approach, we are more focused on the 2D parametric domain instead on 3D geometry. The reason for that lies in our dynamical hierarchical parametric model where new regions may appear and old ones may disappear. Moreover, some parts of the projected geometry cannot be assigned to any regions, and in this case, we use more layers of subparametrizations. The parametric model successfully deals with both situations: connected regions and more layers of sub-parametrizations, but due to faster convergence in optimizations of engineering samples and more simply CAD reversing, the clear boundaries between nature regions are more preferred.

2. PARAMETRIC DOMAIN

In the processes of shape optimization and shape synthesis we have changing geometry, changing topology, changing partitions, and the way of connection between partitions. As we mentioned, the partition creation may be done by segmentation of geometry on 3D triangulation shape. In the beginning, before the initial solution, the segmentation may be applied to the initial 3D geometry with the original triangulated mesh mainly obtained from a 3D scanner (Figure 1 a). After the initial solution in the optimization process, the segmentation can also be applied to the 3D geometry of the parametric model which is smooth geometry whose sharpness depends on how well the parametric model fits the referent geometry (initial geometry, geometric primitives, some mathematical functional, ...). Regardless of that possibility, we do segmentation in the 2D parametric model (Figure 1 b). The benefits of this choice are the usage of fast algorithms from the area of image processing, the matrix topology of parametric values, the easy possibility to make connections between regions (parametric patches) and making more levels in the hierarchical parametric model. Of course, there are some drawbacks. If we do segmentation on the image obtained from parametric projection (Figure 2), there is the possibility that more triangles (all three vertices) have the same pixel positions (discrete x and y coordinates). This drawback is visible as wide white regions between partitions in figure (Figure 4). This problem can be solved by subpixel approaches which are part of future work.



a)



b)

Figure 1. The reference point in the optimization process. a) Initial geometry with included geometric features; b) projection of geometry of a) into the 2D parametric domain. From our point of view, classic segmentation algorithms applied to 3D geometry give more stable and enough sense results. But, in our case where we have the parametric model as a supervisor in making decisions, segmentation on 2D images gives solutions that are completely incorporated into our parametric model. The figure below (Figure. 2) presents the surface of the parametric projection domain (Figure 1 b). In this image, we apply the water shadow algorithm [8] after the iterative distance algorithm.

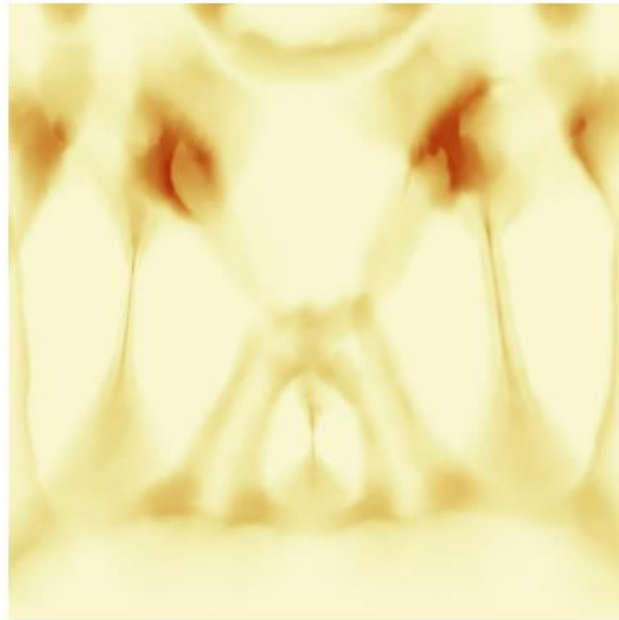


Figure 2. The surface of the parametric projection domain (Figure 1 b).

Despite the subpixel problem, we decided to show the capability of our approach on, in our opinion, not simply geometry. The key part of the algorithm is the usage of the iterative distance function as preparation for the water shadow algorithm. For simplicity, we have used a simple version of the distance function from Algorithm 1. The full versions can be found in [9,10,11]. As can be seen in Figure 3, we applied Algorithm 1 to the image as a projection of the geometry in the parametric 2D domain with two different values of the input parameter ϵ . As a result, in the first row in Figure 3, we get too many components that is the solution we want to avoid. The main goal is to get fewer clear (more convex look) components (second row in Figure 1). Of course, there is a possibility to overdo it with a small value of the ϵ parameter and in that case, we get too few components. The part of the future work is to find how to stop decreasing the parameter ϵ using the derivations of the distance algorithm. We concluded that under the control of the parametric model, i.e., control over what the components should look like and what conditions must be met, it is worth performing segmentation on 2D parameter values.

Algorithm 1. A proposed variation of the distance algorithm

Input:

- Image I .
- Real eps.

Output: result image R .

Begin

Real eps2 = eps / $\sqrt{2}$

for y = 0 to height

for x = width - 1 to 0

 Real SW = eps + $I(x - 1, y - 1)$

 Real SE = eps + $I(x + 1, y - 1)$

 Real S = eps2 + $I(x, y - 1)$

 Real E = eps2 + $I(x + 1, y)$

$R(x, y) = \min \{I(x, y), SW, SE, S, E\}$

end

end

for y = height to 0

for x = 0 to width

 Real NW = eps + $I(x - 1, y + 1)$

 Real NE = eps + $I(x + 1, y + 1)$

 Real N = eps2 + $I(x, y + 1)$

 Real W = eps2 + $I(x - 1, y)$

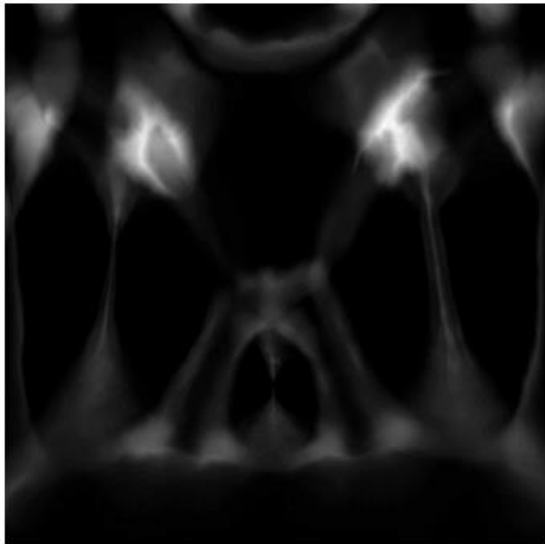
$R(x, y) = \min \{I(x, y), NW, NE, N, W\}$

end

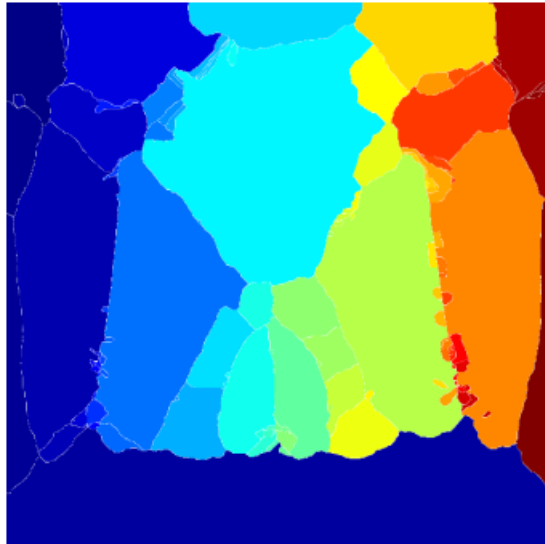
end

return R .

End



a)



b)

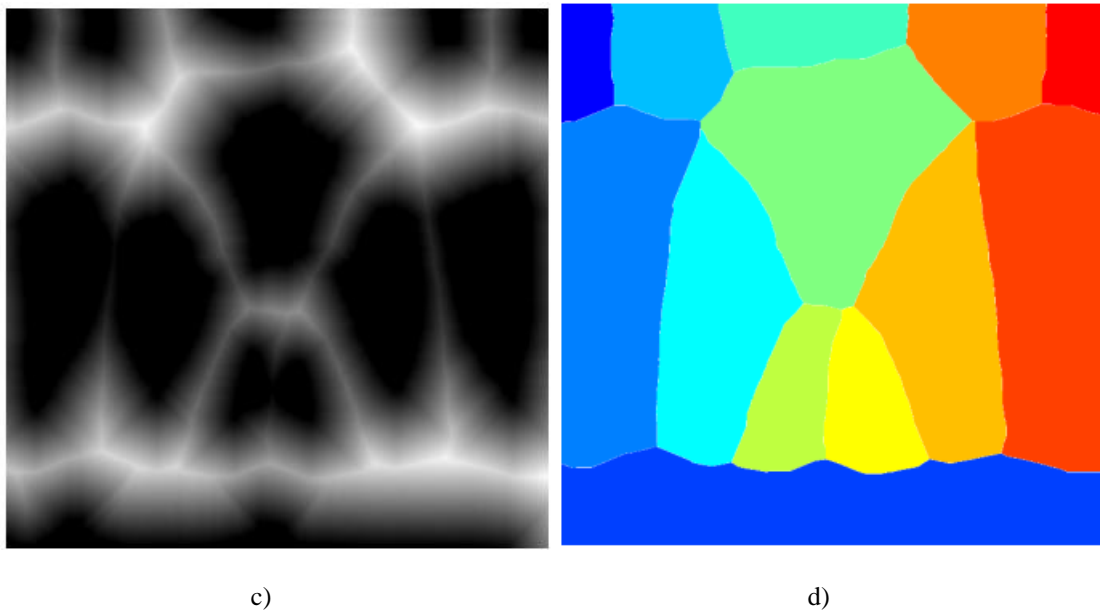


Figure 3. Segmentation of parametric domain of the parametric model. a) the results of the distance algorithm with parameter $\text{eps} = 1e-2$, b) the result of the water shadow algorithm applied on (a); c) the results of the distance algorithm with parameter $\text{eps} = 1e-5$, and (b); the result of water shadow algorithm applied on (c).

The figure above shows the final result of segmentation on the original 3D surface.

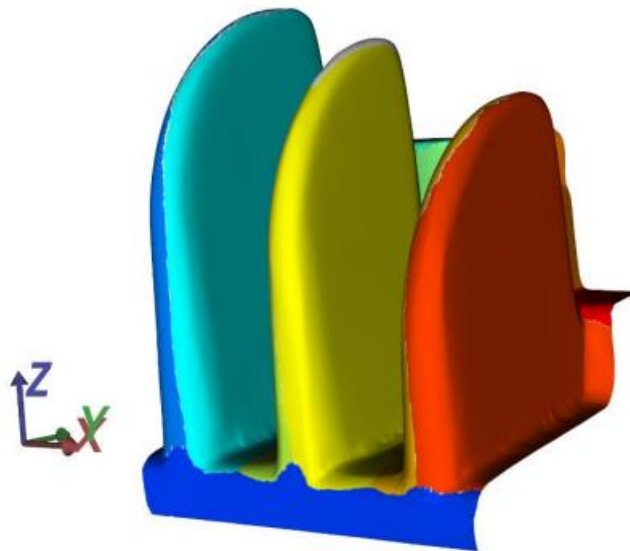


Figure 4. The segmentation result is shown on the initial 3D geometry.

3. CONCLUSIONS

Under parametric model supervision, it's worth performing the surface segmentation indirectly on its parametric domain as a 2D image. Without the control of the parametric model, i.e. the control of how the components should look like and what conditions must be satisfied (convexity), the segmentation presents many more challenges. Otherwise, the "referent" segmentation should be

done directly on the 3D surface. If the segmentation result satisfies the parametric model, the shape optimization and synthesis processes can continue. Otherwise, a local extreme is created, and without the concept of the genetic algorithm, there is no point in continuing with the above processes.

ACKNOWLEDGEMENTS

This work was supported by the Croatian Science Foundation [grant number IP-2018-01-6774]

REFERENCES

- [1] Daniel Mejia, Oscar Ruiz-Salguero, Jairo R. Sánchez, Jorge Posada, Aitor Moreno, Carlos A. Cadavid, Hybrid geometry / topology based mesh segmentation for reverse engineering, *Computers & Graphics*, Volume 73, 2018, pp. 47-58.
- [2] Zhenyu Shu, Sipeng Yang, Haoyu Wu, Shiqing Xin, Chaoyi Pang, Ladislav Kavan and Ligang Liu, 3D Shape Segmentation Using Soft Density Peak Clustering and Semi-Supervised Learning, *Computer-Aided Design*, Vol. 145, 2022
- [3] David George, Xianghua Xie, Yukun Lai and Gary K.L. Tam, A Deep Learning Driven Active Framework for Segmentation of Large 3D Shape Collections, *Computer-Aided Design*, Vol. 144, 2022
- [4] Shuai Luo, Yujie Li, Pengxiang Gao, Yichuan Wang and Seiichi Serikawa, Meta-seg: A survey of meta-learning for image segmentation, *Pattern Recognition*, Vol. 126, 2022
- [5] Dong Wang and Xiao-Ping Wang, The iterative convolution–thresholding method (ICTM) for image segmentation, *Pattern Recognition*, Vol. 130, 2022
- [6] Jordão Bragantini, Alexandre X. Falcão and Laurent Najman, Rethinking interactive image segmentation: Feature space annotation, *Pattern Recognition*, Vol. 131, 2022
- [7] Abdulateef, Salwa Khalid and Mohanad Dawood Salman. A Comprehensive Review of Image Segmentation Techniques. *Iraqi Journal for Electrical and Electronic Engineering*, 2021
- [8] Wang, Bingshu & Chen, C., Local Water-Filling Algorithm for Shadow Detection and Removal of Document Images, 2020. *Sensors*. 20. 6929. 10.3390/s20236929.
- [9] Maurer, Calvin, Rensheng Qi, and Vijay Raghavan, A Linear Time Algorithm for Computing Exact Euclidean Distance Transforms of Binary Images in Arbitrary Dimensions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 2, February 2003, pp. 265-270.
- [10] T. Schouten, E. van den Broek, Fast exact Euclidean distance (FEED) transformation, *Proceedings of the 17th International Conference on Pattern Recognition*, 2004., ICPR 2004, vol.3, IEEE pp. 594-597.
- [11] Lakshitha Dantanarayana, Gamini Dissanayake, Ravindra Ranasinge, C-LOG: A Chamfer distance based algorithm for localisation in occupancy grid-maps, *CAAI Transactions on Intelligence Technology*, 2016., Volume 1, Issue 3

THE PROBLEM SOLVER: A MOBILE PLATFORM TO MEDIATE TEENAGER FAMILY RELATIONSHIP USING DART AND MACHINE LEARNING

Ziheng Guan¹ and Ang Li²

¹Arcadia High School, 180 Campus Dr, Arcadia, CA 91006

²California State University, Long Beach,
1250 Bellflower Blvd, Long Beach, CA 90840

ABSTRACT

Family conflicts between parents and their children are nothing new and are something experienced by many in such situations [1]. These conflicts can even be exacerbated by cultural differences that exist between the two parties, especially in cases where the parents and child were raised in different countries, cultures and/or generations [2]. This description illustrates my personal experiences of conflict with my parents, which is what inspired me to create this app: The Problem Solver app. The app differs from other methods that could be applied to resolve these conflicts in that it facilitates more direct communication between the two conflicting parties, which would hopefully result in a more rapid and successful conflict resolution [3]. Naturally, there were challenges I faced in the making of the app, but I was eventually able to work through these and build a working product. I will also explore some related works and research into this topic that were helpful in supporting the idea that cultural differences between differently raised generations can have an impact on familial relations [4]. Then, I give a general overview of the system of the app and finally delve into possible limitations of the app and further steps I could take in the development of the app.

KEYWORDS

Machine Learning, Communication, Cultural Differences, Flutter.

1. INTRODUCTION

The background of this topic stems from conflicts that I have experienced at home with my parents. More specifically, my parents and I were raised in very different cultures during our respective childhoods, with my parents having been raised in a traditionally “Eastern” society (China), while I have been raised in the United States (US) since middle school [5]. Thus, I have attended school in the US for much of my life and have been raised in a traditionally “Western” society. Our conflicts originated from occurrences such as disagreements over what amount of studying is an acceptable level. For example, my parents were brought up and taught that studying and working hard is the most important thing one can do, to the point that one should be studying almost anytime one has free time. However, I do not want to study all of the time, as I would prefer to also do other things (playing video games, sports or listening to music) to relax and unwind. From a Western society perspective, studying all the time and never taking a break would be frowned upon or even be seen as concerning [6]. However, in an Eastern society, that would be seen as the “standard” for what one should be doing. As a result, my parents and I have

frequently come into conflict over ideas such as these; this inspired me to seek a solution so my parents and I can work out our issues and communicate better. This idea also extended into relationships outside of just parental conflict, as I realized that the app could also be applied to relationships with friends, siblings and other people in my life. Thus, this is how I envision this application being more generalizable and usable for people in different contexts.

This idea of communicating our problems between the people we are close with is important, as we need to be sure our relationships are good and healthy. More specifically, with a lack of communication, problems and issues in a relationship can build up and build up to the point where they could boil over into an even larger conflict - one that could've been avoided or resolved sooner with better communication [7]. However, even if one knows what they want to communicate regarding an issue or what they would like to say to the conflicting party, it is not always the easiest to directly speak to the conflicting party. That is, sometimes it is hard to make oneself confident enough to bring up an issue, especially when there is fear of punishment or not being listened to. This is where the app provides another unique perspective, being that it allows the users of the app to take time to think about and carefully craft what they are actually trying to communicate instead of essentially having to do it "on the spot." Also, this could potentially allow users to bring up issues that they may have difficulty discussing with someone face to face.

There are not many tools/existing methods that exist to solve issues that present themselves in close parental relationships such as this, but one possible way that a parent could solve this issue (at least through personal experience) is by taking their child to see a specialist like a psychologist. The parents may do this due to their child exhibiting behavior the parents would deem as "unnatural." For example, my parents and I have argued over the amount of time that I need to be using electronics such as iPads, a Chromebook, etc. Using electronics such as these has become increasingly necessary within the past few years, as many classes and assessments have moved online and much of the homework for classes must be either completed and/or submitted online. However, my parents tend to view being on electronics as me being distracted or unproductive, as they don't fully understand how truly everything is done on the computer these days. My parents see me using these devices frequently and seem to assume that there is something mentally wrong with me that is causing me to become very easily distracted or that is causing me to need to be on an electronic device all of the time. From my perspective, clearly, there is nothing mentally wrong, and there is no issue with being on devices so frequently, as that is just the way that the world works today, especially in a school setting.

In such a setting, the parents would likely expect the psychologist to offer a "cure" or an explanation as to why their child is behaving in such a way, believing that the psychologist could solve the issue or prescribe a medicine that would. There is a glaring problem with this "method," however, being that this method would be completely unhelpful and ineffective. That is, since there is no real psychological issue, and since all I am trying to do is to use my electronic devices to complete schoolwork, there is nothing that the psychologist would be able to do to help, as there is no issue to be solved. Perhaps the psychologist would be able to better offer insight into why these familial disagreements are occurring in the first place and offer counseling in that area, but we assume that the parents who are reaching out to the psychologist are convinced that they are firmly in the right and that they do not need any counseling themselves. So, as a result, this method of resolving the issue would be wholly ineffective and perhaps may raise even more tensions between the parents and the child.

There is a somewhat similar app that exists (called the Conflict Coach app), but this app does not offer the same set of services as my app. The conflict coach offers more along the lines of suggestions for defusing tense situations with talking points or suggesting gentle ways to ease into conversations on difficult topics. Thus, there appears to be less direct communication

between the two parties that are experiencing conflict (at least less direct communication through the app itself). While this may be useful for older people or for work conflicts, the Problem Solver app that I built offers more direct communication between the conflicting parties, which may lead to a more efficient problem resolution. Another drawback of the Conflict Coach app is that it is not free, it costs \$28, whereas my app is free.

My tool is called the Problem Solver app, and it is an android-based application written using Flutter that can be used to solve conflicts that arise in familial and other close relationships [8]. The app is purposed as a mobile app for android personal cell phones (hence its availability on the Google Play Store). When the app first launches, the user is greeted with three introductory screens that describe the purpose of the app; that it is for solving problems between parents and their children and can help the user set goals to make that happen. The user can skip these screens or simply scroll through them. After this, the user is prompted to login or to create an account if they don't already have one. After logging in, the user can view their profile to edit account information or they can jump right in to solving their conflicts. To start, the user clicks on the "Create a Conflict" button, which brings up the corresponding page to create a conflict. On this page, the user can give a name to the conflict, describe the conflict and who it involves, set an urgency level from a dropdown list, select a type of conflict from a dropdown list (study, computer time, etc) and select a start date and time for the issue. The Active Conflict Page holds a list of all current active conflicts (those that have not yet been resolved). On this page, one can click on the conflicts in order to edit their contents (name, who it involves, start date, urgency, etc). One can also click on the plus button in order to add a new conflict. The final page is the Old Conflicts page, where the user can view conflicts that have been resolved. Once again, the user can click on the conflicts in order to view their details. Additionally, the user can click on these conflicts to view them again and can also make the conflicts "active" again, which would put that conflict back to the Active Conflicts page.

One of the clear differences between the app and the existing methods discussed previously (e.g psychological counseling) is that this app involves the two conflicting parties communicating with one another directly instead of one of the parties (such as the parents) not really attempting to listen to their child as much and just assuming that their way is the best. As such, there is no "third party" involved in the resolution of the issue. I feel as though this would result in a more direct line of communication between the two conflicting people, and would result in more conflict resolution than if other methods had been used. In my mind, one of the greatest weaknesses of other methods (specifically therapy) is that it would not get to the root cause of the issue, that of the cultural differences that exist across generations. This is where the app becomes very useful, as it allows a safer space for everyone to communicate their issues and how they are feeling about the conflict, which in turn could be very productive towards solving the root cause of some of these issues.

The remainder of the paper will proceed as such: Section 2 details some of the challenges and difficulties I faced in the process of developing the app. Then, Section 3 will describe an overview of the components (each different page) of the app, what they look like (through screenshots), the code that makes up these segments and finally a general overview of how the app handles user data. Finally, Section 4 will explore some related works and studies on this same topic and compare these works and studies to my Problem Solver app, while Section 5 will conclude the essay and wrap up the ideas covered through this writing.

2. CHALLENGES

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

2.1. Finding Resources

One of the most frequent challenges I encountered in the undertaking of this project was that, when I would come across an issue or get stuck on something, I had a difficult time finding resources that could answer my questions. Naturally, I utilized standard resources such as Google search and the Flutter documentation in order to attempt to answer my questions, but those were only useful in a more general sense and did not exactly apply directly to what I was having difficulties with [9]. Specifically, the documentation was somewhat difficult to read, as it is all quite technical and more describes the functionalities of the language and offers few examples. As a result, there were times that I felt that I had hit a dead end on the project. I was able to resolve this issue by making use of other resources that were available to me, such as contacting teachers for help or even asking my friends if they could check my work in case their new perspective was able to offer helpful insight.

2.2. Learning new knowledge

Another challenge that I faced in the building of this app relates to my previous education in computer science. Specifically, I have taken computer science classes during my time in high school, so I have experience with computer science concepts, coding, etc. However, these classes that I have taken are more theoretical in their scope rather than specific to building things like apps, games or other programs. Specifically, computer science classes tend to focus on topics such as algorithms, simple computer communications or just a basic understanding of how a computer works. As a result, I had to teach myself quite a bit about using Flutter, publishing an app on the Google Play Store and Firebase (cloud based user data handling, also Google based). I was able to work my way through this issue by making use of as many online resources as I could, such as YouTube tutorials and examples of how Flutter is used. In cases where these resources were not enough, I took a similar approach as before and sought help through one-on-one time with my teachers as well as consulting with my friends in order to get their input.

2.3. Becoming Distracted

A third challenge that I faced in the development of this app relates somewhat to the issues that I encountered when I felt as though I had hit dead ends on the project. Specifically, I felt it easy to become distracted and become engaged in other things such as playing video games or watching videos on YouTube, especially at times where I felt frustrated with working on the app or felt that I was not making much progress [15]. This would also occasionally extend into my schoolwork as well, which resulted in me occasionally falling behind in classes, resulting in even less time for me to commit to the development of my app. Once I realized I was using my time poorly, I decided to solve this issue by creating a schedule for myself so that I was able to keep up with my schoolwork and set aside time to work on my app as well as time for relaxation.

3. SOLUTION

To take an overview of the entire system, one can see from Figure 1 below that a user would begin by downloading the app from the Google Play store on a supported device (an Android-based mobile phone). Once the user navigates through the three introductory screens (which are shown in interface screenshots below), the user is prompted to either log in or sign up if they do

not already have an account. If the user needs to sign up, they will be taken to a separate page to do so, after which their data will be stored in the Firebase cloud database with the app creator being notified of a new user signing up. Then, the user will be taken to the home page, which is also the Active Conflicts Page (this is where the user would end up if they already had an account and had just logged in). Once the user is at this home screen, they are able to take multiple actions (which Figure 1 compresses into general “User Actions” for simplicity). These actions include: creating a new conflict, viewing/editing an active conflict, viewing/editing old conflicts and viewing/editing user account information. Upon the user making any changes to any conflicts or account information, that respective information will be updated in the Firebase cloud in order to keep user data current.

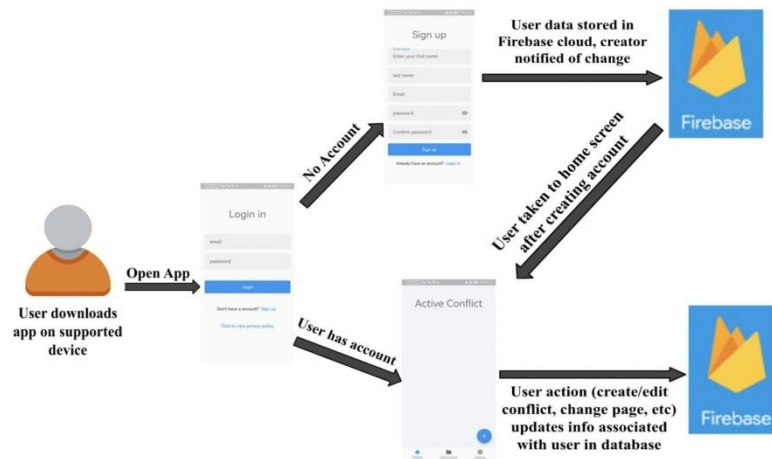


Figure 1. Overview of the solution

For the purposes of this paper, we define the “components” of the app to be each different “page” that the user has the ability to perform actions on. One can also see that, below, we have provided screenshots of each page of the app in order to further clarify what happens on each page. In Screenshots 1, 2 and 3 (SC 1-3), we can see the first three initial pages that greet the user upon the opening of the app. As is suggested by the “Skip” button and the scrolling list at the bottom, the user has the ability to either skip all of these pages or just scroll through them (SC 1-3). Upon passing these pages, the user is prompted to either log in or to create an account if they do not already have one (SC 4 & 5). After successfully logging in, the user comes upon the Home Screen/Active Conflicts screen, from which the user can navigate to other pages such as the Old Conflicts or Settings pages (SC 8 & 9). The user can also use the “Plus” button in the bottom right hand corner to create a new conflict (SC 6 -> SC 7). When there are active conflicts, they will appear on the home page (SC 6) and can be clicked on to edit their contents, bringing up a screen similar to the Create a Conflict page (SC 7) where the user can edit the contents of the conflict. In the Old Conflicts page (SC 8), the user is once again able to click on previously resolved conflicts to view their contents, and can even choose to “restart” the conflict if they feel that a similar issue has already been discussed. Finally, on the User Account page (SC 9), the user is able to see their name, associated email account and can click on buttons that can edit the user’s profile, delete the user’s account or simply log out.

Computer Science & Information Technology (CS & IT)

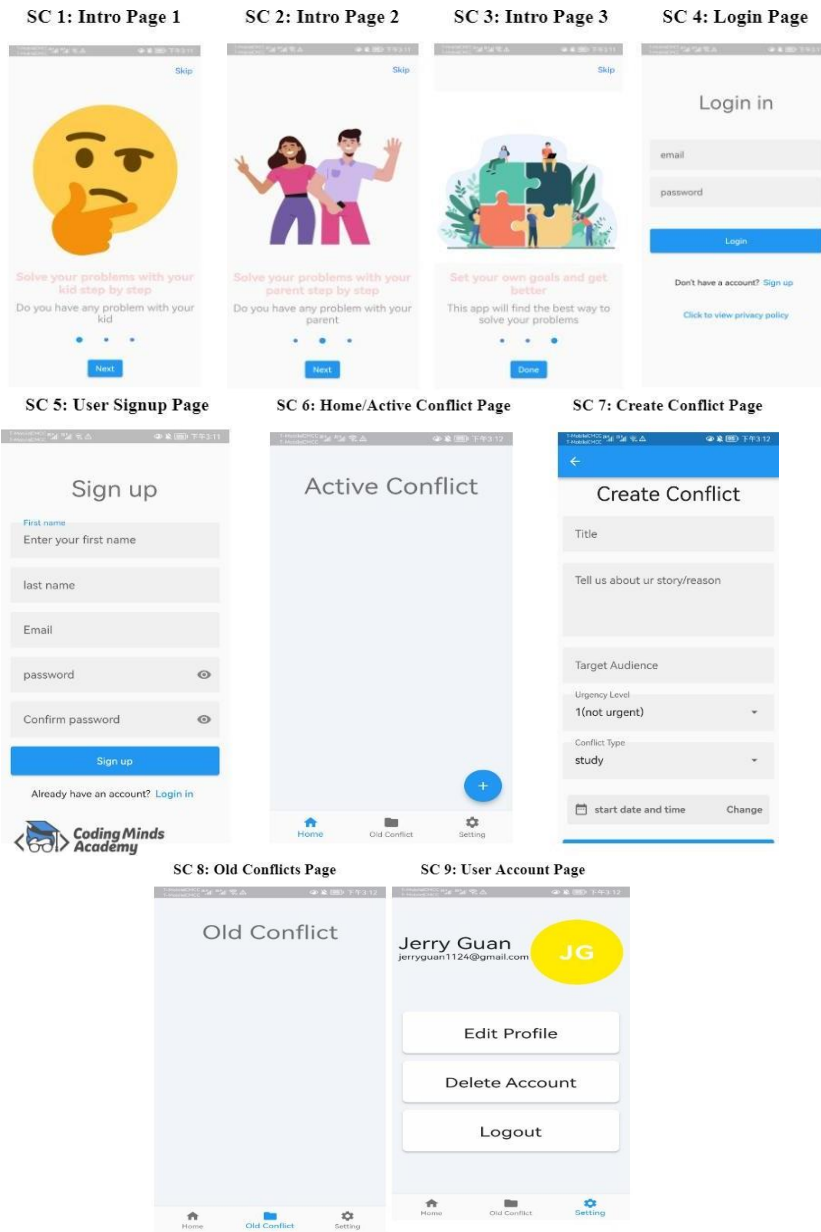


Figure 2. 3 - 4 parts of code screen

```

10 > Screen? (\ home.dart) % HomeScreenState
1 // import 'package:flutter/material.dart';
2 // import 'package:teens_parents_problem_solver/screen/aldget/items.dart';
3 import 'package:teens_parents_problem_solver/screen/pages/dashboard.dart';
4 import 'package:teens_parents_problem_solver/screen/pages/old_conflict.dart';
5 import 'package:teens_parents_problem_solver/screen/pages/settings.dart';
6 import 'package:teens_parents_problem_solver/constant/colors.dart';
7
8 class HomeScreen extends StatefulWidget {
9   HomeScreen({key? key}) : super(key: key); // Prefer declaring const constructors on @immutable classes.
10
11 @override
12 State<HomeScreen> createState() => HomeScreenState();
13 }
14
15 class HomeScreenState extends State<HomeScreen> {
16   int _selectIndex = 0;
17
18   void _onItemTapper(int index){
19     setState() {
20       _selectIndex = index;
21     };
22   }
23
24 @override
25 Widget build(BuildContext context) {
26   return Scaffold(
27     backgroundColor: screenBackgroundColor,
28     bottomNavigationBar: BottomNavigationBar(
29       onTap: _onItemTapper,
30       currentIndex: _selectIndex,
31       items: const [
32         BottomNavigationBarItem(icon: Icon(Icons.home), label: 'Home' ),
33         BottomNavigationBarItem(icon: Icon(Icons.folder), label: 'Old Conflict' ),
34         BottomNavigationBarItem(icon: Icon(Icons.settings), label: 'Setting' ),
35       ],
36     ), // BottomNavigationBar
37     body: [DashboardScreen(), OldConflictScreen(), SettingsScreen()].elementAt(_selectIndex)
38   ); // Scaffold
39 }
40

```

Figure 3. Code of the home screen

```

16 // loginScreen
17 class _loginScreenState extends State<LoginScreen> {
18   final _formKey = GlobalKey<FormState>();
19
20   final _passwordController = TextEditingController();
21   final _emailController = TextEditingController();
22   //bool _confirmPassword = true;
23
24 @override
25 Widget build(BuildContext context) {
26   return Scaffold(
27     backgroundColor: screenBackgroundColor,
28     // SafeArea()
29     body: SafeArea(
30       child: Padding(
31         padding: const EdgeInsets.all(8.0),
32         child: SingleChildScrollView(
33           child: Form(
34             key: _formKey,
35             child: Column(
36               children: [
37                 const SizedBox(
38                   height: 60,
39                 ), // SizedBox
40                 Padding(
41                   padding: const EdgeInsets.all(8.0),
42                   child: Text(
43                     "Login in",
44                     style: Theme.of(context).textTheme.headline4,
45                   ), // Text
46                 const SizedBox(
47                   height: 40,
48                 ), // SizedBox
49                 Padding(
50                   padding: const EdgeInsets.all(8.0),
51                   child: TextFormField(
52                     controller: _emailController,
53                     validator: (value) {
54                       if (value == null || value.isEmpty) {
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```

Figure 4. Code of the login page

```

7 class RegisterScreen extends StatefulWidget {
8   RegisterScreen({Key? key}) : super(key: key); // Prefer declaring const constructors on @immutable classes.
9
10  @override
11  State<RegisterScreen> createState() => _RegisterScreenState();
12 }
13
14 class _RegisterScreenState extends State<RegisterScreen> {
15   final _formKey = GlobalKey<FormState>();
16   final _confirmPasswordController = TextEditingController();
17   final _passwordController = TextEditingController();
18
19   final _firstNameController = TextEditingController();
20   final _lastNameController = TextEditingController();
21   final _emailController = TextEditingController();
22
23   bool _hidePassword = true;
24   bool _hideConfirmPass = true;
25
26   @override
27   Widget build(BuildContext context) {
28     return Scaffold(
29       backgroundColor: screenBackgroundColor,
30       body: SafeArea(
31         child: Padding(
32           padding: const EdgeInsets.all(8.0),
33           child: SingleChildScrollView(
34             child: Form(
35               key: _formKey,
36               child: Column(
37                 children: [
38                   const SizedBox(
39                     height: 40,
40                   ), // SizedBox
41                   Padding(
42                     padding: const EdgeInsets.all(8.0),
43                     child: Text(
44                       "sign up",
45                       style: Theme.of(context).textTheme.headlined,
46                     ), // Text
47                   ), // Padding
48                   const SizedBox(
49                     height: 20,
50                   ), // SizedBox
51                   Padding(
52                     padding: const EdgeInsets.all(8.0),
53                     child: TextFormField(
54                       controller: _firstNameController,
55                       validator: (value) {
56                         if (value == null || value.isEmpty) {
57                           return "first name is required";
58                         }
59                         return null;
60                       },
61                       decoration: const InputDecoration(
62                         labelText: "first name",
63                         hintText: "Enter your first name",
64                       ), // InputDecoration
65                       keyboardType: TextInputType.text,
66                     ), // TextFormField
67                   ), // Padding
68                   Padding(
69                     padding: const EdgeInsets.all(8.0),
70                     child: TextFormField(
71                       controller: _lastNameController,
72                       validator: (value) {
73                         if (value == null || value.isEmpty) {
74                           return "last name is required";
75                         }
76                         return null;
77                       },
78                       decoration: const InputDecoration(
79                         labelText: "last name",
80                         hintText: "Enter your name",
81                       ), // InputDecoration
82                       keyboardType: TextInputType.text,
83                     ), // TextFormField
84                   ), // Padding

```

Figure 5. Code of the register page

4. RELATED WORK

One related work that the co-author of this paper compared my app to is an app called the Love Nudge app, which is something that he uses with his partner in his personal life. The Love Nudge app is based on the idea of the Five Love Languages, a book by Gary Chapman that describes 5 Love Languages as the primary way in which people receive affection from their partners (Chapman, 1992) [10]. As may be apparent from the name and focus topic of the app, however, the Love Nudge app is designed primarily with couples/romantic relationships in mind, not relations between parents and children. Thus, I would say that my app is more useful in the area of general family conflicts rather than only in the limited scope of romantic relationships. The Love Nudge app does hold some similarities to my app, however, in that the Love Nudge app also supports the ability to create conflicts/issues to be resolved, which can be seen by the other user in the app rather than necessitating that the issue be talked about verbally first.

A second related paper I explored was an essay by Preevo and Tamis-LeMonda (2017) titled "Parenting and globalization in western countries: explaining differences in parent-child interactions" [11]. This paper essentially discusses how parents who are the ethnic minority

“differ from majority parents in parenting values, child rearing goals and resources—differences that affect parenting practices and children’s development” (Preevo & Tamis-LeMonda, 2017). In the case of the paper, “ethnic minority parents” were defined to be parents who were of a minority group relative to the majority of other parents raising children in a similar socio-economic background and geographical location. That is, a “minority parent” could be a pair of Asian parents coming from an Asian background raising their child in a community that is primarily made up of white families and/or families coming from a non-Asian background. The most applicable portion of this paper describes traditional values in Asian families stemming back to religious foundations in the values of Confucianism, these being Qian, Chi, Yue, He and Xiao, the last of which translates to “filial piety,” meaning respect for one's parents and other elders. As the authors describe, “One way that filial piety is expressed is through educational success. It is thus unsurprising that pursuing knowledge to improve educational success is an important childrearing goal in Chinese communities” (Preevo & Tamis-LeMonda, 2017). This would tend to support the idea that my parent’s upbringing in a traditionally Chinese society influences their thoughts that I should be studying all of the time and devoting much of my time to schoolwork. Additionally, this article also suggests that cultural differences can cause clashes between these generations, further supporting the idea that no one party is really right or wrong, it can all be a matter of perspective.

A third related work that I explored had to do with conflict resolution in families with adolescents [12]. In the article, Smetana et. al describes how conflict resolution in families with adolescents is a commonly necessary task, and that the ways in which conflicts were resolved and the topics of those conflicts varied due to influence from variables such as age, sex and conflict topic (1991). This relates to my work, as I feel as though my app could open another possible avenue for adolescents and their parents to be able to communicate their issues. This is not to say that the Problem Solver app would be the solution to everyone’s issues. Rather, I feel as though the newer perspective the app provides could play an important role in resolving some issues that were not handled well using traditional methods. Another important point that this article makes is that one of the possible variables that had not been previously accounted for in this area is the topic of the conflict, which seemed to have a significant impact on the outcomes of the conflict resolution in the study (Smetana et. al, 1991). Once again, I feel as though my app may be able to help in such situations, given that the topic of the conflict is one of the necessary components in the creation of a conflict in the app.

5. CONCLUSIONS

As has been discussed throughout this paper, I have designed an app called the Problem Solver app that aims to be of aid in the resolution of conflicts that arise between parents and their children . I wrote the app (using Flutter) to be an android-based mobile app available for free on the Google Play store. My desire to build such an app stems from personal experience, being that I have experienced conflicts with my parents that we had difficulty resolving, inspiring me to try to create a different way to attempt to solve these issues. I also took into account other methods that have/could have been used to solve these same problems and attempted to have my app cover up the weaknesses that the other resolution methods tended to have. I also took slight inspiration from the related works that I researched, specifically the article about conflict resolution in families with adolescents. These works led me to be even more inspired to make a relatively unique product that I felt could offer something new. While we were unable to design a full experiment to determine the effectiveness of the app, I feel as though the app could be quite effective and could help families or other people who have a difficult time expressing what they are struggling with [13].

Some of the limitations that exist in the current version of the app is that we are not exactly sure how effective it would actually be in resolving conflict (at least, we haven't been able to perform an experiment to quantify such a thing). Additionally, the app is only available for android-based mobile devices, which significantly limits the pool of people able to use the app (due to the prevalence of iPhones in today's world). A third limitation one could describe would be that the app itself is somewhat "bare bones," meaning that it does work, run and perform its purpose, but it contains mainly the basic features required (create a conflict, view a conflict, edit profile, etc) without any "Quality of Life" upgrades [14].

Naturally, we would desire to resolve these limitations to create a better working app for all to use. One way we could resolve the first limitation discussed above is by getting more people to use the app for its intended purpose, resulting in more feedback about how well the app worked for people. One could expect to see more usership of the app after running ads for it on sites such as Facebook, Instagram, etc. Solving the android-only issue would be more difficult however, as it would require tinkering with and changing the code and the way the app runs in order to work well on Apple devices. Additionally, the Apple App Store is much more difficult to get an app published on than on the Google Play Store, presenting yet another hurdle to overcome for this limitation. Finally, a much more feasible correcting of a limitation that could be done would be to polish the appearance and features of the app up in order to make it aesthetically pleasing and easier for the user to interact with. This could be accomplished by doing more research into Flutter and how to make things look nicer or how to add new features that users might appreciate (such as connecting with friends who also have the app).

REFERENCES

- [1] Vuchinich, Samuel. "Starting and stopping spontaneous family conflicts." *Journal of Marriage and the Family* (1987): 591-601.
- [2] Leege, David C., et al. "The politics of cultural differences." *The Politics of Cultural Differences*. Princeton University Press, 2009.
- [3] Burlinson, Brant R., and Amanda J. Holmstrom. "Comforting communication." *The international encyclopedia of communication* (2008).
- [4] Harker, Christopher, and Lauren L. Martin. "Familial relations: Spaces, subjects, and politics." *Environment and Planning A* 44.4 (2012): 768-775.
- [5] Westjohn, Stanford A., Holger Roschk, and Peter Magnusson. "Eastern versus Western culture pricing strategy: Superstition, lucky numbers, and localization." *Journal of International Marketing* 25.1 (2017): 72-90.
- [6] Boudon, Raymond. "Education, opportunity, and social inequality: Changing prospects in western society." (1974).
- [7] Barnes, Howard L., and David H. Olson. "Parent-adolescent communication and the circumplex model." *Child development* (1985): 438-447.
- [8] Holla, Suhas, and Mahima M. Katti. "Android based mobile application development and its security." *International Journal of Computer Trends and Technology* 3.3 (2012): 486-490.
- [9] Kuzmin, Nikita, Konstantin Ignatiev, and Denis Grafov. "Experience of developing a mobile application using flutter." *Information Science and Applications*. Springer, Singapore, 2020. 571-575.
- [10] Suriyah, Edwin Adrianta, and Yashinta Levy Septiarly. "Construct validation of five love languages." *Anima Indonesian Psychological Journal* 31.2 (2016): 65-76.
- [11] Prevoo, Mariëlle JL, and Catherine S. Tamis-LeMonda. "Parenting and globalization in western countries: explaining differences in parent-child interactions." *Current opinion in psychology* 15 (2017): 33-39.
- [12] Smetana, Judith G., Jenny Yau, and Shirley Hanson. "Conflict resolution in families with adolescents." *Journal of Research on Adolescence* 1.2 (1991): 189-206.
- [13] Liu, Kaifeng, Zhenzhen Xie, and Calvin Kalun Or. "Effectiveness of mobile app-assisted self-care interventions for improving patient outcomes in type 2 diabetes and/or hypertension: systematic review and meta- analysis of randomized controlled trials." *JMIR mHealth and uHealth* 8.8 (2020):

e15779.

- [14] Omran, Mahamed GH, Andries P. Engelbrecht, and Ayed Salman. "Bare bones differential evolution." *European Journal of Operational Research* 196.1 (2009): 128-139.
- [15] Shaffer, David Williamson, et al. "Video games and the future of learning." *Phi delta kappa* 87.2 (2005): 105-111.

© 2022 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.

AUTHOR INDEX

<i>Aadil Ahamed</i>	01
<i>Ali Skaf</i>	17
<i>Andrijana Ćurković</i>	279
<i>Ang Li</i>	285
<i>Anis Hoayek</i>	103
<i>Annie Dequit</i>	135
<i>Arezki Aberkane</i>	17
<i>Changqing Chen</i>	135
<i>Damir Vučina</i>	279
<i>Domagoj Samardžić</i>	279
<i>Enoc Gonzalez</i>	155
<i>Evelyn RodriguezArgueta</i>	135
<i>Fatema Nafa</i>	115, 135, 155
<i>Fei Wang</i>	239
<i>Gurpreet Kaur</i>	155
<i>Hang Cheng</i>	37
<i>Jannatun Noor</i>	225
<i>Jie Hao</i>	81
<i>Joshua Tian</i>	29
<i>Joyeeta Saha</i>	225
<i>Jun Yu</i>	123
<i>Kaiwen Chen</i>	51
<i>Kamran Alipour</i>	01
<i>Li Shen</i>	123
<i>Liang Zhan</i>	123
<i>Lifang He</i>	123
<i>Longxiao Li</i>	259
<i>Longzhen Zhou</i>	259
<i>Md. Nazrul Huda Shanto</i>	225
<i>Mei Yi Yang</i>	69
<i>Michael Pazzani</i>	01
<i>Michel Kamel</i>	103
<i>Milan Ćurković</i>	279
<i>Mingze Gao</i>	145
<i>Mireille Batton-Hubert</i>	103
<i>Muqing Bai</i>	213
<i>Naiyang Guan</i>	37
<i>Nibula Bente Rashid</i>	225
<i>Nicolò Busetto</i>	195
<i>Nowshin Tasfia</i>	225
<i>Raonak Islam Prova</i>	225
<i>Rodolfo Delmonte</i>	195
<i>Ruize Yu</i>	251
<i>Ryan Kanoff</i>	115
<i>Samir Dawaliby</i>	17

<i>Sateesh Kumar</i>	01
<i>Severine Soltani</i>	01
<i>Shixiong Wang</i>	37
<i>Tiffany Zhan</i>	63, 95
<i>Wenxin Tian</i>	163
<i>Xiaofeng Wang</i>	239
<i>Xu Wang</i>	259
<i>Yingbo Wu</i>	259
<i>Yingzhi Ma</i>	177
<i>Yu Sun</i>	29, 51, 69, 145, 177, 213, 251
<i>Yuanzhi Luo</i>	81
<i>Yuxing Ji</i>	145
<i>Zhaoming Kong</i>	123
<i>Zhenxing Li</i>	239
<i>Ziheng Guan</i>	285
<i>Zusheng Zhang</i>	259