

Ngoc Le Anh · Seok-Joo Koh ·  
Thi Dieu Linh Nguyen · Jaime Lloret ·  
Thanh Tung Nguyen *Editors*

# Intelligent Systems and Networks

Selected Articles from ICISN 2022,  
Vietnam

# **Lecture Notes in Networks and Systems**

**Volume 471**

## **Series Editor**

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences,  
Warsaw, Poland

## **Advisory Editors**

Fernando Gomide, Department of Computer Engineering and Automation—DCA,  
School of Electrical and Computer Engineering—FEEC, University of Campinas—  
UNICAMP, São Paulo, Brazil

Okyay Kaynak, Department of Electrical and Electronic Engineering,  
Bogazici University, Istanbul, Turkey

Derong Liu, Department of Electrical and Computer Engineering, University  
of Illinois at Chicago, Chicago, USA

Institute of Automation, Chinese Academy of Sciences, Beijing, China

Witold Pedrycz, Department of Electrical and Computer Engineering, University of  
Alberta, Alberta, Canada

Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Marios M. Polycarpou, Department of Electrical and Computer Engineering,  
KIOS Research Center for Intelligent Systems and Networks, University of Cyprus,  
Nicosia, Cyprus

Imre J. Rudas, Óbuda University, Budapest, Hungary

Jun Wang, Department of Computer Science, City University of Hong Kong,  
Kowloon, Hong Kong

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose ([aninda.bose@springer.com](mailto:aninda.bose@springer.com)).

More information about this series at <https://link.springer.com/bookseries/15179>

Ngoc Le Anh · Seok-Joo Koh ·  
Thi Dieu Linh Nguyen ·  
Jaime Lloret · Thanh Tung Nguyen  
Editors

# Intelligent Systems and Networks

Selected Articles from ICISN 2022, Vietnam



Springer

*Editors*

Ngoc Le Anh  
Swinburne University Vietnam  
Hanoi, Vietnam

Thi Dieu Linh Nguyen  
Hanoi University of Industry  
Hanoi, Vietnam

Thanh Tung Nguyen  
Vietnam National University  
Hanoi, Vietnam

Seok-Joo Koh  
School of Computer Science  
and Engineering  
Kyungpook National University  
Daegu, Korea (Republic of)

Jaime Lloret  
Integrated Management Coastal  
Research Institute  
Universitat Politecnica de Valencia  
Gandia, Valencia, Spain

ISSN 2367-3370

ISSN 2367-3389 (electronic)

Lecture Notes in Networks and Systems

ISBN 978-981-19-3393-6

ISBN 978-981-19-3394-3 (eBook)

<https://doi.org/10.1007/978-981-19-3394-3>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

# Preface

The International Conference on Intelligent Systems & Network (2022), popularly known as ICISN 2022, was held on March 19–20, 2022, at Swinburne Vietnam, Hanoi, Vietnam. The ICISN 2022 provides an international forum that brings together the researchers as well as the industry practitioners, who are actively involved in the research in fields of intelligent computing, data science, or any other emerging trends related to the theme covered by this conference.

This conference has technical paper sessions, invited talks, and panels organized around the relevant theme. On ICISN 2022, audiences have the opportunity to meet some of the world's leading researchers, learn about some innovative research ideas and developments around the world, and become familiar with emerging trends in science and technology. The ICISN is also a perfect forum and one of the platforms for presenting your own ideas and achievements in front of researchers from multiple countries.

ICISN 2022 received a huge response in terms of the submission of papers around the world. We received papers from various countries outside Vietnam such as India, China, Russia, Australia, New Zealand, Egypt, and South Korea. The organizing committee of ICISN 2022 constituted a strong international program committee for reviewing papers. A double-blind review process has been adopted. The decision system adopted by EasyChair has been employed, and 77 papers have been selected after a thorough double-blind review process. The proceedings of the conference will be published in Lecture Notes in Network Systems (LNNS) by Springer publisher and will be indexed by Scopus. We convey our sincere gratitude to the authority of Springer for providing the opportunity to publish the proceedings of ICISN 2022.

We really appreciate Swinburne Vietnam for agreeing to host the conference and for continuously supporting the organization team during the preparation of the conference. Without their support, this conference would have not been successful.

Our sincere gratitude to all keynote speakers, presenters, session chairs, and high officials in Vietnam for their gracious presence on the campus on the occasion.

We would like to thank the plenary talks by Prof. Dr. Truong Gia Binh, Chairman of FPT Corporation, Dr. Hoang Viet Ha, Director of Swinburne Vietnam, Prof. Dr. Douglas Proctor, Swinburne Australia, Dr. Vijender Kr. Solanki, chair of ICISN 2022.

We would like to thank the keynote speaker—Prof. Lam Kwok Yan, Nanyang Technological University, Singapore; Associate Prof. Rafidah Binti Md Noor, University of Malaya, Malaysia; Associate Prof. Chutiporn Anutariya, School of Engineering and Technology (SET), Asian Institute of Technology (AIT), Thailand—for giving their excellent knowledge in the conference.

We would like to thank Mr. Truong Quoc Hung, The CEO of VinBrain Company, Vingroup, Dr. Le Nhan Tam, Chief Technology Officer (CTO) & Country Technical Manager, IBM, Dr. Tran The Trung, Director of FPT Technology Research Institute and Scientific & FPT Smart Cloud, FPT Group for giving their experience in the conference. We would like to thank the reviewers for completing a big reviewing task in a short span of time.

We would also like to thank the program committee members: Dr. Hoang Viet Ha, Director of Swinburne Vietnam; Dr. Truong Cong Duan, Head of Academic, Swinburne Vietnam; Dr. Le Anh Ngoc, Director of Swinburne Innovation Space, Swinburne Vietnam; Dr. Nguyen Thi Dieu Linh, Deputy of Head of Science and Technology Department, Hanoi University of Industry, Vietnam; Dr. Tran Duc Tan, Vice Dean of Faculty of Electrical and Electronics Engineering, Phenikaa University; Mr. Lai Hong Anh, Head of Admission, Swinburne; the event management support staffs from Swinburne Vietnam: Ms. Dang Thi Phuong, Ms. Tran Huyen Trang, Ms. Le Ngoc Tram, Ms. Huynh Ho Duc Minh for their efforts to make congress success.

We hope that the papers published in ICISN 2022 proceedings will be helpful for the researchers pursuing studies in computer science, information technology, and related areas. Industrial engineers would also find this volume to be a good source of reference.

We are looking forward to seeing you all at the International Conference on Intelligent Systems & Network that will be held on March 19–20, 2023, at the same location—Swinburne University of Technology Vietnam. We hope that ICISN 2023 will be an interesting forum and the perfect platform for presenting your own ideas and achievements.

Ngoc Le Anh  
Seok-Joo Koh  
Thi Dieu Linh Nguyen  
Jaime Lloret  
Thanh Tung Nguyen

# Contents

<b>AIC Algorithm for a Decision to Choose the Convenience Store Chain: Evidence from Vietnam .....</b>	1
Nguyen Thi Ngan and Huy Khoi	
<b>A Simple and Automatic Method to Estimate Lung Volume Based on Thoracic Computed Tomography Images .....</b>	11
Tran Anh Vu, Pham Duy Khanh, Hoang Quang Huy, Nguyen Tuan Dung, and Pham Thi Viet Huong	
<b>Towards Generating RESTful Full Stack Software: A Module Based Domain Driven Design Approach .....</b>	22
Duc Minh Le and Ha Thanh Vu	
<b>Time Series Prediction Based on Consecutive Neighborhood Preserving Properties of Matrix Profile .....</b>	33
Thi Dung Tran, Ngoc Dung Bui, and Siew Ann Cheong	
<b>Fast and Robust Local Texture Information for Fingerprint Segmentation .....</b>	39
Binh T. H. Nguyen, Ngoc Dung Bui, and Xuan Tung Hoang	
<b>Developing Real-time Recognition Algorithms on Jetson Nano Hardware .....</b>	45
Phat Nguyen Huu, Thien Pham Ngoc, and Thanh Le Thi Hai	
<b>Numerical Method of Synthesizing the Automatic Electric Drive System Containing Elastically Linked Elements .....</b>	53
Dang Nguyen Phu, Vu Duc Tuan, Tran Binh Duong, Nguyen Duc Uyen, and Duc-Tan Tran	
<b>A Fuzzy Approximation Supported Model-Free Tracking Control Design for Tower Crane Systems .....</b>	62
Nhu Toan Nguyen, Van-Anh Nguyen, Manh Cuong Nguyen, Danh Huy Nguyen, and Tung Lam Nguyen	

<b>Control Design of Thermoacoustic Generator Systems Using Reinforcement Learning Algorithm .....</b>	71
Duy Tung Le, Trong Tan Do, and Phuong Nam Dao	
<b>Application of Intrinsic Time-Scale Decomposition to Classify EEG Signals Using Machine Learning and Neural Network Methods .....</b>	76
Viet Quoc Huynh, Yen Hua-Thi-Hoang, and Tuan Van Huynh	
<b>Development of Economical and Effective Ultraviolet (UVC) Irradiation Solutions and Equipment for Rapid Decontamination of SARS-CoV-2 on N95 Respirators .....</b>	83
Nguyen Manh Cuong, Vuong Tri Tiep, Nguyen Phan Kien, Hoang ChuDuc, and Grayr K. Aleksanyan	
<b>Sentiment Analysis in Twitter by Resnet-50 and VGG-19 .....</b>	92
Nguyen Quynh Anh and Dao Nam Anh	
<b>Improvement of Self-supervised Depth and Motion Learning with Vision Transformer .....</b>	99
Xuan-Quang Ma, Ngoc-Kien Nguyen, and Van-Tinh Nguyen	
<b>Application of Event-Triggered Sliding Mode Control of 2-DOF Humanoid's Lower-Limb Powered by Series Elastic Actuator .....</b>	110
Anh Khoa Lanh Luu, Van Tu Duong, Huy Hung Nguyen, and Tan Tien Nguyen	
<b>Blockchain in Enterprise Applications: An Introduction .....</b>	121
Hung Ho-Dac, Van Huu Tran, The Bao Nguyen, Hai Vinh Cuong Nguyen, and Van Len Vo	
<b>Backstepping Super-Twisting Based Sliding Mode Controller for a Lower Limb Actuator .....</b>	130
Tri Duc Tran, Van Tu Duong, Huy Hung Nguyen, and Tan Tien Nguyen	
<b>Text Spotting in Vietnamese Documents .....</b>	141
Vu Minh Duc and Tran Ngoc Thang	
<b>An Advanced IoU Loss Function for Accurate Bounding Box Regression .....</b>	151
Ho-Si-Hung Nguyen, Thi-Hoang-Giang Tran, Dinh-Khoa Tran, and Duc-Duong Pham	
<b>Occlusion Robust Face Recognition Based on Mask Learning with Attention Mechanism .....</b>	158
Quan Nguyen Minh, Bang Le Van, Can Nguyen Ngoc, and Viet Dung Nguyen	

<b>A Proposed Business Intelligence Framework for Autonomous and Non-public Higher Education Institutions in Vietnam . . . . .</b>	168
Han Minh Phuong, Pham Minh Hoan, Nguyen Trung Tuan, and Doan Trung Tung	
<b>Fall Detection Using Omnidirectional Camera . . . . .</b>	178
Viet Dung Nguyen, Thi Men Tran, Xuan Bach Nguyen, and Minh Quan Nguyen	
<b>Speed Up Communication for IoT Devices Using Small Form Factor in FPGA Platform . . . . .</b>	185
Trung Dang Tien, Tuan Nguyen Ngoc, Dat Tran Thanh, and Tinh Tran Xuan	
<b>The High-Speed Data Transmission System on Fiber Optic Cable for IoT Equipment Using FPGA Platform . . . . .</b>	194
Tuan Nguyen Ngoc, Dat Tran Thanh, Thang Le Dinh, Tinh Tran Xuan, and Tuyen Tran Van	
<b>Design LMI Sliding Mode Control for the Anti-aircraft Gun Speed Control Loop . . . . .</b>	202
Tuan Nguyen Ngoc, Trung Dang Tien, Tinh Tran Xuan, and Tuyen Tran Van	
<b>Extraction of Characteristic Breast Dimensions of North Vietnamese Girl-Students Using Random Forest Algorithm . . . . .</b>	210
Nguyen Thi Le, Luu Thi Hong Nhung, and Nguyen Nhat Trinh	
<b>Vacant Parking Car Detection and Notification by Using Multi-camera System and Deep Learning . . . . .</b>	221
Dinh Cong Nguyen, Viet Nam Le, Hoang Long Nguyen, Doan Minh Tran, and Van Luyen Cao	
<b>Implementation of Real-Time Human Tracking System Based on Deep Learning Using Kinect Camera . . . . .</b>	230
Van-Truong Nguyen and Duc-Tuan Chu	
<b>An Approach to Evaluate the Reliability of the Face Recognition Process Using Adversarial Samples Generated by Deep Neural Networks . . . . .</b>	237
Dinh Cong Nguyen, Nhan Dong Le, The Cuong Nguyen, Tuong Quan Nguyen, and Van Quan Nguyen	
<b>Multi Deep Learning Model for Building Footprint Extraction from High Resolution Remote Sensing Image . . . . .</b>	246
Ho Trong Anh, Tran Anh Tuan, Hoang Phi Long, Le Hai Ha, and Tran Ngoc Thang	

<b>Optimizing Energy Consumption in Smart Homes Based on OCF Standard Connectivity . . . . .</b>	253
Anam Nawaz Khan, Atif Rizwan, Wenquan Jin, Rashid Ahmed, Anh Ngoc Le, and Dohyeun Kim	
<b>Physical Therapy Exercise Design for Lower Limb Rehabilitation Robot Using Character's Animation Clips . . . . .</b>	260
Minh-Duc Duong, Ngoc-Thuan Ly, Anh-Tu Hoang, Van-Vuong Dinh, and Quy-Thinh Dao	
<b>Real-Time Object Following Based on Visual Information and Online Video Stabilization for Unmanned Aerial Vehicles . . . . .</b>	270
Quoc-Sang Tran	
<b>Mobility Prediction on a Location-Based Social Network Using K Latest Movements of Friends . . . . .</b>	279
Hai Thanh Nguyen, Chi Le Hoang Tran, and Huong Hoang Luong	
<b>A New System for License Plate Recognition in Traffic Violation Scenarios in Vietnam . . . . .</b>	287
Ha Nguyen-Xuan, Dong Hoang-Nhu, Kien Kim-Van, and Tuan Dang-Minh	
<b>Development of a Real-Time Obstacle Detection System on Embedded Computer Based on Neural Network and Multi-camera Streaming . . . . .</b>	298
Quang Khoi Tran, Khanh Hieu Ngo, Anh Huy Le Dinh, Lu Tien Truong, Hai Quan Tran, and Anh Tuan Trinh	
<b>Optimizing Penta-Band Vivaldi Antenna for 5G, WiFi 6, Radar and Satellite Applications . . . . .</b>	309
Thi Thanh Tu Duong, Minh Duc Hoang, and Thi Thu Nga Nguyen	
<b>The Method Calculates of Radial Distance from the Human Body Surface to the Inner Surface of a Women's Shirt . . . . .</b>	320
Do Thi Thuy, Nguyen Thi Thuy Ngoc, and Ngo Chi Trung	
<b>Takagi-Sugeno Fuzzy Control for Stabilizing Nonlinear Inverted Pendulum . . . . .</b>	333
Duc-Binh Pham, Duy-Thi Pham, Quy-Thinh Dao, and Van-Anh Nguyen	
<b>A Comparative Study of Model Selection for High Dimensional Data . . . . .</b>	342
Thanh Trinh	
<b>Extract Forest Height from PolInSAR Image Based on Decomposition Scattering Component and Polarization State Optimization Technique . . . . .</b>	351
Nghia Pham Minh, Cuong Thieu Huu, Ky Le Van, and Dung Duong Quoc	

<b>Fuzzy Portfolio Selection with Flexible Optimization via Quasiconvex Programming . . . . .</b>	360
Tran Thi Thanh Tuo, Truong Tuan Khang, Nguyen Thi Ngoc Anh, and Tran Ngoc Thang	
<b>AN-BEATS for Short-Term Electricity Load Forecasting with Adjusted Seasonality Blocks and Optimization of Block Order . . . . .</b>	369
Nguyen Anh Tuan, Do Thi Thanh Chau, Le Anh Ngoc, and Nguyen Thi Ngoc Anh	
<b>Non-contact Thermometer System for Body Temperature Measurement at Hospitals . . . . .</b>	379
Nguyen Phan Kien, Than Tra My, Tran Anh Vu, and Nguyen Thanh Trung	
<b>Classification of 40 Hand Movements Based on Surface Electromyogram Signal . . . . .</b>	386
Thuy Le – Nguyen Thi, Thuan Nguyet Phan, and Tuan Van Huynh	
<b>An Optimal Path Planning Issue Parameters with Potential Feasible Area Using Evolutions Algorithm . . . . .</b>	394
Trong-The Nguyen, Thi-Xuan-Huong Nguyen, Van-Chieu Do, and Truong-Giang Ngo	
<b>Binning on Metagenomic Data for Disease Prediction Using Linear Discriminant Analysis and K-Means . . . . .</b>	402
Nhi Yen K. Phan and Hai Thanh Nguyen	
<b>Letter Recognition in Hand Sign Language with VGG-16 . . . . .</b>	410
Hai Thanh Nguyen, Linh Thuy Thi Pham, Truc Thuy Nguyen Mai, Trinh Kieu Ngoc Vo, and Tran Thanh Dien	
<b>A New Method for IP Camera Based Face Anti-spoofing Systems . . . . .</b>	418
Ha Nguyen-Xuan, Dong Hoang-Nhu, Anh Nguyen-Duy, and Tuan Dang-Minh	
<b>Vietnamese Spelling Error Detection and Correction Using BERT and N-gram Language Model . . . . .</b>	427
Dong Nguyen Tien, Tuoi Tran Thi Minh, Loi Le Vu, and Tuan Dang Minh	
<b>Clothing Recognition Using Transfer Learning and Segmentation Techniques . . . . .</b>	437
Hai Thanh Nguyen, Thu Anh Thi Nguyen, Khanh Kim Nguyen, Chi Le Hoang Tran, Thai Minh Phan Le, Diem Ngoc Thi Pham, and Dien Thanh Tran	
<b>An Evaluation of the Multi-probe Locality Sensitive Hashing for Large-Scale Face Feature Matching . . . . .</b>	445
Ha Nguyen-Xuan, Dong Hoang-Nhu, and Tuan Dang-Minh	

<b>Determination of Fruit Freshness Using Near-Infrared Spectroscopy and Machine Learning Techniques .....</b>	455
Duy Khanh Ninh, Kha Duy Phan, Chi Khanh Ninh, and Nhan Le Thanh	
<b>Vietnamese Named Entity Recognition on Medical Topic .....</b>	465
Loi Le Vu, Duong Pham Van, Dong Nguyen Tien, Tuoi Tran Thi Minh, and Tuan Dang Minh	
<b>Cooperative Communication in NSA and SA 5G Networks .....</b>	476
Sinh Cong Lam and Duc-Tan Tran	
<b>Cross-Layer Design for Wireless Sensor Networks Using Cooja Tool in Contiki Operating System .....</b>	484
Nguyen Canh Minh, Duc-Tan Tran, Quang-Trung Hoang, Trung Ninh Bui, Gian Quoc Anh, Kolla Bhanu Prakash, and Sayed El-Rabaie	
<b>Mobile Robot Motion Control Using a Combination of Fuzzy Logic Method and Kinematic Model .....</b>	495
Anh-Tu Nguyen and Cong-Thanh Vu	
<b>Examining Topology and Dynamics-Related Modular Robustness in a Large-Scale Signaling Network .....</b>	504
Cong-Doan Truong and Duc-Quynh Tran	
<b>Implementation of SMOTE and VGG-16 for COVID-19 Radiography .....</b>	511
Xuan Tho Dang and Nam Anh Dao	
<b>Research Complex Fuzzy Inference System in Early Warning Credit Risk at Commercial Banks in Viet Nam .....</b>	519
Hong Thi Chu Hai and Trieu Thu Huong	
<b>A Proposed Architecture of Intelligent Operations Center for Smart Cities Development .....</b>	526
Minh Tan Ngo, Ngoc Hai Nguyen, Cong Duan Truong, Duy Hung Tran, and Cong- Doan Truong	
<b>Greedy Detection for ReIQ System .....</b>	535
Trung Tan Nguyen	
<b>Improving Reliability of Laser Sensor in Aviation Equipment with Anti-jamming Encoding Solution .....</b>	541
Thi Nguyen Duc, Trung Dang Tien, Tuyen Tran Van, and Giang Vu Hoang	
<b>Proposing System for Classifying Voucher Images Using Siamese Model .....</b>	550
Phat Nguyen Huu, Nguyet Giap Thi, Hoang Tran Manh, Thanh Le Thi Hai, Hieu Nguyen Trong, and Pha Pham Ngoc	

Contents	xiii
<b>Synthesis of Target Angle Tracking System Combining Kalman Filter Algorithm with Fuzzy Logic Controller . . . . .</b>	557
Nguyen Thi Dieu Linh and Tran Ngoc Duc	
<b>Design of an Adaptive Fuzzy Hierarchical Sliding-Mode Control for Uncertainties Pendubot . . . . .</b>	568
Van-Truong Nguyen and Quoc-Cuong Nguyen	
<b>Adaptive Neural Network Hierarchical Sliding-Mode Control for Pendubot Based Genetic Algorithm Optimization . . . . .</b>	574
Van-Truong Nguyen, Xuan-Thang Vu, and Hai-Binh Giap	
<b>Classification of Malicious Websites Using Feature Based Machine Learning Techniques . . . . .</b>	581
Swetha Mahendrakar ShyamRao, Rakshith Danti, Praveen Narayappa Manjunath, and Muneshwara Mahendrakar ShyamRao	
<b>A Design Model Network for Intelligent Web Cache Replacement in Web Proxy Caching . . . . .</b>	591
Truong Nguyen Xuan, Van Tran Thi, and Lam Ho Khanh	
<b>An Abnormal Heart Rhythm Warning System Based on a Low-Cost Two-Electrode ECG Signal Using Threshold and Neural Network Approaches . . . . .</b>	601
Thuong H. N. Nguyen, Mai Hue T. Truong, Duc M. Le, Ngoc Huynh Tran, Thi Hai Hau Pham, B. Hoang Nguyen, Tuan V. Pham, and Anh Thu T. Nguyen	
<b>Edge AI Implementation for Recognizing Sounds Created by Human Activities in Smart Offices Design Concepts . . . . .</b>	608
Thuong H. N. Nguyen, Quy C. Nguyen, Viet H. H. Ngo, Fabien Ferrero, and Tuan V. Pham	
<b>The Remarkable Enhancement in Fish Freshness Classification on Various Species Using Neural Network on Physiological Characteristics Features . . . . .</b>	615
Anh Thu T. Nguyen, Nguyet Ha T. Tran, Phuong Dieu T. Nguyen, Dat Xuan Ho, Uyen Kim Hoang, Nguyen Ba Hoang, Thuong H. N. Nguyen, and Tuan V. Pham	
<b>Enabling the Smart City Through Agile LP-WAN Network in Da Nang City . . . . .</b>	623
Thao Nguyen Manh, Lic Tran Van, Thanh Ngo Dinh, Thu Nguyen Thi Anh, Huy Le Quoc, Huy Trinh Le, Fabien Ferrero, and Nhan Le Than	
<b>An Effective Beamformer for Interference Mitigation . . . . .</b>	630
Luyen Tong, Cuong Nguyen, and Duy Le	

<b>Support Vector Machine Used for Gender Classification . . . . .</b>	640
Meena Tiwari and Ashish Mishra	
<b>Relay Selection for Enhancing the Performance of Full-Duplex System with Imperfect Transceiver Hardware . . . . .</b>	650
Ba Cao Nguyen, Le The Dung, Tran Manh Hoang, Xuan Nam Tran, and Taejoon Kim	
<b>Heart Disease Prediction Using Soft Computing Methods . . . . .</b>	660
N. Veena, S. Mahalakshmi, T. Ragunthar, Ambika Rani Subash, and K. Vinutha	
<b>Enhancing Inflammatory Bowel Disease Diagnosis Performance Using Chi-Squared Algorithm on Metagenomic Data . . . . .</b>	669
Hai Thanh Nguyen, Huong Hoang Luong, Trong Thanh Tran, Ngoc Van Nguyen, and Khoi Dinh Nguyen	
<b>Author Index . . . . .</b>	679

# About the Editors

**Dr. Ngoc Le Anh** is a Director of Swinburne Innovation Space, Swinburne University of Technology (Vietnam). He received his B.S in Mathematics and Informatics from Vinh University and VNU University of Science, respectively. He received a master's degree in Information Technology from Hanoi University of Technology, Vietnam. He obtained a Ph.D. in Communication and Information Engineering from the School of Electrical Engineering and Computer Science, Kyungpook National University, South Korea, in 2009. His general research interests are embedded and intelligent systems, communication networks, the Internet of things, image/video processing, AI, and big data analysis. He served as a keynote speaker, conference chair, book editor, and reviewer of the international conferences and journals (Email: [nle@swin.edu.au](mailto:nle@swin.edu.au)).

**Dr. Seok-Joo Koh** received B.S. and M.S. degrees in Management Science from Korea Advanced Institute of Science and Technology (KAIST), Korea, in 1992 and 1994, respectively. He received Ph.D. degree in Industrial Engineering from KAIST, Korea, in 1998. From August 1998 to February 2004, he worked for the Protocol Engineering Center in Electronics and Telecommunication Research Institute (ETRI). Since March 2004, he has been with the school of Computer Science and Engineering (CSE) at the Kyungpook National University (KNU) as a professor (<http://cse.knu.ac.kr>). In KNU, he has also been working for Software Education Center as the director. (<http://swedu.knu.ac.kr>).

He has so far published over 70 international papers. His current research interests include the Internet of things (IoT) and IoT-related protocols and services, such as in-vehicle infotainment (IVI) services. He has also been interested in the newly emerging communication protocols (such as SCTP, CoAP, and QUIC) and the Internet mobility management and multicasting.

He has also been participating in the International Standardization as an editor (and published over 20 international standards) in the ITU-T SG13, SG20, IEC TC100, and ISO/IEC JTC1/SC6 and SC41.

**Dr. Thi Dieu Linh Nguyen** is Deputy Head of Science and Technology Department, Hanoi University of Industry, Vietnam (HaUI). She has more than 20 years of academic experience in electronics, IoT, smart garden, and telecommunication. She received Ph.D. in Information and Communication Engineering from Harbin Institute of Technology, Harbin, China, in 2013; she has authored or co-authored many research articles that are published in journals, books, and conference proceedings. She is editor for many books and also a board member for the International Journal of Hyperconnectivity and the Internet of Things (IJHIoT) IGI Global, USA; Information Technology Journal; Mobile Networks and Application Journal and some other reputed journals and international conferences.

She has attended many conferences as keynote speaker, conference chair, guest honor; and chaired many technical events in different universities in Vietnam.

She can be contacted at [nguyen.linh@hau.edu.vn/nguyenlinh79.hau@gmail.com](mailto:nguyen.linh@hau.edu.vn/nguyenlinh79.hau@gmail.com).

**Dr. Jaime Floret** received his B.Sc.+M.Sc. in Physics in 1997, his B.Sc.+M.Sc. in electronic Engineering in 2003 and his Ph.D. in telecommunication engineering (Dr. Ing.) in 2006. He is a Cisco Certified Network Professional Instructor. He worked as a network designer and administrator in several enterprises. He is currently Associate Professor in the Polytechnic University of Valencia. He is the Chair of the Integrated Management Coastal Research Institute (IGIC) since January 2017 and he is the head of the “Active and collaborative techniques and use of technologic resources in the education (EITACURTE)” Innovation Group. He is the director of the University Diploma “Redes y Comunicaciones de Ordenadores” and he has been the director of the University Master “Digital Post Production” for the term 2012–2016. He was Vice-chair for the Europe/Africa Region of Cognitive Networks Technical Committee (IEEE Communications Society) for the term 2010–2012 and Vice-chair of the Internet Technical Committee (IEEE Communications Society and Internet society) for the term 2011–2013. He has been Internet Technical Committee chair (IEEE Communications Society and Internet society) for the term 2013–2015. He has authored 14 books and has more than 600 research papers published in national and international conferences, international journals (more than 300 with ISI Thomson JCR). He has been the co-editor of 54 conference proceedings and guest editor of several international books and journals. He is editor-in-chief of the “Ad Hoc and Sensor Wireless Networks” (with ISI Thomson Impact Factor), the international journal “Networks Protocols and Algorithms”, and the International Journal of Multimedia Communications. Moreover, he is Associate editor of “Sensors” in the Section sensor Networks, he is advisory board member of the “International Journal of Distributed Sensor Networks” (both with ISI Thomson Impact factor), and he is IARIA Journals Board Chair (8 Journals). Furthermore, he is (or has been) associate editor of 46 international journals (16 of them with ISI Thomson Impact Factor). He has been involved in more than 500 Program committees of international conferences, and more than 160 organization and steering committees. He has led many local, regional, national and European projects. He was the chair of the Working Group

of the Standard IEEE 1907.1 from 2013 to 2018. Since 2016 he is the Spanish researcher with highest h-index in the TELECOMMUNICATIONS journal list according to Clarivate Analytics Ranking. He has been general chair (or co-chair) of 61 International workshops and conferences. He is IEEE Senior, ACM Senior and IARIA Fellow.

**Assoc. Prof. Dr. Thanh Tung Nguyen** is Director of the Interdisciplinary Research Department on Information Systems and Computer Engineering at Vietnam National University, Hanoi. He graduated in Information Technology from Royal Melbourne Institute of Technology (RMIT), Ph.D. in Network Systems from Monash University, Australia. He is an expert in the field of wireless sensor networks and information systems. He is a leader of many research projects at State, Ministry and Vietnam National University level and participates in many international conferences in the field of information systems and networking. In his account, more than 40 research articles are listed in reputable places.



# AIC Algorithm for a Decision to Choose the Convenience Store Chain: Evidence from Vietnam

Nguyen Thi Ngan<sup>✉</sup> and Huy Khoi<sup>(✉)</sup>

Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam  
buihuykhoi@iu.edu.vn

**Abstract.** The chapter tests elements influencing a decision to choose the convenience store chain in the context of the Vietnam market. The study's data was compiled by the author from a variety of credible sources, as well as primary data obtained via a survey of 210 Vietnamese customers. The findings of exploratory factor analysis using independent and dependent variables all demonstrate significant convergence of factors in the research model, and the tests all yield the required reliability. Factors discovered through independent variable analysis include Brand (BD), Social influence (SI), Quality Perception (QP), and Price Perception (PP), Distribution density (DD), Responsiveness (RE), and Promotion Perception (PPE). The results of regression analysis showed that factors in the research model could explain 63.73% of the variation in factors affecting shopping decisions at the convenience store chain in the area. A relatively high percentage shows that the suitability of the theoretical model with the actual data is quite good. The chapter uses an optimal choice by the AIC Algorithm for a decision to choose the convenience store chain in the context of the Vietnam market.

**Keywords:** AIC algorithm · Brand · Social influence · Quality Perception · Price Perception · Distribution density · Responsiveness · Promotion Perception · Decision · The convenience store chain

## 1 Introduction

Vietnam has the fastest-growing middle class in Southeast Asia, and the country's retail industry has exploded. Given its exceedingly young populace demographics and high client purchasing conduct, this fashion appears to be destined to remain in the future. Indeed, the upward thrust of e-trade and virtual retail channels is one of the most critical developments to comply within Vietnam: with the aid of 2025, the use of a's e-commerce area maybe second only to Indonesia in Southeast Asia [1]. Along with the development of the economy in recent years, the modern distribution channel, especially the convenience store model, has positively contributed to promoting the consumption of goods - food, at the same time bringing [2]. People add a new choice in deciding where to buy food besides traditional markets or grocery stores. With the advantages of store size and convenient service hours, convenience stores are increasingly present in residential

areas, near schools, to meet the needs of compact and economical shopping. time, but still ensure the high requirements for the quality of the goods[3]. Overcoming the disadvantages of traditional markets as well as supermarkets, convenience stores will likely dominate the retail market shortly. This is an extended arm of businesses and suppliers to bring goods closer to consumers, step by step changing the shopping habits of Vietnamese people [4]. Bui et al. [5] investigate the elements that determine trademark choice behavior among retail chains in Vietnam, as well as the extent to which these factors influence client decision intentions According to the findings, the eight criteria of store image, price perception, risk perception, trademark attitude, trademark awareness, and brand familiarity were strong-minded. In particular, this type of convenience store still has a lot of land for development, with great potential for development. It realizes the importance of how to give consumers more access to safe food products at convenience stores and how to help businesses run the modern retail system. In particular, the convenience store chain captures the key factors that are critical to customers' purchasing decisions scientifically to offer solutions to increase their capacity. Therefore, this paper uses the AIC algorithm for a decision to choose the convenience store chain in the context of the Vietnam market.

## 2 Literature Review

### 2.1 Brand (BD)

Shopping trends of customers are often interested in reputable brands. Because the reputable brand there believes that the product quality will be guaranteed, the quality of service is good [6]. Therefore, the selection of reputable and branded convenience stores to shop for customers is always important when shopping at convenience stores [7].

### 2.2 Social Influence (SI)

Social influence is the perception of an individual, with that individual's key references to whether behavior should or should not be done so social influence impact can be measured through stakeholders. with the consumer, determined by the standard beliefs for the expected performance of the behavior and the personal motivation to comply with that expectation [8]. Social influence is a social determinant of the reflection of social pressure on behavior. These social influences can be cross-cultural and social influences on one's beliefs that others agree to shop for shopping [9].

### 2.3 Quality Perception (QP)

Product quality refers to a customer's view of a product's or service's quality and superiority concerning alternative products and services, the purpose of using the product [10]. The perceived quality of the product will directly affect the purchase decision and customer loyalty for that product. So a quality product will always receive trust from customers [11]. Quality standards: The degree of conformance to the quality specification and the product, in particular, is free of defects. This can be considered as one of the most important requirements of a product because a product with good quality can consumers trust and are willing to use as well as loyal to it and decide to buy [12].

## 2.4 Price Perception (PP)

Price is one of the four key variables of the marketing mix [13]. Price is also one of the factors determining customers' behavior of choosing one or the other. As for convenience stores, it directly affects sales and profits. Therefore, the convenience store chain needs to consider carefully before offering a price for the product so that it is suitable for customers' ability to pay and can compete well with competitors and achieve the best profit [14].

## 2.5 Distribution Density (DD)

The location of a store is a geographical indication of where a business places the convenience store chain to sell products/goods [15]. Customers care about the location as well as the distance and convenience of transportation when coming to the store [16]. Beautiful, well-positioned, well-located, easy-to-find shops often have an advantage over some in an alley area, difficult to see or located on the side of a one-way street. In some cases, this is sometimes critical. Store location has been reviewed by several studies.

## 2.6 Responsiveness (RE)

The convenience store chain provides complete, timely, and accurate information to customers to quickly solve problems, handle complaints effectively and respond to customer requests. During the transaction, when there is any change of information, the store will promptly notify the customer. The store provides complete and accurate information to customers, ready to respond to customers' questions. The store's staff is always attentive during peak hours and attentive service staff for all customers. This increases the Store's responsiveness to customers [17].

## 2.7 Promotion Perception (PP)

This type of impact affects the customer's purchasing decision through the customer's feelings and emotions for the promotion, for the selection of the promotion, even for the dropout. Miss the chance to choose in that promotion [18]. The first positive impact is that when choosing during the promotion, clients will feel happy and confident because they reflect they are clever and wise "shoppers" (because they can find cheaper goods than usual.). Furthermore, the campaign encourages clients to expand and improve their life knowledge and experience. The promotion itself is still measured as interesting, learning promotions, participating in gift programs, and drawing... is a pleasure for many customers. And with the promotional pictures, promotional prices with different colors, promotions also make the difference, diversity, "fun" in the product display [19].

## 2.8 Choosing Decision (CD)

The choosing decision is the final stage of the consumer's choosing behavior [20]. It is the consequence of weighing options based on a balance of requirements and capabilities; the entire benefit or value a consumer receives from a product or service versus the whole

cost they must pay to obtain the product or service; and being affected by relatives, friends, and coworkers; Unexpected situations arise and the risks customers perceive before making purchasing decisions if the products or services provided are worthy and meet previous needs and expectations. They will continue to use that product or service in the future and they will advertise on behalf of the supplier to other customers [21].

### 3 Methodology

Bollen [22] states that for one factor to be assessed, a minimum sample size of 5 samples is required. Table 1 suggests sample traits information.

**Table 1.** Statistics of sample

Characteristics		Amount	Percent (%)
Sex and age	Male	135	64.3
	Female	75	35.7
	Below 20	6	2.9
	21–35	84	40.0
	35–50	98	46.7
	Above 50	22	10.5
Income/month	Below 7 million VND	6	2.9
	7–15 million VND	103	49.0
	Over 15 million VND	101	48.1
Job	Housewife	69	32.9
	Officer	89	42.4
	Student	6	2.9
	Businessman	46	21.9

The sample size is 5: 1 (five observations per one variable) [23]. The information gathered will be encrypted and inserted into Table 2.

**Table 2.** Factor and item

Factor	Code	Item
Brand (BD)	BD1	The convenience store chain is a popular brand
	BD2	The convenience store chain brand name is easy to read, remember, and easy to recognize
	BD3	The convenience store chain is a reputable corporation with long-term operations in the industry
	BD4	The convenience store chain brand makes you feel confident and secure when choosing to shop
Social influence (SI)	SI1	Family members influence you when shopping at the convenience store chain
	SI2	Friends and colleagues influence you when shopping at the convenience store chain
	SI3	Advertising through television, newspapers, and the internet affects me when shopping at the convenience store chain
Quality Perception (QP)	QP1	The convenience store chain specializes in distributing products with reputable brands and clear origins
	QP2	Products sold at the convenience store chain are preserved in sealed, eye-catching, safe, and hygienic packaging
	QP3	The convenience store chain has a team of experts and a food preservation process that ensures product quality
	QP4	All products sold at the convenience store chain are still valid
	QP5	There are no counterfeits for products sold at the convenience store chain
Price Perception (PP)	PP1	The selling price of products at the convenience store chain is currently reasonable
	PP2	The selling price of products at the convenience store chain is commensurate with the quality
	PP3	The selling price of products at the convenience store chain is listed on the package
	PP4	Selling prices at the convenience store chain are competitive compared to other stores
	PP5	The selling price of products at the convenience store chain is in line with my financial ability

(continued)

**Table 2.** (*continued*)

Factor	Code	Item
Distribution density (DD)	DD1	There are many points of sale, convenient and easy to search for me to buy products at the convenience store chain store
	DD2	The convenience store chain's points of sale are located in central, prime locations, populous places, and the center of the market
	DD3	The convenience store chain's chain of points of sale has certain similarities following the brand's image (color design, display, preservation...)
Responsiveness (RE)	RE1	Choosing an essential product at the convenience store chain is easy
	RE2	The convenience store chain meets a variety of items according to my needs
	RE3	Fast and convenient payment time
	RE4	I am provided with all the information you need about the product at the store and on the packaging (production date, expiry date, weight, listing price...) at the convenience store
Promotion Perception (PPE)	PPE1	The promotion applies to all customers who purchase at the convenience store
	PPE2	The convenience store chain has many attractive promotions
	PPE3	Practical promotions for customers
Choosing Decision (CD)	CD1	I feel satisfied when deciding to shop at the convenience store chain
	CD2	I will choose to shop at the convenience store chain first when needed
	CD3	I will continue to shop at the convenience store chain in the future
	CD4	I am always ready to recommend to friends and relatives to shop at the convenience store chain

Factor and Item are coded in Table 2. All study staff and respondents were blinded for the length of the trial. The study participants had no touch with anyone from the outside world.

## 4 Results

### 4.1 AIC Algorithm

The R software used AIC (Akaike's Information Criteria) to choose the most beneficial version. In the theoretical environment, AIC has been used to choose fashions. The AIC

method can also cope with several unbiased variables when multicollinearity develops. AIC may be used as a regression model to estimate one or extra based variables from one or extra independent variables. The AIC is a vital and useful metric for determining a full and easy version. A model with a lower AIC is selected based on the AIC statistics general. When the minimum AIC cost is reached, the exceptional version will come to a cease [24–26]. Every step of the search for the fine version is documented in R reports. With 4 independent variables, the step comes to a halt with  $AIC = -489.36$  for  $CD = f(QP + PP + RE + DD + BR + SI + PPE)$ .

**Table 3.** The coefficients

CD	Estimate	Std. Error	t value	P-value	Decision
Intercept	0.28318				
QP	0.17283	0.03136	5.511	0.000	Accepted
PP	0.10009	0.03421	2.926	0.003	Accepted
RE	0.12514	0.03008	4.161	0.000	Accepted
DD	0.11498	0.03137	3.666	0.000	Accepted
BR	0.16027	0.03242	4.943	0.000	Accepted
SI	0.15354	0.03401	4.515	0.000	Accepted
PPE	0.10218	0.03062	3.337	0.001	Accepted

All variables have a P-value lower than 0.05 [27], so they are associated with Choosing Decision (CD) in Table 3 was influenced Brand (BD), Social influence (SI), Quality Perception (QP), Price Perception (PP), Distribution density (DD), Responsiveness (RE), and Promotion Perception (PPE).

## 4.2 Model Evaluation

**Table 4.** Model test

VIF	QP	PP	RE	DD	BR	SI	PPE
	1.511921	1.347734	1.390825	1.267445	1.461458	1.188374	1.265013
Autocorrelation	Durbin-Watson = 1.7789				Test for autocorrelation		
					p-value = 0.04297		
Model evaluation	R-squared 0.6373			F-statistic 50.69		p-value: 0.00000	

When there may be a substantial degree of correlation among the independent variables in regression fashions, the phenomenon of multicollinearity emerges. When the

VIF coefficient is extra than 10, Gujarati and Porter [28] discovered a few signs of multicollinearity within the version. Table four indicates that the VIF (variance inflation aspect) for unbiased variables is much less than 10 [29], indicating that there may be no collinearity between them. Because the p-price = 0.04297 is less than zero.05, the Durbin-Watson test shows that the version in table four has autocorrelation [30]. Table 4 shows that Brand (BD), Social Influence (SI), Quality Perception (QP), Price Perception (PP), Distribution Density (DD), Responsiveness (RE), and Promotion Perception (PPE) have a 63.73% impact on Choosing Decision (CD). The regression equation is statistically large, consistent with the analyses above [31].

$$\begin{aligned} CD = & 0.28318 + 0.17283QP + 0.10009PP + 0.12514RE + 0.11498DD + 0.16027BR \\ & + 0.15354SI + 0.10218PPE \end{aligned}$$

## 5 Conclusions

This paper showed the AIC algorithm for a decision to choose the convenience store chain in the context of the Vietnam market. It brings convenience and quality products to the Vietnamese market with a completely new model. Open 24 h a day and 7 days a week with specialized client service style - warm, fast, clean, neat and safe environment, convenience store chain brings customers with genuine and diversified products including cold drinks, beer, confectionery, snacks, phone cards, food, personal care products and more... This is an advanced business model that has been developed in many countries around the world, bringing added value to the business as well as the development and penetration of the convenience store chain brand in the Vietnamese market.

## References

1. <https://www2.deloitte.com/content/dam/Deloitte/vn/Documents/consumer-business/vn-cb-vietnam-consumer-retail-2019.pdf>
2. Kim, Y.-E.: A study on performance of retail store: focused on convenience store. *J. Distrib. Sci.* **15**(2), 47–51 (2017)
3. Leonard, L.: Measuring grocery stores service quality in Indonesia: a retail service quality scale approach. *Stud. Sci. Res. Econ. Ed* **26**, 32–44 (2017)
4. Rowley, C., Truong, Q.: *The Changing Face of Vietnamese Management*. Routledge, London (2009)
5. Bui, T.T., Nguyen, H.T., Khuc, L.D.: Factors affecting consumer's choice of retail store chain: empirical evidence from Vietnam. *J. Asian Financ. Econ. Bus.* **8**(4), 571–580 (2021)
6. Aperia, T., Persson, C.: Why Are Swedish Retail Corporate Brands so Reputable? A Comparative Study of Sustainable Brand Equity in Sweden, pp. 1028–1031 (2016)
7. Tran, V.D., Le, N.M.T.: Impact of service quality and perceived value on customer satisfaction and behavioral intentions: evidence from convenience stores in Vietnam. *J. Asian Financ. Econ. Bus.* **7**(9), 517–526 (2020)
8. Howard, D.J.: Introduction to special issue: social influence and consumer behavior. *Soc. Influ.* **7**(3), 131–133 (2012)

9. Dressler, W.W.: Cross-cultural differences and social influences in social support and cardiovascular disease. In: Shumaker, S.A., Czajkowski, S.M. (eds.) Social Support and Cardiovascular Disease, pp. 167–192. Springer, Boston (1994). [https://doi.org/10.1007/978-1-4899-2572-5\\_8](https://doi.org/10.1007/978-1-4899-2572-5_8)
10. Gani, A., Orah, A.N.H.: The effect of product quality, service quality and price on customer satisfaction at Loki Store. In: KnE Social Sciences, pp. 116–128 (2021)
11. Suahily, L., Darmoyo, S.: Effect of product quality, perceived price and brand image on purchase decision mediated by customer trust (study on Japanese brand electronic product). Jurnal Manajemen **21**(2), 179–194 (2017)
12. Lee, E.-J., Shin, S.Y.: When do consumers buy online product reviews? Effects of review quality, product type, and reviewer's photo. Comput. Hum. Behav. **31**, 356–366 (2014)
13. Mullins, J.W., Walker, O.C., Boyd, H.W., Larréché, J.-C.: Marketing Management: A Strategic Decision-Making Approach. McGraw-Hill, New York (2013)
14. Anselmsson, J., Bondesson, N.V., Johansson, U.: Brand image and customers' willingness to pay a price premium for food brands. J. Prod. Brand Manag. **23**, 90–102 (2014)
15. Silver, M.: Store choice, store location and market analysis. J. R. Stat. Soc. Ser. A **152**(2), 275 (1989)
16. Meng, X.: An empirical research on selecting location of chain convenience store, pp. 1–6. IEEE (2008)
17. Handayani, R., Rachmawati, M., Kurniawan, R., Roespinoedji, D.: Building of e-customers' loyalty through image department store in digital era (survey of department store in the city of Bandung). Int. J. Psychosoc. Rehabil. **24**(2), 3078–3086 (2020)
18. Sugiono, E., Widjastutik, S.: The effect of product, price and promotion on purchase decision-mediated by customer satisfaction of Oriflame M3 Network community. Open Access Indonesia J. Soc. Sci. **4**(1), 129–142 (2021)
19. Mullin, R.: Promotional Marketing. Routledge, London (2018)
20. Solomon, M.R., Marshall, G.W., Stuart, E.W.: Study Guide for Marketing: Real People, Real Choices. Prentice Hall, Upper Saddle River (2008)
21. Zhang, J.: Do other customers matter? Examining the impact of other customers in retail/service settings. In: Spotts, H.E. (ed.) Revolution in Marketing: Market Driving Changes. DMSPAMS, p. 5. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-11761-4\\_5](https://doi.org/10.1007/978-3-319-11761-4_5)
22. Bollen, K.A.: Overall fit in covariance structure models: two types of sample size effects. Psychol. Bull. **107**(2), 256 (1990)
23. Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L.: Multivariate Data Analysis, vol. 6. Pearson Prentice Hall, Upper Saddle River (2006)
24. Burnham, K.P., Anderson, D.R.: Multimodel inference: understanding AIC and BIC in model selection. Sociolog. Methods Res. **33**(2), 261–304 (2004)
25. Khoi, B.H.: Factors influencing on university reputation: model selection by AIC. In: Ngoc Thach, N., Kreinovich, V., Trung, N.D. (eds.) Data Science for Financial Econometrics. SCI, vol. 898, pp. 177–188. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-48853-6\\_13](https://doi.org/10.1007/978-3-030-48853-6_13)
26. Mai, D.S., Hai, P.H., Khoi, B.H.: Optimal model choice using AIC Method and Naive Bayes Classification. Proc. IOP Conf. Ser. Mater. Sci. Eng. **1088**, 012001 (2021)
27. Hill, R.C., Griffiths, W.E., Lim, G.C.: Principles of Econometrics. Wiley, New York (2018)
28. Gujarati, D.N., Porter, D.: Basic Econometrics. Mc Graw-Hill International Edition, New York (2009)
29. Miles, J.: Tolerance and Variance Inflation Factor. Wiley StatsRef: Statistics Reference Online (2014)

30. Durbin, J., Watson, G.S.: Testing for serial correlation in least squares regression. III. *Biometrika* **58**(1), 1–19 (1971)
31. Greene, W.H.: *Econometric analysis*. Pearson Education India, New Delhi (2003)



# A Simple and Automatic Method to Estimate Lung Volume Based on Thoracic Computed Tomography Images

Tran Anh Vu<sup>1</sup>(✉), Pham Duy Khanh<sup>1</sup>, Hoang Quang Huy<sup>1</sup>, Nguyen Tuan Dung<sup>2</sup>, and Pham Thi Viet Huong<sup>3</sup>

<sup>1</sup> School of Electronics and Telecommunications, Hanoi University of Science and Technology, Hanoi, Vietnam

[vu.trananh@hust.edu.vn](mailto:vu.trananh@hust.edu.vn)

<sup>2</sup> Radiology Center - Bach Mai Hospital, Hanoi, Vietnam

<sup>3</sup> International School, Vietnam National University, Hanoi, Vietnam

**Abstract.** The CT scan of the chest is a useful tool for diagnosing and analyzing thoracic illnesses. For an automated scan of lung information, a number of techniques have been devised. The majority of these approaches need a substantial dataset with accurate information. We make an effort to support fundamental statistics in order to assist medical practitioners in the healing process. The focus of this study is on image processing methods and algorithms for working with image pixels value arrays. Our technique can segment the lung mask, separate the left and right lungs, and compute the volume of each lung portion as well as the overall lung volume. The average accuracy of the segmentation procedure is 99.18%, with a mean F1-score of 98.31%.

**Keywords:** Lung segmentation · Lung volume · CT image

## 1 Introduction

Lungs are key respiratory organs in humans, and a high degree of accuracy in total lung capacity (TLC) is required to detect symptoms of a ventilatory malfunction and give information for diagnostic and therapeutic purposes [1]. According to the Global Impact of Respiratory Disease, over 65 million people worldwide suffer from chronic obstructive pulmonary disease (COPD), with 3 million dying each year [2]. COPD is the third largest cause of death. Furthermore, in the year 2021, our planet is endangered by an epidemic produced by Coronavirus Disease 2019 (COVID-19), which will spread globally. Therefore, early detection and estimation of lung volume is critical for patients, which helps increase the chance of recovery and survival.

For pulmonary imaging, high-resolution computed tomography (CT) is the gold standard. Initially, we can estimate lung capacity in adults [3–5] and children [6–8] using a number of approaches based on chest radiographs. Then, using chest radiographs and thoracic CT scans, the validity of imaging techniques for assessing lung capacity was

investigated and compared to body plethysmography [9]. Organ volume determination [10] is the general ideal for assessing lung capacity in CT images. The lung margin for each axial section is determined by the observer using the cursor on the CT scanner display, and the cross-sectional area of the region of interest is computed and multiplied by the slice thickness to export the lung volume of each cross section. The total volume of the lungs is then calculated by adding the volumes of all the different sections. In [1], CT scan with spirometry gating give similar report of lung volumes taken by plethysmographic methods in patients with an obstructive defect, while on average significantly lower lung volumes were obtained with the dilution technique. Scoliosis is also known to induce restrictive lung disease, which is characterized by a multifactorial reduction in lung volume [11]. We provide a completely automated approach for extracting lungs from CT scans and distinguishing the left and right halves of the lungs in this study. The overall lung capacity, as well as the left and right lung capacities, are determined. At the same time, using the segmented lung mask from these slides, the three-dimensional lung is rebuilt. We get excellent accuracy results in segmentation and volume calculation by combining several image processing techniques. There are numerous steps in the procedure: Scikit-image techniques, K-means clustering, gray-level thresholding. In addition, a novel features selection method is proposed in this work to improve the accuracy of automatic selection. To add conditions in selecting lung area and surpassing distorted range, segmented results are examined and presented frequently.

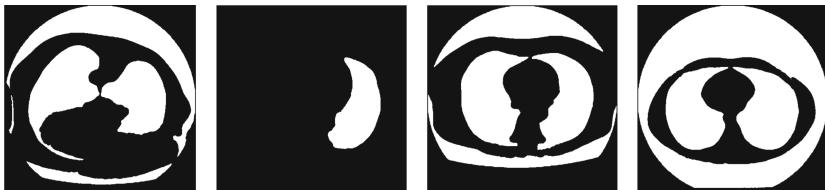
This paper is organized as follows. Section 2 provides detailed description of our method. Section 3 illustrates experimental results, and Sect. 4 concludes the paper.

## 2 Materials and Methods

### 2.1 Databases

#### Lung Mask Segmentation Data

We utilize data from Kaggle to test our segmentation approach. The original 267 TIF pictures with a size of  $512 \times 512$  pixels and their matching 2D binary images with manually segmented were gathered for developing lung segmentation algorithms. However, there are four images that are labelled error with redundant regions and missing half of the lung so that these images are removed from our testing process (Fig. 1).



**Fig. 1.** Four segmented fault images that will not be used.

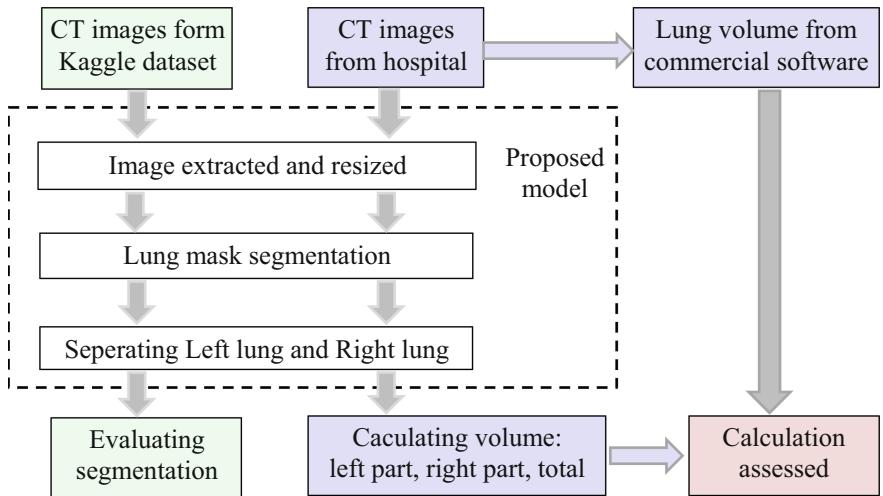
### Lung Volume Dataset

CT volumetric images of the whole chest were chosen for this investigation from a database at Bach Mai Hospital's Radiology Department (Hanoi, Vietnam). This data is made up of scans from eight people who were scanned in the year 2020. The hospital's software provides the volume result of the patients' lungs. In this study, we utilize the total lung, left lung, and right lung volume results from Fig. 2 as the benchmark for evaluating our technique.

	Total	L	LC	LP	R	RC	RP
Vol. [ml]	2784	1183	490	693	1601	761	839

**Fig. 2.** The lung part volume computed by the hospital's software.

## 2.2 Block Diagram



**Fig. 3.** Overall process of the paper's method

The CT scans of the lungs will be examined using the procedure shown in Fig. 3. The DICOM files are read and converted into an array in the first phase, with each element corresponding to a pixel's Hounsfield Unit (HU) value. Then, as described in our earlier study [12], we analyze these arrays to extract the lung mask from original CT images. The left and right halves of the lung mask are then separated. Multiple image processing methods are used to segment the image, and the features selection is generated by repeating our experiment. The whole lung, left, and right sections of the

segmented mask are then utilized to assess our approach and build the three-dimensional model. To evaluate our approach, the computed volume is compared to the data supplied by Bach Mai Hospital. The goal is to create two major approaches for analyzing chest CT images: segmentation, and computing volume, which are the essential information to assist medical practitioners in diagnosing and treating patients.

## 2.3 Extracting and Resampling Original Chest CT Images

### Transform the Data into Pixel Array with HU Value

To begin, the dataset is loaded and stored as NumPy arrays, with each array corresponding to a single image's pixel value. The Pydicom package can read the pixel value, slope, and slice thickness from Dicom pictures, and then the pixel value is converted into the HU value by the function:

$$\text{HU} = \text{pixel\_value} * \text{slope} + \text{intercept} \quad (1)$$

### Resize the Images

The original dataset comprises slice thickness ranging from 1 to 14.9 mm (mm) and pixel size from  $0.645 \times 0.645$  to  $0.789 \times 0.789$  mm. In other words, pixel spacing is defined as the height  $\times$  weight of each pixel in each slice's 2D planar. All of these values are expressed in millimeters (mm). To make it easier to compute the volume, the input DICOM voxel size must be converted to the standard size, which is  $1 \times 1 \times 1$  mm.

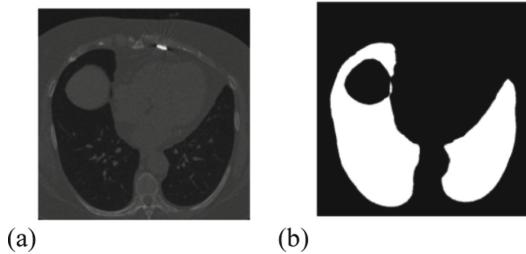
Firstly, we determine the resize factor, which is the original voxel size divided by the usual size ( $1 \times 1 \times 1$ ). After that, we modify the pixel array of the picture using the zoom formula from SciPy image, and the array is zoomed using spline interpolation of the required order. In summary, each patient's entire scan is split into 1 mm thick slices, with each slice divided into  $1 \times 1 \times 1$  mm voxels, or each slice may be thought of as a 2D picture with  $1 \times 1$  mm pixel spacing.

## 2.4 Segmentation of Lung Part

### Extracting Lung Mask

The method for extracting lung mask have been clearly discussed in our previous research [12] and the result of this step is shown in Fig. 4.

To begin, we must choose the appropriate threshold to transform the CT image into a binary image in order to distinguish the foreground (soft tissue/bone) and the background (lung/air). The foreground is presented in dark, while the background is white. The binary image is then labeled to get the information about each region in the image such as: number of regions (marked from index 0), area (number of pixels belonged to each region), row size, column size and boundary box which is the regions range in vertical and horizontal size (min row, min column, max row, max column). The regions which are outside the lung and attached to the edges of image as well as small nodules are then eliminated. To create the completely lung mask, we utilize the fill holes



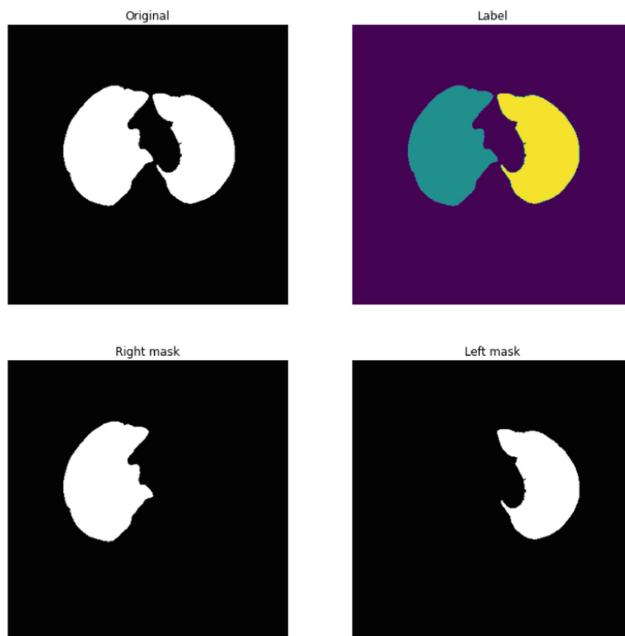
**Fig. 4.** (a) Original thoracic computed tomography image. (b) Lung mask segmented by our method.

image processing approach within the lung. In this research, we have increased the size of retained nodule to eliminate more noise and reduce the radius of filling holes circle to ensure that the lung mask's border is clearly defined and not overextended.

### Left Lung and Right Lung Division

Initially, the segmented lung mask is labeled again by the connected region labeling method. Three scenarios will occur at this stage.

1, *There is no other region accept the background*



**Fig. 5.** Left lung and right lung are not attached and divided directly

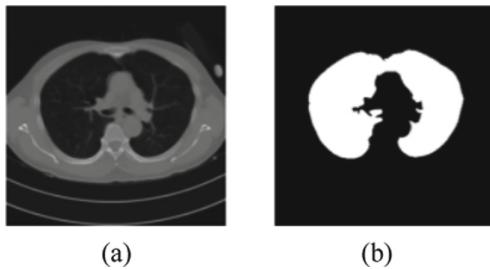
There is no lung mask detected in this situation, so we do not have to process the division step.

## 2, With the lung image, the left lung and right lung regions are clearly differentiated

It's worth noting that this only applies if the labeling phase names precisely three regions: the background, the left lung, and the right lung. As seen in the first left image in Fig. 5, the two sections of the lung are separated. Because the two halves of the lung are not attached, we may use the connected region labeling approach to separate the right and left halves of the lung (Fig. 5(c) and Fig. 5(d)) and then label them with distinct colors as in Fig. 5(b). The rectangular border box of the two regions is compared next, and the region with the smaller min column is designated as the right lung, while the other is designated as the left lung. However, certain small distinct parts of the lung may show early in the scan, increasing the number of identified areas to more than 3. In this example, the right lung mask is defined as an area with a minimum column of less than 4/10 of the total column of the image and a maximum column of less than 7/10. The opposite half of the lung mask is the left mask.

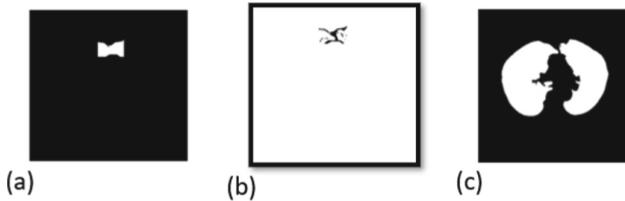
## 3, The lung mask is a fully connected region

It means that we have 2 labeled region: the lung mask and the background. As seen in the first image in Fig. 6, the two sections of the lung are connected.



**Fig. 6.** (a) Orginal CT image. (b) Lung mask segmentation

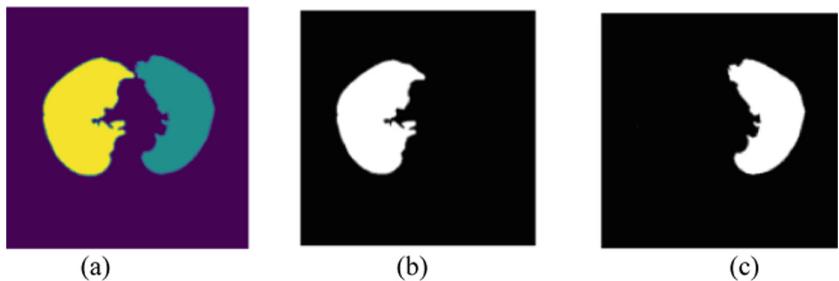
In this situation, we must first locate and highlight the linked region so that it may be easily handled. In our study, we discovered the position of the column, which is the shallowest section of the lung mask. The lung mask's vertical and horizontal ranges are then calculated using the formulas max column minus min column and max row minus min row. Finally, the connected region is determined by the rectangular box with horizontal dimension starting from the min row of the original lung mask to the row index equal to the sum of the min row and 1/10 horizontal range of the lung mask. In terms of the vertical dimension of the connected region box, we receive a 1/10 vertical range of lung mask on both sides of the shallowest lung part's column index. Figure 7(a) shows the connected region of lung mask in Fig. 6.



**Fig. 7.** (a) Connected region of the lung mask. (b) The junction region after applying new threshold. (c) The new lung mask.

According to research [13], the HU value of the junction is much higher than the surrounding section of the lung, thus we need to establish a threshold that will allow us to remove the junction region while keeping the majority of the lung. To determine the threshold, we multiply Fig. 7(a) by the original CT image to obtain the HU value of the linked region, and set the threshold equal to the region's mean HU value. Figure 6(a) and 6(b) are multiplied and a new threshold is used to generate a new lung mask image (Fig. 7(c)). On Fig. 7(c), we apply connected region labeling; if it labels more than two sections, we proceed to the next stage. If there are still two regions labeled, we must lower the HU by a fair amount (in our study, we chose 10 for each drop) and continue the operation until the new lung mask labels at least three regions.

The new lung mask is then labeled, and the division process is repeated as in scenario 2 above, with the outcome shown in Fig. 8.



**Fig. 8.** (a) Labeling new lung mask. (b) Right lung part detection. (c) Left lung part detection.

## 2.5 Measuring Lung Part Volume

Figure 7(c) and 8(b, c) show binary pictures of the mask of the whole lung, left lung, and right lung, with the white region representing the lung portion. The white area is really a collection of pixels with the value 1 and a size of  $1 \times 1$  mm. It indicates that each pixel has a surface area of  $1 \text{ mm}^2$ . As a result, we can use the number of one-value pixels to compute the area of the white mask (in  $\text{mm}^2$ ) on the image. After that, multiply the mask area by the slice thickness to get the volume of the lung portion in a slice. Because all of the slices have been reduced to 1 mm thickness, the volume of the mask

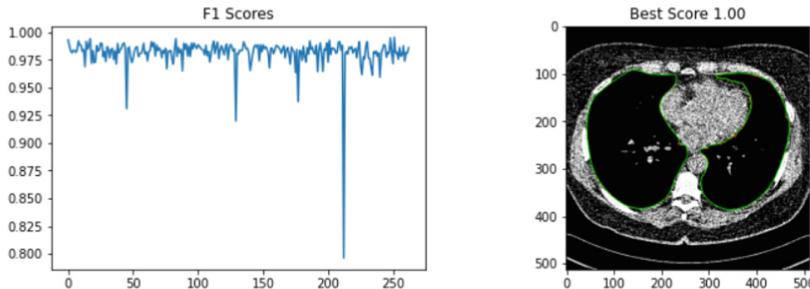
in each slide (in mm<sup>3</sup>) will be equal to the area of the segmented mask in this slice. Our approach calculates the overall volume of the lung, left lung, and right lung by adding the mask volume in each slice.

### 3 Results

#### 3.1 Lung Segmentation

We utilize the Kaggle dataset to compare our results to our prior study and the Kaggle's result to adjust the segmentation stage.

By comparing the pixels value of two images (segmented lung by applying our algorithm and manually segmentation) with the same index we construct the confusion matrix. In the confusion matrix, the true positive value (TP) is the number of pixels value 1 in both ground truth and our segmented image. The number of pixels marked value 0 in ground truth but submitted as 1 by automatic segmentation is counted as false positive (FP). In similarity, true negative (TN) is the situation that pixel value equal to 0 in both images and the case which pixel is predicted as 0 but real value is 1 is confirmed as false negative (FN). These 4 values than be used to calculate the accuracy factors: accuracy, precision, recall and f1-score in individual slices.



**Fig. 9.** Distribution of F1-score of all images and the image achieved the best F1-score result

In comparison to our previous study [12], our segmentation approach has been enhanced and yielded superior results. Figure 9 shows the F1-score distribution across a total of 263 pictures, with an average value of 98.31% compared to 97.89% in previous study and greater than 96.29% F1-scores mean value when using the automated segmentation approach from Kaggle notebooks [14]. The lowest F1 score is 80%, which is greater than the previous research's lowest score (70%). The decrease in F1-score and accuracy appears in the CT scans that the thoracic part is not clearly presented or faded, some of soft tissue and bond part are bounded in the same region in the lung mask. Furthermore, the whole dataset's average accuracy is 99.18%

#### 3.2 Measuring Lung Volume

With 35 Dicom image files corresponding to slice thicknesses ranging from 1 mm to almost 15 mm, we compute the total lung, left lung, and right lung volume of 8 individuals. Among them, our automated volume measurement approach produces excellent

**Table 1.** Lung volume measurement

Patient no.	No. of slices	Slice thickness (mm)	Pixel spacing (mm)	GT_Lung volume (ml)	Lung volume (ml)	GT_Right lung volume (ml)	Right lung volume (ml)	GT_Left lung volume (ml)	Left lung volume (ml)
1	77	5	0.65 × 0.65	4284	4381	2195	2254	2089	2127
	382	1	0.65 × 0.65	4284	4295	2195	2234	2089	2061
2	61	5	0.79 × 0.79	2784	2828	1601	1637	1183	1191
	66	5	0.79 × 0.79	2784	2894	1601	1680	1183	1214
	302	1	0.79 × 0.79	2784	2781	1601	1608	1183	1173
	382	1	0.79 × 0.79	2784	2857	1601	1658	1183	1199
3	67	5	0.70 × 0.70	3886	3900	2041	2041	1845	1859
	333	1	0.70 × 0.70	3886	3864	2041	2019	1845	1845
4	65	5	0.65 × 0.65	3363	3349	1668	1655	1695	1694
	324	1	0.65 × 0.65	3363	3297	1668	1629	1695	1668
5	66	5	0.73 × 0.73	2946	2900	1511	1483	1434	1417
	328	1	0.73 × 0.73	2946	2750	1511	1394	1434	1356
6	78	5	0.65 × 0.65	5222	5239	2800	2815	2421	2424
	390	1	0.65 × 0.65	5222	5191	2800	2745	2421	2446
7	68	5	0.65 × 0.65	3136	3122	1737	1728	1399	1394
	337	1	0.65 × 0.65	3136	3073	1737	1704	1399	1381
8	76	5	0.70 × 0.70	6457	6488	3600	3617	2857	2871

(continued)

**Table 1.** (continued)

Patient no.	No. of slices	Slice thickness (mm)	Pixel spacing (mm)	GT_Lung volume (ml)	Lung volume (ml)	GT_Right lung volume (ml)	Right lung volume (ml)	GT_Left lung volume (ml)	Left lung volume (ml)
	379	1	0.70 × 0.70	6457	6413	3600	3578	2857	2835

results with original 1 and 5 mm slice thicknesses. Table 1 compares the results of our technique to the ground truth (GT) data from the hospital's software.

According to Table 1, our method produces results that deviate by less than 1% from the ground truth value over a half-dozen scans, with two outcomes that are identical to the dataset. 117 ml (7.74% ground truth value) of patient number 5 is the most intradividual. Furthermore, our method's average absolute error for total lung volume is 1.49%, right lung volume is 1.97%, and left lung measurement is 1.21%.

**Table 2.** Calculated Lung volumes distribution over dataset

Lung volume	GT_Total (ml)	Total (ml)	GT_Right lung (ml)	Right lung (ml)	GT_Left lung (ml)	Left lung (ml)
Mean	3874	3868	2084	2082	1789	1786
Standard deviation	1222	1227	679	681	1695	1681
Median	3363	3323	1737	1716	559	563

Table 2 shows the mean, standard deviation, and median values of our approach and the hospital's software computed from Table 1. In both techniques, the mean total left lung capacity (1789 and 1786 ml) is found to be 0.86 times lower than the mean total right lung volume (2084, 2082 ml). The difference between our finding and the ground truth in total lung volume is 6 ml (0.15% mean), 2 ml (0.10% mean) in right lung volume, and 3 ml (0.17% mean) in left lung volume. Furthermore, with differences ranging from 2 to 40 ml, the standard deviation and median values of two techniques are almost identical.

## 4 Discussion

Our segmentation method performs well with both datasets, according to continuous learning and testing. Compare to our previous research, the noise reduction is improved by increasing the size of retained nodule and we also reduce the radius of filling holes circle to ensure that the lung mask's border is clearly defined and not overextended. The data from Kaggle is utilized to test our segmentation approach. The results achieved

an average accuracy of 99.18% indicating a high degree of reliability of the method. We also use real data of 8 patients in the hospital as measured from the CT device. The calculation results for these 8 patients were compared with the results from the commercial software that came with the device, showing an error of less than 2%. The same approach may be used to segment and construct a 3D model of the pulmonary trachea, which aids in the identification of lung such as COPD. In addition, since the whole research process is processing pixel value array, it has potential to be improved for handling different types of medical image such as: MRI, SPECT and PET.

**Acknowledgments.** This research is funded by the Hanoi University of Science and Technology (HUST) under project number T2021-PC-017.

## References

1. Tantucci, C., et al.: Methods for measuring lung volumes: is there a better one? *Respiration* **91**, 273–280 (2016)
2. The Global Impact of Respiratory Disease – Second Edition. Sheffield, European Respiratory Society (2017)
3. Hurtado, A., Fray, W.W.: Studies of total pulmonary capacity and its subdivisions: correlation with physical and radiological measurements. *J. Clin. Invest.* **12**, 807–823 (1933)
4. Bamhard, H.J., et al.: Roentgenographic determination of total lung capacity: a new method evaluated in health, emphysema and congestive heart failure. *Am. J. Med.* **28**, 51–60 (1960)
5. Barrett, W.A., et al.: Computerized roentgenographic determination of total lung capacity. *Am. Rev. Respir. Dis.* **113**, 239–244 (1976)
6. Hiller, E.J., Kirkpatrick, J.A., Huang, N.N.: Radiographic determination of total lung capacity in persons with cystic fibrosis. *J. Pediatr.* **78**, 435–440 (1971)
7. Campbell, S.C.: Estimation of total lung capacity by planimetry of chest radiographs in children 5 to 10 years of age. *Am. Rev. Respir. Dis.* **127**, 106–107 (1983)
8. White, K.S., et al.: Determination of functional residual capacity from digital radiographs of the normal neonatal chest: studies in a rabbit model. *AJR* **156**, 1209–1214 (1991)
9. Schlesinger, A.E., et al.: Estimation of total lung capacity from chest radiography and chest CT in children: comparison with body plethysmography. *AJA* **165**, 151–154 (1995)
10. Schlesinger, A.E., et al.: Volume of the spleen in children as measured on CT scans: normal standards as a function of body weight. *AJR* **160**, 1107–1109 (1993)
11. Daghighi, A., Tropp, H.: Computed tomography lung volume estimation and its relation to lung capacities and spine deformation. *J. Spine Surg.* **5**(1), 132–141 (2019)
12. Vu, T.A., Khanh, P.D., Huy, H.Q., Thanh, H.T., Dung, N.T., Huong, P.T.V.: An improved automatic lung segmentation algorithm for thoracic CT image based on features selection. In: Tran, D.-T., Jeon, G., Nguyen, T.D.L., Lu, J., Xuan, T.-D. (eds.) ICISN 2021. LNNS, vol. 243, pp. 206–214. Springer, Singapore (2021). [https://doi.org/10.1007/978-981-16-2094-2\\_26](https://doi.org/10.1007/978-981-16-2094-2_26). Paper 61
13. Lee, Y.J., Lee, M., Kim, N., Seo, J.B., Park, J.Y.: Automatic left and right lung separation using free-formed surface fitting on volumetric CT. *J. Digit. Imaging* **27**(4), 538–547 (2014). <https://doi.org/10.1007/s10278-014-9680-5>
14. Kaggle Expert: 2D&3D Lung segmentation. <https://www.kaggle.com/azaemon/2d-3d-lung-segmentation>



# Towards Generating RESTful Full Stack Software: A Module Based Domain Driven Design Approach

Duc Minh Le<sup>1</sup>(✉) and Ha Thanh Vu<sup>2</sup>

<sup>1</sup> Department of Information Technology, Swinburne Vietnam, FPT University,  
Hanoi, Vietnam  
[duclm20@fpt.edu.vn](mailto:duclm20@fpt.edu.vn)

<sup>2</sup> Design and Consulting Centre, MobiFone, Hanoi, Vietnam  
[ha.vuthanh@mobilene.vn](mailto:ha.vuthanh@mobilene.vn)

**Abstract.** Domain-driven design (DDD) has widely been used to develop RESTful software in a range of programming language platforms. The use of code generators in these technologies helps significantly increase productivity and achieve large scale software reuse. However, there have been no work that address incremental construction of RESTful full stack software (RFS) at both the module and software levels. In this paper, we propose a generative module-based method for RFS to bridge this gap. We characterise an essential RFS and realise this in a module-based DDD software architecture, named MOSA<sup>R</sup>. Our method takes as input a software configuration and the required software assets and automatically generates a module-based RFS software. We present algorithms for the frontend and backend generation functions. The backend software consist of web service modules. The frontend software consists of modules that are designed with single-page views. Each view can be nested to reflect the containment tree of the module. We demonstrate by implementing the generators for two popular platforms (React and Spring Boot). The evaluation shows that the generators are able to support large software.

**Keywords:** Domain-driven design · Software framework · System modelling · System configuration

## 1 Introduction

The rapid evolution of modern object-oriented programming languages (OOPLs) over the past two decades has greatly contributed to the rise of domain-driven design (DDD) [3, 10, 16] as a dominant force in the software industry. The basic DDD's tenet is clear: the core (a.k.a “heart”) of software is the domain model and, thus, effectively constructing this model is a central issue. The addition of the *annotation*<sup>1</sup> construct in OOPL has brought new modelling capabilities [7, 12] that help ease the development of both domain model and software.

<sup>1</sup> What Java calls *annotation* is *attribute* in C#.

DDD has widely been used to develop web-service-based software in OOPs (e.g. C# [17] and Java [11]). In particular, a number of software frameworks (most notably ApacheIsIs [1], OpenXava [5] and jHipster [9]) have been developed to make RESTful [4] software development in DDD more productive. The use of code generators in these frameworks helps significantly increase productivity and achieve large scale software reuse. However, there have been very few methods that focus on RESTful full stack software (RFS) and none of these methods support incremental software generation at both the module and software levels. In this paper, we propose a generative module-based method to construct RFS to bridge this gap. We first characterise RFS and realise this in a software architecture, named **MOSA<sup>R</sup>**, which we extend from our previously-developed module-based MOSA architecture for DDD. Our method takes as input a software configuration, written in an annotation-based domain specific language, and the required software assets and generates a module-based RFS software as the output. The backend software consist of web service modules. The frontend software consists of modules that are designed with single-page views. Each view is nested to support the module's containment tree. We presents the algorithms for two generator functions for the front- and back end software. We demonstrate our method by an implementation of the generators in two popular programming platforms: React (frontend) and Spring Boot (backend). Our evaluation shows that the generators are scalable to support large software.

The rest of the paper is structured as follows. Section 2 explains a motivating example and presents some key background concepts about MOSA. Section 3 discusses the characteristics of RFS and presents MOSA<sup>R</sup>. Section 4 discusses our RFS generator functions. Section 5 discusses tool support and evaluation. Section 6 reviews the related work and Sect. 7 concludes the paper.

## 2 Background on MOSA Architecture

In this section, we present a motivating software example named COURSEMAN and a number of key concepts about the MOSA architecture.

**Example: CourseMan.** To illustrate the concepts presented in this paper, we adopt the software example, named course management (COURSEMAN), that we used in previous works [7,8]. This example is scoped around a domain model whose elements are represented by the following fundamental UML [13] meta-concepts: class, attribute, operation, association, association class and generalisation. The COURSEMAN's domain model consists of six domain classes and associations between them. Class **Student** represents the domain concept Student. Class **CourseModule** represents the CourseModules that are offered to students. Class **Address** represents the addresses where Students live as they undertake their studies. The study arrangement involves enrolling each **Student** into one or more **CourseModules**. Class **Enrolment** represents the enrollment records.

**MOSA Architecture Model.** MOSA [6,7] is a module-based, MVC software architecture for DDD, which consists of three layers: domain model (the core),

module and software. A **domain model** essentially consists of a set of domain classes and the associations between them. We define this model by an aDSL named DCSL. This language expresses the design space of a domain class in terms of *state space* and *behaviour space*. The state space, in particular, consists of all the domain fields. A software is incrementally constructed from its modules. A module is a ‘little’ MVC-based software, whose model is a domain class. Modules are instantiated from a *module class*. A module’s structure is constructed from a *rooted containment tree*, whose non-root nodes are *descendant modules* that own the domain classes associated (directly or indirectly in an *association chain*) to the module’s owned domain class.

We adapted the *software product-line development* (SPD) approach [2] to generatively construct software and its modules from their configurations. We use *annotation-based domain-specific languages* (aDSLs) [12] to express these configurations. To generate the module class we defined an aDSL, named MCCL, to express *module class configuration* (MCC). The MCC is automatically generated from a domain class and then customised to satisfy the required module design specification. In this paper, we use the term MCC synonymously to module class. To generate a software class from the modules we defined another aDSL, named SCCL, to express the *software configuration classes* (SCCs). Among the essential properties of an SCC is **modules**, which specifies the set of MCCs of the modules that make up the software class.

For example, the aforementioned COURSEMAN domain model results in six MCCs, each of which owns one domain class. The view of **ModuleStudent**, for instance, has 2 view fields that represent the two domain fields: **Student.id** and **name**. In addition, this view contains two subviews of the following level-1 nodes of **ModuleStudent**’s containment tree: **ModuleAddress** and **ModuleEnrolment**.

### 3 MOSA<sup>R</sup>: Extending MOSA for RFS Software

Given that REST-style architecture [4] is the underpinning architecture of the modern Web. A **RESTful full stack (RFS) software** is a web-based software whose frontend (FE) and backend (BE) fully comply with a REST-style architecture. Figure 1 presents our proposed MOSA<sup>R</sup> architectural model, which extends MOSA for RFS. MOSA<sup>R</sup> uses the same core domain model layer as MOSA. However, the module and software layers are restructured to support frontend, backend and RESTful API.

Featurewise, in this paper we will focus on the following essential component categories of FE and BE that were reported by Taivalsaari et al. [15]:

1. FE components: web frontend components that are constructed in HTML/CSS/JavaScript, using a frontend framework (e.g. Angular, React, Vue)
2. BE components: database-enabled domain-logic-processing components that are constructed in at least one popular backend frameworks
3. RESTful API components: to enable REST-style communication between FE and BE

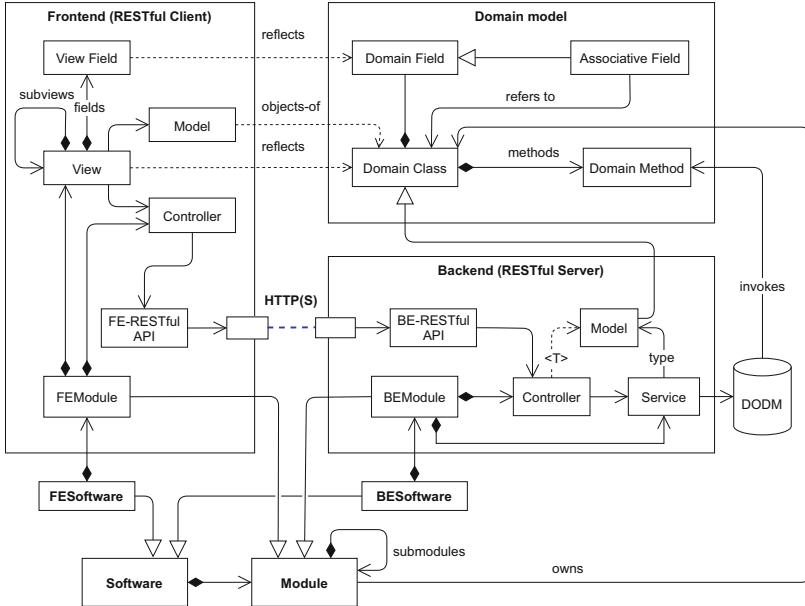


Fig. 1. The MOSA<sup>R</sup> architecture model.

4. Independent platforms: the FE and BE components can be constructed and deployed to different language platforms

**FrontEnd Software.** **FESoftware** is a subtype of **Software** that consists in a set of FE modules. Conceptually, an **FEModule** is a subtype of **Module** that uses frontend technologies to construct the **View**, **Model** and **Controller** components. In modern frontend frameworks, **FEModule** consists of a **Model** component, which is a *mirror representation* of the domain class's state space. That is it consists in the same core structure as the state space, but is expressed in terms of a frontend language. In MOSA<sup>R</sup>, **FEModule**'s **Model** captures part of the **View**'s state that pertains to the domain objects manipulated by the user on the view. Conceptually, **Controller** is a cohesive set of event-handling methods that are invoked when certain events concerning **Model** are fired. As shown in Fig. 1, **Controller** uses a frontend RESTful API to connect to the backend.

**Single Page View.** We adapted the **single page view (SPV)** [14] design approach (which is popular among FE frameworks) to design the **FEModule**'s **View**. In this design, **View** consists of view components which are ‘micro modules’. A **view component** constructs the view elements, handles the controller functionality related to these elements and maintains the domain objects that form part of the view state. More formally, the **View** of a module  $m$  is the tuple  $\langle c', V_I, \mathcal{V}_C \rangle$ , where  $c'$  is the front-end representation of  $m$ 's domain class.  $V_I$  is the **index** or ‘home’ component presented to the user. This component’s content

is constructed from other components specified in the set  $\mathcal{V}_C$ . Each component in this set is responsible for providing one view aspect. Within the scope of this work, we consider two basic view aspects: form ( $V_F$ ) and browser ( $V_B$ ).  $V_F$  is the **form** component that presents an input form for the user to create a new domain object or to edit an existing one.  $V_B$  is the **browser** component that basically presents a table of the existing domain objects for the user to browse. Both  $V_F$  and  $V_B$  allow the user to select an object to view its details on  $V_F$ . It also allows the user to delete objects. We will discuss in Sect. 4.1 an algorithm for generating this type of view.

**BackEnd Software.** **BESoftware** is a subtype of **Module** that consists in a set of BE modules. An **BEModule** is a subtype of **Module** that uses backend technologies to handle the domain-specific logic (in response to **FEModule**'s requests). In the context of DDD and REST, **BEModule**'s functionality consists in a set of web services, which are called upon by the controller component. Unlike **FEModule**, however, **BEModule** typically does not include the **View**.

We thus define a **BEModule**  $m$  as the tuple  $\langle c, \mathcal{C}, \mathcal{S} \rangle$ , where  $c = m$ 's domain class,  $\mathcal{C}$  is the set of **Controllers** of the modules in  $m$ 's containment tree and  $\mathcal{S}$  is the set of **Services** that constitute  $m$ 's functionality. **Controller** basically acts as a dispatcher that forwards requests to the relevant **Services** to handle. This pattern also applies to **Controllers** of the descendant modules that participate in the containment tree of a **BEModule**. Similar to MOSA, both **Controller** and **Service** are parameterised with the domain class of the module, so that they can handle the domain objects of this class. In  $\text{MOSA}^R$ , the set  $\mathcal{S}$  contains at least the repository service [3], which uses the MOSA's DODM component to manage objects and their storage.

**RFS Configuration.** We extend SCC to define  $\text{SCC}^R$  for  $\text{MOSA}^R$ .  $\text{SCC}^R$  includes an annotation, named **RFSDesc**, which specifies the configuration for RFS generation.

**Listing 1.1.** An illustration of **RFSDesc**'s core structure

---

```

1 @RFSDesc(
2   domain = "courseman",      // software domain name
3   feLangPlatform = LangPlatform.REACT, // target frontend language platform
4   beLangPlatform = LangPlatform.SPRING, // target backend language platform
5   outputPath = "src/example/java" // output path for the generated components
6 )

```

---

In this work, we define **RFSDesc** in terms of 4 basic properties. Other properties would be added in the future to accommodate the configuration of new  $\text{MOSA}^R$ 's features. Listing 1.1 gives an example that illustrates **RFSDesc**. The embedded comments provide descriptions of the properties.

## 4 RFS Generation

In this section, we discuss an adaptation of our template-based software generation method [8] for RFS. A unique feature of our method is to use the same core domain model as the input and to incrementally and automatically generate

the FESoftware and BESoftware from their modules. We denote by  $\text{SCC}^R$  the set of all the  $\text{SCC}^R$ 's concerning a software domain. Further, we denote by  $\mathcal{A}$  a Cartesian product of sets of domain and software assets that are referenced (directly or indirectly) by an  $\text{SCC}^R$ . Among the **required assets** include the domain model and the FE and BE component templates.

## 4.1 Generating FrontEnd Software

We adopt the SPV design for FE (see Sect. 3) and define the FE generation function:  $\text{FEGen}: \text{SCC}^R \times \mathcal{A} \rightarrow \text{FESoftware}$ . This takes as input an  $\text{SCC}^R$  with the required assets  $A \in \mathcal{A}$  and generates as the output an **FESoftware** consisting of **FEModules**, whose views are designed with SPV.

---

**Alg. 1: FEGen: FESoftware as a set of single-page FEModules**

---

```

1  input :  $s: \text{SCC}^R, A \in \mathcal{A}$ 
2  output:  $W_F: \text{FESoftware}$ 
3   $W_F \leftarrow \emptyset$                                 /* the output FESoftware */
4  foreach  $m_c \in \text{mccs}(s)$  do
5       $V_F \leftarrow \text{genViewForm}(m_c, A)$            /* Form view */
6       $V_B \leftarrow \text{genViewBrowser}(m_c, A)$         /* Browser view */
7       $V_I \leftarrow \text{genViewIndex}(V_F, V_B, A)$        /* Index view */
8       $W_F \uplus \{(V_I, V_F, V_B)\}$ 
9  return  $W_F$ 
10 Function  $\text{genViewForm}(m_c: \text{MCC}, V_F: \text{ViewForm}, A: \mathcal{A})$ 
11    $V_F \leftarrow \text{loadTemplate}(\text{"view_form"}, A)$     /* init form template */
12    $P \leftarrow \emptyset$                                 /* view components set */
13    $c \leftarrow m_c.\text{domainClass}$ 
14   foreach  $f \in c.\text{fields}$  do
15      $f_d \leftarrow \text{fieldDef}(m_c, f)$ 
16     if  $\text{isNonAssoc}(f)$  then /* generate view field */
17        $v \leftarrow \text{genViewField}(f, f_d, A)$  ;  $P \uplus \{v\}$ 
18   else /* generate subview */
19      $v \leftarrow \text{genSubView}(f, f_d, A)$  ;  $P \uplus \{v\}$ 
20   rewrite( $V_F$ , "title",  $m_c.\text{title}$ ) ; rewrite( $V_F$ , "view.comps",  $P$ )
21   return  $V_F$ 

```

---

Algorithm 1 presents the algorithm that realises function **FEGen**. The **foreach** loop in the top half of the algorithm creates, for each MCC ( $m_c$ ), an **FEModule** as a single-page component consisting in three views: form view ( $V_F$ ), browser view ( $V_B$ ) and the index view ( $V_I$ ). The three views are generated at lines 3–5 using pre-defined view templates. Function **genViewBrowser** is a simple adaptation of **genViewForm**, while function **genMainView** basically involves rewriting a **ViewIndex** template using the source codes of the two component views. The pseudocode for **genViewForm** is presented in the bottom half of Algorithm 1. It basically implements a template-based code generation logic. After loading the template at line 9, its **foreach** loop iterates over the domain fields of the module's domain class ( $c$ ) to create a view component ( $v$ ) for each field. This can be either a normal field (line 15) or a subview of a descendant module (line 17). After all the view components have been created, they are inserted into  $V_F$  (by function **rewrite** at line 18), replacing the template variable “**view.comps**”.  $V_F$ 's title is also rewritten (line 18) using title in  $m_c$ .

### Form: Student

The form consists of several input fields and buttons. At the top right are three buttons: 'Main' (blue), 'Browse' (light blue), and 'Delete' (red). To the right of these are two search input fields: one containing 'S2021' and another with a magnifying glass icon. Below these are five data entry fields: 'Id' with value 'S2021', 'Name' with value 'Le Minh Duc', 'Gender' with value 'Male', 'Dob' with value '10/12/2000', and an 'Address' field divided into 'Address ID' (containing '1') and 'Address' (containing '1-Hanoi'). To the right of the address fields are two buttons: 'Unlink' (red) and a dropdown arrow. At the bottom left is a green button labeled 'Form: Enrolment' with a dropdown arrow.

**Fig. 2.** The generated form view of `ModuleStudent`

For example, Fig. 2 shows the rendered form view of `ModuleStudent`, whose source code is generated by function `FEGen` for the React.js platform<sup>2</sup>. Due to space restriction, we do not include the source code in this paper. Refer to our extended paper for details.

## 4.2 Generating BackEnd Software

We define the BE generation function:  $\text{BEGen}: \text{SCC}^R \times \mathcal{A} \rightarrow \text{BESoftware}$ , which takes as input an  $\text{SCC}^R$  with the required assets  $A \in \mathcal{A}$  and generates a **BESoftware** as the output. Algorithm 2 presents the algorithm of this function. The main **foreach** loop creates a **BEModule** from each MCC  $m_c$  of the input  $\text{SCC}^R$ . Line 3 uses the function `genRepoServiceType` to generate a repository service class (using a template available in  $A$ ) and adds it to a services set ( $\mathcal{S}$ ). Calls to other functions can be added here to generate additional services.

---

<sup>2</sup> <https://reactjs.org>.

**Alg. 2:** BEGen: BESoftware as a set of BEModules

---

```

1 input :  $s : \text{SCC}^R, A \in \mathcal{A}$ 
2 output:  $W_B : \text{BESoftware}$ 
3  $W_B \leftarrow \emptyset$  /* the output BESoftware */
4 foreach  $m_c \in \text{mccs}(s)$  do
5    $S \leftarrow \text{genRepoServiceType}(m_c, A); \mathcal{S} \leftarrow \{S\}$  /* RESTService */
6    $C \leftarrow \text{genControllerType}(m_c, A); \mathcal{C} \leftarrow \{C\}$  /* RESTController */
7    $T \leftarrow \text{containmentTree}(m_c)$ 
8   if  $T \neq \emptyset$  then /* gen. controllers of descendant mods in T */
9     foreach  $m_d \in T \wedge m_d \neq T.\text{root}$  do
10     $C_d \leftarrow \text{genDescendantControllerType}(m_c, m_d, A)$ 
11     $\mathcal{C} \uplus \{C_d\}$ 
12    $c \leftarrow m_c.\text{domainClass}$ 
13    $W_B \uplus \{(c, \mathcal{C}, \mathcal{S})\}$ 
14 return  $W_B$ 

```

---

Similarly, line 4 uses the function `genControllerType` to generate the corresponding controller class for  $m_c$  and adds it to the controller set ( $\mathcal{C}$ ). This set may also contain the controller classes of the descendant modules ( $m_d$ ) in the containment tree  $T$  of  $m_c$ . These controllers ( $C_d$ ) are generated (if not already) by the function `genDescendantControllerType` in the nested **foreach** loop that starts at line 7.

The three aforementioned generator functions basically involve creating a subtype of a pre-defined service or controller type and overriding the default methods already implemented in the supertypes. For CRUD functionality, this results in full generation.

**Listing 1.2.** The controller of ModuleStudent

---

```

1 @RestController()
2 @RequestMapping(value = "/students")
3 public class StudentController extends DefaultRestfulController<Student> {
4   @GetMapping() // paginated list of objects
5   public Page getEntityListByPage(PagingModel arg0) {
6     return super.getEntityListByPage(arg0);
7   }
8   @PostMapping() // Create object
9   public Student createEntity(@RequestBody() Student arg0) {
10    return super.createEntity(arg0);
11  }
12  @GetMapping(value = "/{id}")
13  public Student getEntityById(Identifier arg0) {
14    return super.getEntityById(arg0);
15  }
16  @PatchMapping(value = "/{id}") // Update object
17  public Student updateEntity(Identifier arg0, @RequestBody() Student arg1) {
18    return super.updateEntity(arg0, arg1);
19  }
20  @DeleteMapping(value = "/{id}") // Delete object
21  public void deleteEntityById(Identifier arg0) {
22    super.deleteEntityById(arg0);
23  }
24 }

```

---

For example, Listing 1.2 shows the code of the Controller of `ModuleStudent` which is generated by `BEGen` for the Spring Boot<sup>3</sup> platform. This Controller handles the RESTful requests from the FE module shown earlier in Fig. 2. To conserve space, we show in the listing only the essential part of the generated code. The complete source code is given in our extended paper. The

<sup>3</sup> <https://spring.io/projects/spring-boot>.

controller class is tagged with `@RestController`, which is treated by Spring Boot as RESTful controller component. Similarly, the service class is tagged with `@Service`. The CRUD requests are mapped to the methods that are presented in the listing with explanatory comments. Note that both the controller and service classes support a method to handle the pagination request. This is used by the frontend’s browser view to obtain a fixed-sized collection of objects to present to the user.

## 5 Tool Support and Evaluation

We implemented our method as a module in our previously-developed software framework, named JDA [8]. The implementation is available in a GitHub repository<sup>4</sup> and is explained in more detail in our extended paper, accessible from this repository. Our implementation supports the React platform for the frontend and Spring Boot platform for the backend.

**Performance Analysis.** We analyse the time complexities of two generator functions (`FEGen` and `BEGen`) to show that they run in **quadratic time**. This is scalable to support software with large domain models. Due to space restriction, we do not present the proofs in this paper. Refer to our extended paper for details.

## 6 Related Work

Our work in this paper is positioned at the intersections among three areas: SPD, DDD and RFS development.

**SPD.** Generative RFS development is a specialisation of the traditional SPD [2] for two components: frontend and backend software generation. Our work is unique in the use of a core domain model and a module-based software configuration to automatically generate both the frontend and backend. We support fullstack configuration whereby the frontend and backend are expressed in different language platforms. The generated RFS conforms to the MOSA<sup>R</sup> architecture which we also defined in the paper.

**DDD.** Our work in this paper contributes to a current research theme [7,8] in enhancing the DDD method [3,10,16] with module-based software generation capabilities, especially using annotation-based DSLs [12]. Our contribution extends our most recent work [7] with an extended software architecture for RFS (MOSA<sup>R</sup>) and a software generation method for this type of software. Comparing to recent DDD frameworks [1,5,9], our method is unique in using an incremental, module-based generation method. In this, software modules are first generated from the domain model and then combined to generate the software.

**RFS Development.** Our method is similar to three DDD software frameworks ApacheIsIs [1], OpenXava [5] and jHipster [9] in the support for RFS

---

<sup>4</sup> <https://github.com/jdomainapp/jda-mosar>.

software generation. Although these frameworks provide more RFS functionality than currently provided in our method, they do not follow an incremental module-based software generation approach as in ours. Consequently, developers of these frameworks cannot configure the module and software structure and behaviour before generation. In our method, developers can configure a module using MCCL and the software using SCCL. Further, this can be performed for both the frontend and backend software subsystems of RFS.

## 7 Conclusion

In this paper, we proposed a solution to address the problem of generative, module-based RESTful full stack software construction. We adopted our incremental, module-based DDD approach and MOSA architecture to define an architecture named MOSA<sup>R</sup> for RFS. We extended our work on generative software development in MOSA to perform RFS generation in MOSA<sup>R</sup>. The enhanced generator includes two components: one for the frontend software and the other for the backend one. Both software subsystems can be configured and executed by different programming platforms. We defined an RFS software configuration to address the core requirements of RFS generation in MOSA<sup>R</sup>. We presented two algorithms for the frontend and backend software generation functions. The frontend function, in particular, generates the frontend modules following the popular single-page component design. We evaluated both algorithms to show that they are scalable to generate large software.

In future work, we plan to enhance our implementation to support other frontend and backend technologies. We would also like to extend our design further to tackle other full stack challenges [15] posed by modern Internet technologies, including micro-services architecture and cloud and serverless computings.

**Acknowledgement.** The authors would like to thank Binh Hai Do for his physical design and implementation of the first prototype of the RFS generation functions using React.js and Spring Boot. These were carried out in Spring 2021 as part of his graduation thesis project at the Faculty of IT, Hanoi University.

## References

1. Apache-S.F: Apache Isis (2021). <http://isis.apache.org/>
2. Apel, S., Batory, D.S., Kästner, C., Saake, G.: Feature-Oriented Software Product Lines - Concepts and Implementation. Springer, Heidelberg (2013). <https://doi.org/10.1007/978-3-642-37521-7>
3. Evans, E.: Domain-Driven Design: Tackling Complexity in the Heart of Software. Addison-Wesley Professional, Boston (2004)
4. Fielding, R.T., Taylor, R.N.: Principled design of the modern Web architecture. ACM Trans. Internet Technol. **2**(2), 115–150 (2002)
5. Paniza, J.: OpenXava (2021). <http://openxava.org/>
6. Le, D.M., Dang, D.H., Nguyen, V.H.: On domain driven design using annotation-based domain specific language. Comput. Lang. Syst. Struct. **54**, 199–235 (2018)

7. Le, D.M., Dang, D.H., Nguyen, V.H.: Generative software module development for domain-driven design with annotation-based domain specific language. *Inf. Softw. Technol.* **120**, 106–239 (2020)
8. Le, D.M., Dang, D.H., Vu, H.T.: jDomainApp: a module-based domain-driven software framework. In: Proceedings of 10th International Symposium on Information and Communication Technology (SOICT), pp. 399–406. ACM (2019)
9. Raible, M.: JHipster - Full Stack Platform for the Modern Developer! (2021). <https://www.jhipster.tech/>
10. Millett, S., Tune, N.: Patterns, Principles, and Practices of Domain-Driven Design. Wiley, New York (2015)
11. Nair, V.: Practical Domain-Driven Design in Enterprise Java: Using Jakarta EE, Eclipse MicroProfile, Spring Boot, and the Axon Framework, 1st edn. Apress, New York (2019)
12. Nosál', M., Sulír, M., Juhár, J.: Language composition using source code annotations. *Comput. Sci. Inf. Syst.* **13**(3), 707–729 (2016)
13. OMG: Unified Modeling Language version 2.5 (2015). <http://www.omg.org/spec/UML/2.5/>
14. Scott, E.: SPA Design and Architecture: Understanding Single Page Web Applications, 1st edn. Manning, Shelter Island (2015)
15. Taivalsaari, A., Mikkonen, T., Pautasso, C., Systä, K.: Full stack is not what it used to be. In: Brambilla, M., Chbeir, R., Frasincar, F., Manolescu, I. (eds.) ICWE 2021. LNCS, vol. 12706, pp. 363–371. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-74296-6\\_28](https://doi.org/10.1007/978-3-030-74296-6_28)
16. Vernon, V.: Implementing Domain-Driven Design, 1st edn. Addison-Wesley Professional, Upper Saddle River (2013)
17. Zimarev, A.: Hands-On Domain-Driven Design with .NET Core: Tackling complexity in the heart of software by putting DDD principles into practice. Packt Publishing, Birmingham (2019)



# Time Series Prediction Based on Consecutive Neighborhood Preserving Properties of Matrix Profile

Thi Dung Tran<sup>1</sup>, Ngoc Dung Bui<sup>2</sup>✉, and Siew Ann Cheong<sup>3</sup>

<sup>1</sup> Campus in Ho Chi Minh City, University of Transport and Communications, 450-451 Le Van Viet Stress, Tang Nhon Phu A Ward, Thu Duc City, Ho Chi Minh City, Viet Nam  
dungtt\_ph@utc.edu.vn

<sup>2</sup> University of Transport and Communications, 3 Cau Giay Street, Ha Noi, Viet Nam  
dnbui@utc.edu.vn

<sup>3</sup> Nanyang Technological University, 21 Nanyang Link, Singapore 637371, Singapore  
cheongsa@ntu.edu.sg

**Abstract.** Scientists are recently paying attention to the time series prediction due to various practical applications. Some of the suggested methods include self-regression model, artificial neural network model, linear prediction model, and moving average model. Yet they remain obstacles such as time consuming, lack of practical examples, and complication. Our recent work tried to solve these problems by a modified approach in which a time series prediction used matrix profile with adjacent pairs. Also, the optimal prediction parameters have been suggested for better prediction. The application of Consecutive Neighborhood Preserving (CNP) properties helps improve accuracy and reduce calculation time experimentally.

**Keywords:** Time series · Motif · Time series forecast · Consecutive neighborhood preserving

## 1 Introduction

Time series forecasting is quickly becoming an indispensable need for in the face of the explosion on human activities information. Prediction will provide the necessary inputs for planning and it can be said that, without prediction science, outlines of our future intentions would not be convincing. Thus far, time series forecasting is heavily utilized for stock price prediction [1], petroleum business forecast [2], university enrollment forecast [3], and forecasting population growth and change [3].

Many time series prediction methods have been proposed by researchers in recent years. In 1979, Bruce Bowerman, Richard O Connell, and Anne Koehler used statistical approaches to highlight the importance of forecasting in marketing, finance, personnel management, production scheduling, process control, and strategic management [4]. Support vector machines (SVMs) are introduced for the prediction of financial time-series by Kim in 2003 [5]. In 2007, Lora used the closest number of neighbor techniques

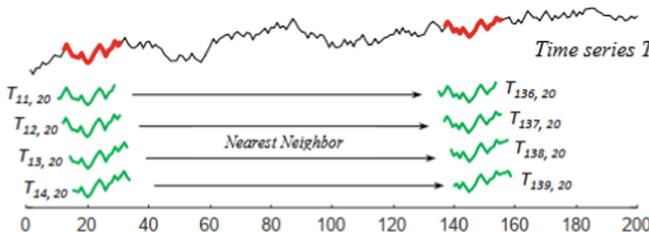
to forecast data [6]. In 2009, Jiang proposed a method of forecasting stock time series based on motif information [7]. Yolcu et al. [8] proposed a combined robust technique for fuzzy time series in 2017, examining how the approaches' prediction performance is impacted by the outlier. Silva et al. present a Non-Stationary Fuzzy Time Series (NSFTS) approach in 2020 [9], which uses time-varying parameters tailored from the data distribution.

A novel approach for time series prediction is the matrix profile method. Matrix profile is a distance vector of pairs of motifs, or adjacent pairs, proposed by Dr. Keogh's Lab [10]. Matrix profile has implications for time series motif discovery, classification, visualization, clustering, rule discovery, etc. [10]. However, there are no reports on using matrix profile for prediction. This is the method of finding the nearest neighbor of each subsequence in time series. Based on the closest proximity property, one can make predictions about the continuing values in the time series. Beside matrix profile, we compute the spread of amplitude to make a more accurate prediction. The method tested on the neuroscience data set gave better results in terms of accuracy and timing.

## 2 Methodology

### PreSCRIMP Algorithm

The index of the subsequence in the time series  $T$  is  $[1, 2, 3, \dots, n, m + 1]$ .  $I$  is the matrix profile index. The matrix indexes, by definition, have intervals of consecutive values. Consecutive Neighborhood Preserving (CNP) refers to the index values that correspond to consecutive values in  $I$  as well as the nearest neighbors. According to the CNP, if  $i$  and  $j$  are neighbors, the high probability that  $i + 1$  is likewise a neighbor of  $j + 1$ . Figure 1 shows the subsequences 11, 12, 13, and 14 that are closest to the 136, 137, 138, and 139.



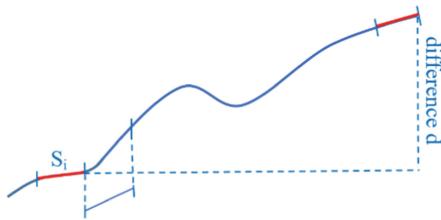
**Fig. 1.** Consecutive neighborhood preserving properties [11]

The objective is to present the preSCRIMP preprocessing technique, which uses the CNP property to find a matrix profile that approximates the run time significantly faster than the SCRIMP algorithm.

The matrix profile will be refined using the SCRIMP algorithm after executing the preSCRIMP algorithm until the algorithm converges. We can stop the SCRIMP++ algorithm at any moment, rather than having to wait for it to finish properly, and the result will be an estimated Matrix profile and the motif of this time series. Based on the results

of SCRIMP++ algorithm, we can use matrix profile to perform the prediction problem. Once the Matrix profile has been calculated, we can see the most similar string pairs. Then, based on the CNP property in the preSCRIMP algorithm, we can predict the sequence appearing behind the last series in the time series, which is also the time series needed to predict.

However, because the sub-chains have different amplitudes in different positions, we need to put prediction time series with the amplitude closed to the amplitude of the actual one to make the prediction time series. To perform the normalization of the amplitude, we do as follow. First, we subtract the last point in the time series from the last point of the neighboring time series to have a value of the deviation of these 2 points. Then, we find the predicted series and normalize the amplitude by summing up every point in the next series of the neighboring series and the corresponding deviation, as shown in Fig. 2.



**Fig. 2.** Estimate the predicted value by normalization of the amplitude

### 3 Experimental Results

The neuroscience dataset we used in this work were data sets about human neuroscience from [12], which is the supporting webpage for the ICDM paper. In this experiment, with the neuroscience dataset, we propose the following prediction parameters: prediction\_string\_length = subsequence\_length/4. The experimental on the proposed method will be implemented in two cases.

+ Case 1: Time series length: 200 points, subsequence length: 40 points, prediction string length: 10 points.

**Table 1.** Results of case experiment 1.

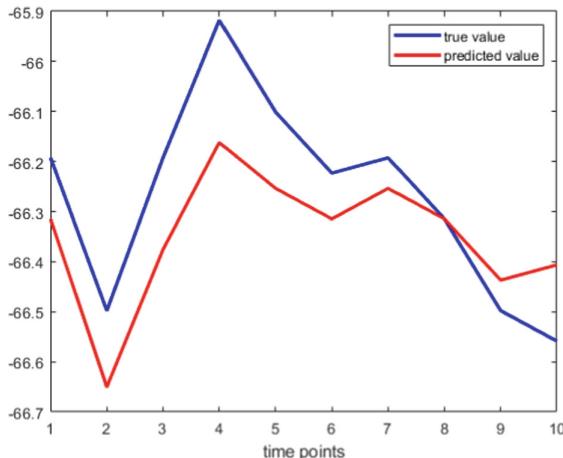
Points	Real value	Predicted value	Difference
201	-66.1926	-66.3147	0.1221
202	-66.4978	-66.6504	0.1526
203	-66.1926	-66.3757	0.1831

(continued)

**Table 1.** (continued)

Points	Real value	Predicted value	Difference
204	-65.9180	-66.1621	0.2441
205	-66.1011	-66.2537	0.1526
206	-66.2231	-66.3147	0.0916
207	-66.1926	-66.2537	0.0611
208	-66.3147	-66.3147	0
209	-66.4978	-66.4368	0.061
210	-66.5588	-66.4063	0.1525

The value discrepancy between the anticipated and real series is not too great, according to the data in Table 1. The largest difference in this case is approximately 0.24. Figure 3 shows the difference of the actual series (blue) and the predicted series (red). In Fig. 3, most of the time, the predicted values increase and decrease linearly, but less than the real values.



**Fig. 3.** Results of experimental series of predictions in case 1 (blue: true time series, red: predicted time series).

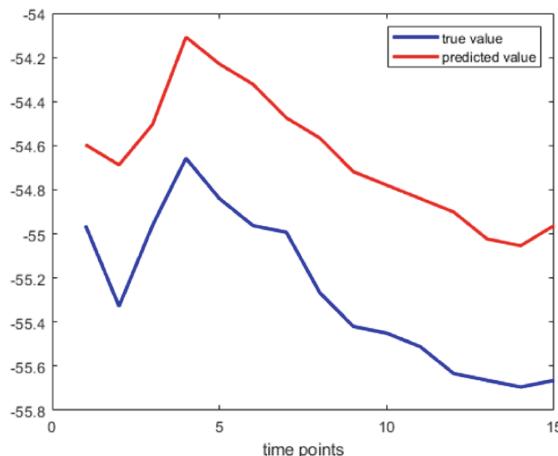
+ Case 2: Time series length: 500 points, subsequence length: 60 points, predicted string length: 15 points.

According to the results in Table 2, the difference between predicted series and actual series is not too large. The largest difference is 0.7. Figure 4 shows the visualization of the true (blue) and predicted values (red). The predicted values increase and decrease linearly compared to the real values, but the predicted value is larger than the real value.

Through two experimental cases with different time series length, substring length and predictive string length, we got different prediction results. The different time series

**Table 2.** Results of case experiment 2

Points	Real value	Predicted value	Difference
501	-54.9622	-54.5959	0.3663
502	-55.3284	-54.6875	0.6409
503	-54.9622	-54.5044	0.4578
504	-54.6570	-54.1077	0.5493
505	-54.8401	-54.2297	0.6104
506	-54.9622	-54.3213	0.6409
507	-54.9927	-54.4739	0.5188
508	-55.2673	-54.5654	0.7019
509	-55.4199	-54.7180	0.7019
510	-55.4504	-54.7791	0.6713
511	-55.5115	-54.8401	0.6714
512	-55.6335	-54.9011	0.7324
513	-55.6641	-55.0232	0.6409
514	-55.6946	-55.0537	0.6409
515	-55.6641	-54.9622	0.7019

**Fig. 4.** Results of experimental series of predictions in case 2 (blue: true time series, red: predicted time series)

length helps us evaluating the predicted results in the cases of more or less data points. The subsequence length is also changed in 2 cases. Changing the subsequence length will also affect the prediction. If the subsequence length is too large, the neighboring subsequence of the last subsequence may have many overlaps, causing the predicted

result to be difference. If the subsequence length is too small, the CNP property can be affected to give predictive results. In order to choose a suitable subsequence length, it is dependent on both the length of the time series and the data collection. The prediction is based on the CNP property, so the larger the predicted time series length, the result will no longer be accurate. The predicted time series length should be less than the subsequence length.

## 4 Conclusion

Through the experimental cases, it shows that the predicted data are approximately equal to the real value. The up and down patterns in the predicted results over time are comparable to the true value. Since then, the prediction method of finding motifs using matrix profiles has a relatively good result and can be applied in practice. The next development step may be on the different subsequence to predict the time series rather than considering a subsequence.

## References

1. Mehtab, S., Sen, J.: A Time series analysis-based stock price prediction using machine learning and deep learning models. *Int. J. Bus. Forecast. Mark. Intell. (IJBFMI)* **6**, 272–335 (2020)
2. Yáñez, I.L., Sheremetov, L., Ponomarev, A.: Time series forecasting: Applications to the upstream oil and gas supply chain. In: 7th IFAC Conference on Manufacturing Modelling, Management, and Control, vol. 46, no. 9, pp. 957–962 (2013)
3. Chen, Y.A., Li, R., Hagedorn, L.S.: Undergraduate international student enrollment forecasting model: an application of time series analysis. *J. Int. Stud.* **9**(1), 242–261 (2019)
4. Bowerman, B.L.O., Connell, R.T.: *Time Series and Forecasting*. Duxbury Press, North Scituate (1979)
5. Kim, K.J.: Financial time series forecasting using support vector machines. *Neurocomputing* **55**(1–2), 307–319 (2003)
6. Lora, A.T., Santos, J.M.R., Expósito, A.G., Ramos, J.L.M.: Electricity market price forecasting based on weighted nearest neighbors techniques. *IEEE Trans. Power Syst.* **22**(3), 1294–1301 (2007)
7. Jiang, Z.Q., Xie, W.J., Xiong, X., Zhang, W., Zhang, Y.J., Zhou, W.X.: Trading networks, abnormal motifs and stock manipulation. *Quant. Financ. Lett.* **1**(1), 1–8 (2013)
8. Yolcu, O.C. Lam, H.K.: A combined robust fuzzy time series method for prediction of time series, *NEUCOM* **18265**, S0925-2312(17)30552-0 (2017)
9. Silva, P.C.L., Junior, C.A.S., Alves, M.A., Silva, R., Cohen, M.W., Guimarães, F.G.: Forecasting in non-stationary environments with fuzzy time series. *Appl. Soft Comput.*, S1568-4946(20)30763-8 (2020)
10. The UCR Matrix Profile Page. <https://www.cs.ucr.edu/~eamonn/MatrixProfile.html>. Accessed 12 Sept 2021
11. Zhu, Y., Yeh, C.C.M., Zimmerman, Z., Kamgar, K., Keogh, E.: Matrix Profile XI: SCRIMP++: time series motif discovery at interactive. In: IEEE International Conference on Data Mining (ICDM) (2018)
12. [https://sites.google.com/site/scrimpplusplus/?fbclid=IwAR1pUTxCZMKtrYlpK7HVxsHRXg1ONdrXcwRWnIqxx1n2YBdVU8aQ\\_hQxxpc](https://sites.google.com/site/scrimpplusplus/?fbclid=IwAR1pUTxCZMKtrYlpK7HVxsHRXg1ONdrXcwRWnIqxx1n2YBdVU8aQ_hQxxpc)



# Fast and Robust Local Texture Information for Fingerprint Segmentation

Binh T. H. Nguyen<sup>1</sup>, Ngoc Dung Bui<sup>2</sup> , and Xuan Tung Hoang<sup>3</sup>

<sup>1</sup> Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam  
nth.binh@hutech.edu.vn

<sup>2</sup> University of Transport and Communications, Hanoi, Vietnam  
dnbui@utc.edu.vn

<sup>3</sup> VNU University of Engineering and Technology, Hanoi, Vietnam  
tunghx@vnu.edu.vn

**Abstract.** Fingerprint segmentation is an essential step in fingerprint recognition systems. The task of the fingerprint segmentation is to isolate the background and foreground from the fingerprint image. There are many existing works on this topic. However, most existing algorithms have a trade-off between accuracy and processing time. This paper proposes a method using local texture information for the fingerprint segmentation which has high accuracy while the processing time is much reduced. The proposed method of local texture information uses the set of three block pixels features, including mean, variance, and Gabor filter. An optimal linear classifier is trained for the classification per block. This method is tested on the FVC2000 database, and the experimental results demonstrate the accuracy of the method.

**Keywords:** Fingerprint segmentation · Local texture information · Linear classification

## 1 Introduction

Fingerprint segmentation plays an important role in fingerprint recognition systems. This step is usually placed at the beginning of the recognition pipeline for improving quality of extracted features from fingerprint's images [1, 2]. It segments a fingerprint image into a foreground image and a background image. The former contains important information of a fingerprints called the minutia, which help in distinguishing between fingers [2, 3]. The latter contains unimportant and noisy information and is usually discarded.

Fingerprint segmentation methods can be categorized into pixel-based and block-based approaches [2, 4, 5]. In pixel-based methods, pixels of a fingerprint image will be classified into either foreground or background [2, 5]. Instead of classifying each pixel into foreground and background, block-based ones divide the fingerprint image into blocks of pixels first, then the classification is applied to each block rather than each pixels. According to Msiza and his colleagues [4], block-based methods are considered faster than pixel-based ones since they perform foreground-background segmentation at block level only.

Wu et al. [6], classified fingerprint segmentation methods into unsupervised and supervised ones. The unsupervised methods use block features such as the local histogram of ridge orientation, gray-level variance, magnitude of the gradient of each image block and Gabor feature. The supervised methods use some features such as coherence, average gray level, variance, Gabor response and use linear classifier for classification. Wu et al. also confirmed that supervised approach is more accurate but computationally more expensive.

Bazen et al. [7] suggested a supervised and pixel-based approach. The method used coherence, local mean and variance as features, and a linear classification algorithm to reduce computational time. Gabor features with a block-based approach was used in [8, 9]. The authors also compare the accuracy of their algorithm with the algorithm using variance, and the results are almost the same. In [10], a short time Fourier transformation analysis was applied for fingerprint segmentation. Factorized directional bandpass method was suggested in [11].

Thai et al. [12] presented a novel method that divides a fingerprint into three layers: a cartoon image, a texture image, and a noise image. Foreground of the fingerprint image is then extracted through binarization and morphology operation. The authors compared their method with five other important fingerprint segmentation methods presented in [6–8, 10, 11]. The comparison results showed that it outperforms the competitors in terms of error rate.

More recent works in fingerprint segmentation, e.g. [5, 13, 14], focus on using state-of-the-art machine learning techniques such as random forest and deep neural network. Suitable machine learning techniques can help improving fingerprint segmentation in terms of accuracy. However, these methods are usually computational hungry, and not suitable for systems with low resources.

In this paper, we present a novel block-based method for segmenting fingerprint images using local texture information. The local texture information used for segmentation is extracted from mean, variance and Gabor filters. Linear classification is then used to classify image blocks and a morphology-based post-processing is applied to improve accuracy. Our experiment results show that the proposed method is efficient in terms of both accuracy and complexity.

In the remaining of this paper, a detailed description of the proposed method is presented in Sect. 2; experiments and their results are shown in Sect. 3; and finally Sect. 4 closes this paper with conclusions.

## 2 Proposed Method

### 2.1 Feature Extraction

**Mean and variance:** When a finger is scanned by a finger sensor, center part of the finger that touches the sensor forms the foreground image with black and white lines representing finger's ridge-valley structure. The finger's edge that does not touch the finger sensor forms the background image, which is usually brighter [15]. As a result, mean gray value of the foreground is usually lower than that of the background. Also, since the foreground contains black lines while the background is a white region with

noise, variance of the foreground is also higher than variance of the background [16]. The mean  $\mu_b$  and variance  $\sigma_b$  of a  $w \times w$  block is defined as below:

$$\mu_b = \frac{1}{w \times w} \sum_{x=0}^{x=w-1} \sum_{y=0}^{y=w-1} I(X + x, Y + y) \quad (1)$$

$$\sigma_b = \frac{1}{w \times w} \sum_{x=0}^{x=w-1} \sum_{y=0}^{y=w-1} (I(X + x, Y + y) - \mu_b)^2 \quad (2)$$

where  $w$  is the block size ( $w = 8$  in our experiments),  $I(X, Y)$  is the intensity value of the pixel  $(X, Y)$ .

- Gabor features: Gabor filter has been used for fingerprint quality measure [6] as well as fingerprint segmentation [9]. The even symmetric form of Gabor filter is given by:

$$h(x, y, \theta, f, \sigma_x, \sigma_y) = \exp \left[ -\frac{1}{2} \left( \frac{x_\theta^2}{\sigma_x^2} + \frac{y_\theta^2}{\sigma_y^2} \right) \right] \times \cos(2\pi f x_\theta) \quad (3)$$

$$x_\theta = x \cos \theta + y \sin \theta$$

$$\text{where } y_\theta = -x \sin \theta + y \cos \theta$$

Observations show that, in a small fingerprint image block, all ridges will have similar orientations. It means that applying a Gabor filter to that block will result in high response if the filter has similar direction to that of the block. Utilizing this characteristic, we define eight Gabor filters corresponding to 8 directions  $\frac{\pi(k-1)}{8}$ , for  $k = 1 \dots 8$ , and apply those filters to each fingerprint image block. If a block produces high responses, it belongs to the foreground. Otherwise, it is a background image block.

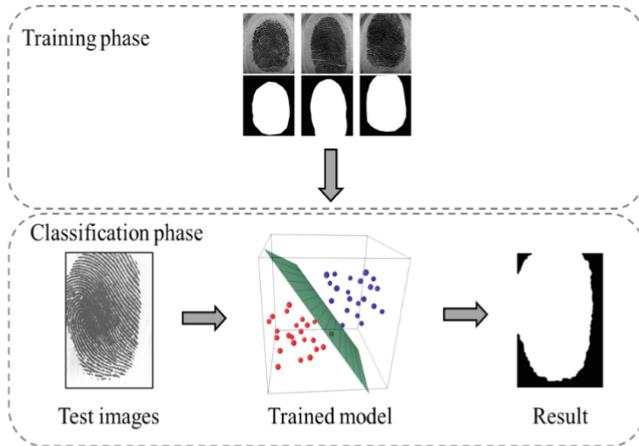
The magnitude of the Gabor feature is calculated by:

$$G(X, Y, \theta, f, \sigma_x, \sigma_y) = \left| \sum_{x=-\frac{w}{2}}^{\frac{w}{2}-1} \sum_{y=-\frac{w}{2}}^{\frac{w}{2}-1} I(X + x, Y + y) h(x, y, \theta, f, \sigma_x, \sigma_y) \right| \quad (4)$$

where  $w = 8$ ,  $\theta_k = \frac{\pi(k-1)}{8}$  ( $k = 1, \dots, 8$ ),  $\sigma_x = \sigma_y = 4$ . In our experiments,  $f$  is estimated by using the algorithm described in [17].

## 2.2 Linear Classification

Our segment process includes two phases which are training and classification as shown in Fig. 1. In training phase, we used 30 images from the FVC2000 database as the training data. Each image is manually processed to produce a segmented image that helps separating foreground and background. In the segmented images, white pixels are corresponding to foreground pixels, and black pixels are corresponding to background pixels. Each training image is processed into an array of  $8 \times 8$  blocks, each has a feature vector  $x_k$  containing mean, variance, and Gabor features and a label  $v$  of either 0 or 1 for background and foreground, respectively. Using all feature vectors and labels calculated



**Fig. 1.** Overall process of the proposed algorithm

from training images, weight vector  $w$  and bias  $b$  of the linear classification shown in Eq. 5 can be obtained.

$$v = w^T x_k + b \quad (5)$$

The classification phase accepts a grayscale fingerprint image as its input. After dividing the input image into  $8 \times 8$  blocks, feature vectors are constructed by calculating mean, variance and Gabor features of those blocks. We use feature vectors and Eq. (5) to calculate  $v$  which is used to classify this block. If  $v$  is greater than 0, this block is classified as foreground. Otherwise, it is classified as background. The result of the classification step is then post-processing to obtain the final segmentation.

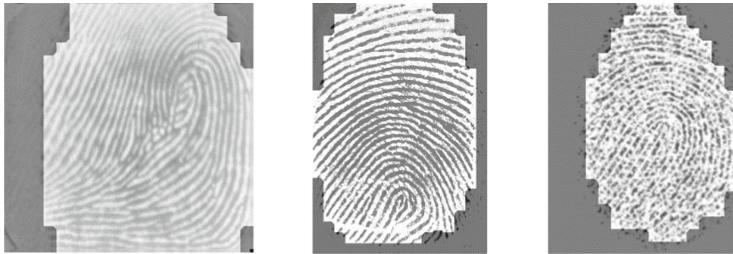
Post processing: After classifying each block as foreground or background, we will get a binary image, in which white pixels are foreground, black pixels are background. This binary image is blocky at borders separating bright foreground and dark background regions. To eliminate such blocky noise, morphological closing and morphological opening operations are applied to smooth the border.

### 3 Experimental Results and Discussions

The proposed method was experimented with 4 databases of FVC2000. In each database, 30 images were selected for training, the remaining 770 images were used to evaluate the proposed algorithm. All experiments were performed using MATLAB 2018 in a Corei5 CPU 2.50 GHz machine. Figure 2 shows the segmentation results of three sample fingerprint images in FVC2000's DB1.

To evaluate the results of the proposed segmentation algorithm, we calculate the fraction  $P_{error}$  between misclassified pixels and the total number of pixels in an image:

$$P_{error} = \frac{\text{Number of pixels that were misclassified}}{\text{Total number of pixels}}$$



**Fig. 2.** Segmentation results of three sample fingerprint images

The measure found in our method is 5.1% compared to 6.8% using features mean, variance and coherence in [7] for the DB2. The overall results in four databases are shown in Table 1. We can see from the table that average error rate of the algorithm is as low as 6.8%. This shows that our method can achieve high accuracy.

**Table 1.** Results of classification on FVC2000 Database

FVC2000 DB	Weight vector $w = [w_0 \ w_1 \ w_2 \ w_3]$	Perror
DB1b	(7.7533 -0.0726 -0.0004 0.0632)	0.07312
DB2b	(7.9106 -0.0794 0.0059 0.0820)	0.05149
DB3b	(4.8047 -0.0458 0.0129 -0.2557)	0.07348
DB4b	(13.6946 -0.1348 0.0013 -0.0158)	0.07341
Average error rate		0.0679

The proposed method is also fast, lightweight and easy to implement. We observed the average segmentation time during experiments is as low as 0.012 s. This makes the algorithm suitable for not only servers and workstations but also embedded systems. There are several reasons explaining the advantages. First of all, the algorithm is block-based and uses linear classification only. Secondly, the features are fixed mean and variance and eight direction Gabor filter, so that it does not need many iterations. And finally, this method uses relatively small training data set but can still produce fair accuracy as shown in our experiment results.

## 4 Conclusion

This paper has introduced a new algorithm for fingerprint segmentation. The method exploits local texture information including the mean, variance, and Gabor filters for classifying image blocks with linear classification, which is considered lightweight in computation. The performance evaluation proved the accuracy of the algorithm. In addition, the algorithm is easy to implement and has low cost in computational complexity. In the future, due to the fast speed and also ability to be easily applied in many systems of this algorithm, this is expected to be applied in fingerprint recognition systems.

## References

1. Babatunde, I.G., Kayode, A.B., Charles, A.O., Olatubosun, O.: Fingerprint image enhancement: segmentation to thinning. (*IJACSA*) *Int. J. Adv. Comput. Sci. Appl.* **3**(1), 15–24 (2012)
2. Jomaa, D., Dougherty, M., Fleyeh, H.: Segmentation of low quality fingerprint images. In: 2010 International Conference on Multimedia Computing and Information Technology (MCIT) (2010)
3. Bana, S., Kaur, D.D.: Fingerprint recognition using image segmentation. *Int. J. Adv. Eng. Sci. Technol.* **5**(1), 012–023 (2011)
4. Msiza, I.S., Mathekga, M.E., Nelwamondo, F.V., Marwala, T.: Fingerprint segmentation: an investigation of various techniques and a parameter study of a variance-based method. *Int. J. Innov. Comput. Inf. Control (IJICIC)* **7**(9), 5313–5326 (2011)
5. Wan, G.C., Li, M.M., Xu, H., Kang, W.H., Rui, J.W., Tong, M.S.: XFinger-Net: pixel-wise segmentation method for partially defective fingerprint based on attention gates and U-net. *Sensor* **20**(16), 4473 (2020)
6. Wu, C., Tulyakov, S., Govindaraju, V.: Robust point-based feature fingerprint segmentation algorithm. In: Lee, S.-W., Li, S.Z. (eds.) *ICB 2007. LNCS*, vol. 4642, pp. 1095–1103. Springer, Heidelberg (2007). [https://doi.org/10.1007/978-3-540-74549-5\\_114](https://doi.org/10.1007/978-3-540-74549-5_114)
7. Bazen, A.M., Gerez, S.H.: Segmentation of fingerprint images. In: ProRISC 2001: Workshop on Circuits, Systems and Signal Processing, Veldhoven, The Netherlands, pp. 276–280 (2001)
8. Shen, L., Kot, A., Koo, W.: Quality measures of fingerprint images. In: Bigun, J., Smeraldi, F. (eds.) *AVBPA 2001. LNCS*, vol. 2091, pp. 266–271. Springer, Heidelberg (2001). [https://doi.org/10.1007/3-540-45344-X\\_39](https://doi.org/10.1007/3-540-45344-X_39)
9. Alonso-Fernandez, F., Fierrez-Aguilar, J., Ortega-Garcia, J.: An enhanced Gabor filter-based segmentation algorithm for fingerprint recognition systems. In: Proceedings of the 4th International Symposium on IEEE Image and Signal Processing and Analysis, ISPA, pp. 239–244 (2005)
10. Chikkerur, S., Govindaraju, V., Cartwright, A.N.: Fingerprint image enhancement using STFT analysis. In: Singh, S., Singh, M., Apte, C., Perner, P. (eds.) *ICAPR 2005. LNCS*, vol. 3687, pp. 20–29. Springer, Heidelberg (2005). [https://doi.org/10.1007/11552499\\_3](https://doi.org/10.1007/11552499_3)
11. Thai, D.H., Huckemann, S., Gottschlich, C.: Filter design and performance evaluation for fingerprint image segmentation. *PLoS ONE* **11**(5), e0154160 (2016)
12. Thai, D.H., Gottschlich, C.: Global variational method for fingerprint segmentation by three-part decomposition. *IET Biom.* **5**(2), 120–130 (2016)
13. Ezeobiejesi, J., Bhanu, B.: Latent fingerprint image segmentation using deep neural network. In: Bhanu, B., Kumar, A. (eds.) *Deep Learning for Biometrics. ACVPR*, pp. 83–107. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-61657-5\\_4](https://doi.org/10.1007/978-3-319-61657-5_4)
14. Sankaran, A., Jain, A., Vashisth, T., Vatsa, M., Singh, R.: Adaptive latent fingerprint segmentation using feature selection and random decision forest classification. *Inf. Fusion* **34**, 1–15 (2017)
15. Chen, X., Tian, J., Cheng, J., Yang, X.: Segmentation of fingerprint images using linear classifier. *EURASIP J. Adv. Signal Process.* **2004**(4), 1–15 (2004). <https://doi.org/10.1155/S1110865704309194>
16. Singh, V. K., Mathai, K.J., Singh, S.: Fingerprint segmentation: optimization of a filtering technique with Gabor filter. In: 2014 Fourth International Conference on Communication Systems and Network Technologies, pp. 823–827 (2014)
17. Hong, L., Wan, Y., Jain, A.: Fingerprint image enhancement: algorithm and performance evaluation. *IEEE Trans. PAMI* **20**, 777–789 (1998)



# Developing Real-time Recognition Algorithms on Jetson Nano Hardware

Phat Nguyen Huu<sup>1(✉)</sup>, Thien Pham Ngoc<sup>1</sup>, and Thanh Le Thi Hai<sup>2</sup>

<sup>1</sup> School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam

[phat.nguyễnhuu@hust.edu.vn](mailto:phat.nguyễnhuu@hust.edu.vn), [thien.pn172833@sis.hust.edu.vn](mailto:thien.pn172833@sis.hust.edu.vn)

<sup>2</sup> School of Engineering Physics, Hanoi University of Science and Technology, Hanoi, Vietnam

[thanh.lethihai@hust.edu.vn](mailto:thanh.lethihai@hust.edu.vn)

**Abstract.** Today, the field of artificial intelligence (AI) is increasingly developing with the development of the 4.0 revolution. The applicability for the social life of AI and deep learning is indeed enormous. Convolutional neural network (CNN) models have many advantages for object detection and classification problems. In the paper, we present the results of real-time object recognition algorithms on Jetson Nano hardware. We perform the algorithm to recognize and deploy on GPU with largest optimal rate as 76.26%. The results show that Mobilenetv2 and YOLOv3 models are the most optimal for object recognition with the processing time of 50 and 51 milliseconds, respectively.

**Keywords:** TensorRT · Jetson Nano · AIoT · Object detection · Deep learning

## 1 Introduction

Today, the field of artificial intelligence (AI) is increasingly developing with the development of the 4.0 revolution. The applicability for the social life of AI and deep learning (DL) is indeed enormous. Convolutional neural network (CNN) models have shown accuracy up to 99% for object detection and classification problems. However, it is necessary to have powerful hardware to train networks in the best way to be able to study this field. Besides, the issues relating to deploying these models down to embedded systems are also quite complicated. We are not able to use extremely expensive servers for products that prioritize the economy. Therefore, we have developed object recognition algorithms such as YOLOv3, YOLOv3v4, SSD-Mobilenetv2 on Jetson Nano embedded system. The results are satisfactory processing speed (50 milliseconds) in the paper.

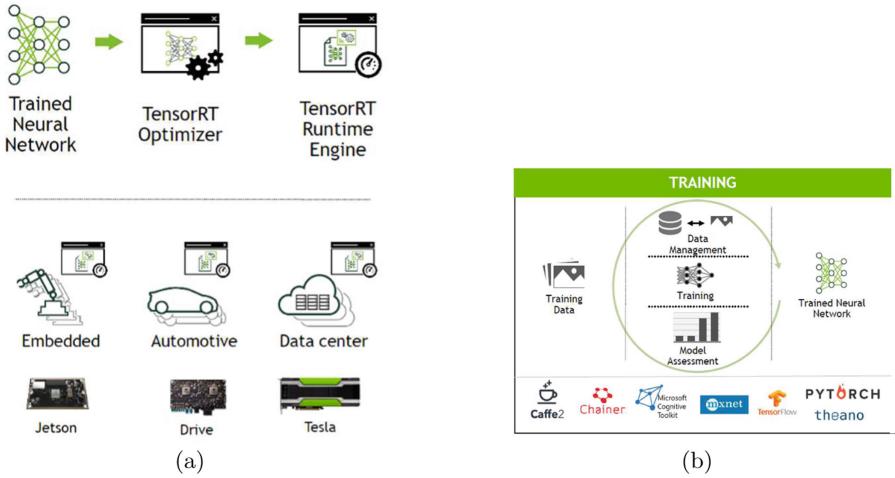
The rest of the paper includes four parts and is organized as follows. Section 1 will discuss the related work. In Sect. 2, we present the proposal system. Section 3 will evaluate the proposed system and analyze the results. In the final section, we give conclusions and future research directions.

## 2 Related Work

The implementation of recognition algorithms on low hardware in recent years is a matter of great attraction of companies such as Google, Intel, NVIDIA. The companies have developed frameworks that make it easier to implement algorithms on hardware. Intel has launched the OpenVINO toolkit to increase the ability to implement CNN algorithms on processors from 6<sup>th</sup> to 11<sup>th</sup> generation or Intel HD Graphics [5,9,12]. This makes it easier to implement the on-chip algorithm from intel and can be deployed on a Raspberry Pi embedded computer to connect an additional Intel Neural Compute Stick 2. However, the downside of this toolkit is that it only supports computing support on intel chips and the speed is not outstanding because the CPUs have very few cores. If you want to increase speed, you must add an Intel Neural Compute Stick 2 or a dedicated VPU. Google provides two types of hardware with TPU to help implement the CNN algorithm on hardware at high speed. These are Coral Dev Board [2,8] and USB Accelerator [1,16]. Besides, Google is also the developer of the Tensorflow framework - a framework that supports CNN training. Therefore, combining a framework and its Dev Board to deploy DL models on embedded computers will give the same results as Intel. However, this makes it impossible for us to interfere with physics and we only develop algorithms at the application layer. Finally, NVIDIA is also researched in the field of AI because it is a supplier of GPUs with extremely high computing capabilities. Everyone uses NVIDIA GPUs while training models since there are CUDA Cores. This is a mandatory requirement while using frameworks Tensorflow, PyTorch, etc. to train models on GPU. The number of cores is thousands of CUDA Cores on one GPU or several hundred Tensor Cores to support Tensor-level computation. The company has released versions of Jetson Kit [4,6] and TensorRT framework [7,17] with relatively cheap prices for embedded computers and processing speed no less than the above brands. In the paper, we present the results of real-time algorithm development on Jetson Nano. This is the NVIDIA Dev kit. We have developed object recognition algorithms such as YOLOv3, YOLOv4, SSD-MobileNetV2 based on [10,11,15] for the kit. These are the most commonly used state-of-the-art (SOTA) models.

## 3 Developing System

First, we present an overview of model implementation stages as shown in Fig. 1. As shown in Fig. 1(a), there are three stages to perform the model including training, optimizing using TensorRT, and creating TensorRT runtime engine. We can then use the engine to deploy on Jetson embedded computers, self-driving cars, or data centers. Using TensorRT improves latency, throughput, performance, and memory usage. However, TensorRT also has disadvantages. It is possible not only to train but also to deploy the model. Besides, the model training also relies on libraries like Tensorflow, Pytorch, etc.



**Fig. 1.** (a) Model deployment stages and (b) training phase.

### 3.1 Training Mode

The steps to train the model are as shown in Fig. 1(b) including:

1. Creating training and test datasets,
2. Using libraries that support model training to build a CNN network (e.g. Tensorflow, Pytorch, Keras)
3. Imputing data to train the model to learn their features
4. Evaluating the model and improve accuracy.

### 3.2 Optimizing Model

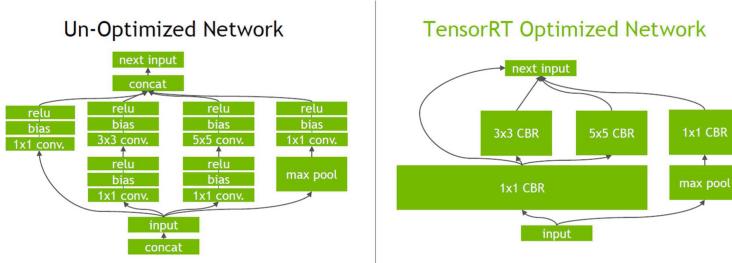
The optimization methods include

1. Layer and Tensor fusion

This method is to optimize the number of hidden layers making model lighter while maintaining accuracy. As shown in Fig. 2, we can see that this optimization is absolutely important. From very heavy models we can optimize to very light models, enough for embedded computers to run in real time.

2. Weights and activation precision calibration

Due to the limitation of the GPU's ability to compute data type FP32 (32-bit real number), we need to care about the data type of parameters of network to optimize the number of classes if we want fast computing speed. The data types are shown in Table 1.



**Fig. 2.** Network model before and after optimization.

### 3. Kernel auto tuning

This method helps to define the cores on the GPU that the model can execute on. TensorRT helps identify and deploy on GPU cores in the best way. That leads to the highest execution speed of the model. However, this causes the problem that we are not able to optimize the model on many different GPUs.

#### 4. Dynamic Tensor memory.

In normal mode, the CUDA kernel runs one thread per GPU. All CUDA kernels execute serially if there are no optimizations. However, it is possible to execute parallel using TensorRT by launching independent CUDA kernels for separate threads.

**Table 1.** Computational data types.

Precision mode	Dynamic range
FP32	$-3.4 \times 10^{38} \div +3.4 \times 10^{38}$
FP16	$-65504 \div +65504$
INT8	$-128 \div +127$

### 3.3 Developing TensorRT

First, we present an overview of the model deployment stages as shown in Fig. 3. We will convert to network definition from trained network. In this step, we have two ways to perform it. Firstly, we use TensorRT API to pass each layer of CNN. In a second way, we use an intermediate model that is open neural network exchange (ONNX) [13,14]. Then we use the API to analyze this model. From there, we get the Network definition. However, this is not easy. Since TensorRT does not support all calculations finding in all current models, we need to build a plug-in that TensorRT can define and make CUDA Cores perform them. In [3], a function is added in the YOLOv4 network as follows

$$f(x) = x \times \tanh(\ln(1 + e^x)). \quad (1)$$

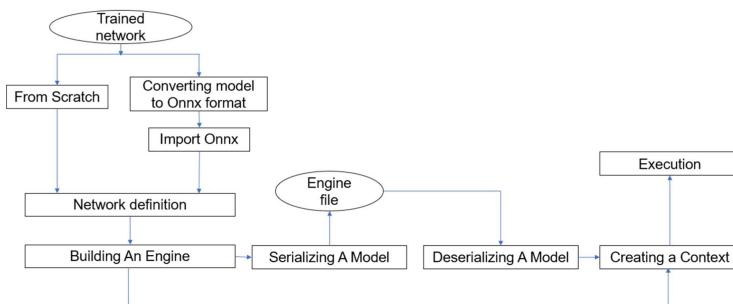
We create a device-appropriate engine from network definition. Since each type of GPU has a different memory organization and processing method, this can only be performed on the device to be deployed. When the engine is created, we will store it in a machine to use next time. Therefore, we do not need to create another engine. When we have the engine, we just create context on GPU. The goal is to configure GPU 1 context. Therefore, GPU will understand and run according to this context. The downside of this implementation is that it can only be used on devices with NVIDIA GPUs. TensorRT instances can only support a handful of math operations since building a separate plug-in for each model has to perform. In addition to the above disadvantages, this method has many advantages such as fast processing speed, support for streaming to cloud/host by CODEC suites on Jetson Nano kit, and multi-threading due to NVIDIA deep-stream support.

## 4 Result

We first evaluate the networks before and after being optimized for hardware implementation. In the experiment, we perform to recognize many objects at the same time. The results are shown in Table 2. In Table 2, we see that ResNet-152 network has the largest optimal rate with 76.26%.

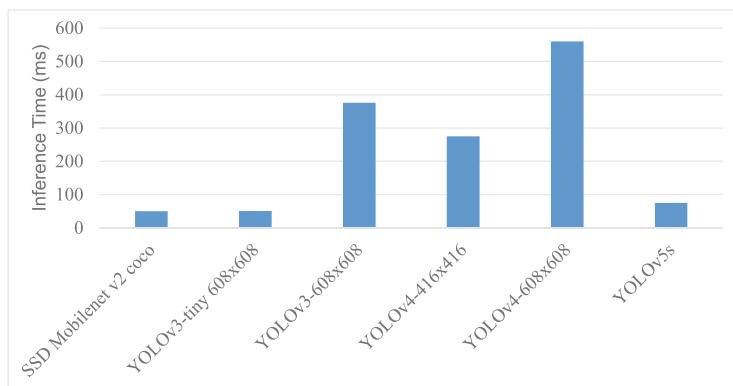
**Table 2.** Results after optimizing layers.

Model	Number of layer	Number of layer after optimizing	Optimization rate(%)
VGG19	43	27	38.21
Inceptionv3	309	113	63.43
ResNet-152	670	159	76.26

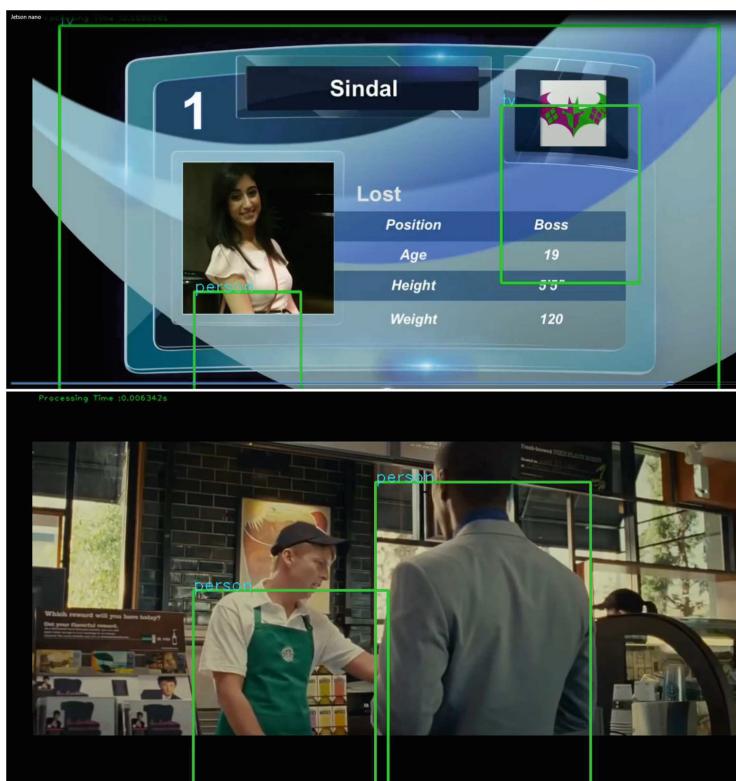


**Fig. 3.** Overview of developing models.

Secondly, we perform into one object recognition model on Jetson Nano based on Table 2. The results are shown in Figs. 4 and 5. In Fig. 4, we see that the mobilenet SSD model has the fastest execution time and the YOLOv4 model has the longest time. This is understandable since the number of network layers



**Fig. 4.** Results time processing of object recognition model.



**Fig. 5.** An example of object recognition model.

of the mobilenet SSD model is almost completely reduced to be able to perform on low-profile hardware. Therefore, it is often implemented for applications that require real-time and low accuracy.

## 5 Conclusion

In this paper, we implement DL models on real Jetson Nano hardware. The results show that Mobilenetv2 and YOLOv3 models are the most optimal with an object recognition time of 50 milliseconds. This is the first step for performing other research projects. Therefore, we can apply for smart home projects and deploy DL models on self-driving cars by combining the knowledge of image processing and embedded programming in the future.

## References

1. Accelerator, C.U.: Get started with the USB accelerator (2021). <https://coral.ai/docs/accelerator/get-started>
2. Board, C.D.: Get started with the dev board (2021). <https://coral.ai/docs/dev-board>
3. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.: Yolov4: optimal speed and accuracy of object detection, pp. 1–18, April 2020
4. Cui, H., Dahoun, N.: Real-time stereo vision implementation on Nvidia Jetson TX2. In: 2019 8th Mediterranean Conference on Embedded Computing (MECO), pp. 1–5 (2019)
5. Deval, A., Ananthakrishnan, A., Forbell, C.: Power management on 14 nm intel core m processor, pp. 1–3, April 2015
6. DEVELOPER, N.: Jetson modules (2021). <https://developer.nvidia.com/embedded/jetson-modules>
7. DEVELOPER, N.: NVIDIA TensorRT (2021). <https://developer.nvidia.com/tensorrt>
8. Fried, L.: Making machine learning arduino compatible: a gaming handheld that runs neural networks. IEEE Spectr. **56**(8), 14–15 (2019)
9. Graphics, I.H.: System requirements (2021). <https://software.intel.com/content/www/us/en/develop/tools/openvino-toolkit/system-requirements.html>
10. Iyer, R.V., Ringe, P.S., Bhensdadiya, K.P.: Comparison of YOLOV3, YOLOV5s and MobileNet-SSD V2 for real-time mask detection. Int. Res. J. Eng. Technol. (IRJET) **8**(07), 1156–1160 (2021)
11. Lang, L., Xu, K., Zhang, Q., Wang, D.: Fast and accurate object detection in remote sensing images based on lightweight deep neural network. Sensors **21**, 5460 (2021)
12. Le Duc, H., et al.: Hardware implementation of all digital calibration for undersampling tiadc. In: 2015 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 2181–2184 (2015)
13. Lin, W.F., et al.: ONNC: a compilation framework connecting ONNX to proprietary deep learning accelerators. In: 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), pp. 214–218 (2019)
14. ONNX: Build model (2021). <https://onnx.ai/get-started.html>

15. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788 (2016)
16. Tsiktsiris, D., Ziouzios, D., Dasygenis, M.: A portable image processing accelerator using FPGA. In: 2018 7th International Conference on Modern Circuits and Systems Technologies (MOCAST), pp. 1–4 (2018)
17. Xu, R., Han, F., Ta, Q.: Deep learning at scale on NVIDIA V100 accelerators. In: 2018 IEEE/ACM Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS), pp. 23–32 (2018)



# Numerical Method of Synthesizing the Automatic Electric Drive System Containing Elastically Linked Elements

Dang Nguyen Phu<sup>1(✉)</sup>, Vu Duc Tuan<sup>2</sup>, Tran Binh Duong<sup>3</sup>, Nguyen Duc Uyen<sup>4</sup>, and Duc-Tan Tran<sup>5(✉)</sup>

<sup>1</sup> Le Quy Don Technical University, Hoang Quoc Viet Street, Hanoi, Vietnam  
npdang@lqdtu.edu.vn

<sup>2</sup> University of Transport Technology, Nguyen Trai Street, Hanoi, Vietnam  
tuanvd@utt.edu.vn

<sup>3</sup> VietNam Paper Corporation, No 25A Ly Thuong Kiet Road, Hanoi, Vietnam

<sup>4</sup> VOV College No. 1, No. 36 Quy Luu Street, Phu Ly city, Hanam, Vietnam

<sup>5</sup> Faculty of Electrical and Electronic Engineering, Phenikaa University, Hanoi, Vietnam  
tan.tranduc@phenikaa-uni.edu.vn

**Abstract.** The article proposes and investigates an approach to the solution problem of synthesizing electric drive systems containing linked elastic elements, with the main content: modeling elastic conveyor under as the form transfer function, construction of a system structure diagram has considered the influence of the conveyor, analyze and established a regulator synthesis algorithm by real interpolation method, recommend overshoot calibration methods of the system synthesized. At the end of the article, we build an automatic program to synthesize according to the proposed algorithm and present the results for a specific calculation model.

**Keywords:** Automatic control system · Regulator synthesis · Objects with distributed parameters · Elastic link element

## 1 Introduction

Automated electric drive systems with multiple active motors linked by elastic element that incident in terms of speed are essential components in most production lines, mechanical processing equipment (such as continuous rolling mill systems), paper production lines, optical thin-film handling equipment, cable sheathing line, spinning machine, canning lines, automatic robots in automobile assembly lines, production of electronic components, etc. The transfer function describing elastic element will take the following form:

$$W_{dt}(s) = f\left(s, e^{\frac{A(s)}{B(s)}}, \sqrt{s}, \cos(s), \sin(s), sh(s), ch(s), \dots\right), \quad (1)$$

it contains not only the argument as  $s$  with the linear system but also the inertial and intangible components (function of  $s(\sqrt{s}, \cos(s), \sin(s), sh(s), ch(s), \dots)$ ) [1–4].

For such a drive system that enhances performance controllability, then the influence of the elastic conveyor linking the motors needs to be considered. The transfer function has complexity making it much more challenging to synthesize the regulator. Some studies have modeled and constructed controllers for multi-motor drive systems containing elastic conveyors [5–8]. The works [9–11] proposed designing tension-controlled multi-motor drive systems controllers based on nonlinear models. The mathematical tools used [9–11] are implemented based on fuzzy theory and neural networks, requiring designers to have a lot of experience. Sustainable methods of calibrating the regulator based on example standard  $H^\infty$  can lead to solutions that do not guarantee sustainability. This limitation also exists in methods using the integral asymptote criterion between the desired and synthesized system [12, 13]. To satisfy the complex linkages in the multi-motor systems [14–16], the system parameters need sufficient information to apply LQG method. This paper proposes a solution to synthesize and calibrate the regulator based on the real interpolation method with a simple procedure, which reduces the computational volume and preserves the specific properties, characteristics, and effects of conveyors on the system [17, 18].

## 2 Material and Methods

### 2.1 Modeling of an Electric Drive System Containing an Elastically Linked Element

The general structure of electric drive system, containing elastically linked elements is shown in Fig. 1. The conveyor is surveyed as a form closed-loop associated with the motor represented by the equations [2, 3]:

$$L_t[u(x, t)] = \rho(x) \frac{\partial^2 u(x, t)}{\partial t^2} - E \frac{\partial^2 u(x, t)}{\partial x^2} = f(x, t); 0 \leq x \leq l, t \geq 0, \\ \rho(x) > 0, E > 0 \quad (2)$$

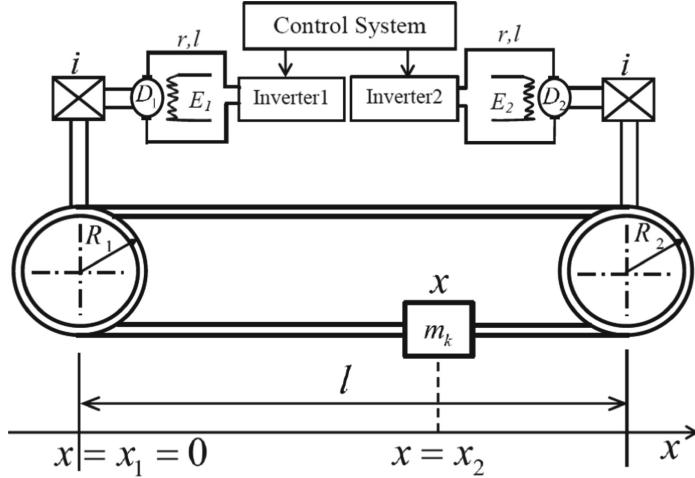
with initial and boundary conditions:

$$\begin{cases} u(x, t)|_{t=0} = u_0(x); \frac{\partial u(x, t)}{\partial t}|_{t=0} = u_1(x) \\ u(x, t)|_{x=0} = u(x, t)|_{x=l}; \frac{\partial u(x, t)}{\partial x}|_{x=0} = \frac{\partial u(x, t)}{\partial x}|_{x=l} \end{cases}, \quad (3)$$

in which:  $L_t$  - differential operator;  $u(x, t)$  - displacement of the point on the elastic link element with coordinates  $x$  at the  $t$ ;  $E = const$  - the elastic modulus of the element under investigation;  $u_0(x)$ ,  $u_1(x)$  - displacement and displacement rate of the elastic element section at coordinates  $x$  and time  $t = 0$ ;  $f(x, t)$  - input impacts in space, time;  $\rho(x)$  - the material density of the elastic element in x-axis, computed by:

$$\rho(x) = \rho_l + \sum_{i=1}^n m_i \delta(x - x_i) \quad (4)$$

with:  $\rho_l = const$  is the density of elastic conveyor when unloaded.



**Fig. 1.** Active two-motor drive system structure linked by elastic conveyors

In the survey system, two mass components are associated with the motor and conveyor:  $m_1$  - the mass of the drive part associated with the motor, concentrated at the point  $x = x_1 = 0$ , also  $m_2$  - the mass of the conveyor is concentrated at the point  $x = x_2$  ( $n = 2$ ).

Based on the research results in [2, 3], we get the mathematical model of transfer function representing the relationship between the force at a point of the coordinate  $x_1 = 0$  and mass  $m_1$  ( $F_d(0, s)$ ) with the velocity at the point of the coordinate  $x_2$  and mass  $m_2$  ( $V_\lambda(x_2, s)$ ):

$$W_{dt}(s) = W(x_2, 0, s) = \frac{V_\lambda(x_2, s)}{F_d(0, s)} = \frac{q \cdot shs \cdot ch\lambda \cdot s}{sh^2 s + \mu_1 \mu_2 s^2 (ch^2 s - ch^2 \lambda \cdot s) + (\mu_1 + \mu_2) s \cdot sh2s} \quad (5)$$

where,  $\mu_1 = \frac{\eta_1}{l} = \frac{m_1}{m_k}$ ,  $\mu_2 = \frac{\eta_2}{l} = \frac{m_2}{m_k}$ ,  $m_k = \rho_l l$  - the mass of conveyor;  $q = \frac{1}{2a}$  - the transmission coefficient of the elastic element,  $\lambda = 1 - \frac{2x_2}{l}$  - the output space coordinates of the system. The calculation is done similarly for input-output coordinates:  $x_1 = \xi = 0$ ;  $x_2 = 0$ , we will get the transfer function between the force on the active rewinder ( $F_d(0, s)$ ) and its velocity:  $V_d(0, s)$ :

$$W_{dt}^*(s) = W(0, 0, s) = \frac{V_d(0, s)}{F_d(0, s)} = \frac{q [sh2s + s\mu_2(ch^2 s - ch^2 \lambda \cdot s)]}{[sh^2 s + (\mu_1 + \mu_2)s \cdot sh2s + \mu_1 \mu_2 s^2 (ch^2 s - ch^2 \lambda \cdot s)]} \quad (6)$$

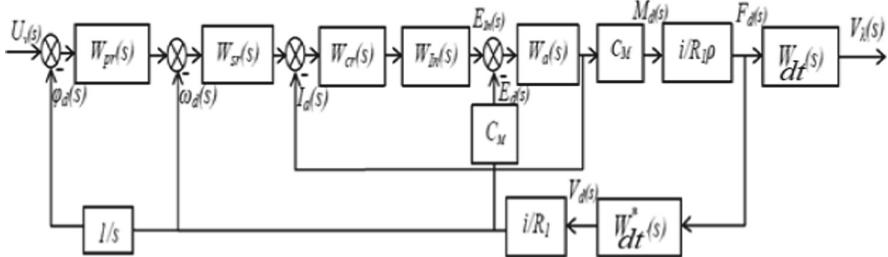
Thus, the structural diagram representing the electric drive system taking into account the elastically linked conveyor (6) is shown in Fig. 2 in which  $W_{pr}(s)$ ,  $W_{sr}(s)$ ,  $W_{cr}(s)$  - The transfer function of the rotation angle, speed, and current regulators;  $W_{In}(s)$  - The transfer function of the inverter:  $W_{In}(s) = \frac{k_{In}}{T_{In}s + 1}$ ;  $k_{In}$ ,  $T_{In}$  - The gain and time constant of the inverter;  $W_a(s)$  - The transfer function describing the motor armature:  $W_a(s) = \frac{1}{R_a(T_a s + 1)}$ . The influence of the elastic conveyor on the motor motion through a transfer function  $W_{dt}^*(s)$  that is fed into the feedback circuit of the speed control loop [3].

## 2.2 The Basis for Synthesizing the Regulator by the Interpolation Method

At each loop of the system shown in Fig. 2, the aim is to find a regulator of the form:

$$W_i(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \dots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0} (m \leq n, i = 1 \div 3) \quad (7)$$

and the feedback coefficients  $K_i (i = 1 \div 3)$  satisfy the conditions of each round and system of quality criteria:



**Fig. 2.** Structure of electric drive system

$$\begin{cases} \sigma_R - \Delta\sigma \leq \sigma_S \leq \sigma_R + \Delta\sigma \\ t_{ST}^S \leq t_{ST}^R \end{cases} \quad (8)$$

or

$$\begin{cases} \sigma_R - \Delta\sigma \leq \sigma_S \leq \sigma_R + \Delta\sigma \\ t_{ST}^S \rightarrow t_{ST}^{min} \end{cases}, \quad (9)$$

with,  $\sigma_R$  - the required overshoot;  $\sigma_S$  - the synthesized overshoot;  $\Delta\sigma$  - allowable overshoot erroneous between the desired and the synthetic system;  $t_{ST}^R$  - the required excessive time of the desired system;  $t_{ST}^S, t_{ST}^{min}$  - The transient time and the minimum achievable transient time of the system are summed up, respectively.

Determining the desired transfer function of a closed system  $W_D^C(s)$  that meets the required quality criteria  $\sigma_R, t_{ST}^R$  is the first thing to do to solve this problem. The method developed in [17], allowing to receive a transfer function  $W_D^C(s)$  has the form

$$W_D^C(s) = \frac{\frac{\alpha_1}{2}s + 1}{\alpha_0 s^2 + \alpha_1 s + 1} H^\infty; \alpha_0 = \frac{\left[ \ln\left(\frac{H^{\max}}{H^\infty} - 1\right) \right]^2}{\frac{9}{(t_{ST}^R)^2} \left\{ \left[ \ln\left(\frac{H^{\max}}{H^\infty} - 1\right) \right]^2 + \pi^2 \right\}}; \alpha_1 = \frac{6\alpha_0}{t_{ST}^R}, \quad (10)$$

where  $H^\infty, H^{\max}$  - the steady-state value and the corresponding maximum of the transient characteristic are determined based on  $\sigma_R$  and the desired static mode.

Then, establish the synthesis equation representing the relationship between the desired transfer function  $W_D^C(s)$  and the regulators ( $W_{pr}(s), W_{sr}(s), W_{cr}(s)$ ), feedback

coefficient ( $K_i$ ) and the transfer function of the fraction remains constant in each round ( $W_i^{UN}(s)$ ,  $i = 1 \div 3$ ), relating to engine ( $W_a(s)$ ), inverter ( $W_{In}(s)$ ), elastic conveyor ( $W_{dt}^*(s)$ ), ...:

$$W_D^C(s) \cong F \left[ W_{pr}(s), W_{sr}(s), W_{cr}(s), K_1, K_2, K_3, W_1^{UN}(s), W_2^{UN}(s), W_3^{UN}(s) \right] \quad (11)$$

where:  $W_1^{UN}(s)$ ,  $W_2^{UN}(s)$ ,  $W_3^{UN}(s)$  - the transfer function of the constant part within the current, speed, and corresponding rotational angle (Fig. 2), in the form:

$$W_1^{UN}(s) = W_{In}(s) * W_a(s); W_2^{UN}(s) = \frac{C_m \cdot i^2}{R_1^2 \rho} \cdot W_{dt}^*(s); W_3^{UN}(s) = \frac{1}{s} \quad (12)$$

Interpolation method is widely used in the engineering [20–22]. We applied the real interpolation to convert the Eq. (11) to the form with real arguments [18]:

$$W_D^C(\delta) \cong F \left[ W_{pr}(\delta), W_{sr}(\delta), W_{cr}(\delta), K_1, K_2, K_3, W_1^{UN}(\delta), W_2^{UN}(\delta), W_3^{UN}(\delta) \right]. \quad (13)$$

We choose the distribution rule of interpolated nodes  $\delta_i$ , e.g., according to the uniform distribution regulation:

$$\delta_i = i\delta_1, i = 1 \div \eta \quad (14)$$

with  $\eta$  – the interpolate the number of nodes, or establishment the value of the nodes  $\{\delta_i\}_\eta$  to coincide with the zero points of the Chebyshev polynomials [19]:

$$\delta_i = \frac{1 + x_i}{1 - x_i} a, i = \overline{1, \eta}. \quad (15)$$

where:  $a$  – The real parameter is used to correct the composite error, and  $\{x_i\}_\eta$  is the zero points of a Chebyshev polynomial of the first order of degree  $\eta$  ( $T_\eta(x) = 0$ ), determined by the relation [19]:

$$T_0(x) = 1; T_1(x) = x; \dots; T_{\eta+1}(x) = xT_\eta(x) - \frac{1}{4}T_{\eta-1}(x); x \in [-1, 1] \quad (16)$$

The following is to carry out calculate the numerical features  $\{W_D^C(\delta_i)\}_\eta$ ,  $\{W_{pr}(\delta_i)\}_\eta$ ,  $\{W_{sr}(\delta_i)\}_\eta$ ,  $\{W_{cr}(\delta_i)\}_\eta$ ,  $\{W_1^{UN}(\delta_i)\}_\eta$ ,  $\{W_2^{UN}(\delta_i)\}_\eta$ ,  $\{W_3^{UN}(\delta_i)\}_\eta$  corresponding to the selected set of nodes. Then, we solve the system of equations:

$$\left\{ W_D^C(\delta_i) \right\}_\eta = \left\{ F[W_{pr}(\delta_i), W_{sr}(\delta_i), W_{cr}(\delta_i), K_1, K_2, K_3, W_1^{UN}(\delta_i), W_2^{UN}(\delta_i), W_3^{UN}(\delta_i)] \right\}_\eta \quad (17)$$

to compute the coefficients  $K_i$  and the unknown ones. To ensure that the system (17) has a unique solution, the value  $\eta$  must be equal to the required coefficients of the regulator.

The feedback coefficients  $K_i$  of each loop can be determined from the static mode under considering:

$$K_i < \frac{1}{H_i^\infty - \Delta H_i} \simeq \frac{1}{H_i^\infty}, \quad i = 1 \div 3, \quad (18)$$

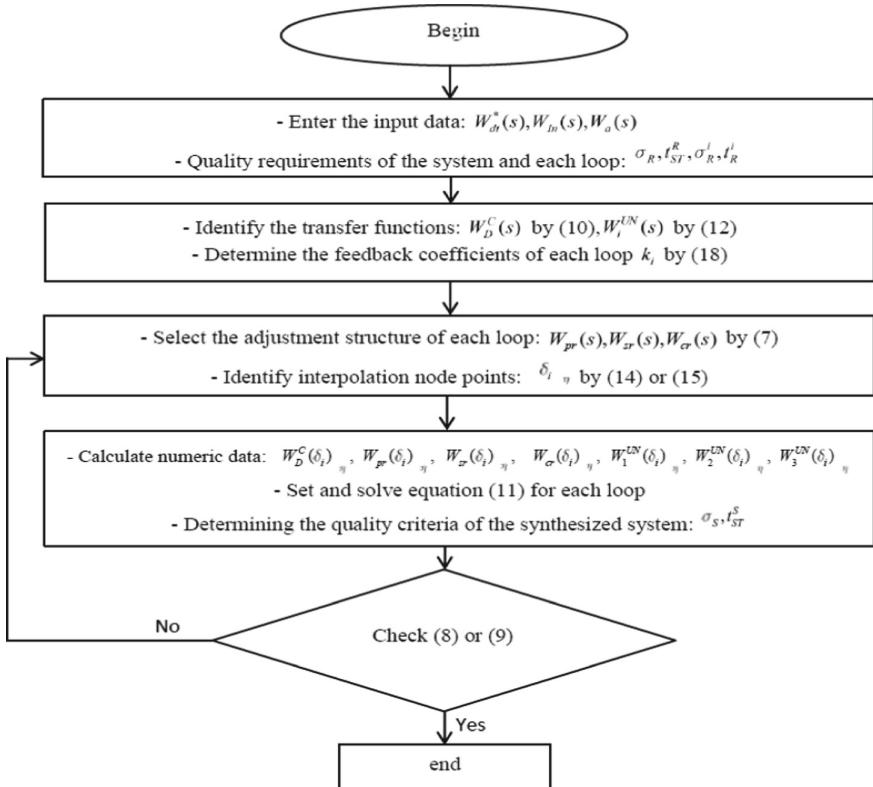
with,  $\Delta H_i$  - systematic error in the established mode in loop  $i$ , while the nodes are uniformly distributed, the first node can be determined by:  $\delta_1 = \frac{3.5}{t_K^R}$  [2].

The next stage lies in determining the quality criteria and tuning the synthesized system. We used the synthesized system calibration to achieve overshoot according to the condition (8) include [18]: Changing the interpolation nodes point  $\{\delta_i\}_\eta$  and/or different structures of the regulator ( $m, n$ ); Achieving the desired overshoot by varying the required settling time ( $t_{ST}^R$ ); Using weight functions of the particular form ( $\omega(t)$ ); Achieving a given overshoot based on generalized desired transfer functions.

### 3 Results

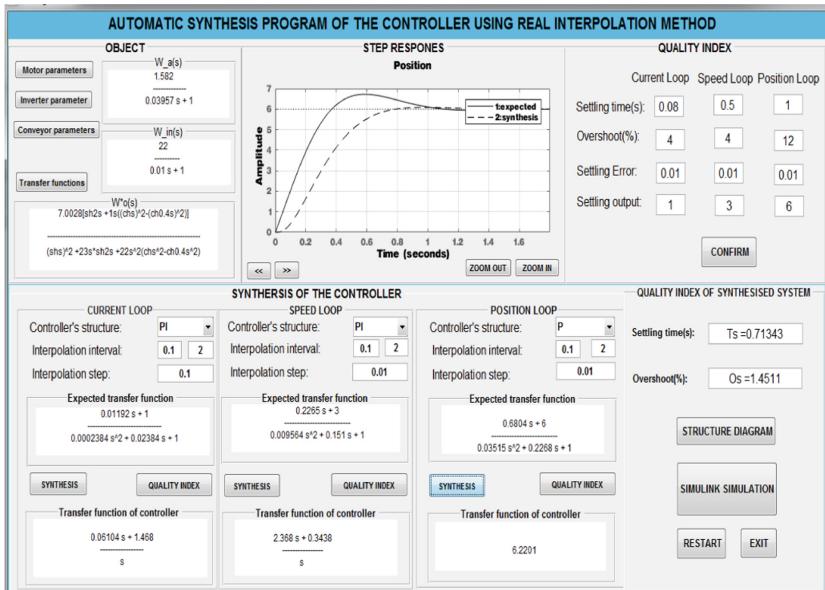
#### 3.1 Setup the Simulation Program

Synthesis of the electric drive system according to the algorithm diagram (Fig. 2) by the proposed method is presented in Fig. 3.



**Fig. 3.** Set of interpolation operations of a synthesis flow diagram

Table 1 summarizes a summary of regulators with desired system quality indicators with the motor, inverter, and conveyor parameters for the purpose of proving the algorithm and program are correct. Motor: Rated power of the electric motor  $P_r = 4.5 \text{ kW}$ ; rated armature voltage  $U_a = 220 \text{ V}$ ; rated speed  $n_r = 1000 \text{ rpm}$ ; rated armature current  $I_a = 25.2 \text{ A}$ ; armature coil resistance  $R_a = 0.632 \Omega$ ; resistive coil resistance  $R_f = 560(\Omega)$ ; motor inertia torque  $J = 2.8 \text{ kg.m}^2$ . Inverter: output voltage  $U_m = 230(V)$ , control voltage  $U_k = 10(V)$ , switching frequency  $f = 500(\text{Hz})$ ,  $K_{In} = 22$ ;  $T_{In} = 0.01 \text{ s}$ . Elastic conveyor with  $q = \frac{1}{2a} = 7$ ,  $\lambda = 1 - \frac{2x_2}{l} = 0, 4$ ,  $\mu_1 = \frac{m_1}{m_k} = 11$ ,  $\mu_2 = \frac{m_2}{m_k} = 0$ . The GUI is shown in Fig. 4.



**Fig. 4.** GUI for automatically synthesizing the regulators

### 3.2 Simulation Results

Table 1 shows the results in each current loop. The transition processes in the multi-loop control system of the regulators met the quality requirements. This proves that the approach proposed in the paper is feasible in synthesizing regulators of electric drive systems containing elastic conveyors, without the use of tension sensors [9–11] or state observers as in previous studies [14–16]. With the proposed simulation and calculation results, it has been proved that the original model describing the conveyor (6), can operate directly without any significant difficulties.

**Table 1.** The parameters of synthesized regulators

Loops	Regulators	Quality criteria					
		Desired system			Synthesized system		
		$t_{ST}^R(s)$	$\sigma_R(\%)$	<i>Settling output</i>	$t_{ST}^S(s)$	$\sigma_S(\%)$	<i>Settling output</i>
Current	$\frac{0.06103s+1.468}{s}$	0.08	4	1 (A)	0.0766	4.0095	1 (A)
Speed	$\frac{2.368s+0.3438}{s}$	0.5	4	3 (rad/s)	0.3881	1.0027	3 (rad/s)
Rotation angle	6.2201	1	12	6 (rad)	0.7134	1.4511	6 (rad)

## 4 Conclusion

In this paper, we have detailed the approach of the research method for the purpose of synthesizing the regulator of the automatic electric drive system. The transfer functions are modeled by the algorithm flowchart, in which the performance of the proposed algorithm is checked through simulation. With the hope that multi-motor automatic electric drive systems can be applied in real production conditions, the algorithm needs to be further improved.

## References

1. Рапопорт Э.Я. Анализ и синтез систем автоматического управления с распределенными параметрами. М. Высш. шк., 2005. – 292 с
2. В.М. Терехов, О.И. Осимов. Системы управления электроприводов. Москва: Издательский центр “Академия”, 2006. – 304с
3. Рассудов Л.Н. Электроприводы с распределенными параметрами механических элементов. Л.: Энергоатомиздат, Ленингр. Отд-ние, 1987. – 144 с
4. Alahakoon, S.: Digital motion control techniques for electrical drives–stockholm, p. 310 (2000)
5. Subaru, H., Chong, S.H., Hee, W.K.: Investigation of model parameter variation for tension control of a multi motor wire winding system. IEEE (2015)
6. Shi, T., Liu, H., Geng, Q., Xia, C.: Improved relative coupling control structure for multi-motor speed synchronous driving system. IET Electr. Power Appl. **10**(6), 451–457. The Institution of Engineering and Technology (2016)
7. Baumgart, M.D., Pao, L.Y.: Robust Lyapunov-based feedback control of nonlinear web-winding systems. In Proceedings of IEEE Conference on Decision and Control, June 2003
8. Abjadi, N.R., Soltani, J., Askari, J., Markadeh, G.A.: Nonlinear sliding-mode control of a multi-motor web-winding system without tension sensor. IET Control Theory Appl. **3**(4), 419–427 (2009)
9. Liu, G., Liu, P., Shen, Y., Wang, F., Kang, M.: Experimental research on decoupling control of multi-motor variable frequency system based on neural network generalized inverse. Supported in part by China ministry fund of education (20050299009) and China Jiangsu Province Nature Science Foud (BK2007094) (2007)

10. Jinmei, L., Xingqiao, L., Guohai, L.: Application of an adaptive controller with a single neuron in control of multi-motor synchronous system. IEEE (2008)
11. Salem, F., Bayoumi, E.H.E.: Robust fuzzy-PID control of three-motor drive system using simulated annealing optimization. J. Electr. Eng. **3**(3), 284–292 (2013)
12. Koc, H., Knittel, D., Mathelin, M.D., Abba, G.: Modeling and robust control of winding systems for elastic webs. IEEE Trans. Control Syst. Technol. **10**(2), 197–208 (2002)
13. Киселев О.Н. Синтез регуляторов низкого порядка по критерию и по критерию максимальной робастности / О.Н. Киселев, Б.Т. Поляк // АиТ. – 1999.- №3. - С. 119–130
14. Glaoui, H., Hazzab, A., Bouchiba, B., Bousserhane, I.K.: Modeling and simulation multi motors web winding system. Int. J. Adv. Comput. Sci. Appl. **4**(2) (2013)
15. Allaoua, B., Laoufi, A., Gasbaoui, B.: Multi-drive paper system control based on multi-input multi-output pid controller. Leonardo J. Sci. **9**(16), 59–70 (2010)
16. Angermann, A., Aicher, M., Schroder, D.: Time-optimal tension control forprocessing plants with continuous moving webs. In: Proceedings of the 35th Annual Meeting—IEEE Industry Applications Society, Rome, October 1999
17. Гончаров В.И. Вещественный интерполяционный метод синтеза систем автоматического управления/ В.И. Гончаров. — Томск: Изд-во ТПУ, 1995. — 108с
18. Гончаров В. И. Синтез электромеханических исполнительных систем промышленных роботов/В. И. Гончаров. — Томск: Изд-во ТПУ, 2002. — 100 с
19. Abutherea, M.A., Lester, D.: Computable function representations using effective Chebyshev polynomial. In: World Academy of Science, Engineering and Technology, pp. 103–109 (2007)
20. Yen, N.T.H., et al.: An improvement for tomographic density imaging using integration of DBIM and interpolation. In: International Conference on Research in Intelligent and Computing in Engineering (RICE), pp. 1–4. IEEE (2018)
21. Tran-Duc, T., Nguyen-Tien, A.: Application of interpolation for DBIM reconstruction of ultrasound tomography. In: International Symposium on Micro-NanoMechatronics and Human Science, pp. 240–243. IEEE (2011)
22. Quang-Huy, T., et al.: Interpolated hybrid DBIM approach for enhanced imaging in ultrasound tomography. Res. Biomed. Eng. (2022). <https://doi.org/10.1007/s42600-021-00192-x>



# A Fuzzy Approximation Supported Model-Free Tracking Control Design for Tower Crane Systems

Nhu Toan Nguyen, Van-Anh Nguyen<sup>(✉)</sup>, Manh Cuong Nguyen, Danh Huy Nguyen,  
and Tung Lam Nguyen

School of Electrical and Electronic Engineering, Hanoi University of Science and Technology,  
Hanoi, Vietnam

[anh.nguyenthivan1@hust.edu.vn](mailto:anh.nguyenthivan1@hust.edu.vn)

**Abstract.** This work proposes model-free controller to deal with problems of load sway phenomenon in a nonlinear tower crane system with unknown components. Because of the swing effect, as well as the indispensable presence of uncertainties, it is really challenging to design an effective model-based controller to ensure tracking performance. Therefore, the proposed controller is constructed based on the adaptive fuzzy approach with the robust Sliding Mode Control (SMC) technique to maneuver this system to reference trajectory and simultaneously suppress payload swing angles. The Lyapunov's stability theory is utilized to demonstrate the system's stability and convergence. The simulation trajectory tracking results show validity and robustness of combined control approach.

**Keywords:** Tower crane · Sliding mode control · Adaptive fuzzy logic control · Lyapunov theory

## 1 Introduction

In recent years, tower cranes are considered as an inevitable hoisting apparatus to transport large-scale cargoes, especially in construction sites, factories, and shipyards. However, many tower crane systems are still manipulated manually, which is difficult to guarantee the high accuracy and efficiency in the operation process. Therefore, an autonomic tower crane system equipped with an effective controller is in great demand.

Regarding a four degrees of freedoms (4-DOF) tower crane system, its fundamental control task is to obtain precise positioning of trolley's translation and jib's rotation, while simultaneously suppress tangential and radical motions of payload below. Nevertheless, tower crane is an underactuated system as fewer control inputs are used to control all independent system states. Besides, its dynamics exhibit high nonlinearity and strong state coupling. The payload normally oscillates considerably while the trolley or jib accelerate or decelerate because of undesirable inertia. In addition, presence of parametric uncertainties and external disturbances are unavoidable, which makes control issues more complicated and challenging.

In order to deal with those problems, tower crane control strategies have been intensively researched. In term of feedforward control techniques, input-shaping schemes are proposed in [1, 2] to eliminate payload swing angles, but they are normally sensitive to external disturbances. Regarding closed-loop control methods, in [3], a nonlinear model predictive control method (NMPC) combined with path-following method is proposed to guarantee the cargo to track the reference path in real time. The advantage of NMPC methods is that the system constraints are effectively solved, but they require system model and computational complexity. Thus, Tuan et al. [4] present two nonlinear controllers including sliding mode control and partial feedback linearization. In [5], an adaptive backstepping SMC is proposed in order to deal with the uncertainties caused by environment interactions. However, these SMC approaches requires the prior knowledge of mathematical dynamic model. Moreover, adaptive control methods are proposed to handle the uncertainties of system parameters [6–8]. Some intelligent control methods are also presented for tower crane system including fuzzy control [9–11], neural network [12, 13]. Because of self-learning capability, these methods can approximate the system uncertainties and disturbances, so the adaptive performance of the system is considerably enhanced.

Inspired by aforementioned control strategies, the adaptive fuzzy SMC controller is designed to achieve accurate trolley/jib tracking trajectories while simultaneously suppress payload oscillation of 4-DOF tower crane system. The proposed method employs a fuzzy system to estimate the unknown components of dynamic model and an adaptive mechanism to present adaptive performance. Therefore, the mathematical information of the system is not required in advance. Besides, Lyapunov method is used to formulate adaptive laws and demonstrate the stability of the system.

## 2 System Modeling

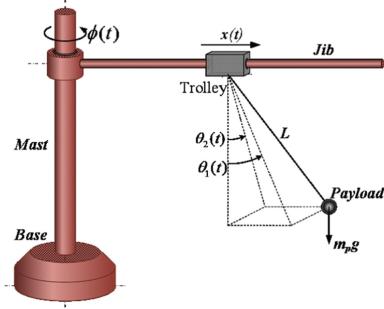
Considering 4-DOF tower crane represented in Fig. 1, three types of motions are involved including translation of trolley  $x$ , rotation of jib  $\phi$ , and two sway motions of payload on cable represented by swing angles  $\theta_1$  and  $\theta_2$ . Therefore, system's state vector is denoted by  $\vartheta = [x, \phi, \theta_1, \theta_2]^T$ . According to [14], the dynamic model is of the following form:

$$\mathbf{M}(\vartheta)\ddot{\vartheta} + \mathbf{N}(\vartheta, \dot{\vartheta})\dot{\vartheta} + \mathbf{D}(\vartheta) = \mathbf{U} \quad (1)$$

where  $\mathbf{M}(\vartheta) \in \mathbf{R}^{4 \times 4}$  is a positive definite inertia matrix,  $\mathbf{N}(\vartheta, \dot{\vartheta}) \in \mathbf{R}^{4 \times 4}$  indicates Coriolis matrices,  $\mathbf{D}(\vartheta) \in \mathbf{R}^{4 \times 1}$  denotes gravity vector,  $\mathbf{U} \in \mathbf{R}^{4 \times 1}$  is control inputs.

Let  $\vartheta_a = [x \ \phi]^T$  and  $\vartheta_u = [\theta_1 \ \theta_2]^T$  be the vectors of actuated states and unactuated states respectively. Therefore, the dynamical Eq. (1) can be separated into two subsystems:

$$\mathbf{M}_{aa}(\vartheta)\ddot{\vartheta}_a + \mathbf{M}_{au}(\vartheta)\ddot{\vartheta}_u + \mathbf{N}_{aa}(\vartheta, \dot{\vartheta})\dot{\vartheta}_a + \mathbf{N}_{au}(\vartheta, \dot{\vartheta})\dot{\vartheta}_u = \mathbf{U}_a \quad (2)$$



**Fig. 1.** Tower crane system.

$$\mathbf{M}_{aa}(\dot{\vartheta})\ddot{\vartheta}_a + \mathbf{M}_{uu}(\dot{\vartheta})\ddot{\vartheta}_u + \mathbf{N}_{ua}(\dot{\vartheta}, \dot{\vartheta})\dot{\vartheta}_a + \mathbf{N}_{uu}(\dot{\vartheta}, \dot{\vartheta})\dot{\vartheta}_u + \mathbf{D}_u(\dot{\vartheta}) = 0 \quad (3)$$

where

$$\mathbf{M}_{aa} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}, \quad \mathbf{M}_{au} = \begin{bmatrix} m_{13} & m_{14} \\ m_{23} & m_{24} \end{bmatrix}, \quad \mathbf{M}_{ua} = \begin{bmatrix} m_{31} & m_{32} \\ m_{41} & m_{42} \end{bmatrix}, \quad \mathbf{M}_{uu} = \begin{bmatrix} m_{33} & m_{34} \\ m_{43} & m_{44} \end{bmatrix},$$

$$\mathbf{D}_u = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

$$\mathbf{N}_{aa} = \begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{bmatrix}, \quad \mathbf{N}_{au} = \begin{bmatrix} n_{13} & n_{14} \\ n_{23} & n_{24} \end{bmatrix}, \quad \mathbf{N}_{aa} = \begin{bmatrix} n_{31} & n_{32} \\ n_{41} & n_{42} \end{bmatrix},$$

$$\mathbf{N}_{au} = \begin{bmatrix} n_{33} & n_{34} \\ n_{43} & n_{44} \end{bmatrix}, \quad \mathbf{U}_a = \begin{bmatrix} u_t \\ u_r \end{bmatrix}$$

More details of these matrices can be found in [14]. Especially,  $u_t$  and  $u_r$  denote the translation force and rotation torque to control the trolley and jib respectively. Rewriting Eqs. (2) and (3) yields:

$$\ddot{\vartheta}_a = \tilde{\mathbf{M}}_a^{-1}(\dot{\vartheta}) \left( -\tilde{\mathbf{N}}_{aa}(\dot{\vartheta}, \dot{\vartheta})\dot{\vartheta}_a - \tilde{\mathbf{N}}_{au}(\dot{\vartheta}, \dot{\vartheta})\dot{\vartheta}_u - \mathbf{D}_a(\dot{\vartheta}) \right) + \tilde{\mathbf{M}}_a^{-1}(\dot{\vartheta})\mathbf{U}_a \quad (4)$$

$$\ddot{\vartheta}_u = \tilde{\mathbf{M}}_u^{-1}(\dot{\vartheta}) \left( -\tilde{\mathbf{N}}_{ua}(\dot{\vartheta}, \dot{\vartheta})\dot{\vartheta}_a - \tilde{\mathbf{N}}_{uu}(\dot{\vartheta}, \dot{\vartheta})\dot{\vartheta}_u - \mathbf{D}_u(\dot{\vartheta}) \right) + \tilde{\mathbf{M}}_u^{-1}(\dot{\vartheta})\mathbf{U}_u \quad (5)$$

where

$$\tilde{\mathbf{M}}_a(\dot{\vartheta}) = \mathbf{M}_{aa}(\dot{\vartheta}) - \mathbf{M}_{au}(\dot{\vartheta})\mathbf{M}_{uu}^{-1}(\dot{\vartheta})\mathbf{M}_{ua}(\dot{\vartheta}); \quad \mathbf{D}_a(\dot{\vartheta}) = -\mathbf{M}_{au}(\dot{\vartheta})\mathbf{M}_{uu}^{-1}(\dot{\vartheta})\mathbf{D}_u(\dot{\vartheta});$$

$$\tilde{\mathbf{M}}_u(\dot{\vartheta}) = \mathbf{M}_{uu}(\dot{\vartheta}) - \mathbf{M}_{ua}(\dot{\vartheta})\mathbf{M}_{aa}^{-1}(\dot{\vartheta})\mathbf{M}_{au}(\dot{\vartheta});$$

$$\tilde{\mathbf{N}}_{aa}(\dot{\vartheta}, \dot{\vartheta}) = \mathbf{N}_{aa}(\dot{\vartheta}, \dot{\vartheta}) - \mathbf{M}_{au}(\dot{\vartheta})\mathbf{M}_{uu}^{-1}(\dot{\vartheta})\mathbf{N}_{ua}(\dot{\vartheta}, \dot{\vartheta});$$

$$\tilde{\mathbf{N}}_{au}(\dot{\vartheta}, \dot{\vartheta}) = \mathbf{N}_{au}(\dot{\vartheta}, \dot{\vartheta}) - \mathbf{M}_{au}(\dot{\vartheta})\mathbf{M}_{uu}^{-1}(\dot{\vartheta})\mathbf{N}_{uu}(\dot{\vartheta}, \dot{\vartheta});$$

$$\tilde{\mathbf{N}}_{ua}(\dot{\vartheta}, \dot{\vartheta}) = \mathbf{N}_{ua}(\dot{\vartheta}, \dot{\vartheta}) - \mathbf{M}_{ua}(\dot{\vartheta})\mathbf{M}_{aa}^{-1}(\dot{\vartheta})\mathbf{N}_{aa}(\dot{\vartheta}, \dot{\vartheta}); \quad \mathbf{U}_u = -\mathbf{M}_{ua}(\dot{\vartheta})\mathbf{M}_{aa}^{-1}(\dot{\vartheta})\mathbf{U}_a;$$

$$\tilde{\mathbf{N}}_{uu}(\dot{\vartheta}, \dot{\vartheta}) = \mathbf{N}_{uu}(\dot{\vartheta}, \dot{\vartheta}) - \mathbf{M}_{ua}(\dot{\vartheta})\mathbf{M}_{aa}^{-1}(\dot{\vartheta})\mathbf{N}_{au}(\dot{\vartheta}, \dot{\vartheta}).$$

### 3 Adaptive Control System Design

#### 3.1 Sliding Mode Control

The primary control objectives are tracking the system's reference trajectories  $\dot{\vartheta}_{ad} = [x_d \ \phi_d]^T$ , while simultaneously eliminating the payload swings,  $\dot{\vartheta}_{ud} = [0 \ 0]^T$ . The error vectors are defined as.

$$\tilde{\vartheta}_a = \dot{\vartheta}_a - \dot{\vartheta}_{ad} = [x - x_d \ \phi - \phi_d]^T \text{ and } \tilde{\vartheta}_u = \dot{\vartheta}_u - \dot{\vartheta}_{ud} = [\theta_1 \ \theta_2]^T \quad (6)$$

where  $\tilde{\vartheta}_a$  and  $\tilde{\vartheta}_u$  denote the error vector of actuated states and unactuated states, respectively. Similar to the SMC design procedure in [4], the vector of sliding function  $s \in \mathbf{R}^2$  is considered as

$$s = \dot{\tilde{\vartheta}}_a + \lambda_a \tilde{\vartheta}_a + \alpha_u \dot{\tilde{\vartheta}}_u + \lambda_u \tilde{\vartheta}_u \quad (7)$$

where  $\lambda_a = \text{diag}(\lambda_{a1}, \lambda_{a2})$ ,  $\lambda_u = \text{diag}(\lambda_{u1}, \lambda_{u2})$ ,  $\alpha_u = \text{diag}(\alpha_1, \alpha_2)$  are opted to ensure the stability of sliding surface. From (6), the sliding function (7) equals:

$$s = \dot{\vartheta}_a + \alpha_u \dot{\vartheta}_u - \dot{\vartheta}_r \quad (8)$$

with  $\dot{\vartheta}_r = \dot{\vartheta}_{ad} + \alpha \dot{\vartheta}_{ud} - \lambda_a (\dot{\vartheta}_a - \dot{\vartheta}_{ad}) - \lambda_u (\dot{\vartheta}_u - \dot{\vartheta}_{ud})$ . Then, the derivative of sliding function can be expressed as

$$\dot{s} = \ddot{\vartheta}_a + \alpha \ddot{\vartheta}_u - \ddot{\vartheta}_r. \quad (9)$$

Substituting (4) and (5) into (9) and yields

$$\dot{s} = \Gamma(\dot{\vartheta}, \dot{\vartheta}) + \Upsilon(\dot{\vartheta}) \mathbf{U}_a - \ddot{\vartheta}_r \quad (10)$$

where

$$\begin{aligned} \Gamma(\dot{\vartheta}, \dot{\vartheta}) &= \tilde{\mathbf{M}}_a^{-1}(\dot{\vartheta}) \left( -\tilde{\mathbf{N}}_{aa}(\dot{\vartheta}, \dot{\vartheta}) \dot{\vartheta}_a - \tilde{\mathbf{N}}_{au}(\dot{\vartheta}, \dot{\vartheta}) \dot{\vartheta}_u - \mathbf{D}_a(\dot{\vartheta}) \right) \\ &\quad + \alpha_u \left( \tilde{\mathbf{M}}_u^{-1}(\dot{\vartheta}) \left( -\tilde{\mathbf{N}}_{ua}(\dot{\vartheta}, \dot{\vartheta}) \dot{\vartheta}_a - \tilde{\mathbf{N}}_{uu}(\dot{\vartheta}, \dot{\vartheta}) \dot{\vartheta}_u - \mathbf{D}_u(\dot{\vartheta}) \right) \right). \\ \Upsilon(\dot{\vartheta}) &= \tilde{\mathbf{M}}_a^{-1}(\dot{\vartheta}) - \alpha_u \tilde{\mathbf{M}}_u^{-1}(\dot{\vartheta}) \mathbf{M}_{ua}(\dot{\vartheta}) \mathbf{M}_{aa}^{-1}(\dot{\vartheta}). \end{aligned}$$

Then, the sliding mode controller is designed as

$$\mathbf{U}_a = \Upsilon^{-1}(\dot{\vartheta}) (\dot{\vartheta}_r - \Gamma(\dot{\vartheta}, \dot{\vartheta}) - \mathbf{K} \text{sign}(s)). \quad (11)$$

#### 3.2 Adaptive Fuzzy Sliding Mode Control

Considering the system parameters are uncertain, therefore  $\Gamma(\dot{\vartheta}, \dot{\vartheta}) \in \mathbf{R}^{2 \times 1}$  and  $\Upsilon(\dot{\vartheta}) \in \mathbf{R}^{2 \times 2}$  in (11) are uncertain. Therefore, the fuzzy system with  $\chi = [s^T \ \dot{s}^T] = [\chi_1 \ \chi_2 \ \chi_3 \ \chi_4]^T$  is employed to approximate these components as follows:

$R^i$ : IF  $\chi_1$  is  $X_1^i$  and ... and  $\chi_4$  is  $X_4^i$  THEN  $z_i$  is  $Z^i$ ,  $i = 1, 2, \dots, h$

where  $\mathbf{s}$  and  $\dot{\mathbf{s}}$  are sliding function and its derivative described in Eqs. (8) and (9),  $h$  indicates number of fuzzy rules. The system output is:

$$z = \frac{\sum_{i=1}^h \prod_{j=1}^4 \mu_{X_j^i}(\chi_j) z_i}{\sum_{i=1}^h \prod_{j=1}^4 \mu_{X_j^i}(\chi_j)} = \boldsymbol{\Phi}^T \boldsymbol{\xi}(\chi) \quad (12)$$

where  $\mu_{X_j^i}(\chi_j)$  is the membership function,  $\boldsymbol{\Phi} = [z_1, \dots, z_h]^T$  is the adjustable parameters

vector,  $\boldsymbol{\xi}(\chi) = [\xi_1, \dots, \xi_h]^T$  with  $\xi_i = \frac{\prod_{j=1}^4 \mu_{X_j^i}(\chi_j)}{\sum_{i=1}^h \prod_{j=1}^4 \mu_{X_j^i}(\chi_j)}$  being fuzzy basis function.

Therefore, the approximation of  $\boldsymbol{\Gamma}(\boldsymbol{\vartheta}, \dot{\boldsymbol{\vartheta}})$  and  $\boldsymbol{\Upsilon}(\boldsymbol{\vartheta})$  in (11) are represented as:

$$\hat{\boldsymbol{\Gamma}}(\chi) = \boldsymbol{\Xi}_{\Gamma}(\chi) \boldsymbol{\Phi}_{\Gamma} \text{ and } \hat{\boldsymbol{\Upsilon}}(\chi) = \boldsymbol{\Xi}_{\Upsilon}(\chi) \boldsymbol{\Phi}_{\Upsilon} \quad (13)$$

where  $\boldsymbol{\Xi}_{\Gamma}(\chi) = \text{diag}(\boldsymbol{\xi}_{\Gamma 1}^T, \boldsymbol{\xi}_{\Gamma 2}^T)$ ,  $\boldsymbol{\Xi}_{\Upsilon}(\chi) = \text{diag}(\boldsymbol{\xi}_{\Upsilon 1}^T, \boldsymbol{\xi}_{\Upsilon 2}^T)$ ,  $\boldsymbol{\Phi}_{\Gamma} \in \mathbf{R}^{(2h) \times 1}$ ,  $\boldsymbol{\Phi}_{\Upsilon} \in \mathbf{R}^{(2h) \times 2}$ . Assuming that  $\chi$ ,  $\boldsymbol{\Phi}_{\Gamma}$  and  $\boldsymbol{\Phi}_{\Upsilon}$  are constrained in the compact set  $\Omega_{\chi}$ ,  $\Omega_{\Gamma}$ ,  $\Omega_{\Upsilon}$  respectively with  $\Omega_{\chi} = \{\chi | \underline{\rho}_{\chi} \leq \chi \leq \bar{\rho}_{\chi}\}$ ,  $\Omega_S = \{\boldsymbol{\Phi}_S | \underline{\rho}_S \leq \boldsymbol{\Phi}_S \leq \bar{\rho}_S\}$ ,  $S = \{\Gamma, \Upsilon\}$  and there exists optimal parameters  $\boldsymbol{\Phi}_{\Gamma}^*$  and  $\boldsymbol{\Phi}_{\Upsilon}^*$  such that

$$\boldsymbol{\Phi}_{\Gamma}^* = \arg \min_{\boldsymbol{\Phi}_{\Gamma}} \left\{ \sup_{\chi \in \Omega_{\chi}} |\boldsymbol{\Gamma}(\chi) - \boldsymbol{\Xi}_{\Gamma}(\chi) \boldsymbol{\Phi}_{\Gamma}| \right\}, \quad \boldsymbol{\Phi}_{\Upsilon}^* = \arg \min_{\boldsymbol{\Phi}_{\Upsilon}} \left\{ \sup_{\chi \in \Omega_{\chi}} |\boldsymbol{\Upsilon}(\chi) - \boldsymbol{\Xi}_{\Upsilon}(\chi) \boldsymbol{\Phi}_{\Upsilon}| \right\}. \quad (14)$$

Define the estimated parameter errors as.

$$\tilde{\boldsymbol{\Phi}}_{\Gamma} = \boldsymbol{\Phi}_{\Gamma} - \boldsymbol{\Phi}_{\Gamma}^* \text{ and } \tilde{\boldsymbol{\Phi}}_{\Upsilon} = \boldsymbol{\Phi}_{\Upsilon} - \boldsymbol{\Phi}_{\Upsilon}^*. \quad (15)$$

The bounded approximation errors for  $\hat{\boldsymbol{\Gamma}}(\chi)$  and  $\hat{\boldsymbol{\Upsilon}}(\chi)$  can be expressed as.

$$\Delta \boldsymbol{\Gamma}(\chi) = \boldsymbol{\Gamma}(\chi) - \hat{\boldsymbol{\Gamma}}(\chi | \boldsymbol{\Phi}_{\Gamma}^*) \text{ and } \Delta \boldsymbol{\Upsilon}(\chi) = \boldsymbol{\Upsilon}(\chi) - \hat{\boldsymbol{\Upsilon}}(\chi | \boldsymbol{\Phi}_{\Upsilon}^*) \quad (16)$$

Throughout this section, the following assumptions are considered:

**Assumption 1:**  $\hat{\boldsymbol{\Upsilon}}(\chi | \boldsymbol{\Phi}_{\Upsilon})$  is invertible.

**Assumption 2:** There exists a positive function  $0 \leq N(\chi) < 1$  and a positive constant  $\varepsilon_{\Upsilon} > 0$  which satisfy  $0 \leq \max_m \left( \sum_{n=1}^2 \left| \Delta \boldsymbol{\Upsilon}(\chi) \hat{\boldsymbol{\Upsilon}}^{-1}(\chi | \boldsymbol{\Phi}_{\Upsilon})_{mn} \right| \right) \leq N_{\Upsilon}(\chi) < 1$  and  $|\lambda_m(\Delta \boldsymbol{\Upsilon}(\chi) \hat{\boldsymbol{\Upsilon}}^{-1}(\chi | \boldsymbol{\Phi}_{\Upsilon}))| \leq \varepsilon_{\Upsilon}$ , where  $\lambda_m(\Delta \boldsymbol{\Upsilon}(\chi) \hat{\boldsymbol{\Upsilon}}^{-1}(\chi | \boldsymbol{\Phi}_{\Upsilon}))$  is the  $m^{th}$  eigenvalue of  $\Delta \boldsymbol{\Upsilon}(\chi) \hat{\boldsymbol{\Upsilon}}^{-1}(\chi | \boldsymbol{\Phi}_{\Upsilon})$ , and  $\Delta \boldsymbol{\Upsilon}(\chi) \hat{\boldsymbol{\Upsilon}}^{-1}(\chi | \boldsymbol{\Phi}_{\Upsilon})_{mn}$  denotes the  $mn^{th}$  element of  $\Delta \boldsymbol{\Upsilon}(\chi) \hat{\boldsymbol{\Upsilon}}^{-1}(\chi | \boldsymbol{\Phi}_{\Upsilon})$ , with  $m = 1, 2$ ,  $n = 1, 2$ .

**Assumption 3:**  $\left| \Delta\Gamma + \Delta\Upsilon\hat{\Upsilon}^{-1}(\vartheta) \left( -\hat{\Gamma}(\vartheta, \dot{\vartheta}) + \ddot{\vartheta}_r \right) \right|_n \leq M_s$ , where  $M_s$  is a positive constant,  $n = 1, 2$ .

**Theorem:** For the uncertain tower crane system (1) with the sliding function (8), the adaptive fuzzy sliding mode controller (AFSMC) given as.

$$\mathbf{U}_a = \hat{\Upsilon}^{-1}(\chi) \left( -\hat{\Gamma}(\chi) + \ddot{\vartheta}_r - \mathbf{K}\text{sign}(\mathbf{s}) \right) \quad (17)$$

and the adaptive parameter adjustment laws as follows.

$$\tilde{\Phi}_{\Gamma}^T = \gamma_{\Gamma} \mathbf{\Xi}_{\Gamma}^T \mathbf{s} \text{ and } \dot{\tilde{\Phi}}_{\Upsilon}^T = \gamma_{\Upsilon} \mathbf{\Xi}_{\Upsilon}^T \mathbf{s} \mathbf{U}_a^T \quad (18)$$

where  $\mathbf{K} = M_s / (1 - \varepsilon_{\Upsilon}) \mathbf{I}_{2 \times 1}$ ,  $\gamma_{\Gamma}$ ,  $\gamma_{\Upsilon}$  are positive adaptation gains, can maneuver the system to equilibrium points.

**Proof:** From (15), (16), and (17), the derivative of sliding function (8) can be obtained as.

$$\begin{aligned} \dot{\mathbf{s}} &= \Gamma + \Upsilon \mathbf{U}_a - \ddot{\vartheta}_r \\ &= \Delta\Gamma + \Delta\Upsilon \mathbf{U}_a - \mathbf{\Xi}_{\Gamma} \tilde{\Phi}_{\Gamma} - \mathbf{\Xi}_{\Upsilon} \tilde{\Phi}_{\Upsilon} \mathbf{U}_a - \mathbf{K}\text{sign}(\mathbf{s}) \end{aligned} \quad (19)$$

The Lyapunov candidate function is chosen as

$$V = \frac{1}{2} \mathbf{s}^T \mathbf{s} + \frac{1}{2\gamma_{\Gamma}} \tilde{\Phi}_{\Gamma}^T \tilde{\Phi}_{\Gamma} + \frac{1}{2\gamma_{\Upsilon}} \tilde{\Phi}_{\Upsilon}^T \tilde{\Phi}_{\Upsilon} \quad (20)$$

Its time derivative is given as

$$\begin{aligned} \dot{V} &= \frac{1}{2} (\mathbf{s}^T \dot{\mathbf{s}} + \dot{\mathbf{s}}^T \mathbf{s}) + \frac{1}{2\gamma_{\Gamma}} \left( \tilde{\Phi}_{\Gamma}^T \dot{\tilde{\Phi}}_{\Gamma} + \dot{\tilde{\Phi}}_{\Gamma}^T \tilde{\Phi}_{\Gamma} \right) + \frac{1}{2\gamma_{\Upsilon}} \text{trace} \left( \tilde{\Phi}_{\Upsilon}^T \dot{\tilde{\Phi}}_{\Upsilon} + \dot{\tilde{\Phi}}_{\Upsilon}^T \tilde{\Phi}_{\Upsilon} \right) \\ &= \frac{1}{2} \left( \mathbf{s}^T \left( \Delta\Gamma + \Delta\Upsilon \mathbf{U}_a - \mathbf{\Xi}_{\Gamma} \tilde{\Phi}_{\Gamma} - \mathbf{\Xi}_{\Upsilon} \tilde{\Phi}_{\Upsilon} \mathbf{U}_a - \mathbf{K}\text{sign}(\mathbf{s}) \right) \right. \\ &\quad \left. + \left( \Delta\Gamma + \Delta\Upsilon \mathbf{U}_a - \mathbf{\Xi}_{\Gamma} \tilde{\Phi}_{\Gamma} - \mathbf{\Xi}_{\Upsilon} \tilde{\Phi}_{\Upsilon} \mathbf{U}_a - \mathbf{K}\text{sign}(\mathbf{s}) \right)^T \mathbf{s} \right) \\ &\quad + \frac{1}{2\gamma_{\Gamma}} \left( \tilde{\Phi}_{\Gamma}^T \dot{\tilde{\Phi}}_{\Gamma} + \dot{\tilde{\Phi}}_{\Gamma}^T \tilde{\Phi}_{\Gamma} \right) + \frac{1}{2\gamma_{\Upsilon}} \text{trace} \left( \tilde{\Phi}_{\Upsilon}^T \dot{\tilde{\Phi}}_{\Upsilon} + \dot{\tilde{\Phi}}_{\Upsilon}^T \tilde{\Phi}_{\Upsilon} \right) \end{aligned} \quad (21)$$

Using adaptive parameter adjustment laws in (18) yields

$$\dot{V} = (\Delta\Gamma + \Delta\Upsilon \mathbf{U}_a - \mathbf{K}\text{sign}(\mathbf{s})) \mathbf{s} \quad (22)$$

Since the Assumptions 2 and 3 are hold, substituting (17) into (22), we have

$$\dot{V} \leq -\frac{M_s}{1 - \varepsilon_{\Upsilon}} \left( \mathbf{I} + \Delta\Upsilon\hat{\Upsilon}^{-1} \right) |\mathbf{s}| + M_s |\mathbf{s}| < 0 \quad (23)$$

Equation (23) indicates that the sliding surface  $\mathbf{s} = 0$  is globally reached in finite time.

**Remark 1.** To guarantee the adaptation parameters  $\Phi_S$ ,  $S = \{\Gamma, \Upsilon\}$  with adaptive adjustment law  $\dot{\Phi}_S$  to be bounded for all  $t \geq 0$  in the constraint set  $\Omega_S$ , the following projection function are used.

$$\dot{\Phi}_S = \begin{cases} 0 & \text{if } \Phi_S > \bar{\rho}_S \text{ and } \dot{\Phi}_S > 0 \\ 0 & \text{if } \Phi_S < \underline{\rho}_S \text{ and } \dot{\Phi}_S < 0 \\ \dot{\Phi}_S & \text{otherwise} \end{cases} \quad (24)$$

**Remark 2.** The sign function in (17) is substituted by a saturation function to remedy the chattering phenomenon around switching surface.

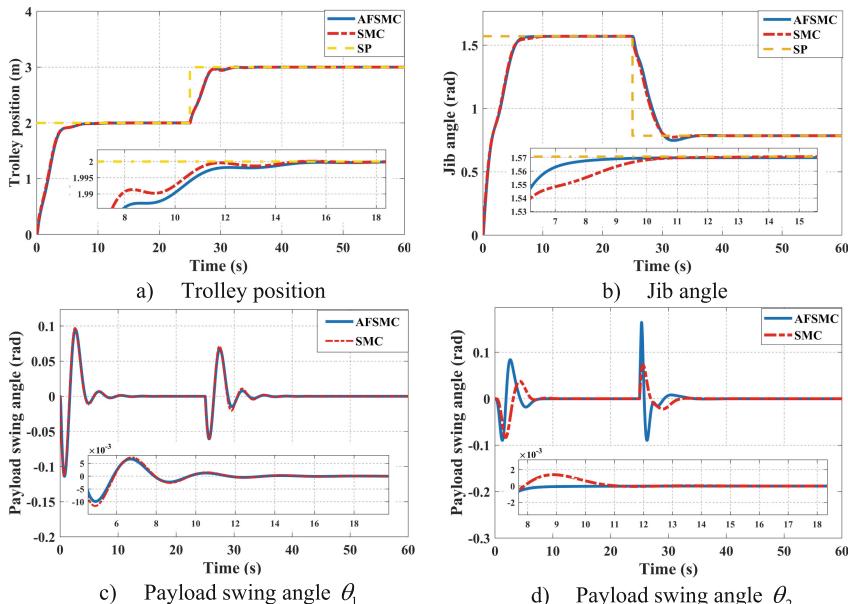
## 4 Simulation Results

Tower crane parameters consisting of the cable length  $l = 3$  m, the trolley and payload mass  $M = 10$  kg,  $m = 5$  kg respectively, and the inertia moment of jib  $J = 15$  kgm<sup>2</sup> are applied for simulation. The membership functions for the fuzzy system are defined as follows.

$$\mu_{P,N}(s) = \exp\left(-\frac{1}{2}\left(\frac{s_i \pm 2}{2}\right)^2\right), \quad \mu_{LP,LN}(\dot{s}) = \exp\left(-\frac{1}{2}\left(\frac{\dot{s}_i \pm 2}{0.5}\right)^2\right)$$

$$\mu_{P,N}(\dot{s}) = \exp\left(-\frac{1}{2}\left(\frac{\dot{s}_i \pm 1.5}{0.5}\right)^2\right), \quad \mu_{SP,SN}(\dot{s}) = \exp\left(-\frac{1}{2}\left(\frac{\dot{s}_i \pm 0.5}{0.5}\right)^2\right),$$

$$i = 1, 2.$$



**Fig. 2.** Tracking trajectories using AFSMC (blue solid line) and SMC (red dash-dotted line) with reference signals (yellow dashed line) (Color figure online)

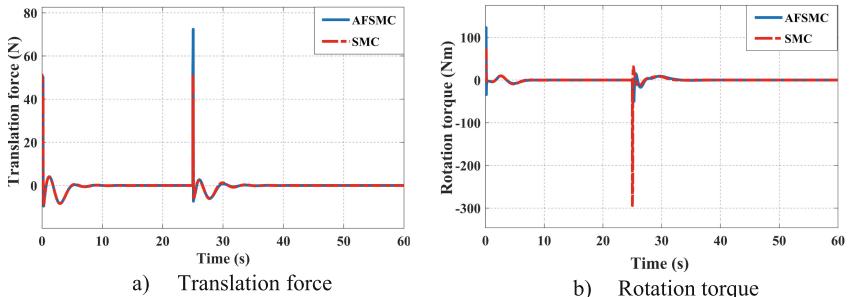
The parameters for proposed controller (21) are chosen as:  $\gamma_{\Gamma} = 10$ ;  $\gamma_{\Upsilon} = 1$ ;  $\lambda_a = \text{diag}(0.5, 0.5)$ ;  $\lambda_u = \text{diag}(-2.5, -2.5)$ ;  $\alpha_u = \text{diag}(0.1, 0.1)$ ;  $\mathbf{K} = \text{diag}(50, 50)$ ;

The constraint sets  $\Omega_{\Gamma}$ ,  $\Omega_{\Upsilon}$  in (24) are chosen as  $\underline{\rho}_{\Gamma 1} = \underline{\rho}_{\Gamma 2} = -200$ ;  $\bar{\rho}_{\Gamma 1} = \bar{\rho}_{\Gamma 2} = 200$

$$\underline{\rho}_{\Upsilon 11} = \underline{\rho}_{\Upsilon 22} = 10; \underline{\rho}_{\Upsilon 12} = 0; \underline{\rho}_{\Upsilon 21} = -10; \bar{\rho}_{\Upsilon 11} = \bar{\rho}_{\Upsilon 22} = 100; \bar{\rho}_{\Upsilon 12} = 10; \bar{\rho}_{\Upsilon 21} = 0;$$

To illustrate the advantages of our proposed controller, we also compare AFSMC with traditional SMC where the system information is fully obtained. Figures 2a and 2b present the trolley position and jib angle tracking process using AFSMC and SMC, respectively.

As can be seen, the two controllers can drive the system to reach the neighborhood of desired point in less than 8 s. The performance of AFSMC is roughly the same as of SMC, however it should be noted that the AFSMC does not require any system information. Besides, the fuzzy system takes time to adapt to the sudden change of setpoint values because its inputs depend on these values. On the other hand, the payload radical and tangential swing angles are also reduced to zeros, depicted in Figs. 2c and 2d, respectively. As compared to SMC, the radical swing angle of AFSMC is lower but its tangential swing angle is higher. The control efforts to translate the trolley and rotate the jib are showed in Fig. 3. Inevitably, the SMC methods usually cause high switching frequency phenomenon in the control signal. Here, control signals of both methods are required to respond rapidly at each time when the references change their values. However, it can be alleviated by setting smoother reference trajectories. The AFSMC requires higher translation force than SMC while the maximum rotation torque of SMC is three times as much as AFSMC.



**Fig. 3.** Control inputs

## 5 Conclusion

This paper presents a model-free fuzzy based sliding mode controller for 4-DOF tower crane system to track the reference trajectory of trolley and jib while simultaneously

suppress payload oscillation. By using adaptive fuzzy system, the controller does not require any system information. The adaptive law is constructed based on Lyapunov method. The simulation results demonstrate the validity and performance of our proposed controller for tower crane system.

**Acknowledgment.** This research is funded by Hanoi University of Science and Technology (HUST) under project number T2021-TT-002.

## References

1. Samin, R.E., Mohamed, Z., Jalani, J., Ghazali, R.: Input shaping techniques for anti-sway control of a 3-DOF rotary crane system. In: 2013 1st International Conference on Artificial Intelligence, Modelling and Simulation, pp. 184–189: IEEE (2013)
2. Lawrence, J., Singhose, W.: Command shaping slewing motions for tower cranes. *J. vibration Acoustics*, vol. 132, no. 1 (2010)
3. Böck, M., Kugi, A.: Real-time nonlinear model predictive path-following control of a laboratory tower crane. *IEEE Trans. Control Syst. Technol.* **22**(4), 1461–1473 (2013)
4. Le, T.A., Dang, V.-H., Ko, D.H., An, T.N., Lee, S.-G.: Nonlinear controls of a rotating tower crane in conjunction with trolley motion. *Proc. Inst. Mech. Eng. Part I: J. Syst. Control Eng.* **227**(5), 451–460 (2013)
5. Bai, W.W., Ren, H.-P.: Horizontal positioning and anti-swinging control tower crane using adaptive sliding mode control. In: 2018 Chinese Control and Decision Conference (CCDC), pp. 4013–4018: IEEE (2018)
6. Zhang, M., Zhang, Y., Ouyang, H., Ma, C., Cheng, X.: Adaptive integral sliding mode control with payload sway reduction for 4-DOF tower crane systems. *Nonlinear Dyn.* **99**(4), 2727–2741 (2020). <https://doi.org/10.1007/s11071-020-05471-3>
7. Chen, H., Fang, Y., Sun, N.: An adaptive tracking control method with swing suppression for 4-DOF tower crane systems. *Mech. Syst. Signal Process.* **123**, 426–442 (2019)
8. Schatz, J., Caverly, R.J.: Passivity-based adaptive control of a 5-DOF tower crane. In: 2021 IEEE Conference on Control Technology and Applications (CCTA), pp. 1109–1114. IEEE (2021)
9. Ramlı, L., Lazım, İ.M., Jaafar, H.I., Mohamed, Z.: Modelling and fuzzy logic control of an underactuated tower crane system. *Appl. Modell. Simul.* **4**, 1–11 (2020)
10. Roman, R.-C., Precup, R.-E., Petriu, E.M.: Hybrid data-driven fuzzy active disturbance rejection control for tower crane systems. *Eur. J. Control.* **58**, 373–387 (2021)
11. Roman, R.-C., Precup, R.-E., Petriu, E.M., Dragan, F.: Combination of data-driven active disturbance rejection and Takagi-Sugeno fuzzy control with experimental validation on tower crane systems. *Energies* **12**(8), 1548 (2019)
12. Hua, H., Fang, Y.: Neural network based adaptive feedback control for tower cranes. In: 2018 IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), pp. 1–5: IEEE (2018)
13. Fasih, S., Mohamed, Z., Husain, A., Ramlı, L., Abdullahi, A., Anjum, W.: Payload swing control of a tower crane using a neural network-based input shaper. *Measurement Control* **53**(7–8), 1171–1182 (2020)
14. Yang, T., Sun, N., Chen, H., Fang, Y.: Observer-based nonlinear control for tower cranes suffering from uncertain friction and actuator constraints with experimental verification. *IEEE Trans. Industr. Electron.* **68**(7), 6192–6204 (2020)



# Control Design of Thermoacoustic Generator Systems Using Reinforcement Learning Algorithm

Duy Tung Le<sup>1</sup>, Trong Tan Do<sup>2</sup>, and Phuong Nam Dao<sup>1(✉)</sup>

<sup>1</sup> Hanoi University of Science and Technology, Hanoi, Vietnam  
nam.daophuong@hust.edu.vn

<sup>2</sup> Center for Micro Electronics and Information Technology, Hanoi, Vietnam

**Abstract.** A novel control design using Reinforcement Learning (RL) algorithm for non-autonomous systems is developed in this article to derive the optimal control law for Thermoacoustic Generator Systems (TAGs). The dynamic programming principle is combined with the computation of partial derivative to achieve the proposed RL scheme. Moreover, to easily derive the proposed method, the model of TAGs is still modified. Finally, it is shown that the numerical studies are carried out to validate the advantage of the proposed controller.

**Keywords:** Thermoacoustic Generator Systems (TAGs) · Reinforcement Learning (RL) algorithm · Non-autonomous systems

## 1 Introduction

Thermoacoustic Generator Systems (TAGs) are the modern solution to provide electricity in the present time [1, 2]. However, because of the disadvantage of time varying essence, it is difficult to find appropriate controller, such as conventional nonlinear control method, etc. As pointed in [3], optimal control is a remarkable solution for many classes of systems because it can obtain not only multi-purposes and minimization of proposed cost function. Still it is impossible fulfill directly optimal control scheme in general situation because the difficulty in solving Hamilton-Jacobi-Bellman (HJB) equation [3–7]. Thus, reinforcement learning algorithm using dynamic programming is discussed in many references with the idea of considering iteration algorithm between estimated optimal controller and estimated Bellman function [3–7]. The traditional approach of Actor-Critic strategy was mentioned in [3, 4, 6] with the purpose of finding learning law of the weight obtaining from minimizing the Hamiltonian function. Moreover, the technique using Kronecker product was utilized in linear systems for developing RL technique. However, almost approaches of RL technique are carried out for autonomous systems with the advantage of eliminating the term of partial derivatives [3–6]. The efforts of solving non-autonomous systems by reinforcement learning algorithm are implemented in [8, 9] by computing the term of partial derivatives after deriving the Dynamic Program principle. However, the remarkable RL technique has not been fulfilled for TAGs yet and this development is validated in this brief.

## 2 Problem Statements

Consider the mathematical model of Thermoacoustic Generator Systems (TAGs) [1, 2] to represented as follows:

$$\begin{cases} \dot{x} = \frac{\gamma\omega x}{A^2 u} \sin(\omega t + \varphi) \\ y = \frac{A^2 \gamma x}{\rho_M A_{ALT}^2} \left[ \frac{(Bl)^2}{R_e + R_L} + R_m \right] Z_{HX}^3 \sin(\omega t + \theta) \end{cases} \quad (1)$$

where  $\rho_M$ ,  $\gamma$ ,  $\omega$ ,  $A$  are the private volume ( $kg/m^3$ ), the adiabatic ratio of the gas, angular frequency ( $s^{-1}$ ) and cross section of Regenerator, respectively. Moreover,  $A_{ALT}$ ,  $\left[ \frac{(Bl)^2}{R_e + R_L} + R_m \right]$ ,  $Z_{HX}$  are diaphragm cross section ( $m^2$ ), total impedance ( $\Omega$ ), equivalent impedance of Heat Exchanger ( $\Omega$ ), respectively, and the angles  $\varphi, \theta$  are displayed as:

$$\begin{aligned} \varphi &= \arctan \left( r_0 \sqrt{\frac{\omega \rho_M}{2\mu}} - 1 \right) \\ \theta &= \arctan \frac{-2 \frac{AC}{B^2+C^2} \left( D + Z_{HX} + \frac{AB}{B^2+C^2} \right)}{\left( D + Z_{HX} + \frac{AB}{B^2+C^2} \right)^2 - \left( \frac{AC}{B^2+C^2} \right)^2} \end{aligned}$$

where  $A = \frac{\omega \rho_M l_{stack}}{A_g}$ ;  $B = 2 - r_0 \sqrt{\frac{\omega \rho_M}{2\mu}}$ ;  $C = 2 \left( -1 + \frac{1}{r_0} \sqrt{\frac{2\mu}{\omega \rho_M}} \right)$ ;  $D = \frac{Z_{HX}^2}{A_{ALT}^2} \left[ \frac{(Bl)^2}{R_e + R_L} + R_m \right]$ .

Based on the control input signal  $u = \frac{T_h - T_c}{T_c}$ , and state variable  $x = p_1$ , output signal  $y = P_{out}$  and defining  $\gamma\omega = M$ ,  $\frac{A^2 \gamma}{\rho_M A_{ALT}^2} \left[ \frac{(Bl)^2}{R_e + R_L} + R_m \right] Z_{HX}^3 = L$ , the state space representation is described as:

$$\begin{cases} \dot{x} = M \frac{x}{u} \sin(\omega t + \varphi) \\ y = Lx \sin(\omega t + \theta) \end{cases} \quad (2)$$

The control purpose is to get the input signals satisfying that minimize the following cost function:

$$J(x_0, t_0, u(t)) = \int_{t_0}^{\infty} L(x(\tau), u(\tau), \tau) d\tau \quad (3)$$

where  $L = (x - x[i])^2 + (y - y[i])^2 + (u - u[i])^2$ .

## 3 Reinforcement Learning Based Control Design for TAGs

The section considers the application of following algorithms in TAG systems under the consideration of non-autonomous systems:

*Algorithm 1:*

*Step 1: Initializing  $V^0(.) = 0$ ,  $i = 0$ .*

*Step 2: Solving the appropriate input signal  $v^i(.)$  from optimization problem*

$$v^i(x(t)) = \arg \min_{u(t) \in U} \left\{ H \left( x(t), V_x^i(x(t), t), u(t), t \right) \right\} \quad (4)$$

with  $V^i \in W$ .

*Step 3: Obtaining the Bellman function  $V^{i+1}(.)$  after solving the equation:*

$$V_\tau^{i+1}(x(\tau), \tau) + H \left( x(\tau), V_x^{i+1}(x(\tau), \tau), v^i(x(\tau), \tau) \right) = 0 \quad (5)$$

with  $V^{i+1}(x_0, \tau_0) = 0$  and  $v^i \in U$ .

*Step 4: Coming back step 2 and 3 until obtaining  $V^*$  and  $v^*$ .*

In order to handle the non-autonomous systems without the knowledge of model, the algorithm 1 is modified to the algorithm 2 as follows:

*Algorithm 2:*

*Step 1: Initializing the input signal  $u_0$  and the initial state  $x_0$ ,  $i = 1$ .*

*Step 2: Check the condition*

$$\left\| \bar{l}^i - \bar{l}^{i-1} \right\| \leq \varepsilon (\varepsilon > 0). \quad (6)$$

- If it satisfies (6) then stop the algorithm;
- If it does not satisfy (6), then implementing the step 4.

*Step 3: Implementing  $j := (j + 1)$ , we obtain  $\bar{w}^{j+1}, \bar{d}^{j+1}, \bar{k}^{j+1}$  after compute that*

$$\left[ \begin{pmatrix} \bar{w}^j \\ \bar{d}^j \\ \bar{k}^j \end{pmatrix}^T, \left( \bar{k}^j \right)^T \right] = \sum_{k=1}^N \left( \int_{t_{k-1}}^{t_k} L \left( x, \left( \bar{l}^{j-1} \right)^T \psi(x), s \right) ds \right) \left( \Theta_k^j \right)^T \left( \sum_{k=1}^N \Theta_k^j \left( \Theta_k^j \right)^T \right)^{-1} \quad (7)$$

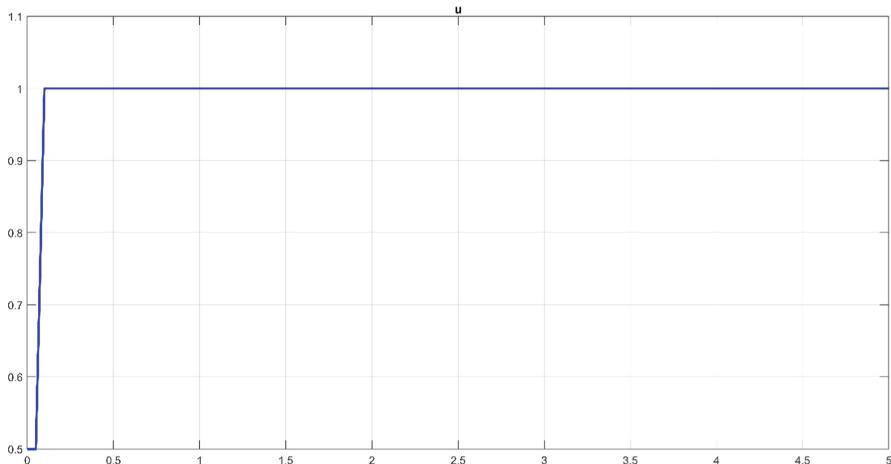
*Solving the optimization problem:*

$$\bar{y}_i = \arg \min_y \left\{ \left( \bar{k}^i \right)^T \varphi \left( x, y^T \psi(x), t \right) + L \left( x, y^T \psi(x), t \right) \right\} \quad (8)$$

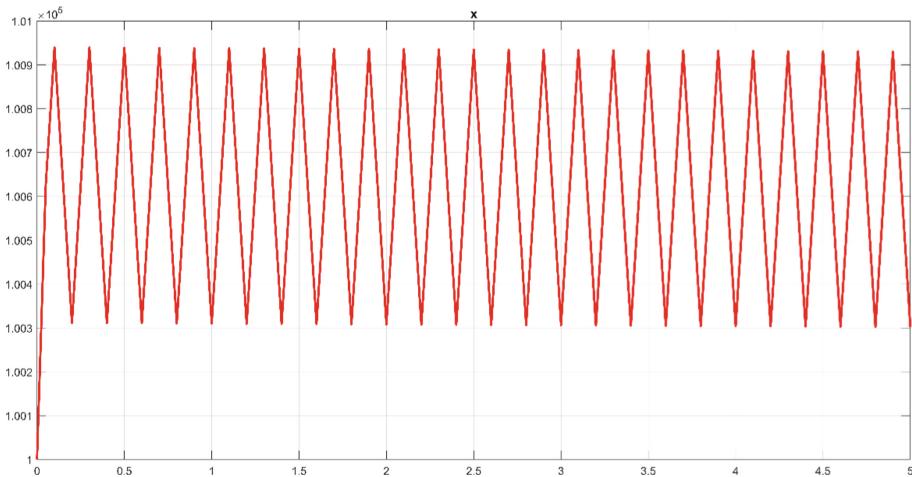
*to obtain  $\bar{y}^{i+1}$ . Coming back step 2.*

## 4 Simulation Studies

The proposed optimal algorithm is validated for TAG model with the sampling time  $t_s = 0.05(s)$ , and the initial state, control input are:  $x_0 = 10^5(Pa)$ ,  $u_0 = 0.5$ . It implies the simulation results to be shown in Fig. 1, 2 with the optimal controllé and appropriate state trajectory.



**Fig. 1.** The optimal controller



**Fig. 2.** The state trajectory under optimal controller

## 5 Conclusion

This paper has studied the reinforcement learning strategy of TAG model subject to non-autonomous systems. First, an appropriate model of TAG has been established, which implies the non-autonomous closed system. Second, by utilizing the reinforcement learning technique for non-autonomous systems, the control system of TAG model is validated through simulation.

**Acknowledgement.** Duy Tung Le was funded by Vingroup Joint Stock Company and supported by the Domestic Master/ PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), Vingroup Big Data Institute (VINBIGDATA), code VINIF.2020.ThS.22.

## References

1. Do, T.T., Le, D.T., Dao, P.N., Bui, M.D., Dao, H.D.: Mathematical modelling of thermoacoustic generator systems and simulation studies. In: Sattler, K.U., Nguyen, D.C., Vu, N.P., Long, B.T., Puta, H. (eds.) Advances in Engineering Research and Application. ICERA 2020. LNNS, vol. 178, pp. 534–540. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-64719-3\\_59](https://doi.org/10.1007/978-3-030-64719-3_59)
2. Backhaus, S., Tward, E., Petach, M.: Traveling-wave thermoacoustic electric generator. *Appl. Phys. Lett.* **85**(6), 1085–1087 (2004)
3. Dao, P.N., Nguyen, H.Q., Ngo, M.D., Ahn, S.J.: On stability of perturbed nonlinear switched systems with adaptive reinforcement learning. *Energies* **13**(19), 5069 (2020)
4. Vu, V.T., Dao, P.N., Loc, P.T., Huy, T.Q.: Sliding variable-based online adaptive reinforcement learning of uncertain/disturbed nonlinear mechanical systems. *J. Control Autom. Electr. Syst.* **32**(2), 281–290 (2021). <https://doi.org/10.1007/s40313-020-00674-w>
5. Dao, P.N., Liu, Y.-C.: Adaptive reinforcement learning strategy with sliding mode control for unknown and disturbed wheeled inverted pendulum. *Int. J. Control. Autom. Syst.* **19**(2), 1139–1150 (2020). <https://doi.org/10.1007/s12555-019-0912-9>
6. Dao, P.N., Do, D.K., Nguyen, D.K.: Adaptive reinforcement learning-enhanced motion/force control strategy for multirobot systems. *Math. Probl. Eng.* 2021 (2021)
7. Wei, Q., Liao, Z., Yang, Z., Li, B., Liu, D.: Continuous-time time-varying policy iteration. *IEEE Trans. Cybern.* **50**(12), 4958–4971 (2019)
8. Sun, T., Sun, X.M.: An adaptive dynamic programming scheme for nonlinear optimal control with unknown dynamics and its application to turbofan engines. *IEEE Trans. Ind. Inform.* **17**(1), 367–376(2020)
9. Sun, J., Zhang, H., Yan, Y., Xu, S., Fan, X.: Optimal regulation strategy for nonzero-sum games of the immune system using adaptive dynamic programming. *IEEE Trans. Cybern.* (2021)



# Application of Intrinsic Time-Scale Decomposition to Classify EEG Signals Using Machine Learning and Neural Network Methods

Viet Quoc Huynh<sup>1,2</sup>, Yen Hua-Thi-Hoang<sup>1,2</sup>, and Tuan Van Huynh<sup>1,2(✉)</sup>

<sup>1</sup> Department of Physics and Computer Science, Faculty of Physics and Engineering Physics, University of Science, Ho Chi Minh, Vietnam

[hvtuan@hcmus.edu.vn](mailto:hvtuan@hcmus.edu.vn)

<sup>2</sup> Vietnam National University, Ho Chi Minh, Vietnam

**Abstract.** In this research, intrinsic time-scale decomposition (ITD) approach was demonstrated as a nonlinear method that could be apply in the EEG signals. Many types of classification algorithms including machine learning and neural networks had been used to estimated the extraction capabilities of ITD method. The results indicated that the classification accuracy using the ITD approach were about 80.0% and had higher result when compare to other methods with the same dataset.

**Keywords:** Intrinsic time-scale decomposition · Neural network · Machine learning

## 1 Introduction

Electroencephalography was the signal generated in the brain due to the communication of a large number of neurons among each other. Based on that, researchers could measure these signals through specialized electrodes.

There were numerous methods to extracted the features of this signals as the frequency or time-domain and nonlinear analysis have been proposed [1–5]. However, the characteristic of the EEG signal were nonlinear. So that nonlinear features characterize this signals more accurately than linear models [6]. There were many signal processing methods were applied including empirical mode decomposition [7], local mean decomposition [8]. But the disadvantage of the empirical mode decomposition method was distorted components and time-consuming process [9]. To overcome these problems, the ITD method was proposed. It extracted proper rotation components (PRCs) [10] by getting temporal information, ITD balances the trade-offs in frequency-domain and time-domain [10].

In this research, ITD method was applied to extract proper rotations of each channel then kurtosis, mean, skewness, average power and standard deviation

feature were calculated. Next, this article used six different classification method as neural network (long short-term memory (LSTM), artificial neural network (ANN)), and machine learning (support vector machine method (SVM), ensemble method (EM), decision tree method (DT), k-nearest neighbors method (k-nn)). The results show that the ITD algorithm could be used to extract features suitable for classify algorithms with an highest average result about 80.00%.

## 2 Materials and Methods

### 2.1 WAY-EEG-GAL [11]

The publishers of WAY-EEG-GAL dataset were Matthew D Luciw and partners [11]. This dataset is intended for testing of approaches for decoding an grasp-and-lift action from the EEG recordings. Twelve healthy volunteers recording the EEG signals. The process used 32 electrodes channels, and the sample rate 500 Hz [11]. The volunteers were instructed to reach the object, grip it with the two finger, raise and hold it until the light go off, then return the object on the surface, and take the hand to first position in each trial [11]. In this research, three labels are established as “Label1”, “Label2” and “Label3” based on the recorded EEG signals when the participant lifts different weights of 165 g, 330 g and 660 g [12].

### 2.2 Feature Extraction

From the data of [11], The Butterworth filter was design with the band-pass in frequency range 3 Hz 70 Hz with the order applied was two to canceled the noise signal. After that, the ITD method was used.

The ITD method process can be summarized by the following steps:

- The local extrema were found from the signal.
- When the operators  $B(t)$  (baseline signal) and  $H(t)$  (PRCs) was calculated, the operator for extraction the baseline is defined as [13]:

$$LI(t) = B(t) = B_k + \frac{((B_{(k+1)} - B_k)}{(I_{(k+1)} - I_k)} * (I(T) - I_k), \tau \in [\tau_k, \tau_{k+1}] \quad (1)$$

where

$$B_{(k+1)} = \beta[I_k + \frac{((\tau_{(k+1)} - \tau_k)}{(\tau_{(k+2)} - \tau_k)}(I_{(k+2)} - I_k)] + (1 - \beta)I_{(k+1)} \quad (2)$$

- The extracted residual signal process using operator  $\Delta$  is defined as [13]:

$$\Delta I(t) \equiv (1 - L)I(t) = I(t) - B(t) \quad (3)$$

Repeat steps one through three until with the new baseline signal the method can get any PRCs.

Since the EEG waveforms were often subdivided into five bandwidths. So that, the first five PRCs were chosen for further feature extraction. These PRCs were extracted the following indices: mean, standard deviation, skewness, kurtosis and average power.

### 3 Classification Methods

For the classification step, all the indices vectors were divided at 80% for the train set 20% for the test set. As the model for the classification, the parameter of each machine learning method were introduced in Tables 1.

**Table 1.** The initial of machine learning method used in this article

Model	Preset	Parameter
k-nn	k-nn fine	No neighbors=1/Euclidean/Equal
	k-nn medium	No neighbors=10/ Euclidean/ Equal
	k-nn coarse	No neighbors=100/ Euclidean/ Equal
	k-nn cosine	No neighbors=10/ Euclidean/ Equal
	k-nn cubic	No neighbors=10/ Euclidean/ Equal
	k-nn weighted	No neighbors=10/ Euclidean/ Squared inverse
SVM	SVM linear	Linear / One-vs-one/ Scale=Automatic
	SVM quadratic	Quadratic/ Ove-vs-one/ Scale=Automatic
	SVM cubic	Cubic/ Ove-vs-one/ Scale=Automatic
	SVM Gaussian Fine	Gaussian/ One-vs-one/ Scale=Automatic
	SVM Medium Gaussian	Gaussian/ One-vs-one/Scale=Automatic
	SVM Coarse Gaussian	Gaussian/ One-vs-one/ Scale=Automatic
DT	Tree fine	Max splits=100/ Gini
	Tree medium	Max splits=20/ Gini
	Tree coarse	Max splits=40/ Gini
EM	Boosted tree	Adaboost
	Bagged tree	Bag
	Subspace k-nn	Subspace
	Rusboosted trees	Rusbossted

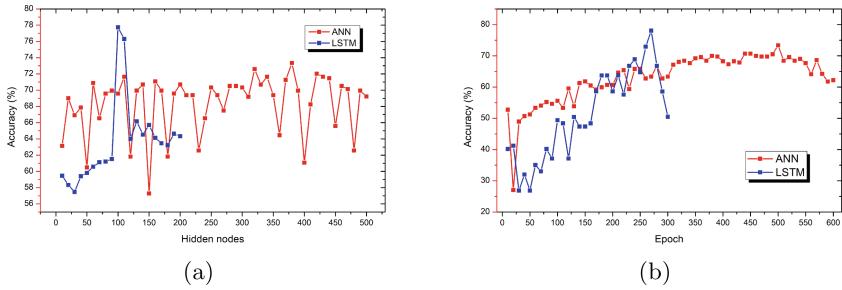
The ANN was then designed with two layers because the processing layers are used to deal with the non-linearity and complexity of the problem. [14].

Next, the LSTM network designed with two layer Additional, the state activation function was Sigmoid function and the gate activation function was Tanh function. The training algorithms and optimization algorithm respectively was the backpropagation and Levenberg-Marquardt algorithm for both models. In this article, the data was divided into train and test set using cross-validation method with the fold equalled to ten.

## 4 Results and Discussions

### 4.1 Results of Performance Using ITD Features Extraction for Each Classification Algorithms

The average classification accuracy of 12 subjects were surveyed through the ANN and LSTM algorithms with different number of nodes and epoch. The results were show in Fig. 1:



**Fig. 1.** The effect of nodes (a) and epoch (b) to the accuracies of ANN and LSTM model

Based on Fig. 1-a, the ANN model had the highest classification when the number nodes reached 380 nodes. For the range of 10 nodes to 350 hidden nodes, the accuracy is fluctuated around from 57.0% to 70.0%. The best classified performance was 380 nodes. Then the accuracy shows signs of decreasing. For the LSTM model, the characteristic was similar in ANN model. The range from 10 to 90 nodes, the classification accuracy was change from 30.0% to 66.0%. After that, the highest accuracy was achieved when reach 100 nodes and the classify performance decreased when the hidden nodes were increased.

In Fig. 1-b, for the ANN model, the number of epochs that best for the classification were 500 epochs and 229 epochs for the LSTM model. It could be seen that the classification accuracy increased as the number of epochs increased because the training process was longer, so the features were trained more accurately.

Base on that, the results proved that the PRC extraction method could be used to train models with not too large and complex architecture. In the next survey, SVM, k-nn, decision tree and ensemble method were used for the classified the EEG signals. The results were presented in Table 2

**Table 2.** Classification results of different types of models

Model	Accuracy (%)	Time (second)	speed
SVM Linear	62.75	1135.80	2400
SVM Cubic	52.57	1126.00	300
SVM Quaratic	60.23	1130.00	340
<b>Fine SVM</b>	<b>64.32</b>	<b>1338.70</b>	<b>250</b>
SVM Medium	55.19	1410.40	260
SVM Coarse	54.06	1562.60	310
<b>Tree Fine</b>	<b>61.92</b>	<b>834.07</b>	<b>3900</b>
Tree Medium	51.67	455.67	3900
Tree Coarse	61.90	1178.32	8800
<b>k-nn Fine</b>	<b>71.47</b>	<b>1820.92</b>	<b>300</b>
k-nn Medium	66.07	1872.53	320
k-nn Coarse	52.63	1928.41	320
k-nn Cosine	65.01	1991.50	330
k-nn Cubic	55.50	3937.82	8
k-nn Weight	67.40	2057.00	350
Boosted Tree	61.81	3518.50	3600
Bagged Tree	62.82	3643.69	3300
<b>Subspace k-nn</b>	<b>77.41</b>	<b>4524.12</b>	<b>24</b>
RusBoosted Tree	50.86	4928.32	1600

As can be seen in Table 2, the fine SVM model gives the most accurate classification and the coarse SVM model has the lowest classification. For the time-training, the SVM medium have the longest training time and the SVM Cubic were the fastest when compare together. This research based on the speed (obs/sec) for the evaluation Therefore, this research would choose the SVM fine model for the optimization process. Just like the survey processed for the SVM model, the model selected here was a tree fine. Because this was the model with the highest accuracy and the time-training could be accepted. After that, k-nn fine, k-nn medium, k-nn coarse, k-nn cosine, k-nn cubic and k-nn weight were used in this research. Then k-nn fine were chosen for best k-nn classification model with the accuracy reach 71.47% and the time-training were 1820 1820 second. Moreover, the results in the time and speed shown that the calculation of each k-nn were similar. Finally, the subspace k-nn has the superiors' classification the EEG signal (77.41%) when compare to other machine learning method. The Rusboosted tree had the lowest classification accuracy and the time consume were the longest.

## 4.2 General Performance of the Method

Many models/algorithms had been developed to recognize emotional states in the last few years. Tianjun Liu and DelingYang used the three branch 3D convolutional neural network (two branch, three branch and four branch) for classification the “LiftOff” event with the 3D data and crop strategy method for 3D image of the signals [15]. Giulia Cisotto et al. used SVM algorithm for the classification and fast Fourier transform (FFT) for the features extraction, the research also used consensus clustering method for the features selection step [16]. An overview of the results was presented in Table 3.

**Table 3.** Classification results comparing with other methods

Author	Method	Accuracy (%)
Giulia Cisotto et al.	SVM (Linear) + consensus	59.10
<b>Our method</b>	<b>DT+ITD</b>	<b>61.9</b>
<b>Our method</b>	<b>SVM+ITD</b>	<b>64.32</b>
Giulia Cisotto et al.	SVM (Radial Basis Function) + consensus	65.70
<b>Our method</b>	<b>k-nn+ITD</b>	<b>71.47</b>
<b>Our method</b>	<b>ANN+ITD</b>	<b>73.72</b>
Tianjin Liu, DelingYang	CNN3D(two branch)	74.70
Tianjun Liu, DelingYang	CNN3D (four branch)	76.90
Tianjun Liu, DelingYang	CNN3D (three branch)	77.10
<b>Our method</b>	<b>Subspace k-nn+ITD</b>	<b>77.41</b>
<b>Our method</b>	<b>LSTM +ITD</b>	<b>80.52</b>

As Table 3, the proposed LSTM model and ensemble method (subspace k-nn) combined with ITD method had the highest classification result (80.52% and 77.41%). Next, the proposed method using ANN and k-nn classification method had better classification performance than SVM and DT method. Moreover, the SVM of the article had higher classification accuracy when compare with the same SVM method of Giulia Cisotto research (5.22% for linear function). This comparison had prove the ITD method was suitable than FFT method. Because each PRC from the ITD algorithm kept all of the original signal’s polar. Addition, the instantaneous frequency and amplitude were well defined so that it overcome the limitations of Fourier transform.

## 5 Conclusion

In this research, our proposed method using the ITD for the features extraction had some positive results. The LSTM+ITD method had the best results (80.52%). The results also proved that the ITD method could be used for the classification lift and grab using EEG signal. For the classification method, the research had survey some of the machine learning classification as SVM, k-nn, DT and EM method. The results shown that the fine SVM (64.32%), fine decision tree (61.92%), fine k-nn (71.47%) and the ensemble method (77.41%) were

suitable method when compare with other type of the same classification algorithm. For the ANN and the LSTM model, the architecture of the models was investigated to find the most suitable accuracy of each model. Finally, the proposed methods compared with the aforementioned approaches, and the results proved that the ITD method was effective for the EEG classification.

**Acknowledgement.** This research is funded by Vietnam National University, Ho Chi Minh City (VNU-HCM) under grant number C2022-18-13.

## References

1. Andrzejak, R.G., et al.: Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. *Phys. Rev. E* **64**(6), 061907 (2001)
2. Samiee, K., Kovacs, P., Gabbouj, M.: Epileptic seizure classification of EEG time-series using rational discrete short-time Fourier transform. *IEEE Trans. Biomed. Eng.* **62**(2), 541–552 (2014)
3. Hassan, A.R., et al.: Epileptic seizure detection in EEG signals using tunable Q factor wavelet transform and bootstrap aggregating. *Comput. Methods Programs Biomed.* **137**, 247–259 (2016)
4. Kafashan, M., et al.: EEG dynamical correlates of focal and diffuse causes of coma. *BMC Neurol.* **17**(1), 1–11 (2017)
5. Lahmiri, S.: Generalized Hurst exponent estimates differentiate EEG signals of healthy and epileptic patients. *Phys. A: Stat. Mech. Appl.* **490**, 378–385 (2018)
6. Wang, C., Hill, D.: Deterministic learning and rapid dynamical pattern recognition. *IEEE Trans. Neural Netw.* **18**(3), 617–630 (2007)
7. Bajaj, V., Pachori, R.B.: Classification of seizure and non-seizure EEG signals using empirical mode decomposition. *IEEE Trans. Inf. Technol. Biomed.* **16**(6), 1135–1142 (2011)
8. Park, C., et al.: The complex local mean decomposition. *Neuro-Comput.* **74**(6), 867–875 (2011)
9. Li, Y., et al.: Rotating machine fault diagnosis based on intrinsic characteristic scale decomposition. *Mech. Mach. Theory* **94**, 9–27 (2015)
10. Frei, M.G., Osorio, I.: Intrinsic time-scale decomposition: time-frequency-energy analysis and real-time filtering of non stationary signals. *Proc. R. Soc. A: Math. Phys. Eng. Sci.* **463**(2078), 321–342 (2007)
11. Luciw, M.D., Jarocka, E., Edin, B.B.: Multi channel EEG recordings during 3,936 grasp and lift trials with varying weight and friction. *Sci. Data* **1**(1), 1–11 (2014)
12. Várszegi, K.: Comparison of algorithms for detecting hand movement from EEG signals. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE (2016)
13. Feng, Z., et al.: Joint amplitude and frequency demodulation analysis based on intrinsic time-scale decomposition for planetary gearbox fault diagnosis. *Mech. Syst. Signal Process.* **72**, 223–240 (2016)
14. Subasi, A., Ercelebi, E.: Classification of EEG signals using neural network and logistic regression. *Computer Methods Programs Biomed.* **78**(2), 87–99 (2005)
15. Liu, T., Yang, D.: A three-branch 3D convolutional neural network for EEG based different hand movement stages classification. *Sci. Rep.* **11**(1), 1–13 (2021)
16. Sarlo, G.L., Holton, K.F.: Brain concentrations of glutamate and GABA in human epilepsy: a review. *Seizure* **91**, 213–227 (2021)



# Development of Economical and Effective Ultraviolet (UVC) Irradiation Solutions and Equipment for Rapid Decontamination of SARS-CoV-2 on N95 Respirators

Nguyen Manh Cuong<sup>1</sup>(✉), Vuong Tri Tiep<sup>1</sup>, Nguyen Phan Kien<sup>2</sup>, Hoang ChuDuc<sup>3</sup>, and Gray K. Aleksanyan<sup>4</sup>

<sup>1</sup> Le Quy Don Technical University, Hanoi, Vietnam

cuongnm@bme.edu.vn

<sup>2</sup> Hanoi University of Science and Technology, Hanoi, Vietnam

<sup>3</sup> VNU University of Engineering and Technology, Hanoi, Vietnam

hoangcd@vnu.edu.vn

<sup>4</sup> Platov South-Russian State Polytechnic University, Novocherkassk, Russia

**Abstract.** N95 disposable respirator masks are of particular importance to the Covid-19 pandemic. The high cost and limited supply of N95 disposable respirators promote research and safe and effective methods of reusing medical masks. The CDC of USA has announced that ultraviolet (UVC) irradiation inactivates SARS-CoV-2, virus, and other microorganisms known to culture on N95, as well as the results affecting mask fit and filterability. In this study, we analyze and evaluate the pathogen inactivation mechanism and the performance of respirators after treatment and perfect the method of ultraviolet irradiation (UVC) to help inactivate of SARS-CoV-2. At the same time, the research team successfully designed, fabricated and tested a semi-automatic system with UVC capacity 0.15 mW/cm<sup>2</sup> at 220 nm that inactivated SARS-CoV-2 (3 log reduction) substances analogues of viruses, and other microorganisms grown on N95. The research results aim to commercialize the system technology (1) to inactivate the SARS-CoV-2 virus through the our application of UVC irradiation at the appropriate wavelength and effective dose, and (2) maintain the suitability and N95 filter rate.

**Keywords:** SARS-CoV-2 · Ultraviolet irradiation · N95 respirator masks

## 1 Introduction

The Covid-19 pandemic is a severe acute respiratory distress syndrome SARS-CoV-2 that causes damage mainly to the lungs of patients [1]. SARS-CoV-2 is the name of the virus strain that is currently a global pandemic that causes lung damage and is the cause of death of more than 5 million people. It is classified in the genus Beta coronavirus of SARS-CoV and the Middle East respiratory syndrome virus (MERS-CoV) [2]. Since 2019, SARS-CoV-2 has developed through four waves and is present in 221 countries,

causing over 486 million cases and over 6.1 million deaths. The high number of cases, and especially the transmission of copy pastes between subjects who have been in close contact with infected patients, raise concerns about emerging infections. Solution or current device to remove SARS-CoV-2 from material surfaces likely to come into contact with humans evaluated by scientific studies virus viability on multiple surfaces over a period of time significantly. For example, SARS-CoV-2 viruses can survive and infect when left on plastic, steel or metal surfaces for 72 h or on cardboard after nearly 1 day [3]. Viral loads of patients infected with SARS-CoV-2 causes of virus release into the environment as determined by RT-PCR analysis of COVID-19 patient rooms, aisles and other locations in medical facilities showing frequent infection Viral RNA on surfaces [4].

However, in some studies, the frequency of SARS-CoV-2 infection on surfaces may be lower than reported. Outside the hospital and clinic environment, accommodation facilities and hotels also have the presence of RNA viruses on table and door surfaces [5]. Viral RNA has also been found on doors and countertops in rooms where people infected with SARS-CoV-2 reside [6]. The WHO has issued guidance and recommendations on surface cleaning or disinfection in areas where SARS-CoV-2 infection may occur. Due to the importance of respiratory protection and the shortage of personal protective equipment, methods, equipment for sterilization and reuse of N95 respirator respirators are also of considerable interest [7]. Ultraviolet light (CUV) as well as other methods have been proposed or tested as a means to disinfect N95 respirators and other personal protective equipment such as plastic masks (PPE) [8].

Chemical and alcohol disinfection of SARS-CoV-2 applies only to small surfaces and is only responsive in finite environments. To disinfect a space of more than 100 m<sup>2</sup> with chemical methods or medical alcohol is a laborious and expensive process and increases the risk of infection [9]. To solve this problem and ensure economic efficiency, UV light has long been considered an effective and direct method for the inactivation of SARS-CoV-2, its analogues, of viruses, and other microorganisms [10]. UV sterilization method has proven advantages and effectiveness as they are less labor intensive to use and do not necessarily require close contact with potentially contaminated surfaces. The UV spectrum is divided into three main groups, UVA (315–400 nm), UVB (280–315 nm), and UVC (100–280 nm), but the inactivation effect of SARS-CoV-2 is mainly concentrated in devices that produce only UVC light [11]. UVC light is the most common UV spectrum produced by commercially available medical devices. However, having devices that can be used in hospital and clinic environments is not yet available.

Our team performed analysis and testing of commercially available UV sterilizers that produce broad pulsed UV light (200–700 nm) in inactivation of SARS-CoV-2 on glass, plastic, metals or N95 respirator materials. The study determined the effectiveness of UV dosing to provide the end user with the ability to quickly determine if an adequate dose of UV light is being delivered to the surface. Scientific research data in Vietnam show that is highly effective in inactivating SARS-CoV-2 on many surfaces, especially on N95 respirators.

## 2 Effect of UV Light on Sav-CoV 2 Virus

UV light has been used widely since the early 20th century in Vietnam to disinfect water, air and surfaces. Since 1940s, it was used to reduce the transmission of tuberculosis in hospitals, by shining the light at the ceiling to disinfect air as it circulated throughout the patient's room. Today, it's used not only in medical area, but in some public bathrooms and airplanes when there are no people in those spaces. Physicists have created more powerful UV LEDs, which they believe medical device manufacturers can incorporate into lightweight devices to kill pathogens like the new coronavirus.

A newly discovered material that conducts electricity and is transparent to UV light could have a use in portable devices for killing SARS-CoV-2. This is the virus that causes COVID-19. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been a worldwide pandemic threat and poses a serious health and economic burden. Contaminated environments, such as personal items and room surfaces, are considered likely to transmit the virus. UV lamp is an effective tool to disinfect and kill viruses, which helps to prevent the spread of respiratory viruses and protect public health in medical facilities, offices and people's house. Ultraviolet C (UVC) has proven to kill bacteria and eliminate environmental pollution. UVC neutralized SARS-CoV-2. Immunoassay analysis for viral spikes and nucleocapsid proteins showed that UVC treatment did not damage viral proteins. Virus particle morphology remained intact even when the virus completely lost its infectivity after UVC irradiation, as observed by transmission electron microscopy. In contrast, UVC irradiation-induced genomic damage was determined using the newly developed long reverse transcriptase-quantitative polymerase chain (RT-qPCR) assay, but not conventional RT-qPCR.

To effectively use UV light, it is necessary to determine the effectiveness of different UV wavelengths in killing pathogens, especially SARS-CoV-2. Recent studies have shown the effectiveness of ultraviolet light to aid in the control of the global COVID-19 pandemic.

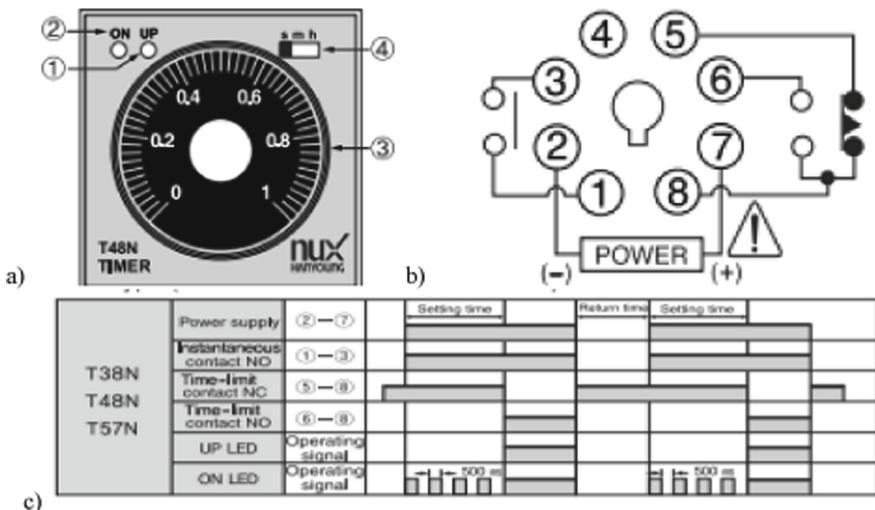
The SARS-CoV-2 virus strain can be effectively inactivated using a range of UVC wavelengths with a wavelength of about 220 nm providing the best sterilization effect. Interestingly, 220 nm radiation has been shown to be safe for human exposure to thresholds that exceed the effective threshold for virus inactivation. Therefore, using ultraviolet lamps and controlling them in real time will play an important and prerequisite role in the disinfection system.

## 3 Design and Manufacture of Sterilization Chamber

### 3.1 Integration of Analog Timing Control Technology

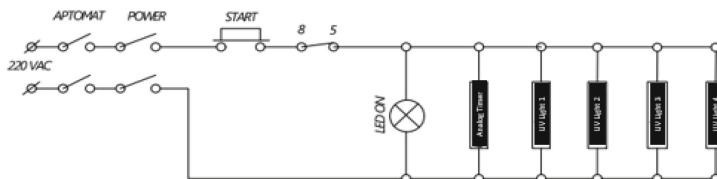
The ATN series timer has 6 output operation modes and 12 set time ranges, so it is often used for applications related to time control. The ATN series has a wide time range from 0.05 s to 100 h, depending on the model. Users can easily set and control the time via analog dial control. Series ATNs are available in 8-pin or 11-pin socket type. Analog Timer T48N, 48.0(W) 48.0(H) mm, Time Range from 60 to 60 h with Time - limit 1c + Constant 1a. Life span of relay: Mechanical (More than 10 million times) or Electrical (More than 100,000 times). Ambient temperature –10 ~ 55 Without freeze.

Power Consumption Less than 4.5 VA (at 240 VAC 60), Less than 1.5 W (at 24 VDC) (Fig. 1).



**Fig. 1.** Analog Timer T48N: a) Front view; b) A type (Time - limit 1c + Instantaneous); c) Timing chart

Function: (1) Output Indication LED (UP LED): The LED after passing setup time will be ON at the same time; (2) The Power Supply Indication LED (ON LED): LED is OFF when timer works. The LED is ON or OFF at the same time when operating time-limit output and its status will be maintained; (3) Time Display: to change time setup together with knob and user can change setup time in the middle of Time Display function; (4) Time unit setup switch (s/m/h) is consisted of Hour(h), Minute(m), Second(s). User can set and use time unit what you need. Available to change it when Power is OFF/ON (Fig. 2).

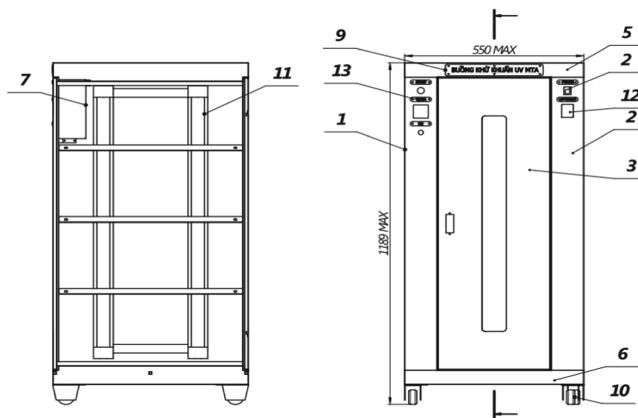


**Fig. 2.** Principle of operation and design of UV emitting system

Working principle: (1) Close APTOMAT and POWER; (2) Press START button pins (2)–(7) (POWER T48N), LED ON, 4 UV lamps are powered -> UV light is on and contacts (1)–(3) are closed, maintaining power supply to when the contact setting time is up (5)–(8) open, disconnect the power LED ON, POWER T48N, 4 UV lamps, contact (1)–(3) open.

### 3.2 Design of Sav-Cov2 Disinfection System

The MTA UV (*MILITARY TECHNICAL INSTITUTE Ultraviolet*) sterilization chamber is designed with a size of 1189 mm × 550 mm, including 13 main parts: (1) The device shell on the left is made of over 300 stainless steel and is welded in groups; (2) The equipment housing on the right is also made of over 300 stainless steel and is welded in groups; (3) The wings of the sterilization chamber are made of stainless steel, fitted with transparent glass to observe inside; (4) The back of the sterilization chamber; (5) The top cover of the sterilization chamber; (6) The base of the sterilization chamber, fitted with a rubber base for electrical insulation, to ensure compliance with medical device standards (Fig. 3);



**Fig. 3.** Design of UV sterilization chamber MTA

(7) Control and sensor module, integrated and mounted inside the sterilization chamber; (8) The operation indicator light of the sterilization chamber; (9): Logo with the nameplate of the Sterilization Chamber; (10) Wheel system to bear force, heat and static electricity to help the process of moving the sterilization chamber take place more convenient and suitable; (11): Vertical support of the sterilization chamber, allowing the mounting of horizontal stainless steel bars to adjust the width of the sterilization chambers. This holder design allows to adjust the distance from the specimen to the disinfection lamp; (12) The system of clocks showing the indicator and operating mode of the sterilization chamber; (13) Relay automatically disconnects to ensure safety during operation; (14) The 220 v–50 Hz power switch provides AC voltage for the sterilization chamber to operate.

## 4 Evaluation of the Radiation Ability and the Virus Killing

### 4.1 Experimental Measurement of Sterilization System Radiation Distribution

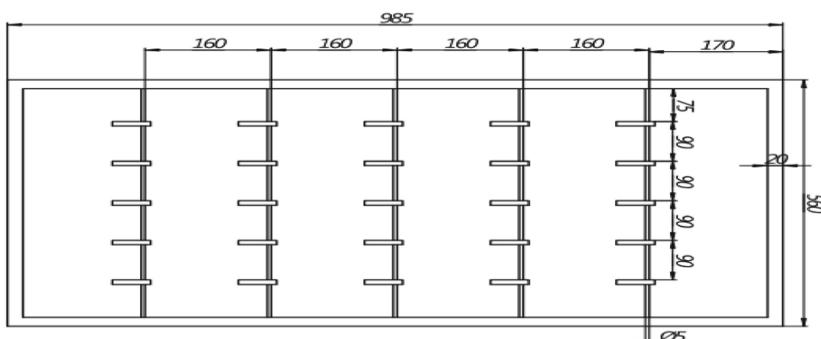
The process of measuring UV intensity in a closed chamber is carried out by direct measurement at 15 points according to the defined matrix. Measurement conditions are

made under normal environmental conditions, the temperature is from 23 °C to 30 °C, Humidity ranges from 65% to 85%. Measuring equipment is used UV energy radiation meter SDL470. The UVA probe records UV radiation measurements at about 365 nm long wavelengths under UVA (black light) source, UVC probe captures short-wave about 254 nm UV irradiance measurements under a UVC light source. Datalogging feature records readings with date and time stamp on control card (included) (Fig. 4).



**Fig. 4.** The MTA UV

Offset-capable device used for zero functionality or to make relative measurements and Data Sampling Rate adjustable from up to about 1 h (1-s increments), store about 100 times. The diagram of measuring points consists of 15 points, divided by a square matrix consisting of 3 cells (1, 2, 3) and 5 rows (A, B, C, D, E). This scheme allows to cover the cross-section of the sterilizer. The measurement process will evaluate the UV radiation intensity, thereby evaluating the affected area of the device (Fig. 5).



**Fig. 5.** Design of 304 stainless steel disinfection cabinet, cell size 20 × 10 cm and stainless-steel pipe size 05 cm

For each measurement site, data were taken 3 times, 5 min apart, and then the average value of the measurements was calculated. The results show the distribution of radiation spectrum in the entire sterilizer (unit in mW/Cm<sup>2</sup>). The lowest measured value is 0.136 mW/Cm<sup>2</sup> at E2 and the highest is 0.261 at B1 (Table 1).

**Table 1.** Measurement of UV intensity according to the distribution of UV chamber MTA.

No	Measuring position	Results (mW/cm <sup>2</sup> )			Mean value
1	A1	0,184	0,186	0,189	0,186
2	A2	0,152	0,158	0,159	0,156
3	A3	0,189	0,191	0,194	0,191
4	B1	0,262	0,257	0,265	0,261
5	B2	0,195	0,187	0,189	0,190
6	B3	0,235	0,239	0,242	0,238
7	C1	0,243	0,246	0,252	0,247
8	C2	0,158	0,162	0,165	0,162
9	C3	0,212	0,221	0,225	0,220
10	D1	0,233	0,239	0,242	0,238
11	D2	0,145	0,152	0,159	0,152
12	D3	0,241	0,247	0,252	0,247
13	E1	0,214	0,217	0,219	0,217
14	E2	0,129	0,137	0,142	0,136
15	E3	0,212	0,217	0,223	0,217

## 4.2 Testing the Process of Killing Bacteria on N95 Masks

The process of testing the bactericidal ability of the system for samples is a plastic dish containing bacteria and a sample of an N95 mask containing bacteria according to the following parameters: Sterile plastic Petri dishes: 10 pieces; Used N95 masks and added bacteria: 20 pieces; Bacterial strain: Pseudomonas aeruginosa (P. aeruginosa): 108 bacteria/ml; Bacterial strain: S. Aureus: 108 bacteria/ml; Special Mueller-Hinton agar, 9 cm<sup>2</sup> area, 4 cm thick, sterile; Aseptic absorbent paper compartment, size 1 × 2 cm: 10 slices; Liquid BHI broth (10 tubes) and Inoculation rods, incubators and other tools...

*Step 1. Prepare for UV irradiation:* Attach the paper plate to the plastic dish, sterilize it with the EO gas system. After sterilization, drop 30 µl of 108 bacteria/ml bacterial suspension onto 1 piece of paper, the quantity is as follows: Pseudomonas aeruginosa (P. aeruginosa) has 5 petri dishes, 3 paper plates each; Staphylococcus aureus (S. aureus) has 5 petri dishes, each with 3 pieces of paper; Attach the paper and hang position in

the cabinet, number and order. Put 4 dishes (1–4) containing strain P. aeruginosa and 4 dishes containing strain S. aureus no (6–9) into the UV MTA cabinet. No. 5 and No. 10 as controls, no projection.

*Step 2. Perform irradiation:* Projection time 15 min, equivalent to  $60 \text{ mJ/cm}^2$ .

*Step 3. Conduct inspection after irradiation:* Number the positions on the agar plate from 1–5; Number the BHI tubes from 1 to 5 with each bacterial strain. Transplant the paper drills at positions 1,2,3,4 on the agar plate. Positions 1,2,3,4 on each agar plate correspond to the numbered positions when irradiated. Position 5 is the control group. Configure the second paper rolls and tubes containing liquid BHI (5 tubes each, numbered from 1 to 5, tube number 5 is the uncirculated tube of the Reference group).

*Step 4. Get measurement results:* The results are shown in the following Table 2:

**Table 2.** Bacterial killing assessment process

No	Bacterial strain	Test tray 1	Test tray 2	Test tray 3	Test tray 4	Test tray 5
1	P. aeruginosa	(–)	(–)	(–)	(–)	(+)
2	S. aureus	(–)	(–)	(–)	(–)	(+)
3	N95 Marks	(–)	(–)	(–)	(–)	(+)

The irradiation dose of  $60 \text{ mJ/cm}^2$  of UV MTA was able to completely kill two types of bacteria P. aeruginosa, S. aureus and N95 Marks with the number of 0.3,107 bacteria/ml.

## 5 Conclusion

The results of measuring and testing the UV-MTA sterilization chamber, number 01-2020, have been assessed by the Department of Standards, Metrology and Quality, the Ministry of Science and Technology on April 15, 2020 to meet the design requirements. The operating power of the device is in the range of  $0.1 \text{ mW/cm}^2$  to  $0.5 \text{ mW/cm}^2$ . Besides, Our evaluation of sterilization efficiency was also tested by the Department of Infection Control, 103 Military Hospital for the bactericidal ability of the UV-MTA sterilization chamber on April 14, 2020. The test results showed that, in all 4 cases of using disinfection chamber and for all 3 types of specimens (P. aeruginosa, S. aureus and even N95 masks) no virus was detected after 18 h. This result is satisfactory when compared with the control specimens (tube number 5 in the experiment) containing the virus.

The UV-MTA sterilization chamber is designed, manufactured and tested for both radiation dose and ability to kill Sars-CoV-2 virus. This is a promising product that can be widely applied in practice in medical facilities in Vietnam. The UV-MTA sterilization chamber demonstrated our research objectives and results: (1) to inactivate the SARS-CoV-2 virus, achieve through our application of UVC irradiation at the both appropriate wavelength and effective dose, and (2) maintain the suitability and filtration efficiency of N95.

## References

1. United States centers for disease control and prevention. Coronavirus disease 2019 (COVID-19). Centers for disease control and prevention. <https://www.cdc.gov/coronavirus/2019-ncov/hcp/ppe-strategy/decontamination-reuse-respirators.html> (2020). Accessed 31 May 2020
2. Institute of Medicine, Board on Health Sciences Policy: Committee on the development of reusable facemasks for use during an influenza pandemic. Reusability of facemasks during an influenza pandemic: facing the flu. National Academies Press, Washington, DC (2006)
3. Ito, A., Ito, T.: Absorption spectra of deoxyribose, ribosephosphate, ATP and DNA by direct transmission measurements in the vacuum-UV (150–190 nm) and far-UV (190–260 nm) regions using synchrotron radiation as a light source. Photochem. Photobiol. **44**(3), 355–358 (1986)
4. Epa US: OAR. UV radiation. <https://www.epa.gov/sunsafety/uv-radiation-1> (2013). Accessed 5 Aug 2020
5. Bolton James, R., Linden, K.G.: Standardization of methods for fluence (UV Dose) determination in bench-scale UV experiments. J. Environ. Eng. **129**(3), 209–215 (2003)
6. Malayeri, A.H., Mohseni, M., Cairns, B., et al.: Fluence (UV dose) required to achieve incremental log inactivation of bacteria, protozoa, viruses and algae. IUVA News. **18**(3), 4–6 (2016)
7. Mills, D., Harnish, D.A., Lawrence, C., Sandoval-Powers, M., Heimbuch, B.K.: Ultraviolet germicidal irradiation of influenza-contaminated N95 filtering facepiece respirators. Am. J. Infect. Control. **46**(7), e49–e55 (2018)
8. Lowe, J.J., Paladino, K.D., Farke, J.D., et al.: N95 filtering facepiece respirator ultraviolet germicidal irradiation (UVGI) process for decontamination and reuse. <https://www.nebraskamed.com/sites/default/files/documents/covid-19/n-95-decon-process.pdf> (2020). Accessed 16 Nov 2020
9. Tseng, C.-C., Li, C.-S.: Inactivation of viruses on surfaces by ultraviolet germicidal irradiation. J. Occup. Environ. Hyg. **4**(6), 400–405 (2007)
10. OSHA: OSHA technical manual, section VIII—Use of respirators. [https://www.osha.gov/dts/osta/otm/otm\\_viii/otm\\_viii\\_2.html#8](https://www.osha.gov/dts/osta/otm/otm_viii/otm_viii_2.html#8). Accessed 22 Apr 2020
11. Ozog, D.M., Sexton, J.Z., Narla, S., et al.: The effect of ultraviolet c radiation against different N95 respirators inoculated with SARS-CoV-2. Int. J. Infect. Dis. **100**, 224–229 (2020)



# Sentiment Analysis in Twitter by Resnet-50 and VGG-19

Nguyen Quynh Anh<sup>(✉)</sup> and Dao Nam Anh

Faculty of Information Technology, Electric Power University, 235 Hoang Quoc Viet, Hanoi,

Vietnam

anhqnq@epu.edu.vn

**Abstract.** This study analyzes the feasibility of applying a number of convolutional neural network models to classify tweeter data. The messages of interest being the text notes of social communication are the objects for sentiment analysis. The study contributes a robust three-stage method for sentiment classification using text data from a social website, such as the Tweeter. Stage one is designed for text features extraction covering initial text processing and statistical examination that facilitates to convert text data to numerical features. In the stage two, implementation of convolution operations in many layers of convolutional neural networks is proposed to produce recognizable features from the numerical features. The use of Resnet-50 and VGG-19 is addressed in the stage allowing to get divergent sets of convolutional features for the same initial social dataset. Finally, classification is provided in the stage three which applies the Support Vector Machines in statistical learning the convolutional features. The description of the method is aligned with the Bayesian inference ensuring the robustness of the method. The results of experiments using a benchmark dataset are reported to show the validity of method.

**Keywords:** Sentiment analysis · Twitter · Resnet-50 · VGG-19 · CNN · SVM

## 1 Introduction

The availability of social media websites for public access has significant attention of end users for information exchange and communication. While this actually ameliorates the interaction of people, it truly strengthens the speed of spreading of all types of information covering both good and bad, true and false. In particular social networking website like Twitter, which has a large volume of messages to be exchanged, would advantage a system proficient of realizing real news or fake news to alert the system in case of disasters. These issues are of considerable interest for observation objectives and raising appropriate actions for dealing with real disasters.

Very often, analysis for identifying true disasters in Twitter messages involves the natural language processing (NLP) techniques, which captures disaster related contents radically from opinions and thoughts of web users. The task of recognizing real disaster

from messages is complicated, in particular for sentences with confusing words particularly used for different interpretation. Traditional disaster Twitter methods would miss true events in such cases.

If we now observe the case of a substantial amount of Twitter messages, desired for high accuracy categorization, we can note that even replacement of a single word in a sentence could lead to wrong classification. In this article, we present a novel reasonable method that combines various concepts and which can improve accuracy for the mentioned problem. It is further demonstrated that an image base registration concept for text features enables convolutional neural networks (CNNs) implemented for text data classification. The representation of Twister data by images is used to get features by implementing Resnet-50 and VGG-19 which are pre-trained CNNs allowing training and testing the data in a consistent machine learning method.

## 2 Related Work

For instance, a single message found on Twitter website is called a tweet. Previous works in the analysis of tweets frequently attempt to extract disaster related contents and analyze them to categorize disaster tweets versus non-disaster tweets. Various methods have been stated in the literature, which can carry out this task.

In relation to manage short tweets, Kopev et al. [1] make use of features including text, emotion sentiment, and color related ones. The text pre-processing like word n-grams and word embedding is applied to capture meaning of tweets. Various classifiers were tested with the extracted features and the best results were issued by SVM and a Hierarchical Attention Neural Network. The simplified convolutional network introduced by Liao et al. [2] relies on a CNN modified for sentence analysis.

In a word sense analysis, where the sense of ambiguous words is marked by unique embedding vector, Chen et al. [3] proposed bi-sense embedding to represent each tweet by two vectors for positive-sense and negative-sense. Here Long Short Term Memory (LSTM), which is based on recurrent neural network, is implemented for detection of sentiments. A deep learning approach is proposed by Gandhi et al. [4] for Twitter sentiment analysis, which categorize tweets to sentiment classes. The detection process starts with the initial feature extraction by analysis for tweet words, stop words and by transforming sentences to numerical vectors.

Resnet-50 is a deep residual network based on ResNet model proposed by He et al. [5]. This is designed with 48 Convolution layers supported by a MaxPool layer and an Average Pool layer. In this case, study with face images, Theckedath et al. [6] try to classify face images into psychological states namely fear, contempt, disgust, angry, sad and happy. To facilitate a deep visual representation Simonyan et al. [7] proposed VGG-19 with small convolution filters. Garcia-Gasulla et al. [8] investigated feature representation using a multiple CNN model including VGG-19. Their work studied the impact of layers for feature extraction in details and connected layers are capable to capture discriminant features.

Here we are interested in the searching a solution of the Tweet disaster that is associated with natural language processing techniques. However, there may be applicability of the CNNs including Resnet-50 and VGG-19 for the mentioned case study. Detailed description of our method is outlined in next section.

### 3 The Method

Looking at the Twitter message classification for *true disaster* or *non-disaster*, we consider the learning objects, which are the tweet message marked by  $m$  and the tweet class noted by  $t$ . From Bayesian point of view [9], common relationship between the message and the class are followed by the Bayes rule:

$$p(t|m) = \frac{p(t, m)}{p(m)} \quad (1)$$

In other words, the probability of  $t = \text{true disaster}$  given a tweet  $m$  is explainable by checking the posterior  $p(t|m)$  with knowing the likelihood  $p(m|t)$  and the prior  $p(t)$ :

$$p(t|m) = \frac{p(m|t)p(t)}{p(m)} \quad (2)$$

As the evidence  $p(m)$  can be dropped, approximation of the posterior distribution is followed:

$$p(t|m) \approx p(m|t)p(t) \quad (3)$$

Hence, the classification of a tweet is realized by the maximum likelihood:

$$t = \operatorname{argmax}_t p(t|m) \quad (4)$$

However, how is likelihood  $p(m|t)$  for (2) estimated? Basing on observed features and class for known tweet in the pass, the likelihood is appropriately derived from training samples  $\hat{m}$ :

$$p(m|t) = \sum_{\hat{m}} p(m|\hat{m})p(\hat{m}|t) \quad (5)$$

We assume the tweet feature values follow Gaussian distribution. This permits to evaluate feature distance  $p(m|\hat{m})$  between test sample and training sample in (5)

$$p(m|\hat{m}) = \frac{1}{\sqrt{2\pi|\delta|}} \exp\left(-\frac{(m - \hat{m})^2}{2\pi\cdot\delta^2}\right) \quad (6)$$

where the standard deviation  $\sigma$  and the mean  $\mu$  are well known with  $n_{\hat{m}}$  - the number of training tweet samples:

$$\delta = \sqrt{\frac{\sum (\hat{m} - \mu)^2}{n_{\hat{m}}}} \quad (7)$$

$$\mu = \frac{\sum \hat{m}}{n_{\hat{m}}} \quad (8)$$

One of the main concerns of learning is the detection of distinct relations between features for interconnection with classification. In this aspect, convolution operation provides a strong support with its calculation for local feature region:

$$m^h(x) = (m * h)(x) = \sum_{dx=-d}^d m(x - dx)h(x) \quad (9)$$

Moving the region to each data point  $x$  the convolution  $h$  uses local neighbors in a limited distance  $d$  to the central  $x$  for applying its calculation (9). So that the tweet feature  $m$  now has convolutional feature  $m^h$ .

The feature distance  $p(m_i | \hat{m}_i^h)$  from (6) can have extended form with the assistance of the convolution:

$$p(m_i^h | \hat{m}_i^h) = \frac{1}{\sqrt{2\pi|\sigma_i|}} \exp\left(-\frac{(m_i^h - \hat{m}_i^h)^2}{2\pi\sigma_i^2}\right) \quad (10)$$

where  $i$  is the note for each feature in general case that we have multiple features for each tweet sample. The standard deviation  $\sigma_i$  and the mean  $\mu_i$  are also estimated for each feature:

$$\sigma_i = \sqrt{\frac{n \sum (\hat{m}_i^h - \mu_i)^2}{n_{\hat{m}_i^h}}}, \mu_i = \frac{\sum \hat{m}_i}{n_{\hat{m}_i}}, \quad (11)$$

The aim of looking at each feature is to see how the feature is distinct for a pair of samples from different classes:  $x_{t_1}$  from *true disaster* class and  $x_{t_2}$  from *non-disaster* one:

$$d_i(x_{t_1} - x_{t_2}) = \frac{1}{\sqrt{2\pi|\sigma_i|}} \exp\left(-\frac{(m_i^h(x_{t_1}) - \hat{m}_i^h(x_{t_2}))^2}{2\pi\sigma_i^2}\right) \quad (12)$$

The sample distance  $d_i(x_{t_1}, x_{t_2})$  is the base for estimation of distinction  $d_i$  responsible for capability of feature  $i$  to show difference of samples from different classes:

$$d_i = \sum_{x_{t_1}, x_{t_2}} d_i(x_{t_1}, x_{t_2}) \quad (13)$$

With the assessment of each feature, the best features can be selected. Actually, the feature selection can help to improve performance in both accuracy and computing time. Finally, we perform maximum likelihood (4) for the convolution base features to classify tweets in test:

$$t(m^h) = \text{argmax}_t p(t|m^h) \quad (14)$$

In order to realize the described learning process, we apply the Support Vector Machines [10] remarked by function  $f^{SVM}(m^h)$  that searches class border in the way of maximizing the width of the gap between the classes.

$$t(m^h) = \begin{cases} t_1 & \text{if } f^{SVM}(m^h) > \epsilon \\ t_2 & \text{otherwise} \end{cases} \quad (15)$$

where  $\epsilon^2$  stands for the width of the border. Below, we formulate the measurement of performance by the accuracy. The possible use of convolution network for the tweet data requires an additional task converting the numerical feature vector to image with the size of  $224 * 224 * 3$  for Resnet-50 and VGG-19. It is designed in the way that the CNNs produce 1000 features for each image.

## 4 Experimental Results

It is important to evaluate described method through testing on real database. As a benchmark database, the Kaggle [11] provides 10,000 tweets collected from the Tweeter social website. The database includes messages containing comments and motions related to disasters, which were labeled by true disaster or non-disaster. We use three stages for analyzing the tweet database:

**A Text Analysis.** In practice, text messages are collected from different sources and appeared in various formats. The pre-processing task is responsible for transforming the messages into standard forms convenient for analysis. However, stop words are remained. When consider frequency of words used in tweets it is distinctly possible to sort the words by their frequency. The most used words need more attention on analysis as their intensive appearance has strong effect on classification.

**B. CNNs Features.** As discussed in Sect. 4, the important step of our method is the transformation of tweets to images, which are suitable for CNNs analysis. The messages have numerical features by NLP tokenization task. Hence, the text features of a tweet are represented by a vector. We apply normalization to the vector so that it can be transformed to image. The image size is  $224 * 224 * 3$ , which supports convolution operations when the Resnet-50 and VGG-19 work on the image. The dimension of the last layer of the CNNs is 1000 providing the same number of features for each tweet.

**C. Classification by SVM.** For cross validation we applied 10 different splits with ratio 50%-50% for datasets in training and testing. The learning model of SVM is implemented for extracted features and accuracy is measured for each split test. Table I displays the averaged accuracy of our classification. The features of Resnet-50 allow getting accuracy of 78.58% while the features of VGG-19 have 75.71%. By feature evaluation and selection, the VGG-19 has accuracy of 77.62% and the Resnet-50 achieved 80.50%. There is a significant enhancement for accuracy by using selected 50% CNNs features. This shows the effectiveness of joined solution (Tables 1 and 2).

**Table 1.** Performance report of accuracy by CNNs features

CNNs model/feature selection	50%	100%
Resnet - 50	80.50%	78.58%
VGG-19	77.62%	75.71%

**Table 2.** Other methods

Method	Database	Accuracy
BERT [12]	CrisisLexT26	67.00%
LSTM CNN [13]	Hurricane Sandy	74.78%
Decision Tree CART [14]	Hurricane Sandy	71.50%
Resnet-50, 50% features (our)	Kaggle	80.50%
VGG-19, 50% features (our)	Kaggle	77.62%

In relation with this tweet disaster Ma [12] used BERT model for categorize tweets. He used database named CrisisLexT6 for experiments, which reported accuracy at 67.00%, see Table II. In addition, Sit et al. [13] studied the tweet disaster problem with other database named the Hurricane Sandy. The authors implemented LSTM and CNNs for classification and reported the accuracy at 74.78%. The Hurricane Sandy database is also used by Goswami et al. [14] in their work with method of Decision Tree CART. Their classification accuracy was shown by 71.50%.

## 5 Conclusion

We have proposed a method for sentiment analysis based on CNNs features. The tweeter messages are analyzed by an integration approach covering natural language processing and pattern recognition algorithms for solving the problem. The case study illustrates that it is realizable for sentiment analysis with text base data, to have solution through image base convolution networks. The features of tweets are the main study object in the article and representation of features are proposed to be transformed in three steps: The use of a text pre-processing feature in numerical vector; the CNNs model of Resnet-50 and VGG-19 enriched the features by multiple convolution operations accessing local areas of the features; the evaluation for feature selection yields finals features for SVM learning. Future work consists of study of extension of the join concept for searching improvement of classification performance for the sentiment analysis.

## References

1. Kopev, D., Atanasov, At., Zlatkova, D., Hardalov, M., Koychev, I.: Tweety at SemEval-2018 Task 2: Predicting Emojis using Hierarchical Attention Neural Networks and SVM (2018)
2. Liao, S., Wang, J., Yu, R., Sato, K., Cheng, Z.: CNN for situations understanding based on sentiment analysis of twitter data. Procedia Comput. Sci. **111**, 376–381 (2017)
3. Y. Chen, J.Y., You, Q., Luo, J.: Twitter Sentiment Analysis via Bi-sense Emoji Embedding and Attention-based LSTM (2018). arXiv.org > cs > [arXiv:1807.07961](https://arxiv.org/abs/1807.07961)
4. G., U.D., M.Kumar, P., C. Babu, . et al. (2021) Sentiment Analysis on Twitter Data by Using Convolutional Neural Network and Long ShortTerm Memory. Wireless Pers Commun
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

6. Theckedath, D., Sedamkar, R.R.: Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN Comput. Sci.* **1**(2), 1–7 (2020). <https://doi.org/10.1007/s42979-020-0114-9>
7. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations (2015)
8. Ferran, D.G.-G., Ferran, P., Armand, P., Armand, V., Suzumura, T.: On the behavior of convolutional nets for feature extraction. *J. Artif. Intell. Res.* **61** (2017)
9. Barber, D., Cemgil, A.T., Chiappa, S.: Bayesian time series models. Cambridge University Press (2011)
10. Cortes, C., Vapnik, V.N.: Support-vector networks (PDF). *Machine Learning*. **20** (3): 273297. CiteSeerX 10.1.1.15.9362 (1995)
11. <https://www.kaggle.com/c/nlp-getting-started/overview>
12. Ma, G.: Tweets Classification with BERT in the Field of Disaster Management (2019)
13. Muhammed, A.S., Caglar, K., Ibrahim, D.: Identifying disasterrelated tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of Hurricane Irma, *International Journal of Digital Earth* (2019)
14. Goswami, S., Raychaudhuri, D.: Identification of Disaster-Related Tweets Using Natural Language Processing: International Conference on Recent Trends in Artificial Intelligence, IOT, Smart Cities & Applications (2020)



# Improvement of Self-supervised Depth and Motion Learning with Vision Transformer

Xuan-Quang Ma<sup>1</sup>, Ngoc-Kien Nguyen<sup>2</sup>, and Van-Tinh Nguyen<sup>2(✉)</sup>

<sup>1</sup> Vingroup Big Data Institute, Hanoi, Vietnam

<sup>2</sup> Hanoi University of Science and Technology, Hanoi, Vietnam

tinh.nguyenvan@hust.edu.vn

**Abstract.** While convolutional neural network can extract well local features of images, it can only process images' global features at low-resolution feature maps which were down-sampled by pooling layers or convolutional layers. Vision transformer, in another hand, can process global features at high resolution features maps by tokenizing high resolution features maps or images and feeding them into the network. Thus, the vision transformer included backbone can process representations at high resolutions and has a global receptive field. The properties allow this kind of backbone to provide finer-grained and more globally coherent predictions compared to fully-convolutional networks. In this work, we propose to use a convolutional network concatenate with a vision transformer module as the backbone for a self-supervised monocular depth estimation pipeline to improve the quality of predicted depth map. In addition, the weights in multi-component loss function in this pipeline were chosen heuristically in previous works, which were not optimal for the problem of estimating dense depth maps. Thus, we propose a weights analyzing-and-choosing method for multi-component loss function to improve quality of depth map prediction. Our work achieves a competitive result among other self-supervised depth estimation methods.

**Keywords:** Depth estimation · Vision transformer · Multi-loss balancing

## 1 Introduction

Per-pixel depth estimation is a low-level task in computer vision that has application in various fields such as autonomous vehicles [1], augmented reality [2], robotics [3], etc. Deep Learning based Dense Depth Estimators in recent years became successful with the development in deep neural networks, it includes self-supervised methods and supervised methods. For supervised methods, the ground truth label of depth can be collected from LiDAR sensors [4], then images are paired with depth maps become the input for depth estimation network [5]. Self-supervised methods do not need ground truth for learning depth, the pipeline setting is also simpler compared to supervised methods, but they still perform worse than supervised methods on benchmark datasets. These weaknesses come from reasons that violate image reconstruction assumptions such as: there are occluded pixels, pixels that do not satisfy the Lambertian assumption,

pixels that belong to moving objects. These pixels inhibit the training process and make network output poor quality depth maps.

Previous methods [6–10] use encoder-decoder architecture to predict depth maps on several scales and use smooth loss [11] to force depths propagation from discriminative regions to textureless regions, which partially solved above problems. However, this propagation has a limited range because convolutional neural network is lack of ability to process global features of image, it can only process global features on low-resolution feature maps. We believe that if the depth network can process image’s global features at high-resolution feature maps, it can reason better about textureless regions as well as regions that belong to move objects and improve depth estimation’s quality.

In recent years, Vision Transformer [12] is emerging as a potential approach for many Computer Vision problems as it has a good ability to process global features of image. AdaBins [13] specified that by integrating local features extracted by convolutional neural network and global features processed by Vision Transformer at high-resolution feature maps helped increase the overall supervised depth estimation framework’s performance.

Inspired from the idea of AdaBins [13], we pose the question of integrating local features extracted from convolutional encoder-decoder architecture and global features extracted from vision transformer in a self-supervised depth estimation pipeline. Previous works [13, 14] found ways to apply vision transformer in supervised setting for problems such as Image Segmentation, Depth Estimation, etc. Self-supervised methods with vision transformer have also been researched in [15, 16, 17] for Image Segmentation, Image Reconstruction, etc. To the best of our knowledge, there isn’t any work had applied vision transformer for self-supervised depth estimation. Therefore, in this work, we propose to use a convolutional encoder-decoder architecture concatenate with vision transformer to estimate dense depth map in self-supervised setting.

In a self-supervised depth estimation pipeline, the usually used loss function composes of multi components e.g. L1 loss, Structural Similarity loss, Smooth loss, etc. Normally, the final objective function is a weighted sum of these components [6, 7, 9]. However, weighting these components is problematic due to several reasons. First, each loss has a different order of magnitudes, which makes some losses suppressed if the total loss is not weighted properly. Second, the contribution of each loss to the overall performance is hard to evaluate. Therefore, in this work, we propose a method of choosing loss weights that find normalized value of each loss component and choose weights based on these normalized values. Extensive experimental results indicate that our proposed weighting method boosts the performance of the monocular depth estimator meaningfully.

In summary, our main contributions are:

1. A self-supervised depth estimation pipeline is proposed that uses convolutional encoder-decoder architecture with vision transformer as the backbone to integrate local features and global features of input images.
2. A loss weights analyzing-and-choosing method for multi-component loss function is proposed to improve the quality of depth map prediction.

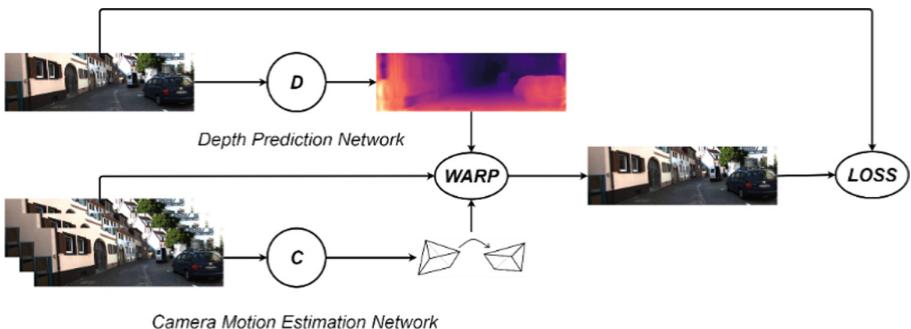
## 2 Related Work

**Self-supervised Depth Estimation Methods.** [6] was one of the first works that used view synthesis as the training signal for depth prediction. Ranjan et al. [7] used geometry constraint to combine the learning of different tasks (Depth Estimation, Pose Estimation, Optical Flow Estimation, Motion Segmentation) into a framework and saw improvements in all tasks. Godard et al. [8] added occlusion masking into its framework by using a minimum photometric loss for adjacent frames instead of a mean photometric loss. Gordon et al. [9] addressed static scene assumption by removing pixels belonging to moving objects with an additional instance segmentation network. Shu et al. [18] avoided zero-gradient problem in textureless regions by representing image by its features and reconstructing the feature representation of image instead of the image itself.

**Multi-loss Weighting Methods.** In previous depth estimation works [6–8], the weight of each component in the total loss was chosen heuristically. Kendal et al. [19] proposed a loss weighting method based on the uncertainty of each loss term. Groenendijk et al. [20] defined a hypothesis that a loss component had been satisfied when its variance had decreased towards zero. They used adaptive weights for loss components and calculated loss's weight based on its coefficient of variation. Chen et al. [21] relied on another hypothesis that the contribution of each loss depended on its gradient magnitudes. They balanced the contribution of each loss term by directly tuning gradient magnitudes. Lee et al. [22] chose loss weights by continuously rebalancing the value of each loss component in total loss.

## 3 Methodology

### 3.1 Self-supervised Depth Estimation Pipeline with Vision Transformer



**Fig. 1.** Self-supervised depth estimation pipeline.

**Overall Pipeline.** The self-supervised depth estimation pipeline as described in Fig. 1 includes 2 networks: Depth Prediction Network ( $D$ ) and Camera Motion Estimation Network ( $C$ ). The inputs of the system are one target image  $I_t$  and two source images  $I_s$ . The  $D$  network inputs target image and output predicted depth map. The  $C$  network inputs target image and source images, it outputs relative camera poses from target image to source images. Based on predicted depth and camera poses, the target image can be reconstructed following the image warping process which will be presented in Sect. 3.2. The target image and reconstructed images are then fed to loss function to produce the training signal for the pipeline.

**Depth Prediction Network.** Our work utilizes AdaBins [13] network architecture which composes of a modern convolutional encoder-decoder and an AdaBins module as the Depth Prediction Network. The encoder-decoder block composes of a modern EfficientNet B5 encoder and a normal upsampling decoder. The decoder outputs a tensor  $x_d \in \mathbb{R}^{h \times w \times C_d}$  instead of a depth map  $d \in \mathbb{R}^{H \times W \times 1}$  ( $H$  and  $W$  are the height and width of input image). The decoded features  $x_d$  is then fed into the AdaBins module. This module includes a vision transformer architecture to perform global statistical analysis at  $x_d$  which is a high-resolution feature map, and then it refines the output to get the final depth prediction.

**Camera Motion Network.** The camera motion network consists of several stacked convolutional layers and followed by adaptive average pooling layers to process image features to get the relative camera motion between frames.

### 3.2 Image Warping Process

**Camera Model.** The camera model that projects a 3D point  $P = (X, Y, Z)$  in camera coordinate to a 2D pixel  $p = (u, v)$  on image plane given the camera intrinsics  $K = (f_x, f_y, c_x, c_y)$ :

$$p = (u, v) = (f_x \frac{X}{Z} + c_x, f_y \frac{Y}{Z} + c_y) \quad (1)$$

Back projecting a 2D pixel  $p = (u, v)$  on image plane to a 3D point  $P = (X, Y, Z)$  given its depth  $D(p)$  follows below equation:

$$P = (X, Y, Z) = D(p) \left( x - \frac{c_x}{f_x}, y - \frac{c_y}{f_y}, 1 \right) \quad (2)$$

**Warping Source Image to Target Image.** A target image  $I_t$  can be reconstructed from a source image  $I_s$  via:

$$\hat{I}_{s \rightarrow t}(p) = I_s(\hat{p}) \quad (3)$$

where  $\hat{p}$  is the corresponding pixel on source image  $I_s$  of a pixel  $p$  on target image  $I_t$ . We need to find  $\hat{p}$  via  $p$ . Given a pixel  $p = (u, v)$  in target image, the corresponding 3D point of it in target image coordinate is:

$$P = D(p) \left( x - \frac{c_x}{f_x}, y - \frac{c_y}{f_y}, 1 \right) \quad (4)$$

The coordinate of  $P$  in source image coordinate is calculated as:

$$P_s = RP + t = (X_s, Y_s, Z_s) \quad (5)$$

where  $R$  and  $t$  are the rotational transformation matrix and transitional transformation vector respectively.

The corresponding pixel  $\hat{p}$  of  $p$  is:

$$\hat{p} = \left( f_x \frac{X_s}{Z_s} + c_x, f_y \frac{Y_s}{Z_s} + c_y \right) \quad (6)$$

$I_s(\hat{p})$  and  $I_t(p)$  should be similar with these assumptions: the corresponding 3D point is static and visible in both frames; it lies on a Lambertian surface; the predicted depth and pose are both correct.

### 3.3 Method of Choosing Loss Function Weights

**Loss Function.** The self-supervised loss function utilized in this work is:

$$L = \lambda_{L1} \mathcal{L}_{L1} + \lambda_{SSIM} \mathcal{L}_{SSIM} + \lambda_s \mathcal{L}_{smooth} \quad (7)$$

where:

$$\mathcal{L}_{L1} = \sum_p |I_t(p) - I_s(\hat{p})| \quad (8)$$

$$\mathcal{L}_{SSIM} = \sum_p 1 - SSIM(I_t(p), I_s(\hat{p})) \quad (9)$$

$$SSIM(x, y) = \frac{(2\mu_x\mu_y+c_1)+(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \quad (10)$$

$$\mathcal{L}_{smooth} = \sum_p \|e^{-\nabla I_t} \nabla d\|^2 \quad (11)$$

The  $SSIM$  index [24] is calculated on several windows of an image,  $x$  and  $y$  are two windows of common size  $N \times N$ , with  $\mu_x$ ,  $\mu_y$  is the average of  $x$ ,  $y$ ;  $\sigma_x^2$ ,  $\sigma_y^2$  is the variance of  $x$ ,  $y$ ;  $\sigma_{xy}$  is the covariance of  $x$  and  $y$ ;  $c_1 = 0.01^2$  and  $c_2 = 0.03^2$  are used to stabilize the division;  $I_t$  and  $I_s$  are target image and source image;  $p$  is a pixel on target image;  $\hat{p}$  is the corresponding pixel of  $p$  on source image;  $\nabla$  is the first derivative along spatial directions;  $d$  is the predicted depth map;  $\lambda_{L1}$ ,  $\lambda_{SSIM}$ ,  $\lambda_s$  are the weighting coefficients between loss terms, which we found the optimal values are  $\lambda_{L1} = 1.0$ ;  $\lambda_{SSIM} = 0.12$ ;  $\lambda_s = 0.21$ .

$L1$  loss  $\mathcal{L}_{L1}$  and  $SSIM$  loss  $\mathcal{L}_{SSIM}$  calculate the difference between target image and reconstructed image. While  $\mathcal{L}_{L1}$  calculates pixel-wise absolute difference between two images,  $\mathcal{L}_{SSIM}$  regards to the difference in luminance, contrast, and structure. Both  $\mathcal{L}_{L1}$  and  $\mathcal{L}_{SSIM}$  can be used individually for calculating the difference between the 2 images, several works [7][8] specified that combining them help improves network performance. Smoothness loss  $\mathcal{L}_{smooth}$ , in another hand, constraints the smoothness of the predicted depth map. It helps avoid zero-gradient problem in textureless regions by allowing gradient to propagate from discriminative regions to low-texture regions.

**Loss Weights Analyzing-and-Choosing Method.** Through experiments, we noticed that the pipeline achieved the best result with weights as SSIM loss, L1 loss and smoothness loss contributed 36%, 63% and 1% respectively in terms of value in total loss. These contributions correspond to  $\lambda_{SSIM} = 0.12$ ,  $\lambda_{L1} = 1.0$ ,  $\lambda_s = 0.21$ . Our method for choosing  $\lambda_{SSIM}$ ,  $\lambda_{L1}$ ,  $\lambda_s$  is described as follows:

- Firstly, we estimate each loss term's contribution in total loss in terms of value.
- Secondly, we conduct experiments on choosing smoothness loss weight.
- Thirdly, we conduct experiments on choosing L1 and SSIM losses weights.

*Estimating Loss Terms' Contributions.* We trained  $D$  network and  $C$  network with  $\lambda_{SSIM} = 1$ ;  $\lambda_{L1} = 1$ ;  $\lambda_s = 1$  to inspect the loss terms' original values without any scaling involved. We trained the networks for 20 epochs until the system converged and plotted each loss term's value distributions in 3 final epochs. From the 3 distributions, we calculated the mean values  $\mu_{SSIM}$ ,  $\mu_{L1}$ ,  $\mu_{smooth}$  of SSIM loss, L1 loss and smoothness loss respectively. We used these mean values for approximating contribution of each component in terms of value in the total loss. In particular:

$$contribution_{L1} = \frac{\lambda_{L1}\mu_{L1}}{\lambda_{L1}\mu_{L1} + \lambda_{SSIM}\mu_{SSIM} + \lambda_s\mu_{smooth}} \quad (12)$$

where  $contribution_{L1}$  is the contribution of L1 loss in terms of value in the total loss. The contributions of SSIM loss and smoothness loss were calculated similarly. By adjusting  $\lambda_{L1}$ ,  $\lambda_{SSIM}$  and  $\lambda_s$  we can manipulate the contribution of each loss term.

*Choosing Smoothness Loss Weight.* We used the Edge Aware Smoothness Loss [11] for our work. This Loss Forces the Network to Predict Smooth Disparity Map according to edges of the input RGB image. Our method for weighting smoothness loss is described as follows:

- We fixed  $\lambda_{L1}$ ,  $\lambda_{SSIM}$  as in Ranjan et al. [7] to find  $\lambda_s$ . Based on mean values of the distributions found above, we adjusted  $\lambda_s$  therefore smoothness loss's contribution to total loss varied from 99% to 0%.
- We realized that when training network from scratch, a larger than 12% of smoothness loss's contribution made network predict trivial solutions, where predicted disparity maps were all the same value at every pixel to minimize smoothness loss.
- The network can learn and predict non-trivial solutions at smoothness loss contribution less than 12%. With the contribution of smoothness loss larger than 5%, the predicted disparity maps were too smooth. The predicted disparity maps could preserve objects and scenes' edges when this proportion was under 5%. The best performance achieved when the smoothness loss contribution was approximate 1% in terms of value in the total loss.

*Choosing L1 loss and SSIM Loss Weights.* Our method of choosing L1 loss and SSIM loss weights is described as follows:

- We fixed the contribution of smoothness loss to 1% - the best value for smoothness loss we found above.
- We adjusted  $\lambda_{L1}$ ,  $\lambda_{SSIM}$  therefore the contribution of L1 loss varied from 99% to 0% and SSIM loss varied from 0% to 99%.
- After several experiments, we saw that the network output the best quality depth maps with SSIM, L1 and smoothness losses accounted for 36%, 63% and 1% respectively in terms of value in total loss, which correspond to  $\lambda_{SSIM} = 0.12$ ,  $\lambda_{L1} = 1.0$ ,  $\lambda_s = 0.21$ .

## 4 Experiments and Discussions

### 4.1 Depth Evaluation

**Evaluation Metrics.** Some popular metrics for evaluating depth are presented as follows:

$$AbsRel : \frac{1}{|D|} \sum_{d \in D} \frac{\|d^* - d\|_1}{|d^*|} \quad (13)$$

$$SqRel : \frac{1}{|D|} \sum_{d \in D} \frac{\|d^* - d\|_1^2}{d^*} \quad (14)$$

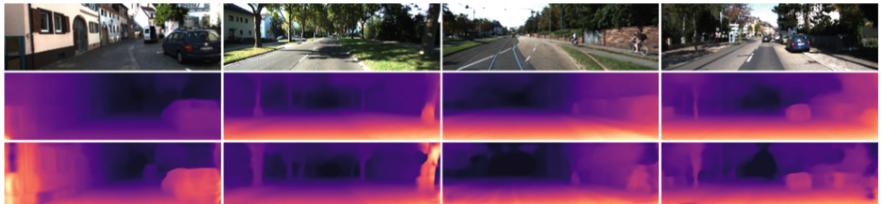
$$\delta_t : \frac{1}{|D|} \left| \left\{ d \in D \mid \max\left(\frac{d^*}{d}, \frac{d}{d^*}\right) < 1.25^t \right\} \right| \quad (15)$$

where  $d^*$  and  $d$  are ground truth depth and predicted depth,  $|D|$  is the number of pixels in the predicted depth map.

On evaluation, the depth value is capped at 80m. Similar to Zhou et al. [6], our predicted depth is multiplied with a scale factor  $\hat{s}$  which is calculated by:

$$\hat{s} = \frac{\text{median}(D_{gt})}{\text{median}(D_{pred})} \quad (16)$$

where  $D_{pred}$  is the predicted depth map and  $D_{gt}$  is the ground truth depth map.



**Fig. 2.** Comparison of depth map's quality between our method (last row) and CC [7] (second row). Our method has finer details on silhouette of trees, humans and on textureless regions like walls' surface.

**Comparison to Other Works.** Table 1 shows performances of several methods on KITTI Eigen et al.’s test split [23] and KITTI Odometry Dataset [4]. Our work achieved a competitive result compared to other self-supervised method. Compared to CC [7], although our result is not as good as CC on KITTI Eigen et al.’s test split [23], we still see that the impact of vision transformer in extracting global features with the clear representation of textureless regions in Fig. 2. CC [7] predicts depth maps on 6 scales to widen effective range of smoothness loss, which derives depths propagation from discriminative regions to textureless regions. In contrast, our work just predicts depth map on one scale but the representation of textureless regions in depth map is still clear, which indicates that global features (the combination representation of the whole image including textureless regions and discriminative regions) are processed properly. It clearly shows that using vision transformer concatenated with convolutional encoder-decoder architecture helps incorporating local features and global features of image, which leads to better depth map predictions.

**Table 1.** Comparison of performances reported on Eigen et al.’s test split [23] and KITTI Odometry Dataset [4].

Method	AbsRel	SqRel	$\delta_1$	$\delta_2$	$\delta_3$	Sequence 09	Sequence 10
SfMLearner [6]	0.208	1.768	0.678	0.885	0.957	$0.021 \pm 0.017$	$0.020 \pm 0.015$
CC [7]	<b>0.140</b>	1.070	0.826	0.941	0.975	<b><math>0.012 \pm 0.007</math></b>	$0.012 \pm 0.008$
Monodepth2 [9]	0.155	<b>0.882</b>	<b>0.879</b>	<b>0.961</b>	<b>0.982</b>	$0.017 \pm 0.008$	$0.015 \pm 0.010$
Ours	0.176	1.467	0.768	0.913	0.962	$0.013 \pm 0.007$	<b><math>0.011 \pm 0.007</math></b>

## 4.2 Odometry Evaluation

**Evaluation Metrics.** The common error metric Absolute trajectory error (ATE) [25] measures the similarity between estimated trajectory and ground truth trajectory, is used in this work for odometry evaluation.

**Comparison to Other Works.** Our results evaluated on KITTI Odometry [4] dataset shown in Table 1. Our work achieves a high result compared to other methods. It is the second-best method on Sequence 09 and the best on Sequence 10. The improvements come from optimal weights chose by our method, where L1 loss, SSIM loss and Smoothness loss cooperate properly to help supervisory signal be more meaningful.

**Table 2.** Results of our work on Eigen et al.’s test split [23] with different choices of loss weights.

Weightings of method	Contribution of L1	Contribution of SSIM	Contribution of smoothness	AbsRel	$\delta_1$
SfMLearner [6]	89%	0%	11%	0.195	0.729
CC [7]	17%	81%	2%	0.189	0.743
Ours	63%	36%	1%	<b>0.176</b>	<b>0.768</b>

**Table 3.** Contribution of each loss term when the networks converged.

Test	Preset contribution			Converged contribution		
	L1	SSIM	Smoothness	L1	SSIM	Smoothness
Test 1	89%	0%	11%	89.36%	0%	10.64%
Test 2	63%	36%	1%	62.55%	36.29%	1.16%
Test 3	17%	81%	2%	16.73%	80.72%	2.55%
Test 4	0%	89%	11%	0%	89.80%	10.20%

### 4.3 Evaluation of Loss Weights Choosing Method

Table 2 compares results on KITTI Eigen et al.’s test split [23] of our framework with our optimal loss weights and loss weights chosen heuristically by other methods. Our choosing method achieves highest result among other heuristic methods.

Table 3 shows the contribution in terms of value of each loss term in total loss when the networks converged. The contribution of each loss term converges to preset parameters when the system converged. This specifies that our weight choosing method is stable and meaningful.

### 4.4 Implementation Details

We train the system with Eigen et al.’s split [23] of KITTI dataset [4]. By removing the static frames, we have 39,810 monocular triplets for training and 4,424 for validation. Intrinsics are the same for all images with the principal point lies in the center of image and the focal length is the mean of all focal lengths in KITTI. For training, the images are scaled to  $256 \times 832$ , batch size is 4 and ADAM optimization is used with learning rate of  $10^{-4}$ . Horizontal flips, random cropping and scaling are used for augmenting data. For all results presented, we trained the system for 20 epochs with each epoch takes approximately 120 min on a Nvidia V100-SMX2 16GB GPU. All the code is publicly available at: <https://github.com/maxuanquang/SfMLearner-ViT>.

## 5 Conclusions

In this work, vision transformer is used with convolutional encoder-decoder architecture in a self-supervised depth estimation pipeline, which concatenates local features and

global features of image to help predicted depth map to be more precise. A new method for choosing loss weights in multi-component loss function is proposed, which increases system's performance and can be widely used for every type of multi-loss function. The depth network achieves a competitive result on benchmark dataset.

## References

1. Menze, M., Geiger, A.: Object scene flow for autonomous vehicles. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015)
2. DeSouza, G.N., Kak, A.C.: Vision for mobile robot navigation: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(2), 237–267 (2002)
3. Newcombe, R.A., Lovegrove, S.J., Davison, A.J.: DTAM: dense tracking and mapping in real-time. In: 2011 International Conference on Computer Vision. IEEE (2011)
4. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE (2012)
5. Lee, J.H., et al.: From big to small: Multi-scale local planar guidance for monocular depth estimation. arXiv preprint [arXiv:1907.10326](https://arxiv.org/abs/1907.10326) (2019)
6. Zhou, T., et al.: Unsupervised learning of depth and ego-motion from video. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017)
7. Ranjan, A., et al.: Competitive collaboration: Joint unsupervised learning of depth, camera motion, optical flow and motion segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2019)
8. Godard, C., et al.: Digging into self-supervised monocular depth estimation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2019)
9. Gordon, A., et al.: Depth from videos in the wild: Unsupervised monocular depth learning from unknown cameras. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2019)
10. Casser, V., et al.: Depth prediction without the sensors: leveraging structure for unsupervised learning from monocular videos. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33. No. 01 (2019)
11. Wang, C., et al.: Learning depth from monocular videos using direct methods. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018)
12. Dosovitskiy, A., et al.: An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint [arXiv:2010.11929](https://arxiv.org/abs/2010.11929) (2020)
13. Bhat, S.F., Alhashim, I., Wonka, P.: Adabins: depth estimation using adaptive bins. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2021)
14. Ranftl, R., Bochkovskiy, A., Koltun, V.: Vision transformers for dense prediction. arXiv preprint [arXiv:2103.13413](https://arxiv.org/abs/2103.13413) (2021)
15. Caron, M., et al.: Emerging properties in self-supervised vision transformers. arXiv preprint [arXiv:2104.14294](https://arxiv.org/abs/2104.14294) (2021)
16. Atito, S., Awais, M., Kittler, J.: Sit: self-supervised vision transformer. arXiv preprint [arXiv:2104.03602](https://arxiv.org/abs/2104.03602) (2021)
17. Li, C., et al.: Efficient self-supervised vision transformers for representation learning. arXiv preprint [arXiv:2106.09785](https://arxiv.org/abs/2106.09785) (2021)
18. Shu, C., et al.: Feature-metric loss for self-supervised learning of depth and egomotion. In: European Conference on Computer Vision. Springer, Cham (2020)

19. Kendall, A., Gal, Y., Cipolla, R.: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018)
20. Groenendijk, R., et al.: Multi-loss weighting with coefficient of variations. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (2021)
21. Chen, Z., et al.: Gradnorm: gradient normalization for adaptive loss balancing in deep multitask networks. In: International Conference on Machine Learning, PMLR (2018)
22. Lee, J.-H., Kim, C.-S.: Multi-loss rebalancing algorithm for monocular depth estimation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12362, pp. 785–801. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-58520-4\\_46](https://doi.org/10.1007/978-3-030-58520-4_46)
23. Eigen, D., Puhrsch, C., Fergus, R.: Depth map prediction from a single image using a multi-scale deep network. arXiv preprint [arXiv:1406.2283](https://arxiv.org/abs/1406.2283) (2014)
24. Wang, Z., et al.: Image quality assessment: from error visibility to structural similarity. In: IEEE Trans. Image Process. **13**(4), 600–612 (2004)
25. Prokhorov, D., et al.: Measuring robustness of Visual SLAM. In: 2019 16th International Conference on Machine Vision Applications (MVA). IEEE (2019)



# Application of Event-Triggered Sliding Mode Control of 2-DOF Humanoid's Lower-Limb Powered by Series Elastic Actuator

Anh Khoa Lanh Luu<sup>3</sup>, Van Tu Duong<sup>1,2,3</sup>, Huy Hung Nguyen<sup>3,4</sup>,  
and Tan Tien Nguyen<sup>1,2,3(✉)</sup>

<sup>1</sup> Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology (HCMUT),  
268 Ly Thuong Kiet, District 10, Ho Chi Minh City, Vietnam

nttien@hcmut.edu.vn

<sup>2</sup> Vietnam National University Ho Chi Minh City (VNU-HCM), Linh Trung Ward, Thu Duc  
District, Ho Chi Minh City, Vietnam

<sup>3</sup> National Key Laboratory of Digital Control and System Engineering, Ho Chi Minh City  
University of Technology, VNU-HCM, Ho Chi Minh City, Vietnam

<sup>4</sup> Faculty of Electronics and Telecommunication, Saigon University, Ho Chi Minh City, Vietnam

**Abstract.** This paper proposes an event-triggered sliding mode control (SMC) scheme combined with a backstepping algorithm for control of 2-DOF humanoid's lower-limb powered by Series elastic actuator (SEA). First, the modelling process for the lower-limb system is implemented by using the Euler-Lagrange theory. With the obtained dynamical equations of lower-limb, the model of the SEA is achieved in both mechanical and electrical perspectives. Then, the event-triggered SMC approach is utilized to ensure the system's stability and eliminate the effect of bounded external disturbance. Next, some assumptions and designed thresholds for the tracking error are given, together with the proof for the convergence of the inter-event time. The backstepping algorithm is applied in the end to determine the needed input control voltage signal. Finally, the results of this research are demonstrated through some simulations in order to prove the efficiency and appropriateness of this method.

**Keywords:** Event-triggered sliding mode control · Series elastic actuator · Humanoid lower-limb

## 1 Introduction

Conventionally, the control input signal from the controller applying to the plant is updated at the sampling instant only and kept unchanged between two consecutive sampling instants. This feature is usually called the *time-triggered control* strategy, which is widely applied in most simple systems due to its massive advantage of simplicity for both programming and system analysis [1]. However, this approach has a significant drawback about the system's sensitivity to the output state when the control input is triggered periodically regardless of the tracking error and the changing rate of the output

variables. Besides, from the communication perspective, the gradual generation and transmission of a control input signal in every instant are usually considered a waste of resources and unproductive, which can also lead to the delay phenomenon for the whole controlled system.

For those reasons, an alternative approach, called *event-triggered control*, has been proposed and analyzed in many research to deal with the downsides of the traditional strategy [2, 3]. By considering stabilization conditions and some desired error limits, the control input's triggering instant is now heavily relied on the system's output states. This approach has shown its remarkable effectiveness in the steady-state of the controlled system, which made it widely accepted and applied worldwide in many scientific research and industrial fields. Recently, event-triggered strategy has gradually been used in spacecraft for attitude control and flying stabilization, where remote control and external disturbance are always the critical problems. In [4, 5], the event-triggered control scheme was applied to reduce the wireless communication pressure between the controller and the aircraft's actuators as well as to give more resources for other vital sensors' feedback signals. Besides, this approach has also proved its appropriateness for the trajectory tracking problem of autonomous vehicles [6, 7], and industrial robots [8, 9]. In those applications, reducing the control signal commutation frequency impacted the decrement of the data broadcasting between the controller and the plant. Also, the updates of the control signal could be suitably tuned for a specific application so as to fulfil particular control requirements.

In this research, the event-triggered SMC is utilized to control the motion of the 2-DOFs humanoid's lower-limb powered by a Series elastic actuator (SEA) in the swing phase. The SEAs have been selected to power the humanoid's joints because of various prominent advantages compared to the traditional stiff actuators [10]. In detail, the SEA can absorb high-frequency shock from the load to protect the physical system, store and release energy thanks to the elastic element, and turn the force control problem into the position control problem with high fidelity. Many reputable functional humanoids have successfully operated with the integration of the SEAs [11, 12], which proved their efficiency in the robot locomotion field.

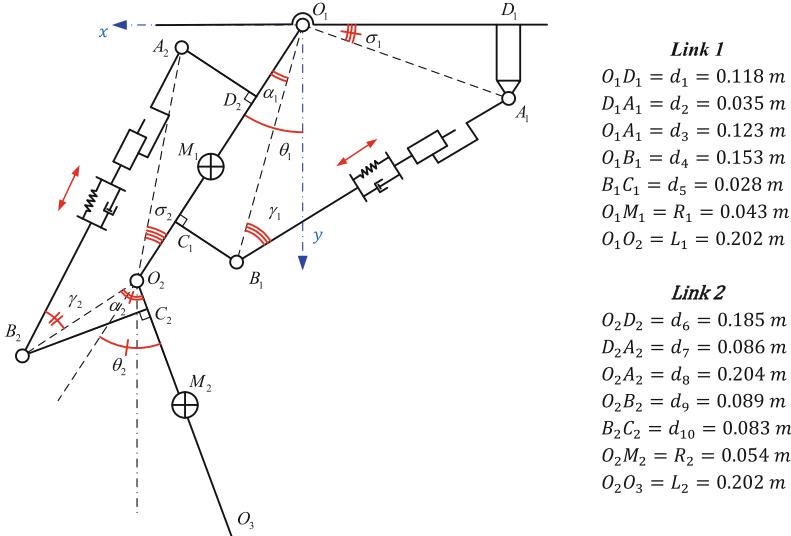
In this paper, the Euler-Lagrange model of the lower-limb system is firstly given together with the overall dynamical equations. Then, the SEA modelling is briefly demonstrated by referring to our previous research [13, 14]. Next, the event-triggered SMC algorithm is utilized to design the controller for the obtained model. Besides, the backstepping technique is applied to determine the desired control input voltage. Finally, some simulations are implemented to evaluate the lower-limb system's response quality and the effectiveness of the event-triggered approach.

## 2 System Modeling

### 2.1 Lower-Limb Modelling

The lower-limb system is demonstrated as a double compound pendulum system with two links powered by two SEAs (assumed as massless compared to the links). The points of  $M_1$  and  $M_2$  are the centre of mass's (COM) positions at each link (as shown in Fig. 1). By making translational movements, the SEAs generate rotational motion of the links

about their joints with rotational angles of  $\theta_1$  and  $\theta_2$ . The angles of  $\alpha_1, \alpha_2, \sigma_1, \sigma_2$  are constant and defined related to the dimensions of each link.



**Fig. 1.** Kinematic diagram and geometric dimensions of the humanoid's lower-limb system.

The moment inertia and total mass of each link are  $I_1, m_1$  and  $I_2, m_2$  respectively. The length of each SEA in every instant (noted as  $L_{S1}, L_{S2}$ ) are functions of the respective rotation angles  $\theta_1$  and  $\theta_2$ . By applying the *cosine rule* for triangles  $O_1A_1B_1$  and  $O_2A_2B_2$ , the lengths of the SEAs are calculated as below (proven in [13]):

$$\begin{aligned} & \left\{ \begin{array}{l} L_{S1} = \sqrt{d_3^2 + d_4^2 - 2d_3d_4\cos(\frac{\pi}{2} - \sigma_1 + \theta_1 - \alpha_1)} \\ L_{S2} = \sqrt{d_8^2 + d_9^2 - 2d_8d_9\cos(\pi - \sigma_2 + \theta_2 - \alpha_2)} \end{array} \right. \\ \Leftrightarrow & \left\{ \begin{array}{l} L_{S1} = \sqrt{d_4^2 + d_3^2 - 2d_4[d_2\cos(\theta_1 - \alpha_1) - d_1\sin(\theta_1 - \alpha_1)]} \\ L_{S2} = \sqrt{d_8^2 + d_9^2 - 2d_9[d_7\sin(\alpha_2 - \theta_2) - d_6\cos(\alpha_2 - \theta_2)]} \end{array} \right. \quad (1) \end{aligned}$$

The dynamical functions of the double compound pendulum system are determined by utilizing the Lagrange method. For the sake of simplicity, the following terms are defined:

$$\begin{aligned} \mathcal{M}_{11} &= m_1R_1^2 + I_1 + m_2L_1^2 + m_2R_2^2 + 2m_2L_1R_2\cos\theta_2 \\ \mathcal{M}_{12} &= m_2R_2^2 + m_2L_1R_2\cos\theta_2 \\ \mathcal{M}_{21} &= m_2R_2^2 + m_2L_1R_2\cos\theta_2 \\ \mathcal{M}_{22} &= m_2R_2^2 + I_2 \\ \mathcal{N}_1 &= -m_1gR_1\sin\theta_1 - m_2gL_1\sin\theta_1 + m_2gR_2\sin(\theta_2 - \theta_1) \end{aligned}$$

$$\begin{aligned}
& -m_2 L_1 R_2 \sin \theta_2 (\dot{\theta}_2^2 - 2\dot{\theta}_1 \dot{\theta}_2) - B_1 \dot{\theta}_1 \\
\mathcal{N}_2 = & -m_2 g R_2 \sin(\theta_2 - \theta_1) - m_2 L_1 R_2 \dot{\theta}_1^2 \sin \theta_2 - B_2 \dot{\theta}_2
\end{aligned} \tag{2}$$

Then, the dynamical functions for the rotation angles of  $\theta_1$  and  $\theta_2$  are:

$$\begin{cases} \mathcal{M}_{11} \ddot{\theta}_1 - \mathcal{M}_{12} \ddot{\theta}_2 = \tau_{S1} + \mathcal{N}_1 + d_1 \\ \mathcal{M}_{22} \ddot{\theta}_2 - \mathcal{M}_{21} \ddot{\theta}_1 = \tau_{S2} + \mathcal{N}_2 + d_2 \end{cases} \tag{3}$$

In which,  $\tau_{S1}$ ,  $\tau_{S2}$  are the control torques created by the SEAs at the link  $O_1 O_2$  and  $O_2 O_3$ ;  $B_1$ ,  $B_2$  are the viscous coefficients at joint  $O_1$  and  $O_2$ ; and  $d_1$ ,  $d_2$  are the bounded disturbance torques acted about joint  $O_1$  and  $O_2$ .

Furthermore, the torques generated by the SEAs can be calculated as:

$$\begin{cases} \tau_{S1} = -k_1 \Delta_1 d_4 \sin(\gamma_1) \\ \tau_{S2} = -k_2 \Delta_2 d_9 \sin(\gamma_2) \end{cases} \tag{4}$$

In which,  $k_1$  and  $k_2$  are the spring stiffness coefficients,  $\Delta_1$  and  $\Delta_2$  are the spring's deformations,  $\gamma_1$  and  $\gamma_2$  are varied angles (shown in Fig. 1) of respective joints.

By using Eq. (3) and Eq. (4), the dynamical equations are simplified as:

$$\begin{cases} \mathcal{M}_{11} \ddot{\theta}_1 - \mathcal{M}_{12} \ddot{\theta}_2 = -k \Delta_1 d_4 \sin(\gamma_1) + \mathcal{N}_1 + d_1 \\ \mathcal{M}_{22} \ddot{\theta}_2 - \mathcal{M}_{21} \ddot{\theta}_1 = -k \Delta_2 d_9 \sin(\gamma_2) + \mathcal{N}_2 + d_2 \end{cases} \tag{5}$$

## 2.2 SEA Modelling

By referring to our previous research in [13], the dynamical equations for the ball screw nut's motions inside the SEAs at two joints are given as:

$$\begin{cases} \ddot{r}_1 = U_{v1} - 48\dot{r}_1 \\ \ddot{r}_2 = U_{v2} - 48\dot{r}_2 \end{cases} \tag{6}$$

In which  $r_1$ ,  $U_{v1}$  and  $r_2$ ,  $U_{v2}$  are the nut's displacement and input voltage of each joint, respectively.

Next, it is necessary to identify the dynamical characteristic of the end effector point of the SEA. Let  $r_{B1}$  and  $r_{B2}$  respectively be the linear displacements of the SEA's end effector at each joint. Define  $\Delta_1 = r_{B1} - r_1$  and  $\Delta_2 = r_{B2} - r_2$ , the dynamical equation of the whole SEAs, with voltage input and internal spring deformation output is:

$$\begin{cases} \ddot{\Delta}_1 = \ddot{r}_{B1} + 48\dot{r}_1 - U_{v1} \\ \ddot{\Delta}_2 = \ddot{r}_{B2} + 48\dot{r}_2 - U_{v2} \end{cases} \tag{7}$$

Finally, to simplify the obtained dynamical equations, the following matrices are defined:

$$\boldsymbol{\theta} \triangleq \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \quad \boldsymbol{v}_x \triangleq \begin{bmatrix} \Delta_1 \\ \Delta_2 \end{bmatrix} \quad \boldsymbol{M} \triangleq \begin{bmatrix} \mathcal{M}_{11} & -\mathcal{M}_{12} \\ -\mathcal{M}_{21} & \mathcal{M}_{22} \end{bmatrix}$$

$$\begin{aligned} \mathbf{r} &\triangleq \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \quad \mathbf{f}_x \triangleq \mathbf{M}^{-1} \begin{bmatrix} \mathcal{N}_1 \\ \mathcal{N}_2 \end{bmatrix} \quad \mathbf{g}_x \triangleq \mathbf{M}^{-1} \begin{bmatrix} -k_1 d_4 \sin(\gamma_1) & 0 \\ 0 & -k_2 d_9 \sin(\gamma_2) \end{bmatrix} \\ \mathbf{F}_R &\triangleq \begin{bmatrix} \ddot{r}_{B1} \\ \ddot{r}_{B2} \end{bmatrix} \quad \mathbf{U}_v \triangleq \begin{bmatrix} U_{v1} \\ U_{v2} \end{bmatrix} \quad \mathbf{d} \triangleq \begin{bmatrix} \frac{-d_1}{k_1 d_4 \sin(\gamma_1)} \\ \frac{-d_2}{k_2 d_9 \sin(\gamma_2)} \end{bmatrix} \end{aligned} \quad (8)$$

As a result, sets of Eq. (5) and Eq. (7) are rewritten in the matrix form as:

$$\begin{cases} \ddot{\theta} = \mathbf{f}_x + \mathbf{g}_x(v_x + d) \\ \ddot{\Delta} = \mathbf{F}_R + 48\dot{r} - \mathbf{U}_v \end{cases} \quad (9)$$

### 3 Controller Design

First, let's define  $\mathbf{x}_1 \triangleq \theta$ ;  $\mathbf{x}_2 \triangleq \dot{\theta}$  as the state variables. The first dynamic equation in Eq. (9) can be then rewritten as:

$$\begin{cases} \dot{\mathbf{x}}_1 = \mathbf{x}_2 \\ \dot{\mathbf{x}}_2 = \mathbf{f}_x + \mathbf{g}_x(v_x + d) \end{cases} \quad (10)$$

Next, define the following matrices:

$$\begin{aligned} \mathbf{X} &\triangleq [\mathbf{x}_1 \mathbf{x}_2]^T \quad \mathbf{X}_d \triangleq [\mathbf{x}_d \dot{\mathbf{x}}_d]^T \quad \tilde{\mathbf{X}} \triangleq [\tilde{\mathbf{x}}_1 \tilde{\mathbf{x}}_2]^T \\ \tilde{\mathbf{x}}_1 &\triangleq \mathbf{x}_1 - \mathbf{x}_d \quad \tilde{\mathbf{x}}_2 \triangleq \mathbf{x}_2 - \dot{\mathbf{x}}_d \quad \boldsymbol{\epsilon} \triangleq [\tilde{\mathbf{X}} \mathbf{X}_d]^T \triangleq \begin{bmatrix} \tilde{\mathbf{x}}_1 & \mathbf{x}_d \\ \tilde{\mathbf{x}}_2 & \dot{\mathbf{x}}_d \end{bmatrix}^T \end{aligned} \quad (11)$$

In which,  $\mathbf{X}$  is the state variables,  $\mathbf{X}_d$  is the desired angle and angular velocity trajectories,  $\tilde{\mathbf{X}}$  is the tracking error, and  $\boldsymbol{\epsilon}$  is an extended error matrix utilized in the later event-triggered control process.

#### 3.1 Event-Triggered SMC

Before implementing the controller design work, the following assumptions must be considered:

- *Assumption 1:* The time derivative of desired trajectories  $\dot{\mathbf{x}}_d$  is a function of  $\mathbf{x}_d$ , which means  $\dot{\mathbf{x}}_d = \mathbf{f}_d(\mathbf{x}_d)$
- *Assumption 2:* The functions  $\mathbf{f}_x$ ,  $\mathbf{g}_x$ , and  $\mathbf{f}_d$  are Lipschitz in a compact set.
- *Assumption 3:* The following functions are bounded by positive constants:  $\|\tilde{\mathbf{X}}\| \leq r$ ,  $\|\mathbf{X}_d\| \leq h$ ,  $\|\mathbf{d}\| \leq d_0$ ,  $\|\mathbf{g}_x\| \leq g_0$ , in which  $r$  is a designed value.

Then, the sliding function vector is given as follow:

$$s = c\tilde{\mathbf{x}}_1 + \tilde{\mathbf{x}}_2 \quad (12)$$

In which,  $c$  is a positive constant. The control matrix  $\mathbf{v}_x$  is chosen as follow:

$$\mathbf{v}_x = \mathbf{g}_x^{-1}[-f_x + \ddot{x}_d - c\tilde{x}_2 - \rho \text{sign}(s)] \quad (13)$$

where  $\rho$  is a constant that  $\rho \geq g_0 d_0 + \eta$  ( $\eta > 0$ ). Next, the Candidate Lyapunov Function (CLF) for the first dynamical equation is proposed as below:

$$V_x = \frac{1}{2}\mathbf{s}^T \mathbf{s} \quad (14)$$

Taking the first-time derivative of Eq. (14), it yields:

$$\dot{V}_x = \mathbf{s}^T [c\tilde{x}_2 + f_x + \mathbf{g}_x(\mathbf{v}_x + \mathbf{d}) - \ddot{x}_d] \leq -\eta\sqrt{1+c^2}\|\tilde{\mathbf{X}}\| \quad (15)$$

Let  $t_i$  be the  $i^{th}$  triggered instant,  $t$  is the period between two consecutive triggered instants, which means  $t \in [t_i; t_{i+1})$ . The difference between  $\mathbf{e}(t_i)$  at a triggered instant and  $\mathbf{e}(t)$  are called the event-triggered error:

$$\mathbf{e}(t) = \mathbf{e}(t_i) - \mathbf{e}(t) \quad (16)$$

Consequently, the control law at the triggered instant is inferred as:

$$\mathbf{v}_x(t_i) = \mathbf{g}_x^{-1}(t_i)[-f_x(t_i) + \ddot{x}_d(t_i) - c\tilde{x}_2(t_i) - \rho \text{sign}(s(t_i))] \quad (17)$$

**Noted:** For simplified expression, the afterwards value of variables at the feedback sampling instant  $t$  would be kept in original form (e.g.  $\mathbf{x}_1$  instead of  $\mathbf{x}_1(t)$ ).

As the event-triggered law now controls the lower-limb system in Eq. (17), the first derivative of the CLF along the trajectories of Eq. (13) is rewritten as:

$$\begin{aligned} \dot{V}_x &= \frac{\partial V_x}{\partial \tilde{\mathbf{X}}} \frac{d\tilde{\mathbf{X}}}{dt} = \frac{\partial V_x}{\partial \tilde{\mathbf{X}}} \left[ \begin{array}{l} \dot{\tilde{x}}_1 \\ f_x + \mathbf{g}_x(\mathbf{v}_x + \mathbf{d}) - \dot{\tilde{x}}_d \end{array} \right] + \frac{\partial V_x}{\partial \tilde{\mathbf{X}}} \left[ \begin{array}{l} 0 \\ \mathbf{g}_x(\mathbf{v}_x(t_i) - \mathbf{v}_x) \end{array} \right] \\ &\leq -\eta\sqrt{1+c^2}\|\tilde{\mathbf{X}}\| + (1+c^2)\|\tilde{\mathbf{X}}\|L\|\mathbf{e}\| \quad (L > 0) \end{aligned} \quad (18)$$

Consequently, the system is ensured to be stable when the following that is called the event-triggered condition, holds:

$$\|\mathbf{e}\| < \frac{\eta}{L\sqrt{1+c^2}} \quad (19)$$

Since the inter-event time  $T_i$  features the Zeno phenomenon, it is essential to determine the lower bound of  $T_i$ . Considering the first time-derivative of tracking error matrix, it yields:

$$\dot{\tilde{\mathbf{X}}} = \dot{\mathbf{X}} - \dot{\mathbf{X}}_d = \begin{bmatrix} x_2 \\ f_x \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{g}_x \end{bmatrix}(\mathbf{v}_x + \mathbf{d}) - \begin{bmatrix} \dot{x}_d \\ \ddot{x}_d \end{bmatrix} \quad (20)$$

Taking the  $2-norm$  of Eq. (20), it can be obtained:

$$\|\dot{\tilde{\mathbf{X}}}\| = \left\| \begin{bmatrix} x_2 \\ f_x \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{g}_x \end{bmatrix}(\mathbf{v}_x(t_i) + \mathbf{d}) - \begin{bmatrix} \dot{x}_d \\ \ddot{x}_d \end{bmatrix} \right\|$$

$$\leq \lambda(r + h + d_0 + \|e\|) \quad (\lambda > 0) \quad (21)$$

From the relations defined in Eq. (16) and Eq. (21), the first time-derivative of event-triggered error matrix  $e$  can be calculated as:

$$\frac{d\|e\|}{dt} \leq \|\dot{e}\| \leq \|\dot{\tilde{X}}\| \leq \lambda(r + h + d_0 + \|e\|) \quad (22)$$

Consequently, from differentiating inequation in Eq. (22), the inter-event time lower bound is determined as:

$$T_i \geq \frac{1}{\lambda} \ln \left( r + h + d_0 + \frac{\eta}{L\sqrt{1+c^2}} \right) > 0 \quad (23)$$

### 3.2 Backstepping Process

Consider the first augmented dynamic system:

$$\begin{cases} \dot{\tilde{X}} = \begin{bmatrix} \mathbf{x}_2 \\ f_x + g_x d \\ \dot{z}_1 = v_1 \end{bmatrix} \\ \begin{bmatrix} 0 \\ g_x \end{bmatrix} z_1 \end{cases} \quad (24)$$

As the control law  $v_x(t_i)$  is proven to be able to ensure the stability of the overall system, the control goal now is to drive  $z_1$  follow the desired value of  $v_x(t_i)$ . First, the augmented CLF is proposed as:

$$V_1 = V_x + \frac{1}{2}(v_x(t_i) - z_1)^T(v_x(t_i) - z_1) \quad (25)$$

The first time-derivative of  $V_1$  yields:

$$\begin{aligned} \dot{V}_1 = & s^T [c\tilde{x}_2 + f_x + g_x(v_x(t_i) + d) - \ddot{x}_d] \\ & + s^T g_x(z_1 - v_x(t_i)) + (v_x(t_i) - z_1)^T(\dot{v}_x(t_i) - v_1) \end{aligned} \quad (26)$$

To stabilize the proposed extended system, the first pseudo-control law is:

$$v_1 = k_{p1}(v_x(t_i) - z_1) + \dot{v}_x(t_i) - (g_x)^T s \quad (k_{p1} > 0) \quad (27)$$

Through the backstepping design method, the second extended system is proposed as:

$$\begin{cases} \dot{X}_1 = f_1(X_1) + g_1(X_1)z_2 \\ \dot{z}_2 = f_2(X_2) + g_2(X_2)U_v \end{cases} \quad (28)$$

where:  $X_1 \triangleq \begin{bmatrix} X \\ z_1 \end{bmatrix}$ ,  $f_1(X_1) \triangleq \begin{bmatrix} \mathbf{x}_2 \\ f_x + g_x d \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ g_x \end{bmatrix} z_1$ ,  $g_1(X_1) \triangleq \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ ,  $X_2 \triangleq \begin{bmatrix} X_1 \\ z_2 \end{bmatrix}$ ,  $f_2(X_2) \triangleq F_R + 48\dot{r}$ ,  $g_2(X_2) \triangleq -1$ .

The  $X_1$ -system is assumed to be stable by the control law  $v_1$ . Then, to make  $z_2$  reach the value of  $v_1$ , the 2<sup>nd</sup> augmented CLF is proposed as:

$$V_2 = V_1 + \frac{1}{2}(\mathbf{v}_1 - \mathbf{z}_2)^T(\mathbf{v}_1 - \mathbf{z}_2) \quad (29)$$

Take the first time-derivative of Eq. (29), it yields:

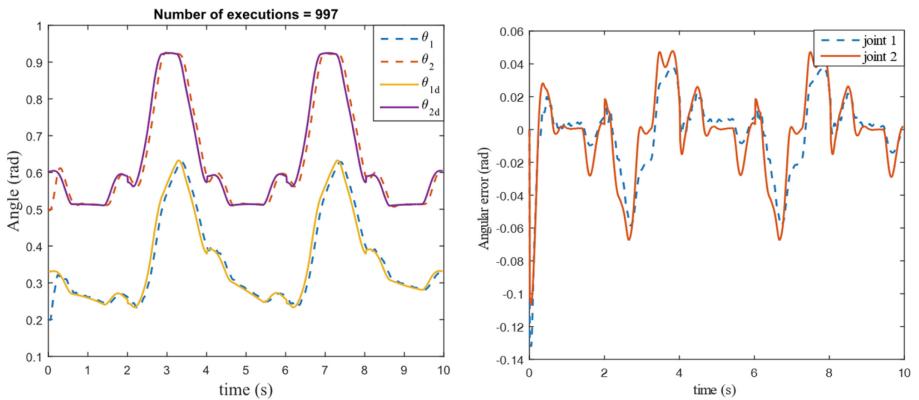
$$\begin{aligned} \dot{V}_2 &= \dot{V}_x + (\mathbf{v}_x(t_i) - \mathbf{z}_1)^T(\dot{\mathbf{v}}_x(t_i) - \mathbf{v}_1) + (\mathbf{v}_x(t_i) - \mathbf{z}_1)^T(\mathbf{v}_1 - \mathbf{z}_2) \\ &\quad + (\mathbf{v}_1 - \mathbf{z}_2)^T[\dot{\mathbf{v}}_1 - \mathbf{f}_2 - \mathbf{g}_2 \mathbf{U}_v] \end{aligned} \quad (30)$$

In order to prove that the error term  $(\mathbf{v}_1 - \mathbf{z}_2)$  is bounded, the first time-derivative of  $V_2$  must be proven to be  $\leq 0$ . Then, the control input signal of the second extended system, which is also the input voltage of the SEA, is chosen as:

$$\mathbf{U}_v = \mathbf{g}_2^{-1}[-\mathbf{f}_2 + k_{p2}(\mathbf{v}_1 - \mathbf{z}_2) + \dot{\mathbf{v}}_1 + (\mathbf{v}_x(t_i) - \mathbf{z}_1)](k_{p2} > 0) \quad (31)$$

## 4 Simulation and Analysis

In this section, a simulation is conducted to compare the response quality between traditional SMC and event-triggered SMC. Generally, the remote-control process for the robotics system suffers significant delay due to massive data transmission and computer resource consumption. Besides, the low sampling time control signal might become redundant since the heavy physical structure of the robot can hardly respond quickly enough to every single updated demand. Therefore, the event-triggered SMC would be applied to deal with the above downsides.



**Fig. 2.** The rotational angles response of the lower-limb system (left) and the tracking errors (right) when being controlled by traditional SMC

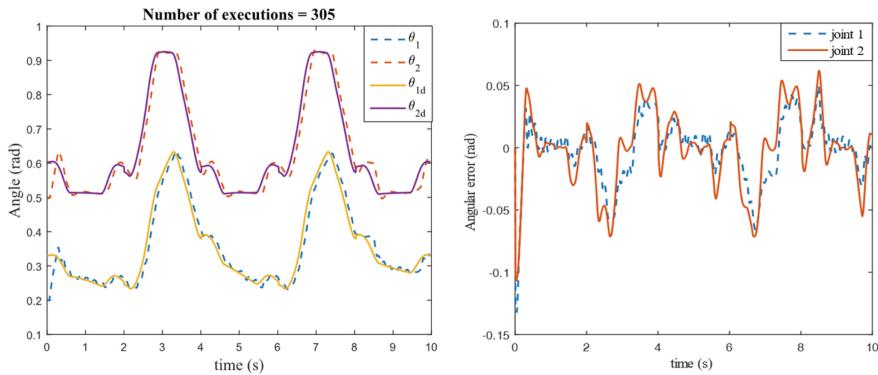
The desired trajectories of rotation angle are derived from our previous research in [13]. The period of one walking cycle in simulation is 10 s and the sampling time of

output state is 10 ms. The physical parameters of the lower-limb system are:  $m_1 = 1.07$  kg,  $m_2 = 0.89$  kg,  $I_1 = 0.028$  kgm $^2$ ,  $I_2 = 0.002$  kgm $^2$ ,  $K_{s1} = 20000$  N/m,  $K_{s2} = 20000$  N/m,  $B_1 = 0.3$  Nms/rad,  $B_2 = 0.3$  Nms/rad.

First, the lower-limb system is controlled with the traditional SMC strategy. A simple tuning process is applied to find out the appropriate controller's gain values, which is  $c = 30$ ,  $\rho = 20$ ,  $k_{p1} = 3$ ,  $k_{p2} = 5$ . The resulted responses and tracking errors are demonstrated in Fig. 2.

Overall, the tracking process shows a constant delay compared to the desired trajectories due to the inertia of the lower-limb system and the integration of the backstepping process. The maximum error at both joints is about 0.07 rad and mainly occurs at the peaks of the graphs where the links suddenly change their rotational direction. The number of execution instants triggered by traditional SMC is 997, which is almost the same as the number of output samples taken in one walking cycle.

Next, an event-triggered SMC strategy is applied for the humanoid's lower-limb system. The controller gain values utilized in the traditional SMC are kept unchanged. The control law of the event-triggered SMC is designed with the bounding coefficients of:  $r = 0.2$ ,  $\eta = 0.3$ ,  $L = 0.85$ .

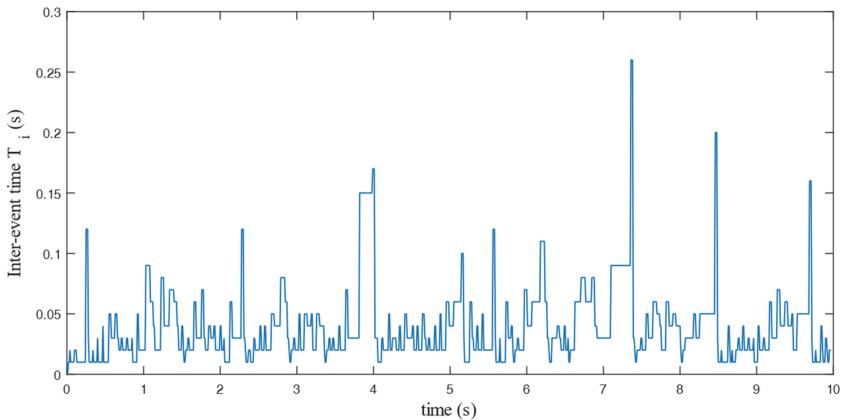


**Fig. 3.** The angles response of the lower-limb system (left) and the tracking errors (right) when being controlled by event-triggered SMC

The results illustrated in Fig. 3 prove the almost equal quality of response from the event-triggered SMC compared to the traditional one. In specific, besides the delay characteristic from the system itself, the tracking errors in the event-triggered SMC case are slightly larger, about 0.075 rad at the peak of the graph. However, in 90% of the other part of the graph, the tracking errors of both control strategies are the same. Moreover, it should be noted that the number of executions in event-triggered SMC, with only 305 triggered instants, is significantly lower than that of traditional strategy. This number clearly shows the advantage of event-triggered SMC when it can help reduce the total transmission time by more than threefold, which saves up much more computer resources than the traditional SMC.

As shown in Fig. 4, the highest inter-event time can be up to 0.25 s and mainly maintains at about 0.05 s, which proves for the efficiency of the proposed controller,

compared to the constant 0.01 s inter-event time of traditional SMC. This means the controller only generates and sends the new control signal every 0.05 s in average while the sampling time in the whole system is 0.01 s, which significantly reduces the load of data needed to transmit in the entire process and considerably save up the calculation time. In the periods when the control signal is kept unchanged, the lower-limb system is ensured to maintain the tracking process with the tracking error  $\|\tilde{X}\|$  lower than the designed value  $r$ . The trigger instant occurs whenever  $\|\tilde{X}\| > r$  or when the Lyapunov stability standard of the system is violated.



**Fig. 4.** The inter-event time of the humanoid's lower-limb system when being controlled by traditional SMC

## 5 Conclusion

In summary, the event-triggered SMC strategy has clearly shown its advantages when applied in the complex physical system with considerable inertia like lower-limb powered by SEAs. With the remarkable potential of reducing overall execution time, saving up resources for transmission, and consuming less energy during operation, the event-triggered approach in general and event-triggered SMC, in particular, would be appropriate for practical applications. In the future, a comprehensive practical experiment will be conducted to further evaluate the efficiency and optimize the performance of this control strategy.

For further research intention, the Event-triggered SMC algorithm will be utilized to control the whole lower body of the humanoid robot with 6 DOFs powered by the SEAs during a walking cycle. The entire process would be controlled and managed remotely between the humanoid and the master control device through a wireless transmission protocol, which may lead to some delay phenomenon as for the distant control and for the high complexity of the system itself. In that case, the event-triggered SMC would be an appropriate choice to optimize the amount of data transmitted to the humanoid and considerably save up consumed energy.

**Acknowledgements.** This research is supported by DCSELAB and funded by Vietnam National University Ho Chi Minh City (VNU-HCM) under grant number TX2022-20b-01. We acknowledge the support of time and facilities from Ho Chi Minh University of Technology (HCMUT), VNU-HCM for this study.

## References

1. Mustafa, A., et al.: Event-Triggered sliding mode control for trajectory tracking of nonlinear systems. *IEEE/CAA J. Autom. Sin.* **7**(1), 307–314 (2020)
2. Gupta, S.C.: Increasing the sampling efficiency for a control system. *IEEE Trans. Automat. Contr.* **8**(3), 263–264 (1963)
3. Tabuada, P.: Event-triggered real-time scheduling of stabilizing control tasks. *IEEE Trans. Automat. Contr.* **52**(9), 1680–1685 (2007)
4. Wu, B., Shen, Q., Cao, X.: Event-triggered attitude control of spacecraft. *Adv. Sp. Res.* **61**(3), 927–934 (2018)
5. Zhang, C., Wu, J., Dai, M.Z., Li, B., Wang, M.: Event- and self-triggered control of attitude coordination to multi-spacecraft system. *Aircr. Eng. Aerosp. Technol.* **92**(7), 1085–1092 (2020)
6. Zhang, K., Su, R., Zhang, H., Tian, Y.: Adaptive resilient event-triggered control design of autonomous vehicles with an iterative single critic learning framework. *IEEE Trans. Neural Networks Learn. Syst.* (2021)
7. Hung, N.T., Pascoal, A.M.: Cooperative path following of autonomous vehicles with model predictive control and event triggered communications. *IFAC-PapersOnLine* **51**(20), 562–567 (2018)
8. Kumar, N., Porwal, A., Singh, A.R., Naskar, R., Purwar, S.: Event triggered control of robot manipulator. In: 2019 6th International Conference on Signal Processing and Integrated Networks, SPIN 2019, pp. 362–366 (2019)
9. Semeniuta, O., Falkman, P.: Event-driven industrial robot control architecture for the Adept V+ platform. *PeerJ Comput. Sci.* **5**(7), e207 (2019)
10. Pratt, G.A., Williamson, M.M., Dillworth, P., Pratt, J., Wright, A.: Stiffness isn't everything. *Exp. Robot.* IV, pp. 253–262, December 1997
11. Radford, N.A., et al.: Valkyrie: NASA's first bipedal humanoid robot. *J. F. Robot.* **32**(3), 397–419 (2015)
12. Paine, N., et al.: Actuator control for the NASA-JSC valkyrie humanoid robot: a decoupled dynamics approach for torque control of series elastic robots. *J. F. Robot.* **32**(3), 378–396 (2015)
13. Truong, K.D., et al.: Design of series elastic actuator applied for humanoid. In: International Conference on Advanced Mechatronic Systems, ICAMechS, vol. 2020, pp. 23–28, December 2020
14. Lanh, L.A.K., et al.: Design of robust controller applied for series elastic actuators in controlling humanoid's joint (2021)



# Blockchain in Enterprise Applications: An Introduction

Hung Ho-Dac, Van Huu Tran, The Bao Nguyen, Hai Vinh Cuong Nguyen,  
and Van Len Vo(

Thu Dau Mot University, Ho Chi Minh, Binh Duong, Vietnam  
`{hunghd, huutv, baont, cuongnhv, lenvv}@tdmu.edu.vn`

**Abstract.** Recently, blockchain has rapidly been developed and applied in various fields, not only in cryptocurrencies, including but not limited to digital exchanges, the internet of things, supply chains. Blockchain primitively provides privacy, security, and transparency that are needed to be consolidated in centralized applications. Most enterprise applications have been developed and deployed using a centralized approach which makes them inherent weaknesses of centralized applications. In this work, we propose an approach to apply blockchain in enterprise applications with multi-party smart contracts integration. This integration helps enterprise applications improve privacy, security, and transparency. We also propose a procedure to evaluate the suitability of blockchain for enterprise applications to avoid abuse and irrationality. The analysis demonstrates that choosing the feasible modules to deploy on the blockchain improves the reliability of enterprise applications.

**Keywords:** Blockchain · Enterprise applications · Multi-party smart contracts

## 1 Introduction

Blockchain is one of the emerging technologies of this decade. Although there are no formal standardization organizations for blockchain so far, the undeniable fact is that blockchain is being applied in many fields and has great influence [1]. A blockchain essentially is a distributed network of nodes with specific storage, communication, and consensus mechanism to store synchronous copies of the ledger. This technical architecture has helped blockchain have features that promise to overcome the limitations of current systems.

Most of today's enterprise applications are deployed towards a centralized system approach. Enterprises fully own the application as well as the information related to the application by deploying on dedicated servers or cloud servers and taking full control. This approach suffers from a number of privacy, security, and transparency issues including but not limited to information that can be leaked by an unauthorized actor, information that can be stolen or modified, information that may be concealed or tampered. Meanwhile, blockchain strongly supports privacy, security, and transparency [2, 3]. Therefore, we propose an approach to integrate blockchain into enterprise applications to overcome current limitations. In addition, from an architectural perspective, we

also examine the feasibility of software architecture when integrating blockchain as well as the procedure of evaluating the blockchain compatibility with enterprise applications.

The study is organized as follows: Sect. 2 presents related work; Sect. 3 presents the blockchain foundation; Sect. 4 presents the blockchain-based enterprise applications architecture and the procedure of evaluating compatibility; Sect. 5 presents the compatibility evaluation experiments; Sect. 6 presents conclusions and directions for future research.

## 2 Related work

Currently, there are many studies demonstrating the application of blockchain in various fields including but not limited to education [4], human resource management [5], Internet of Things [6], healthcare [7], entrepreneurship and innovation [8], cybersecurity [9].

Related to the field of enterprise applications, many studies focus on digital content and business processes. Traditional content is rapidly being changed by digital content due to the digital revolution. Hasan et al. [10] proposed a trusted, decentralized, and secure Proof of Delivery method/framework for digital content trade and delivery. Their approach leverages blockchain, Ethereum smart contracts, and the InterPlanetary File System. The authors also conducted a security study to demonstrate that the smart contract code is safe from known attacks and devoid of security vulnerabilities, defects. In another study, Kinomoto et al. [11] presented a blockchain-based anonymized dataset distribution platform. The platform is built from peers who function as data brokers, data receivers, and transaction verifiers. Blockchain is utilized to record all anonymized dataset transactions between a data broker and a data receiver. The authors created a platform prototype and evaluated it. In another study, a new Multimedia Blockchain framework based on digital watermarking that handles trustworthy digital content delivery mechanisms was proposed by Bhowmik et al. [12]. A cryptographic hash that incorporates transaction histories and an image hash that retains retrievable original media content make up the unique watermark information. In the field of crypto currency, Sun et al. [13] proposed a multi-blockchain-based digital currency (MBDC) model for central bank digital currency. MBDC employs a multi-blockchain architecture to enhance scalability and process payment. The blockchain is built and maintained by the central bank, commercial banks, and other authorities under this approach. The authors also illustrate the model's performance in terms of scalability and transaction execution speed, both theoretically and practically.

The application of blockchain in the business processes implementation is not only desirable but also improvable due to basic characteristics of blockchain. FIDES [14], a framework for increasing the trustworthiness of business operations by leveraging blockchain was introduced by Johng et al. provide. A systematic and reasonable approach with a clear and formal representation is a major technical achievement in FIDES. Another contribution is a preliminary reference process model for diagnosing difficulties in AS-IS business processes related to stakeholder trust concerns, investigating TO-BE business process alternatives that utilize blockchain, and selecting one of the alternatives. Carminati et al. [15] investigated how to maintain data confidentiality during the

execution of business processes even in the presence of an untrusted data source. The authors presented a technique in which they convert a BPEL into a confidential smart contract, encrypting sensitive data while keeping the smart contract executable. They've also experimented with the overhead that our method necessitates. Jhala et al. [16] proposed a method for coordinating collaborative engineering projects. Trust management is critical in such initiatives. When many parties are participating in a project, it is critical that no one party monopolizes the project and introduces patches that the other parties are unaware of. Furthermore, no one should be able to reveal the whole substance of the work to a third party. Collaborators would publish their modifications, and they would only be accepted to the blockchain once everyone agreed. Their suggested approach ensures the project's safe distribution as well as contributors' non-repudiation.

Most studies focus on the application of blockchain in specific fields. From a perspective of software architecture, our approach considers a combination of digital content and business process for enterprise application and the corresponding infrastructure. On the other hand, we also investigate the key factor to determine whether a module should be on-chain or off-chain.

### 3 Blockchain Foundation

Blockchain enables safe data storage via an exceedingly complicated encryption mechanism. In a blockchain, data that existed in previous blocks are not modified, data is only added to the current block. Each block contains information about its creation time, transaction data, and is connected to the previous block [17]. There is no way to modify data once it has been accepted by the network. Data is stored in an immutable manner using timestamps, public audit, and consensus mechanisms [18]. Thus blockchain can protect data against tampering and fraud.

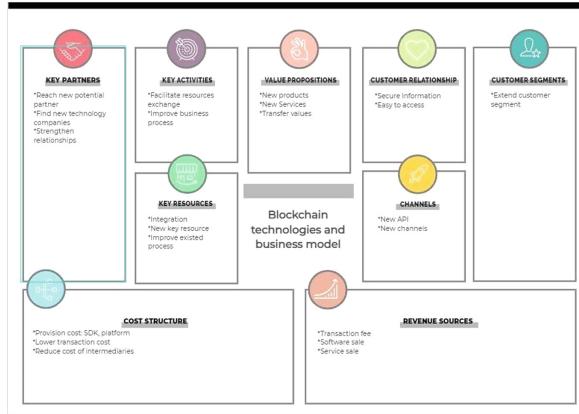
The first blockchain network was introduced in 2009. It is known as Blockchain 1.0. The concept revolved around payment and how it could be used to generate cryptocurrency. Next year, its second version was released. It introduced smart contracts and many financial services for applications. The next generation of Blockchain which is well-known as Blockchain 3.0, it's designed for applications that can be decentralized related to health care, IoT, business, and smart city [19]. Recently, Blockchain 4.0 has been introduced. It is built to be suitable for Industry 4.0 based applications. It makes use of smart contracts, which do away with the need for paper contracts and regulate the network by consensus.

Based on access level, blockchain can be divided into two types: private and public. Public blockchains are those in which anybody can be a participant. Anyone can join a public blockchain without revealing their identity. A native cryptocurrency is usually used in public blockchains, and the consensus is commonly based on "proof of work" (PoW) and economic incentives. Bitcoin and Ethereum are two examples of this type of blockchain. On the other hand, the private blockchain runs a blockchain among a set of identified and known participants. A well-known of this type is IBM's Hyperledger.

## 4 Blockchain-Based Enterprise Application Architecture

### 4.1 Blockchain Technologies and Their Impact on Business Model

Blockchain technology is an opportunity as well as a challenge for the business enterprise. Businesses or organizations that want to apply blockchain in their field need to consider many factors such as how their customer segment changes with the implementation of blockchain; what architect they should choose and whether they can compete with other businesses regardless of the costs incurred when building or transforming their enterprise architecture. Vida J. Morkunas and their partner published [15] How blockchain technologies impact your business model (2019) in which they discussed how Blockchain might potentially impact traditional business models. As we can see in Fig. 1, using Business Model Canvas [20] the authors showed that each of the nine essential elements can be affected by Blockchain technology: Expanding their Customer segments by accessing to a target market that was previously not reachable; Influencing customer value by reducing middle layers; Improving their Business Channel with new API, SDKs; Automating the process to communication in Customer relationships; stabilizing revenue source with new type of fundraising and services; remodeling Key resources and activities by improving in verification, Documentation and audit; strengthening and reaching new key partnerships and last but not least reducing intermediaries costs and transaction cost consider both time and money as factors.



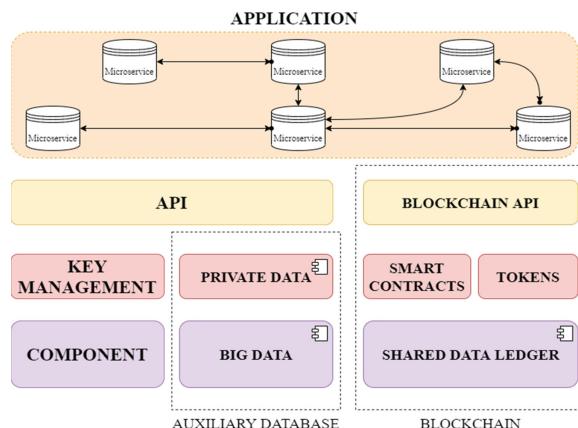
**Fig. 1.** Blockchain technology affects nine elements of the business model canvas

### 4.2 Blockchain Based Software Architecture

From an architectural perspective, blockchain is a complex network-based software component, with the ability to provide data storage, computing services, and communication services. Not only that, some of its features such as immutability, decentralized, distributed ledgers or consensus can be integrated into a practical architecture. Our proposed model integrates blockchain into traditional software architecture Fig. 2, using

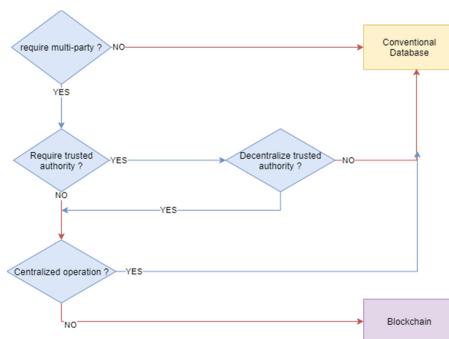
blockchain ledger, smart contracts and making use of their computational power to participate in the consensus mechanism ensuring that any changes to a licensed transaction are legal in addition to the ability to automatically archive all transactions that have occurred in a particular blockchain.

Our proposed architect allocated Blockchain to the following components: Storage, Communication Mechanism and Asset Management and Control Mechanism. The details are presented as follows.



**Fig. 2.** Blockchain based application in a software architecture

In its early days, Blockchain was used as a public ledger which stored all the transactions that have occurred in the system. Later, the technology concept was generalized to a distributed ledger which can be used to include transactions that weren't cryptocurrency. In our proposed architecture, data is added into contract storage which contains the information of the transfer value. After the transaction is included in the blockchain, the data can be publicly accessible by all the participants within the network. When compared to the traditional centralized data storage model or cloud Services, blockchain does not



**Fig. 3.** Process to evaluate the suitability of blockchain technology

need to put their trust in any individual. Users can also monitor or even join the node that contains the relevant blockchain.

Blockchain has many applications, but it also has many limitations. When considering choosing between a blockchain database and a traditional database, we have to evaluate the suitability of blockchain technology against a specific scenario. Our research proposes a procedure (Fig. 3) as a template, hence when considering the possibility that the system is required to support multi-party, the block chain will be priority over conventional databases. The architecture of blockchain is not optimal when applied on systems that only serve independent individuals, so for a system that does not need to serve different parties, applying blockchain is redundant and inefficient. Another factor when considering the adoption of blockchain technology is whether a Trusted Authority is required and whether the Trusted Authority needs to be decentralized. Depending on the function of the system, if a centralized Trusted Authority is required to perform the system's tasks safely and securely, the conventional database option will be preferred.

In our architecture, blockchain is used as a communication component between components of a software architecture. Components in the application layer use transactions to transmit information to the blockchain and use queries to retrieve the said data in case of need. In addition, the transactions in the system are also cached to avoid the network being congested while processing queries. In the network, smart contracts which are created by submitting transactions with the code of the contracts to the blockchain network, can also be used to coordinate computation between different components of an architecture.

Following the latter generation of blockchain (Ethereum) our smart contracts and data structures are much more expressive and flexible. Using the concept of tokenization combined with smart contracts, the control of cryptographic token can be implemented as a multiway process. The ownership of corresponding assets can be validated when an asset is represented using a cryptographic token and then it can be traced back when said user wants to recover their own asset.

## 5 Experiment

We implement our experiments on Dell XPS 13 9370 Laptop, Intel(R) Core(TM) i5-8250U CPU @ 1.60 GHz 1.80 GHz, and 8 GB RAM and Ubuntu 20.0.4 OS installed. We build two blockchain networks: Hyperledger Fabric v2.1.0 represents private blockchains, and Ethereum v1.2.1 represents private blockchains. We then implement three smart contracts such as Open, Inquiry, and Transfer, to perform basic transactions in the bank, where:

- Query provides a transaction for querying account information.
- Open provides transactions for opening a bank account.
- Transfer provides transfers between accounts.

In experiment #1, we evaluate the overall performance of Hyperledger, a typical enterprise blockchain. We use 1000 transactions each type with a transmission rate of 100 transactions per second. The results of latency and throughput are detailed in

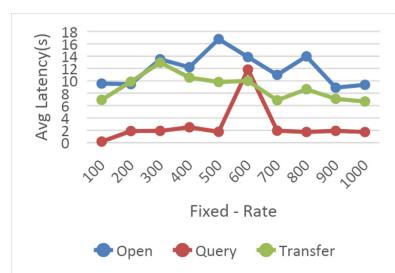
**Table 1.** We discovered that the Open and Query transactions complete 100% of transactions, while Transfer completed only 44% of transactions due to conflicts with Query transactions at a time. Whereas TPS of Query, Open is 100, 55.3, and 65.5 respectively. And the average Latency of Query, Open, and Transfer transactions is 0.18(s), 9.56(s), and 6.92(s). Additionally, we also conduct spectral study of transmission rate from 100 to 1000 in 100 increments to determine detailed results of Hyperledger performance. The results are shown in Figs. 4, 5, and 6. With 300 transactions per second transfer rate with a low profile server, the average latency is less than 10 s per transaction, which is acceptable in banking transactions. Besides, with a server with a strong configuration in the bank, the transaction efficiency can be maximized and the transaction latency will be much lower.

**Table 1.** Hyperledger Fabric v2.1.0 performance stats, with 1000 transactions per type and fixed transmission rate 100 transactions per second.

Name	Succ	Fail	Send Rate (TPS)	Max Latency (s)	Min Latency (s)	Avg Latency (s)	Throughput (TPS)
open	1000	0	109.1	12.30	1.43	9.56	55.3
query	1000	0	100.5	0.72	0.01	0.18	100.0
transfer	441	559	104.4	9.90	1.06	6.92	65.5



**Fig. 4.** Statistics of successful transactions with Open, Query, Transfer transaction types with different fixed rates.

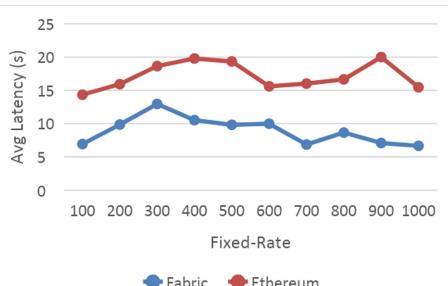


**Fig. 5.** Statistics of Avg. Latency with Open, Query, Transfer transaction types with different fixed rates

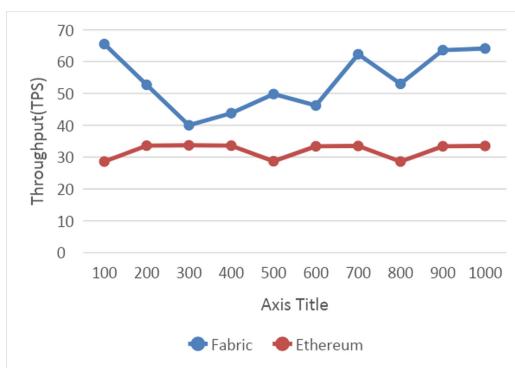
In experiment #2, we compared Hyperledger with Ethereum on latency and throughput metrics to discover the difference between the two blockchains. We found that Hyperledger's average latency and throughput are better than Ethereum's, but the difference is small. Detailed results are presented in Figs. 7 and 8. The evaluation and comparison of two typical representatives of the private blockchain and the public blockchain have shown the feasibility of applying blockchain to enterprise applications.



**Fig. 6.** Statistics of throughput (TPS) with different fixed rates for Open, Query, Transfer transaction types



**Fig. 7.** Average latency statistics per transfer transaction with different fixed-rates.



**Fig. 8.** Throughput statistics (TPS) with different fixed-rate

## 6 Conclusion

The application of blockchain to enterprise applications will help overcome their current limitations. The results carried out in the study show that the feasibility of applying blockchain to enterprise applications is completely feasible. In addition, using the right architecture based on the on-chain and off-chain component identification evaluation process will help optimize costs for enterprise applications.

**Acknowledgments.** This research is funded by Thu Dau Mot University under grant number DT.21.1-077. Also, we would like to thank Institute of Engineering - Technology (TDMU) for the support of time and facilities for this study.

## References

1. Tapscott, D., Tapscott, A.: Blockchain revolution. Penguin Publishing Group, New York (2016)

2. Al Amiri, W., et al.: Privacy-preserving smart parking system using blockchain and private information retrieval. In: 2019 International Conference on Smart Applications, Communications and Networking (SmartNets). IEEE (2019)
3. Vu, T.X., Chatzinotas, S., Ottersten, B.: Blockchain-based content delivery networks: content transparency meets user privacy. In: 2019 IEEE Wireless Communications and Networking Conference (WCNC). IEEE (2019)
4. Sun, H., Wang, X., Wang, X.: Application of blockchain technology in online education. *Int. J. Emerg. Technol. Learn.* **13**(10), 252 (2018)
5. Onik, M.M.H., Miraz, M.H., Kim, C.-S.: A recruitment and human resource management technique using blockchain technology for industry 4.0. In: Smart Cities Symposium 2018. IET (2018)
6. Kumar, N.M., Mallick, P.K.: Blockchain technology for security issues and challenges in IoT. *Procedia Comput. Sci.* **132**, 1815–1823 (2018)
7. Farouk, A., et al.: Blockchain platform for industrial healthcare: vision and future opportunities. *Comput. Commun.* **154**, 223–235 (2020)
8. Chen, Y.: Blockchain tokens and the potential democratization of entrepreneurship and innovation. *Bus. Horiz.* **61**(4), 567–575 (2018)
9. Fernandez-Carames, T.M., Fraga-Lamas, P.: A review on the application of blockchain to the next generation of cybersecure industry 4.0 smart factories. *IEEE Access* **7**, 45201–45218 (2019)
10. Hasan, H.R., Salah, K.: Proof of delivery of digital assets using blockchain and smart contracts. *IEEE Access* **6**, 65439–65448 (2018)
11. Lee, R. (ed.): SERA 2017. SCI, vol. 722. Springer, Cham (2018). <https://doi.org/10.1007/978-3-319-61388-8>
12. Bhowmik, D., Feng, T.: The multimedia blockchain: a distributed and tamper-proof media transaction framework. In: 2017 22nd International Conference on Digital Signal Processing (DSP). IEEE (2017)
13. Sun, H., et al.: Multi-blockchain model for central bank digital currency. In: 2017 18th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT). IEEE (2017)
14. Johng, H., et al.: Using blockchain to enhance the trustworthiness of business processes: a goal-oriented approach. In: 2018 IEEE International Conference On Services Computing (SCC). IEEE (2018). <https://doi.org/10.1109/scc.2018.00041>. Accessed 4 Sept 2021.
15. Carminati, B., Rondanini, C., Ferrari, E.: Confidential business process execution on blockchain. In: 2018 IEEE International Conference on Web Services (ICWS). IEEE (2018)
16. Jhala, K.S., Oak, R., Khare, M.: Smart collaboration mechanism using blockchain technology. In: 2018 5th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2018 4th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom). IEEE (2018)
17. Bodkhe, U., et al.: Blockchain for industry 4.0: a comprehensive review. *IEEE Access* **8**, 99 (2020)
18. Mingxiao, D., Xiaofeng, M., Zhe, Z., Xiangwei, W., Qijun, C.: A review on consensus algorithms of blockchain. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff Center, Banff, Canada, 5–8 Oct 2017
19. Vora, J., et al.: Bheem: A blockchain-based framework for securing electronic health records. In: 2018 IEEE Globecom Workshops (GC Wkshps). IEEE (2018)
20. Osterwalder, A., Pigneur, Y.: Business model generation: a handbook for visionaries, game changers, and challengers. Wiley, Hoboken, NJ (2013)



# Backstepping Super-Twisting Based Sliding Mode Controller for a Lower Limb Actuator

Tri Duc Tran<sup>3</sup>, Van Tu Duong<sup>1,2,3</sup>, Huy Hung Nguyen<sup>3,4</sup>, and Tan Tien Nguyen<sup>1,2,3(✉)</sup>

<sup>1</sup> Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology (HCMUT),  
268 Ly Thuong Kiet, District 10, Ho Chi Minh City, Vietnam  
nttien@hcmut.edu.vn

<sup>2</sup> Vietnam National University Ho Chi Minh City, Linh Trung Ward, Thu Duc District, Ho Chi  
Minh City, Vietnam

<sup>3</sup> National Key Laboratory of Digital Control and System Engineering (DCSELAB), HCMUT,  
268 Ly Thuong Kiet, District 10, Ho Chi Minh City, Vietnam

<sup>4</sup> Faculty of Electronics and Telecommunication, Saigon University, Ho Chi Minh City, Vietnam

**Abstract.** In recent years, passive compliant series elastic actuators (SEA) have been utilized as an alternative for traditional actuators in humanoid robots to improve their performance. Despite its advantages over traditional actuators, the SEA mechanism may be sensitive to perturbation, a design of a robust controller which addresses the aforementioned problem of the SEA's mechanics is crucial. Super twisting algorithm (STA) is well known for its robustness against perturbation. Therefore, a backstepping super-twisting-based sliding mode is implemented to control a commercialized humanoid robot (UXA-90)'s lower leg in this paper. First, a dynamic model for a single leg of UXA-90 is taken from our previous study. Secondly, we divided the aforementioned dynamic system into subsystems by utilizing the backstepping technique. Then a control law based on the STA is synthesized to control the subsystems. Finally, a simulation is carried out in MATLAB and the results are compared to our previous study to verify the effectiveness of the proposed controller.

**Keywords:** Super twisting algorithm · Sliding mode · SEA · Humanoid robot

## 1 Introduction

Recently, bio-inspired robots have been used widely in various fields: medical, military, entertainment. Biped robots or humanoid robots, in particular, are a unique class of bio-inspired robots, which have been extensively researched since the introduction of the world's first recognized humanoid robot WABOT. Despite many successes in mimicking human motion, the task of replicating human motion naturally still poses a challenging problem.

From the perspective of designing, actuators are one of the most important aspects to realize the natural movement of the humanoid robot. DC motors are commonly utilized since they feature high precision. However, the stiffness property of DC motors can cause damage to the motors if humanoid robots are subjected to external forces. Therefore, the solution to reduce the damage done to the actuator has been shifted to the development of a hybrid actuator-compliant actuator, Series Elastic Actuator (SEA) [1] is a notable example of this design.

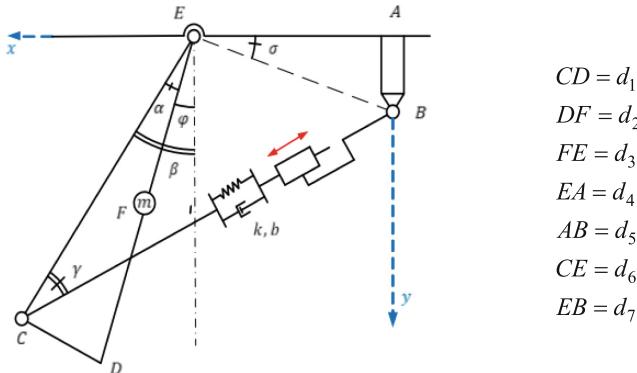
So far in our previous studies [2, 3], two types of controller have been proposed for the SEA-inspired humanoid robot's lower limb: PID and traditional sliding mode control (SMC). The PID can track the desired trajectory well in the case where there are no disturbances. While the SMC can perform well in dealing with uncertainties and disturbances, the chattering phenomenon is not well addressed in the study. This phenomenon is the source of instability [4], which in turn leads to catastrophic damage of the SEA's mechanism.

One of the approaches is to utilize the high-order sliding mode, which can be considered an enhancement of the traditional SMC. If it is appropriately used, HOSM can remove the chattering effect [5, 6]. A special case for the HOSM is the super-twisting algorithm (STA), which is the second-order sliding mode (SOSM) and has been proven to be well suited for the system with a sliding variable that has a relative degree of one [7]. An interesting feature of the STA is that it can force the trajectory of the system to the sliding surface in finite time without the knowledge of the derivative of the sliding variable [8].

In this paper, a novel backstepping super-twisting-based sliding mode control is proposed to reduce the chattering phenomenon and solve the problem of perturbation. To develop the controller for the dynamic recursive structural system of the SEA taken from our previous study, a backstepping technique is implemented. With the utilization of the backstepping technique, the complex system of the SEA is separated into two second-order subsystems, giving us a reduced-order system. With the new reduced-order system, a sliding mode controller utilizing the STA is synthesized to drive SEA's angle into the desired sliding manifold. Finally, simulation results in MATLAB are given to validate the effectiveness of the proposed controller.

## 2 Dynamic Modeling

The model of the SEA-inspired humanoid robot's leg is taken from our previous work [3] and represented in Fig. 1 below.

**Fig. 1.** SEA leg system

The SEA dynamic equation for Fig. 1 is as follows [3]:

$$\begin{cases} \ddot{\varphi} = f(\varphi, \dot{\varphi}) - g(\varphi)\Delta \\ \ddot{\Delta} + \omega^2\Delta = U_{eq} - \frac{F_R}{m} \end{cases} \quad (1)$$

where

$$f(\varphi, \dot{\varphi}) = \frac{-mgd_3 \sin \varphi - B\dot{\varphi} + \tau_D}{md_3^2}, \quad g(\varphi) = \frac{kd_6d_7 \sin(\varphi + \sigma + \frac{\pi}{2} - \alpha)}{md_3^2 L(\varphi)} \quad (2)$$

$$L(\varphi) = \sqrt{D + 2d_6[d_5 \sin(\varphi - \alpha) + d_4 \cos(\varphi - \alpha)]}, \quad D = d_4^2 + d_5^2 + d_6^2$$

$$F_R = \frac{mgd_3 \sin(\varphi)}{d_6d_7 \sin(\varphi + \sigma + \frac{\pi}{2} - \alpha)} L(\varphi), \quad \omega = \sqrt{\frac{k}{m}}$$

The description for each symbol in Eq. (1) and Eq. (2) is shown in Table 1.

**Table 1.** List of symbols

Symbol	Description	Units
$\varphi$	Absolute joint's angle	rad
$\alpha, \sigma$	Fixed angles in lower limb model	rad
$L(\varphi)$	SEA's total length	m
$\tau_D$	System's perturbation	Nm
$B$	Damping coefficient of the joint	Nms/rad
$k$	Stiffness coefficient of the spring	N/m
$F_R$	Reaction force	N
$m$	SEA's point mass	kg

(continued)

**Table 1.** (continued)

Symbol	Description	Units
$g$	Gravitational acceleration	$m/s^2$
$\Delta$	Spring's deformation	$m$
$U_{eq}$	Controlled voltage of DC motor	$V$

### 3 Controller Design

The main objective is to synthesize a robust controller that utilizes the control input defined as  $U_{eq}$  for the DC motor to drive  $\varphi$  to the pre-defined trajectory in the presence of the system's perturbation  $\tau_D$ .

Let us define the state variables as follows:

$$X \triangleq [x_1 x_2]^T = [\varphi \dot{\varphi}]^T \quad (3)$$

$$x_3 \triangleq \Delta$$

$$x_4 \triangleq \dot{\Delta}$$

The following set of equations describe the system of Eq. (1):

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= f(x_1, x_2) - g(x_1)x_3 \\ \dot{x}_3 &= x_4 \\ \dot{x}_4 &= -\omega^2 x_3 + U_{eq} - \frac{F_R}{m} \end{aligned} \quad (4)$$

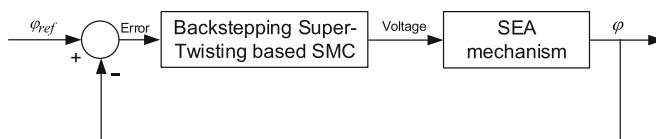
The reference trajectories are denoted as  $\varphi_d \triangleq x_{1d}$ ,  $\dot{\varphi}_d \triangleq x_{2d}$ . The controller design process is described through the following procedure:

Step 1: The variable  $x_3$  is considered as the first virtual control input, a control law is derived to drive  $x_1$  and  $x_2$  to the reference trajectory  $x_{1d}$  and  $x_{2d}$ .

Step 2: The obtained virtual control input  $x_3$  from is considered as the new reference signal for the third equation while  $x_4$  is considered as the second virtual control input.

Step 3: The process is repeated until  $U_{eq}$  is found.

The flow chart of the system is shown in Fig. 2.

**Fig. 2.** Flowchart of the system

### 3.1 Sliding Surface Design

Define  $u_1 \triangleq x_3$ , the first two equations in Eq. (4) can be rewritten as:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = f(x_1, x_2) - g(x_1)u_1 \end{cases} \quad (5)$$

In order to apply the STA, the system must have a relative degree one [9]. Therefore, in order to utilize the STA, the sliding surface of Eq. (5) is defined as:

$$\sigma_1 \triangleq c_1(x_1 - x_{1d}) + \dot{x}_1 - \dot{x}_{1d} \quad (6)$$

where  $c_1$  is chosen constant.

Taking the first derivative of Eq. (6), gives:

$$\dot{\sigma}_1 = c_1(x_2 - x_{2d}) + f(x_1, x_2) - g(x_1)u_1 - \dot{x}_{2d} \quad (7)$$

The first virtual control input  $u_1$  based on STA to stabilize Eq. (7) is given as:

$$u_1 = g(x_1)^{-1} \left( \alpha_1 |\sigma_1|^{\frac{1}{2}} \text{sign}(\sigma_1) + \int_0^t \beta_1 \text{sign}(\sigma_1) + \frac{-mgd_3 \sin \varphi - B\dot{\varphi}}{md_3^2} - \dot{x}_{2d} + c_1(x_2 - x_{2d}) \right) \quad (8)$$

where  $\alpha_1$  and  $\beta_1$  are chosen constants.

Substituting Eq. (8) into Eq. (7), it yields:

$$\dot{\sigma}_1 = -\alpha_1 |\sigma_1|^{\frac{1}{2}} \text{sign}(\sigma_1) - \int_0^t \beta_1 \text{sign}(\sigma_1) + \frac{\tau_D}{md_3^2} \quad (9)$$

Consider the third equation in Eq. (4) by defining  $u_2 \triangleq x_4$ , it can be rewritten as:

$$\dot{x}_3 = u_2 \quad (10)$$

The sliding surface and its first derivative for Eq. (10) is defined as:

$$\begin{aligned} \sigma_2 &\triangleq x_3 - u_1 \\ \dot{\sigma}_2 &= u_2 - \dot{u}_1 \end{aligned} \quad (11)$$

The second virtual control input  $u_2$  is given as:

$$u_2 = -\alpha_2 |\sigma_2|^{\frac{1}{2}} \text{sign}(\sigma_2) - \int_0^t \beta_2 \text{sign}(\sigma_2) + \dot{u}_1 \quad (12)$$

Substituting Eq. (12) into Eq. (11), it yields:

$$\dot{\sigma}_2 = -\alpha_2 |\sigma_2|^{\frac{1}{2}} \text{sign}(\sigma_2) - \int_0^t \beta_2 \text{sign}(\sigma_2) \quad (13)$$

where  $\alpha_2$  and  $\beta_2$  are chosen constants.

To synthesize the final control law  $U_{eq}$ , let us define the sliding surface and its first derivative for the third subsystem of Eq. (4) as follow:

$$\begin{aligned}\sigma_3 &\triangleq x_4 - u_2 \\ \dot{\sigma}_3 &= -\omega^2 x_3 + U_{eq} - \frac{F_R}{m} - \dot{u}_2\end{aligned}\quad (14)$$

The control input to stabilize Eq. (14) is given as:

$$U_{eq} = -\alpha_3 |\sigma_3|^{\frac{1}{2}} sign(\sigma_3) - \int_0^t \beta_3 sign(\sigma_3) + \omega^2 x_3 + \frac{F_R}{m} + \dot{u}_2 \quad (15)$$

Substituting Eq. (15) into Eq. (14), it yields:

$$\dot{\sigma}_{eq} = -\alpha_3 |\sigma_3|^{\frac{1}{2}} sign(\sigma_3) - \int_0^t \beta_3 sign(\sigma_3) \quad (16)$$

It can be easily seen that if  $\sigma_1, \sigma_2, \sigma_3$  tend to zero, then the whole system in Eq. (4) will be stable. In the next section, the stability of the system is analyzed in detailed.

### 3.2 Stability Analysis

Consider a system in the following form:

$$\begin{aligned}\dot{x}_1 &= -\alpha |x_1|^{\frac{1}{2}} sign(x_1) + x_2 + \rho(x, t) \\ \dot{x}_2 &= -\beta sign(x_1)\end{aligned}\quad (17)$$

where  $x_i$  are scalar state variables;  $\alpha$  and  $\beta$  are the chosen gains;  $\rho(x, t)$  is a bounded perturbation.

It is appealing to consider a Lyapunov function in the quadratic form [10] for Eq. (17) as follows:

$$V = \boldsymbol{\varepsilon}^T \mathbf{P} \boldsymbol{\varepsilon} \quad (18)$$

where  $\boldsymbol{\varepsilon}^T = [\varepsilon_1, \varepsilon_2] = [|x_1|^{\frac{1}{2}} sign(x_1), x_2] \in \mathbb{R}^{2 \times 1}$ ;  $\mathbf{P} = \mathbf{P}^T > 0$  is a symmetric positive definite matrix;  $\mathbf{P} \in \mathbb{R}^{2 \times 2}$ .

In order to analyze the robust stability, assume that  $\rho$  satisfies the following condition:

$$\begin{aligned}\omega(\rho, \boldsymbol{\varepsilon}) &= (\rho - \mathbf{L}_1^T \boldsymbol{\varepsilon})(\mathbf{L}_2^T \boldsymbol{\varepsilon} - \rho) \\ &= [\rho \ \boldsymbol{\varepsilon}^T] \begin{bmatrix} -1 & \frac{1}{2}(\mathbf{L}_1^T + \mathbf{L}_2^T) \\ \frac{1}{2}(\mathbf{L}_1 + \mathbf{L}_2) & -\mathbf{L}_2 \mathbf{L}_1^T \end{bmatrix} \begin{bmatrix} \rho \\ \boldsymbol{\varepsilon} \end{bmatrix} \geq 0\end{aligned}\quad (19)$$

where  $\mathbf{L}_1^T, \mathbf{L}_2^T$  are lower bound and upper bound of  $\rho$  respectively.

$\mathbf{L}_1^T, \mathbf{L}_2^T$  are considered, in particular, to be symmetric, that is  $\mathbf{L}_1^T = -\mathbf{L}_2^T$  and have the following form:  $\mathbf{L}_i^T = g_i [1 \ 0]$ .

From Eq. (19), it follows immediately that:

$$\bar{\omega}(\rho, \boldsymbol{\varepsilon}) = \theta\omega(\rho, \boldsymbol{\varepsilon}) \geq 0, \forall \theta \geq 0 \quad (20)$$

By representing  $\bar{\omega}(\rho, \boldsymbol{\varepsilon})$  in the quadratic form, we have:

$$\bar{\omega}(\rho, \boldsymbol{\varepsilon}) = [\rho \ \boldsymbol{\varepsilon}^T] \begin{bmatrix} -\theta & \mathbf{S} \\ \mathbf{S}^T & \mathbf{R} \end{bmatrix} \begin{bmatrix} \rho \\ \boldsymbol{\varepsilon} \end{bmatrix} \geq 0 \quad (21)$$

where  $\mathbf{R} = -\theta \mathbf{L}_2 \mathbf{L}_1^T$ ;  $\mathbf{S} = \frac{1}{2}\theta(\mathbf{L}_1^T + \mathbf{L}_2^T)$ .

Taking the first derivative of  $\boldsymbol{\varepsilon}$ , it yields:

$$\dot{\boldsymbol{\varepsilon}} = \begin{bmatrix} \frac{1}{2|x_1|^{\frac{1}{2}}} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} \quad (22)$$

It is noted that  $\dot{x}_2 = -2\beta\varepsilon_1' \varepsilon_1$ , then Eq. (22) can be rewritten as:

$$\dot{\boldsymbol{\varepsilon}} = \varepsilon_1' (\mathbf{A}\boldsymbol{\varepsilon} + \mathbf{B}\rho) \quad (23)$$

where  $\mathbf{A} = \begin{bmatrix} -\alpha & 1 \\ -2\beta & 0 \end{bmatrix}$ ;  $\mathbf{B} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ;  $\varepsilon_1' = \frac{1}{2|x_1|^{\frac{1}{2}}}$ .

Utilizing Eq. (23), the first-order derivative of Lyapunov function is obtained as:

$$\dot{V} = \varepsilon_1' [\boldsymbol{\varepsilon}^T \ \rho] \begin{bmatrix} \mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} & \mathbf{P} \mathbf{B} \\ \mathbf{B}^T \mathbf{P} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon} \\ \rho \end{bmatrix} \quad (24)$$

Equation (24) and Eq. (21) give us the following inequality:

$$\begin{aligned} \dot{V} &\leq \varepsilon_1' \left\{ [\boldsymbol{\varepsilon}^T \ \rho] \begin{bmatrix} \mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} & \mathbf{P} \mathbf{B} \\ \mathbf{B}^T \mathbf{P} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon} \\ \rho \end{bmatrix} + \bar{\omega}(\rho, \boldsymbol{\varepsilon}) \right\} \\ &= \varepsilon_1' [\boldsymbol{\varepsilon}^T \ \rho] \begin{bmatrix} \mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} + \mathbf{R} \mathbf{P} \mathbf{B} + \mathbf{S}^T & \mathbf{P} \mathbf{B} \\ \mathbf{B}^T \mathbf{P} + \mathbf{S} & -\theta \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon} \\ \rho \end{bmatrix} \end{aligned} \quad (25)$$

Suppose that there exists a symmetric and positive definite matrix  $\mathbf{P} = \mathbf{P}^T > 0$ , positive constant  $\theta > 0$  and  $\gamma > 0$  so that the following condition is satisfied:

$$\begin{bmatrix} \mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} + \gamma \mathbf{P} + \mathbf{R} \mathbf{P} \mathbf{B} + \mathbf{S}^T & \mathbf{P} \mathbf{B} \\ \mathbf{B}^T \mathbf{P} + \mathbf{S} & -\theta \end{bmatrix} \leq 0 \quad (26)$$

Equation (26) can be expressed equivalently in the form of Algebraic Riccati Inequality as follows:

$$\mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} + \gamma \mathbf{P} + \mathbf{R} \mathbf{P} \mathbf{B} + \mathbf{S}^T \theta^{-1} (\mathbf{B}^T \mathbf{P} + \mathbf{S}) \leq 0 \quad (27)$$

From Eq. (26) and Eq. (27), one can derive the following expression:

$$\begin{bmatrix} \mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} + \mathbf{R} \mathbf{P} \mathbf{B} + \mathbf{S}^T & \mathbf{P} \mathbf{B} \\ \mathbf{B}^T \mathbf{P} + \mathbf{S} & -\theta \end{bmatrix} \leq -\gamma \mathbf{P} \quad (28)$$

By utilizing Eq. (25) and Eq. (28), it yields:

$$\dot{V} \leq -\varepsilon'_1 \gamma V = -\frac{1}{2|x_1|^{\frac{1}{2}}} \gamma V \quad (29)$$

Recall the Rayleigh-Ritz inequality and apply it for Eq. (18):

$$\lambda_{min}\{P\} \boldsymbol{\varepsilon}_2^2 \leq \boldsymbol{\varepsilon}^T \mathbf{P} \boldsymbol{\varepsilon} \leq \lambda_{max}\{P\} \boldsymbol{\varepsilon}_2^2 \quad (30)$$

where  $\boldsymbol{\varepsilon}_2^2 = \varepsilon_1^2 + \varepsilon_2^2 = |x_1| + x_2^2$  and note that the following inequality is satisfied:

$$|x_1|^{\frac{1}{2}} \leq \boldsymbol{\varepsilon}_2 \leq \frac{\sqrt{\boldsymbol{\varepsilon}^T \mathbf{P} \boldsymbol{\varepsilon}}}{\lambda_{min}\{P\}} = \frac{V^{\frac{1}{2}}}{\lambda_{min}^{\frac{1}{2}}\{P\}} \quad (31)$$

With Eq. (29), Eq. (31), one can deduce that:

$$\dot{V} \leq -\frac{1}{2} \lambda_{min}^{\frac{1}{2}}\{P\} \gamma V^{\frac{1}{2}} \leq 0 \quad (32)$$

Equation (32) showed that the trajectory of the dynamic system in Eq. (17) converges to zero in finite time. This automatically leads to the convergence to zero of the sliding surface in Eq. (9), Eq. (13), Eq. (16) in finite time.

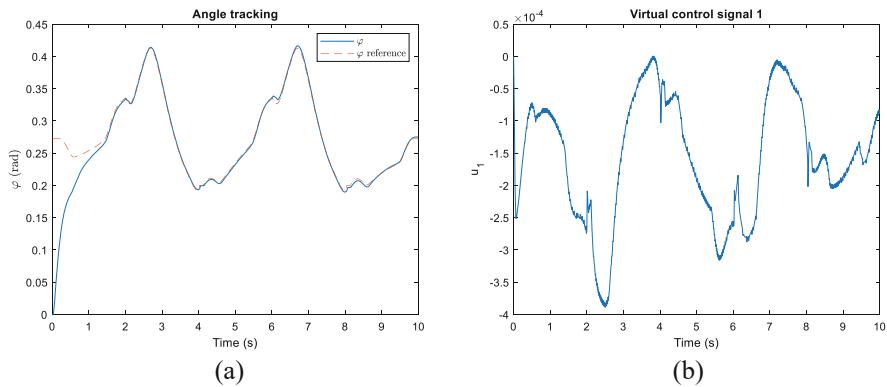
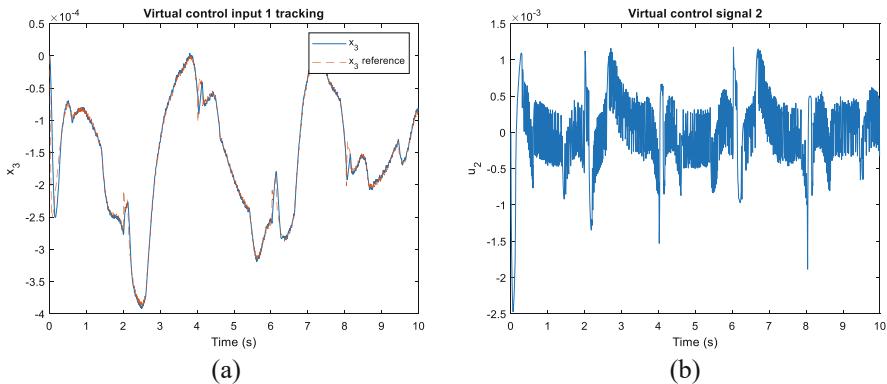
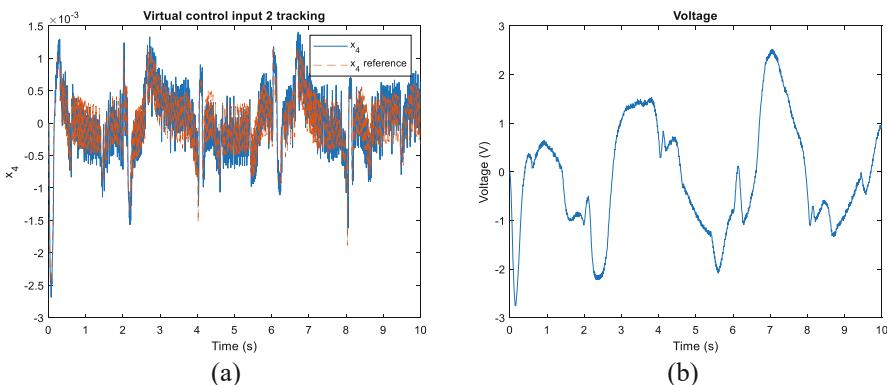
## 4 Simulation and Discussion

In this section, a simulation is carried out in MATLAB and the results are compared to our previous control scheme's results to show the advantages of the STA. The system perturbation  $\tau_D$  is chosen to be  $0.1\sin(2t)$ , the simulation time is 10(s) and the sampling time is  $10^{-3}(s)$ .

The physical parameters are given as:  $d_1 = 0.028$ ,  $d_2 = 0.0525$ ,  $d_3 = 0.0525$ ,  $d_4 = 0.035$ ,  $g = 9.8$ ,  $B = 0.5$ ,  $m = 1.5$ ,  $k = 20000$ .

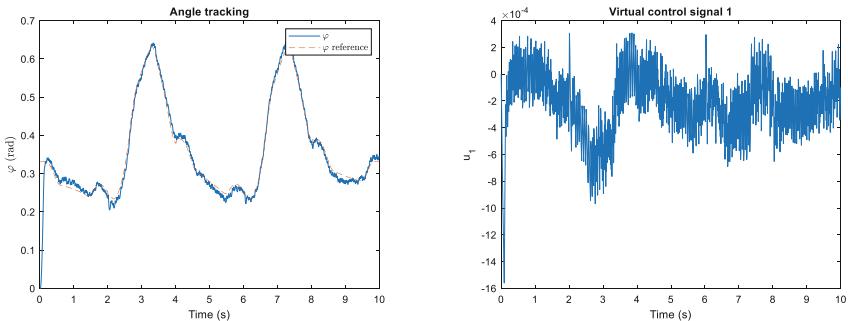
The tracking results of Eq. (5), (10), (12) and their control signals are shown in Fig. 3, 4, 5, respectively. Figure 3a, 4a, 5a show the tracking of the state variables while Fig. 3b, 4b, 5b show the control signals. The solid lines in Fig. 3a, 4a, 5a represent the actual state variables whereas the dash lines represent the reference signals.

In the first subsystem (Fig. 3), the reference signal  $\varphi$  is pre-calculated [11]. The trajectory of  $\varphi$  in Fig. 3a converges to the reference signal approximately after 1.5 seconds. While the virtual control input 1 (Fig. 3b) is still affected by slight chattering, its trajectory can be considered smooth since the maximum amplitude of the trajectory is  $4 \times 10^{-4}$ . In the second subsystem (Fig. 4), Fig. 4a shows that the reference trajectory is tracked closely. However, at the time of 2 seconds and 8 seconds, the controller cannot track the reference signal closely due to the abrupt change in the amplitude in a small time interval. In the third subsystem (Fig. 5), the final control signal – voltage (Fig. 5b) is still affected by the chattering phenomenon due to the accumulated chattering in the previous subsystems.

**Fig. 3.** Simulation result of the first subsystem**Fig. 4.** Simulation result of the second subsystem**Fig. 5.** Simulation result of the third subsystem

In our previous study [3], the controller was a backstepping based traditional sliding mode. Despite the tracking of the reference angle was achieved successfully in the simulation, it is well-known that the use of the discontinuous *sign* function eventually led to the undesirable chattering phenomenon. As a consequence, the final control input eventually suffered from the chattering phenomenon heavily – which is impractical for real-life application.

To further elaborate the disturbance rejection and chattering elimination properties, a simulation utilizing the traditional sliding mode is shown in Fig. 6. The simulation time is 10(s), the sampling time is  $10^{-3}$ (s).



**Fig. 6.** Simulation result for the first subsystem using traditional sliding mode

Figure 6 shows that by implementing the traditional sliding mode, the trajectory of the angle  $\varphi\varphi$  tracked the reference signal well. However, the virtual control signal is affected by the chattering phenomenon significantly due to the use of the *sign* function, thus making the angle  $\varphi\varphi$  non-smooth compared to the result in Fig. 3. This phenomenon is even worse with the perturbation  $\tau_D$  introduced to the system.

In general, the proposed controller tracked the reference trajectories closely. The control signal in each subsystem still suffers from the chattering phenomenon, albeit to a lesser extent due to the implementation of the STA. The utilization of STA has also nullified the external perturbation  $\tau_D$ , thus guaranteeing the robustness of the whole system.

## 5 Conclusion

A novel backstepping super-twisting-based sliding mode controller is proposed to control a SEA-inspired humanoid robot's lower limb. The backstepping method effectively decomposed the system into three subsystems, making the controlling process simpler. It is shown that the STA has effectively reduced the chattering phenomenon while rejecting the unknown disturbance. As a result, one of the SEA's weaknesses is alleviated and protecting the SEA from potential damages.

**Acknowledgments.** This research is supported by DCSELAB and funded by Vietnam National University Ho Chi Minh City (VNU-HCM) under grant number TX2022-20b-01. We acknowledge the support of time and facilities from Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for this study.

## References

1. Pratt, G.A., Williamson, M.M.: Series elastic actuators. In: IEEE International Conference on Intelligent Robots and Systems, vol. 1, pp. 399–406 (1995)
2. Truong, K.D., et al.: A study on series elastic actuator applied for human-interactive robot. In: ICAMechS, December 2020, pp. 7–12 (2020)
3. Luu, A.K.L., et al.: Design of robust controller applied for series elastic actuators in controlling humanoid's joint (2021). <http://arxiv.org/abs/2107.08610>
4. Edwards, C., Sarah, S.K.: Sliding Mode Control, Theory and Applications. (1998)
5. Levant, A.: Higher-order sliding modes, differentiation and output-feedback control. Int. J. Control **76**(9–10), 924–941 (2003)
6. Levant, A.: Principles of 2-sliding mode design. Automatica **43**(4), 576–586 (2007)
7. Shtessel, Y., et al.: A novel adaptive-gain supertwisting sliding mode controller: methodology and application. Automatica **48**(5), 759–769 (2012)
8. Cheng, C.C., Lee, J.J.: Design of adaptive terminal super-twisting controllers for nonlinear systems with mismatched perturbations. Int. J. Control **94**(8), 2021–2031 (2021)
9. Chalanga, A., et al.: Implementation of super-twisting control: super-twisting and higher order sliding-mode observer-based approaches. IEEE Trans. Industr. Electron. **63**(6), 3677–3685 (2016)
10. Moreno, J.A.: A linear framework for the robust stability analysis of a generalized super-twisting algorithm. In: 2009 6th International Conference on Electrical Engineering, Computing Science and Automatic Control, CCE 2009 (2009)
11. Nguyen, X.T., et al.: Controlling center of mass in humanoid robot using sliding mode control. In: ICAMechS, December 2020, pp. 17–22 (2020)



# Text Spotting in Vietnamese Documents

Vu Minh Duc<sup>(✉)</sup> and Tran Ngoc Thang

School of Applied Mathematics and Informatics, Hanoi University of Science  
and Technology, Dai Co Viet, Hanoi 100000, Vietnam  
[20141202@student.hust.com.vn](mailto:20141202@student.hust.com.vn), [thang.tranngoc@hust.edu.vn](mailto:thang.tranngoc@hust.edu.vn)

**Abstract.** Text spotting in images is a common task in the scientific world and the research to apply in practice is an important problem. In this paper, we propose a system for recognizing text documents both printed text and handwriting text in Vietnamese. We use the text detection module and text recognition module for the main section. In addition, we also provide some features in preprocessing module and post-processing module to correct the error of prediction, improve accuracy especially for Vietnamese handwritten. The proposed system was trained in a huge, various type data with English, Vietnamese language and the best model get good results in many evaluation datasets. The experimental results in the recognition module show that our system achieves 2.94% of character error rate (CER), which helps improve CER of another previous approach on the VNOnDB-Word dataset of the Vietnamese online handwritten text recognition competition (HANDS-VNOnDB or VOHTR2018). These results are better than all results compare to other recognition systems in the competition.

**Keywords:** Text spotting system · Scene text recognition · Scene text detection · Vietnamese handwritten recognition · Optical character recognition (OCR)

## 1 Introduction

Optical character recognition (OCR) is a procedure that converts images to editable text files. It includes two main tasks: text detection operation and text recognition operation. There are many advantages from reading text with a high level of accuracy prediction of information in an image like processing of OCR faster, recreating the document layout, etc. [1]. So that, OCR is the best tool for implementing a paperless approach in the business and organization. It can work many languages like: English, Chinese, Japanese,... However there are few methods for solving with Vietnamese. The writing of Vietnamese language has various fonts, complex backgrounds, many writing styles from many writers (with handwritten). In many cases, text recognition cannot be resolved without a post-processing module for correcting the language's error of the text (Fig. 1).



**Fig. 1.** Examples of multi-type text.

In this paper, we propose a pipeline of a text recognition system that includes 4 modules: preprocessing, text detector, text recognizer, post processing. It addresses the problems of the current text recognition pipeline with real-world data in Vietnamese language. Moreover, we re-implement many top models such as Differentiable Binarization (DB) [2], Character Region Awareness for Text Detection (CRAFT) [3] for detector, deep-text-recognition-benchmark (CLOVA\_AI) [4] for recognizer. One additional benefit of our approach is that we can apply with both printed and handwritten text documents. We test the text detection module on SROIE 2019 dataset and text recognition module on two datasets: Vietnamese Online Handwriting Database (VONDB) and Handwriting OCR for Vietnamese Address (Cinnamon) for Vietnamese language.

In summary, the main contributions of this paper are summarized as follows:

1. We propose that the pipeline system for recognizing printed and handwritten text.
2. The result of the best model in the recognizer module outperforms in text recognition task of VOHTR2018 and Cinnamon competition in word level.

## 2 Related Work

In recent years, the study and research of text detection and text recognition have developed faster with the advancement of deep learning algorithms. In this section, we give a brief overview of text detection and recognition method and current text spotting models.

### 2.1 Text Detection

The function of text detection is to detect or segment the area of the image which covered the text. Traditional text detection methods use localization and verification procedures. Text localization methods get features from low-level information such as color, gradient based on texture or region. Text verification aims at verifying the text candidate regions as text or non-text. Most verification methods are prior knowledge, support vector machine (SVM), and conditional random fields (CRFs), and recently approach by convolution neural

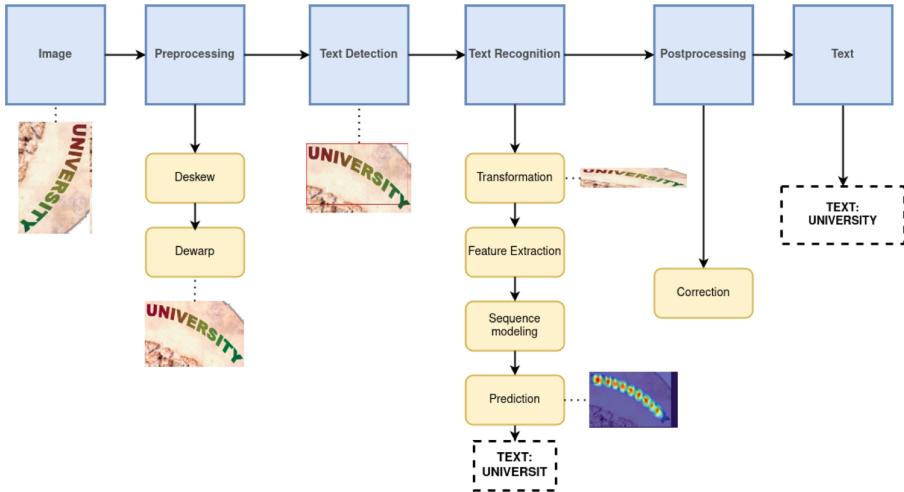
network (CNN). These methods have been replaced by deep learning models with better accuracy and a simplified pipeline. In the early stage, most techniques in scene text detection were inspired and adopted by general object detection. The newer methods can be categorized as regression-based methods, instance segmentation-based methods. Modern methods can directly predict oriented bounding boxes or quadrilaterals by single-shot or two-stage models.

## 2.2 Text Recognition

Text recognition is a task that recognizes the text context (character) in the detected area. It is the main task in the text spotting system, which predicts text results. Tradition text recognition methods use hand-crafted features, such as MSER, stroke width transform, histogram of oriented gradients (HOG). It often uses SVM or KNN model at classification step. Recently most researches have used deep learning based on encoder-decoder frameworks. We can divide text recognition methods into two main categories, Connectionist Temporal Classification (CTC) and attention-based methods (Attn). Text recognizers typically use CNN to extract features and RNN (Recurrent Neural Network) to generate sequences. Most of the methods from two above categories use a spatial transformer network (STN) for rectifying input text (straight line by arbitrary-shaped texts). The input image is transformed then fed to a full convolution network to encode the image to features. CTC methods use RNN to decode features into an output sequence by using RNN-stacked architectures on top of a CNN. A symbol is predicted by each feature column. It removes duplicate symbols for the final prediction. With attention methods, RNN is used to calculate hidden state, extract image features (context) and predict the symbol per step by prediction at the previous step.

## 2.3 Text Spotting System

A text spotting system is a computer vision process, that is given a text image, and output is string sequences of it. It often includes text detection and text recognition module (two stages model). The modern methods are approached to jointly text detection task and text recognition task based on shared information, called end-to-end model. However, they need more data and much of them are required about fully annotated images for training. By the time, the recognizer is converged but the detector is usually overfit. Therefore, most text spotting systems use the two independent stages for changing and updating models easily. The text spotting system, that we propose, uses text detection and text recognition module separately and adds two modules preprocessing and post-processing for improving accuracy.



**Fig. 2.** Overall architecture of text spotting system.

### 3 Methodology

With the proposal system, we divided the system into many components as presented in Fig. 2. The architecture has four main parts:

1. **Preprocessing:** The input image is given to resize, denoise, deskew.
2. **Text detection:** Using the output image from preprocessing module, we put it here. The module will get the bounding boxes and confidence score of it in word or line level (almost word level). Then, it crops the image to many sub-images based on a co-ordinary bounding box and sends them to the text recognition module.
3. **Text recognition:** In this module, input is cropped image from the text detection module. We use a text recognizer model with four stages include transformation stage, feature extraction stage, sequence modeling stage, and prediction stage. After finishing that progress, the output is multiple texts with a confidence score of text recognition.
4. **Post processing:** Select top-3 confidence results from output text, combine it with the dictionary of language and multiple data collections by domain for creating multi-candidate (called correction module). We will measure and get the best candidate based on our scoring mechanism. The output text of this module is the final result.

#### 3.1 Preprocessing Module

Firstly, the system applies computer vision algorithms to improve image quality. While scanning a document or taking a photo, the image might be slightly askew certain angle with horizontal so that correcting the skew is needed. Secondary,

we denoise the image by removing small dots or patches that have high intensity than the rest of the image. This is important for minimizing errors and more accurate predictions. Finally, the module will resize the input image to fit the dimension size for the text detection module.

### 3.2 Text Detection Module

In this module, we choose two state-of-the-art models which are DB [2] and CRAFT [3] for evaluation. Techniques in Character level are new emerging these years, yet they are powerful and effective. CRAFT is a method there, that predicts the location of each character and incorporates them into a whole text line using the affinity between characters. The network produced Gaussian Heatmaps for each character prediction. With the DB method, the model is based on segmentation. The key step in the segmentation-based method is the post-processing part. In this step, the segmentation result is converted into a text box or text area. The text detection using DB will simplify post-segment processing steps and can set an adaptive threshold to improve network performance. The DB method shows better results in the accuracy and speed of many testing datasets. So that, we use two models of text detection with separate purposes. Moreover, the text detection the proposal system use is word level. In addition, because of using text spotting system for both printed and handwritten text, we add more two tricks for a better result:

1. We need to expand bounding boxes after prediction since the image data is very enormous and several languages such as Vietnamese have many marks above or below the character so that text detection is very difficult without expanding the box.
2. Using non-maximum suppression (NMS) to combine multiple stacked bounding boxes to one bounding box make optimize time and remove redundant boxes.

### 3.3 Text Recognition Module

According to [4], text recognition module is perform in four stages:

1. **Transformation stage (Trans.)**: normalizes and rectifies the curve text to a horizontal text by using the STN model.
2. **Feature extraction stage (Feat.)**: use CNN, like VGG and ResNet as a backbone and extract visual feature representation from the image.
3. **Sequence modeling stage (Seq.)**: the contextual information in the sequence characters is captured by converting visual features.
4. **Prediction stage (Pred.)**: decode output character sequence from features and get text result. This is done by a CTC or attention decoder.

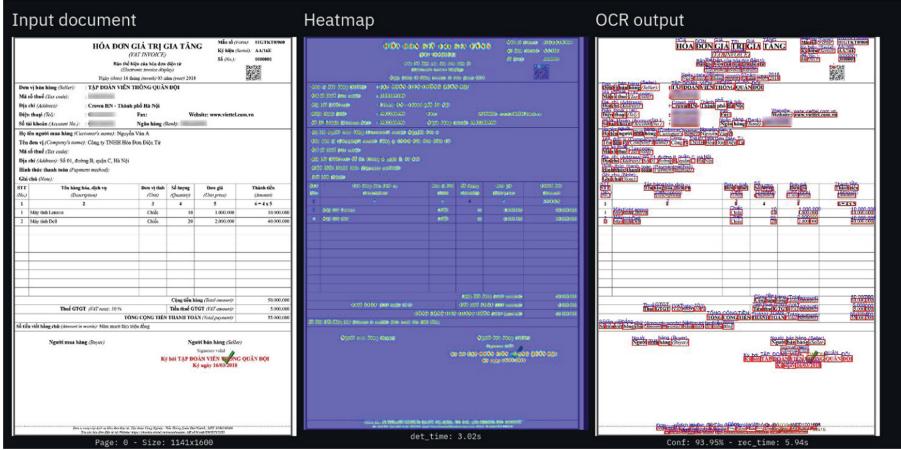
For our experiment, we adopt two widely-used models for the text recognition benchmark: SCATTER [5], and CLOVA\_AI [4]. We also increase the default max length sequence parameter and hidden state to predict the long text (received from the text detection module).

### 3.4 Post Processing Module

Our approach is rather simple, that is implemented like BeamSearch. We use the top-3 highest accuracy from the text recognition module, combined with a common word dictionary of this language and a list of words collected by domain of document. We assign weights in order of priority like the output from text recognition results, common word dictionary, list of domain's word. We sort from high to low score all match candidates go through process there called “scoring mechanism”. Finally, the candidate with a high score will be returned as a corrected result.

## 4 Experiments and Results

In this section we empirically demonstrate the effectiveness and benefit of our proposed system (Fig. 3).



**Fig. 3.** Some pictures about system

### 4.1 Dataset

we use 7 text datasets for training data in our experiments, 2 Vietnamese datasets (VNOnDB and Cinnamon Handwriting dataset - Cin) and one English dataset benchmark chosen to evaluation our method as Table 1

**Table 1.** Dataset for training and testing

Dataset (Year)	Image Num (Total - train/val/test)	Text Num (Total)	Orientation	Language	Feature
MJSynth	9 million	90K words	Multi-oriented	English	Wild scene text
SynthText	858,750	7,266,866 words	Multi-oriented	English	Wild scene text
IAM	1539 documents	13353 textlines 115320 words	Horizontal	English	Handwriting text
CVL	311 documents	101069 words	Horizontal	German/English	Handwriting text (Only use English)
ORAND CAR 2014	11700 - 5000/0/5000	-	Horizontal	English	Handwriting text (Digit string of bank check)
NIST	810000	-	Horizontal	English	Handwriting text (Digit, upper, lower text field)
SROIE [6]	626/-/361 page	-	Horizontal	English	Scanned text
VNOnDB [7]	total: 1146 para- graphs line-level: 4433/1229/163 word-level: 66991/18640/25115	-	Horizontal	Vietnamese	Handwriting
Cin [8]	line-level: 1822/- /548 *word-level: 23567/-/5511	-	Horizontal	Vietnamese	Handwriting (Address)

\*The word-level of Cin dataset: is labeled annotation by our team and publicly available at [https://github.com/hust-ailab/cinnamon\\_word.dataset](https://github.com/hust-ailab/cinnamon_word.dataset).

## 4.2 Implementation Details

We implement our model in PyTorch framework. All experiments are conducted on a server with CPU: AMD Ryzen 9 3900x, GPU: 1 RTX 2080Ti and 1 RTX 2080 GPU, RAM: 64 GB memory.

**Model and Training Strategy:** With the text detection module, we keep all configurations in the base model. The text recognition module is trained from scratch using architecture: TPS-ResNet-BiLSTM-Atten from CLOVA\_AI. We only change the max length character to 300, the number of fiducial points of TPS-STN to 200 and the hidden size of Long short term memory (LSTM) to 512 so that model can recognize the long text and a large amount of Vietnamese alphabet. Other parameters keep default like the original paper.

**Dataset:** Despite having many above datasets, we still need to gen more synth text datasets and augment some images, such as blurring, perspective transformations, gaussian noise and random add them to training images. Because

amount of Vietnamese data is a little in training dataset. Our team also collect and annotate both printed and handwritten data to overcome the data imbalance.

**Plan Training:** The system is trained based on weakly supervised learning. Firstly, we train the system with synthetic datasets. After that, using it like pretrain model for other datasets with annotated labels at word and line level. At last, we train with the real images to output the finally model.

## Evaluation Metric

- Text detection: In this module, the result is evaluated by three metrics: Precision, Recall, Hmean (like task 1 in Sroie 2019 competition [6]):

$$Precision = \frac{\text{Total number of correctly predicted text regions}}{\text{Total number of predicted text regions}} \quad (1)$$

$$Recall = \frac{\text{Total number of correctly predicted text regions}}{\text{Total number of text regions in dataset}} \quad (2)$$

$$Hmean = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$

- Text recognition: We use character error rate (CER) on two benchmark datasets. They are calculated based on the Normalized Edit Distance (NED). Given a predicted string A and the ground truth string B, CER is specified by:

$$CER_{A,B} = \frac{\text{Levenshtein}(A,B)}{\max(\text{len}(A), \text{len}(B))} \quad (4)$$

Evaluation based on character error rate is average all CER on whole dataset with Edit distance computed on word level.

## 4.3 Comparison with Other Approaches

We report full results of our experiments in Table 2, Table 3 and Table 4. With the text detection module, we achieve a result better than top 23 in ranking competition table (Fig. 4). The best model text recognition is outperformed in text recognition task on VNonDB-word dataset with other approaches. With the Cin dataset, by using post processing module we achieve good prediction on Handwriting address (Fig. 5). Furthermore, when we add this module to OCR flow, the inference still keep balance between speed and result (6.24 FPS, acc: 2.92%) with base flow (6.67 FPS, acc: 3.10%).

**Table 2.** SROIE dataset (task 1)

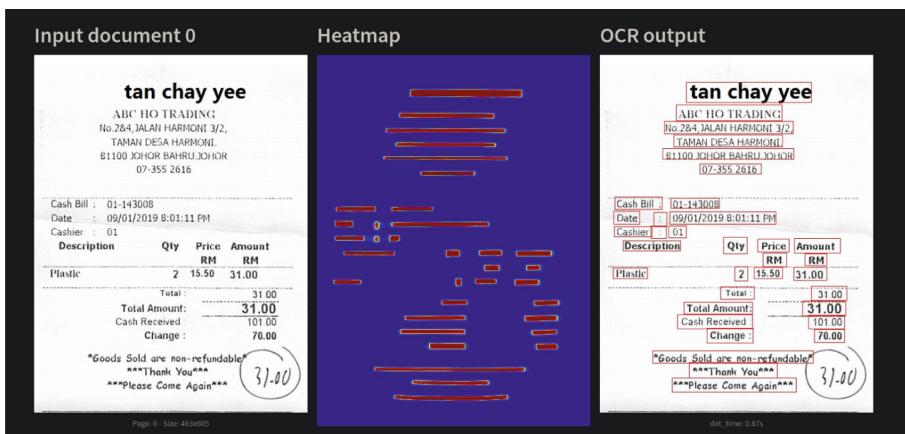
Method	Precision	Recall	Hmean	FPS
Top 1: BreSee OCR	<b>0.9915</b>	<b>0.9940</b>	<b>0.9928</b>	–
Top 23: IFLYTEK-textDet_v3	0.9377	0.9589	0.9481	–
<b>CRAFT</b>	0.9118	0.8969	0.9043	6.69
<b>DB</b>	0.9325	0.9661	0.9490	10.35

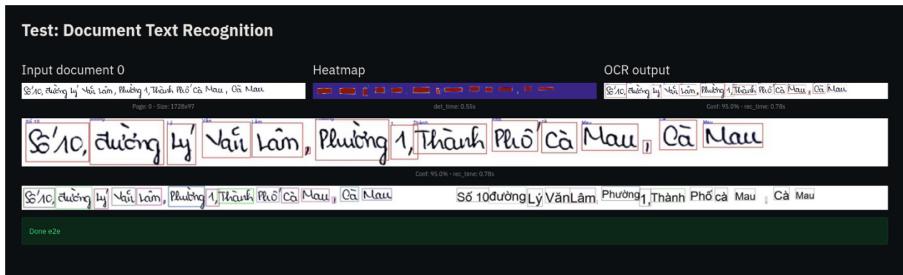
**Table 3.** VNOnDB-word dataset

Method	Corpus	CER (%)
Anh. model with DenseNet encoder [9]	None	4.10
SCATTER (reimplement)	None	7.45
<b>CLOVA_AI</b>	None	<b>2.94</b>

**Table 4.** Cinnamon handwriting address dataset

Method	CER (%)		FPS
	Word	Line	
SCATTER (reimplement)	8.16	7.22	5.00
CLOVA_AI	6.52	3.10	6.67
<b>CLOVA_AI + post processing</b>	<b>4.36</b>	<b>2.92</b>	6.24

**Fig. 4.** The text detection module.



**Fig. 5.** The system recognizes handwritten line text

## 5 Conclusion

In this work, we have proposed an text spotting system with many note for recognizing text in image quickly and efficiently. Moreover, we also compare multiple model in text recognition with two Vietnamese handwriting datasets and get the outperform in VNonDB-word with the best model. We hope our system inspires the developers for applying OCR in real-world with multi-languages, and will serve as the tool for researcher who want to develop their ideas faster.

## References

1. Artificial Intelligence for Automated Pricing Based on Product Descriptions. SAST, Springer, Singapore (2022). <https://doi.org/10.1007/978-981-16-4702-4>
2. Liao, M., Wan, Y., Yao, C., Chen, K., Bai, X.: Real-time scene text detection with differentiable binarization. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 11474–11481 (2020). doi: 10.1609/aaai.v34i07.6812
3. Baek, Y., Lee, B., Han, D., Yun, S., Lee, H.: Character region awareness for text detection, pp. 9357–9366 (2019). doi: 10.1109/CVPR.2019.00959
4. Baek, J., et al.: What is wrong with scene text recognition model comparisons? Dataset and model analysis, pp. 4714–4722 (2019). doi: 10.1109/ICCV.2019.00481
5. Litman, R., Anschel, O., Tsiper, S., Litman, R., Mazor, S., Manmatha, R.: SCATTER: selective context attentional scene text recognizer, pp. 11959–11969 (2020). doi: 10.1109/CVPR42600.2020.01198
6. Sroie 2019 (2019). <https://rre.cvc.uab.es/?ch=13&com=introduction>
7. VNonDB (2018). <https://sites.google.com/view/icfhr2018-vohtr-vnondb>
8. Cinnamon AI marathon (2018). <https://ktmt.uit.edu.vn/2017-10-26-10-00-02/student-activities/990-cinnamon-ai-marathon-2018>
9. Le, A.D., Nguyen, H.T., Nakagawa, M.: End to end recognition system for recognizing offline unconstrained Vietnamese handwriting. [arXiv: abs/1905.05381](https://arxiv.org/abs/1905.05381) (2019)



# An Advanced IoU Loss Function for Accurate Bounding Box Regression

Ho-Si-Hung Nguyen<sup>1</sup>, Thi-Hoang-Giang Tran<sup>1()</sup>, Dinh-Khoa Tran<sup>1</sup>,  
and Duc-Duong Pham<sup>2</sup>

<sup>1</sup> University of Science and Technology, The University of Danang, Danang, Vietnam

{nhshung, tthgiang, tdkhoa}@dut.udn.vn

<sup>2</sup> FTECH Co., Ltd., Hanoi, Vietnam

duongpd@ftech.ai

**Abstract.** Bounding box regression (BBR) is one of the important steps for object detection. To improve accuracy of recognition ability between true object and prediction object, many researches have developed loss functions for BBR. In existing researches, some main drawbacks can be shown: (*i*) both IOU-based loss functions and  $l_n - norm$  are inefficient enough to detect the object; (*ii*) the loss functions ignore the imbalance issues in BBR when the large number of anchor boxes have overlaps with the target boxes; (*iii*) the loss functions own redundant parameters which lead to extend training process. To address these problems, this paper is proposed a new approach by using an Advanced IoU loss function. Three geometric factors including overlap area, distances and side length are considered in the proposed function. The proposal focuses on the overlap area to improve accuracy for object detection. By this way, the proposal can relocate anchor box for covering the ground truth in the training process and optimize anchor boxes for object detection. The proposal is tested on VOC Pascal and MS COCO dataset. The experimental results are compared to existing IoU models and show that the proposal can improve accuracy level for Bounding box regression.

**Keywords:** Advanced IoU · Bounding box regression · Object detection · Loss function · Geometric factors

## 1 Introduction

In recent years, using deep learning to object detection and instance segmentation have got many extraordinary achievements. Some applications can be mentioned as visual tracking [7], face recognition [2], face detection [8], etc. Based on the rapid development of deep learning, many state-of-the-art object detection models had been successfully studied. Depending on the number of detector module to create candidate bounding boxes, the basic solutions include one-stage [14], two-stage [4] detector. Two-stage detector is restricted by a small

amount of bounding box proposal using a category-independent method such as R-CNN [5], Fast-RCNN [4] and etc. On the other hand, one-stage detector takes advantage of the densely pre-defined candidate boxes (anchors) to achieve high inference speed. From pieces of information in AP-loss article [1] point that one-stage detector generally computes much faster than two-stage detectors. So that, they could be easily scrutinized about notable holes in the accuracy.

One of the problems that one-stage detector can solve is the imbalance between the foreground and background regions. To decrease this imbalance, Zhang et al. [15] built a novel single-shot based detector (RefineDet). This model bases on a feed-forward convolutional neural network that produces a fixed number of bounding boxes and the scores indicating the presence difference in classes of those bounding boxes. As a result, a non-maximum suppression produced can reduce imbalance between the foreground and background regions. Another modern object detection is YOLOv5 [11]. The structure of the model consists of two parts: the backbone part containing the bottleneckCSP, SPP and other modules. Based on this structure, the characteristic information is well maintained. In addition, this structure can prevents background noise inside the foreground in the training process.

Besides that, the operation of Intersection over Union (IoU) mainly contributes to stabilizing the foreground-background issue. To improve the ability of bounding box regression, researchers have been given the concept of Intersection over Union (IoU) [13]. IoU is the ratio of the intersection and union of two bounding boxes. A lot of researches have proposed IoU models to improve the accuracy for object detection such as Complete IoU (CIoU) [16], Distance IoU (DIoU) [16], Generalized IoU (GIoU) [13], IoU+ [12] and etc. In our study, we also propose a new IoU, named Advanced IoU(AIoU), to concentrate on the space in the true bounding box and enhance geometric factors of bounding box regression into the inference of deep models for object detection. The AIoU is a promising solution to improve accuracy level for bounding box regression.

The rest of this paper is organized as follows. The Sect. 2 is devoted to describe the related works and proposes an AIoU loss function by taking into account complete geometric factors. The next section gives experimental results tested on MS COCO and PASCAL VOC dataset. After that, the evaluation is proposed based on the analysis of the experimental section. Finally, the conclusion deduced from this work is presented in the last section.

## 2 Related Work and AIoU Loss Function

### 2.1 The Overlap Area

In recent years, the overlap area in BBR is usually thought of as IoU loss function [12].

$$L_{iou} = 1 - IoU \quad (1)$$

This function faces much trouble in training regression such as optimization, instability, box regression speed, and so on. Lu, Z. et al. [10] had figured out a

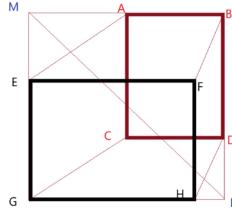
new highlight formula to fix these issues as follows:

$$Co = \frac{B^{pd} \cap B^{gt}}{B^{gt}} \quad (2)$$

$$L_{Co} = \gamma * IoU + (1 - \gamma) * Co \quad (3)$$

where  $\gamma$  is a positive trade-off parameter, which determines the weight of intersection over union and intersection over ground true. From the study of Lu, Z. et al. [10],  $\gamma$  should be set in  $[0.75, 0.8, 0.85]$  to get a good result of validation period in the training duration. In our study,  $\gamma$  has been defaulted as 0.8.

## 2.2 The Distances



**Fig. 1.** Schematic diagram of corner distance

The Lo-IoU [10] and ClIoU [17] were proposed to optimize the route between the prediction box and the true box. The problem of these proposals is how decrease the distances between two vertices of two boxes and change the length of edges of predicted boxes in order to close to the length of true boxes gradually. To solve this problem, a distance loss function is proposed as below:

$$L_{distances} = \frac{EA^2 + BF^2 + DH^2 + CG^2}{4MN^2} \quad (4)$$

where  $AE$ ,  $BF$ ,  $CG$ ,  $DH$  are the distance between each corner of two boxes and  $MN$  is the diagonal length of the smallest enclosing box covering two boxes as Fig. 1. It should be noted that black rectangle is defined the real box and the red rectangle is the prediction box

## 2.3 The Side Length

The ClIoU given by Zheng, Z. et al. [17] is a great technique to regress scale of  $w^{gt}$  and  $w$  toward 1 as well as  $h^{gt}$  and  $h$  toward 1. The formula was defined as:

$$V = \frac{4}{\pi^2} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (5)$$

The trade-off parameter helped normalize  $\mathbf{V}$  to  $[0, 1]$ , which was defined as:

$$\alpha = \begin{cases} 0, & \text{if } IoU < 0.5; \\ \frac{V}{(1-IoU)+V}, & \text{if } IoU \geq 0.5; \end{cases} \quad (6)$$

The aspect ratio is a partial improvement of CIoU [17] over DioU [16]. When  $IoU \geq 0.5$ ,  $\alpha V$  will contribute to improve the accuracy level of BBR.

## 2.4 The Proposed AIoU Loss Function

To improve the accuracy level of BBR, we proposed an AIoU loss function. Three geometric factors are considered in the proposed function including the overlap area, the distances and the side length. The formula was defined as:

$$L_{AIoU} = 1 - L_{co} + L_{distance} + V\alpha \quad (7)$$

$$= 1 - (\gamma IOU + (1 - \gamma)Co) + \frac{EA^2 + BF^2 + DH^2 + CG^2}{4MN^2} + V\alpha \quad (8)$$

According to theory provided by Kosub, S [6],  $1 - L_{co}$  is defined as a part of AIoU. This part plays the key role in the classification task. In that way, the calculation process is going to increase the percentage of union between the ground true area and predicted area. The proposed  $L_{distance}$  presented in Subsect. 2.2 is used to obtain more well performance.  $L_{distance}$  not only adjusts the rate between each edge length of the true box and predicted box, but also the calculation obtains a considerably improved localization accuracy without a complex operation. Finally, the side length  $V\alpha$  taken from Subsect. 2.3 helps to improve the accuracy level for BBR.

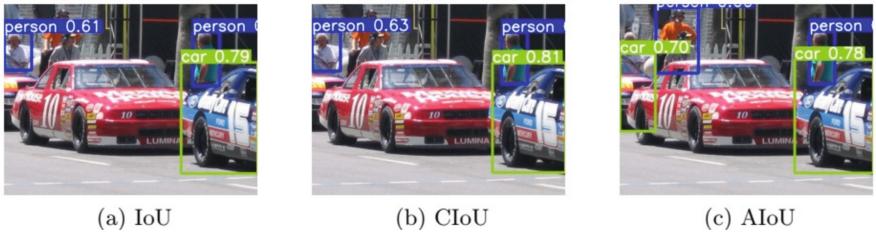
The stability level of  $L_{AIoU}$  can be checked by technique requirements. Each of parts must follow properties: (1)  $0 \leq 1 - L_{co} \leq 1$ , (2)  $0 \leq L_{distance} < 1$ , (3)  $0 \leq V\alpha \leq 1$ ,  $V\alpha$  is normalized to  $[0, 1]$  according to the proving in CIoU [17]. Hence, it confidently ensures that the value of  $L_{AIoU}$  always limits in range of  $[0, 3]$ . So that, this evidence points out the good stability of the formula.  $L_{AIoU}$  significantly reduces risks of overwhelmed computation.

## 3 Experimental Results

In this section, the experimental datasets are two popular benchmarks, they are MS COCO [9] and PASCAL VOC dataset [3]. PASCAL VOC dataset has 20 classes. We use VOC2012 containing 11,530 for train/val. MS COCO is large image recognition/classification, object detection, segmentation, and captioning dataset. We use the COCO train-2017 split (115k images) for training and report the ablation studies on the 2017 split (5k images). The dataset has 80 different object categories with 5 captions per image.

### 3.1 Base Model

**YOLOv5:** YOLOv5 is selected for the base model due to its highlighting quality. Model is trained with default parameters for three stages Train > Val > Test in turn. We choose version S of YOLOv5 due to the simple structure of the model, with little BottleneckCSP lead to decrease time-consuming in the training process. The results are collected after 12 iterations, 0.6 confidence score for VOC PASCAL dataset. The comparison results of three demos are demonstrated in Fig. 2.



**Fig. 2.** Comparison examples between AIoU and other models

Table 1 presents the obtained results when the proposed AIoU is compared to the existing models for object detection. The primary baseline of these determinants had been applied on the two validation datasets of MS COCO and PASCAL VOC. The experimental results implemented prove that the proposed AIoU loss is better accuracy than the others as shown in Table 1. The performance is significantly improved for object categories with a large ratio of background to object pixel as Fig. 2. It should be noted that although accuracy is improved thanks to the proposed AIoU, its training time is not higher than other models as provided by Table 2.

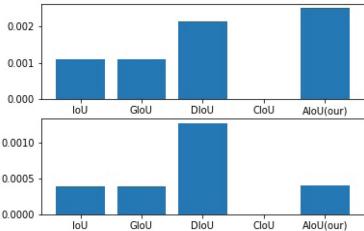
**Table 1.** Accuracy ability of different IoU models

Method	$mAP_{0.5}(VOC)$	$mAP_{0.5:0.95}(VOC)$	$mAP_{0.5}(COCO)$	$mAP_{0.5:0.95}(COCO)$
IoU	0.78031	0.50398	0.53934	0.34755
GIoU	0.77705	0.50671	0.53934	0.34755
DIoU	0.77744	0.50297	0.54037	0.34844
CIoU	0.78301	0.50892	0.53823	0.34716
AIoU(our)	<b>0.78388</b>	<b>0.50916</b>	<b>0.54074</b>	<b>0.34757</b>

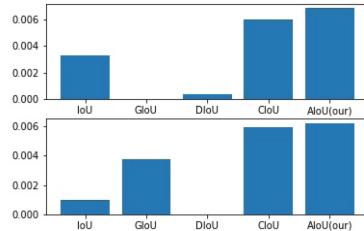
The Fig. 3 has shown that the mAP score of AIoU models got extraordinary ability at the final iteration. Because the AIoU is more outstanding than other IoU models at mAP score after a permanent training time.

**Table 2.** Training time of different IoU models during 12 iterations

<i>Method</i>	<i>time<sub>minute</sub>(VOC)</i>	<i>time<sub>minute</sub>(COCO)</i>
IoU	74.5	535.8
GIoU	74.3	530.2
DIoU	74	529
CIoU	75.4	545
AIoU(our)	73.4	540.4



(a) MS COCO



(b) PASCAL VOC

**Fig. 3.** The change in mean average precision (mAP) of each method in tests with the MS COCO and the PASCAL VOC datasets.

## 4 Conclusions

In this paper, we propose an AIoU loss function to enhance geometric factors. The proposed loss function deeply focuses on overlap area to get the most outstanding foreground. The AIoU loss function simultaneously considers three geometric factors including the overlap area, the distances and the side length. Based on the proposal, the ratio of length and width side are changed at the same time in training process of deep learning model. It should be noted that the purpose of training process is to find and classify a variable number of objects on an image. By this way, the prediction box reaches to the true box fastly. The experimental results confirm that the proposed AIoU loss significantly contribute toward object detection in term of accuracy. One drawback of the proposed model is running time. In the future, we hope to reduce the running time by combining the distances and the side length to create a more simple IoU model.

**Acknowledgment.** This work was supported by The University of Danang, University of Science and Technology, code number of Project: T2020-02-38.

## References

- Chen, K., Lin, W., See, J., Wang, J., Zou, J., et al.: AP-loss for accurate one-stage object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**(11), 3782–3798 (2020)

2. Deng, J., Guo, J., An, X., Zhu, Z., Zafeiriou, S.: Masked face recognition challenge: the insight face track report. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1437–1444 (2021)
3. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results (2012). <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html>
4. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)
5. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)
6. Kosub, S.: A note on the triangle inequality for the Jaccard distance. Pattern Recogn. Lett. **120**, 36–38 (2019)
7. Li, W., Xiong, Y., Yang, S., Deng, S., Xia, W.: SMOT: Single-shot multi object tracking. arXiv preprint [arXiv:2010.16031](https://arxiv.org/abs/2010.16031) (2020)
8. Li, X., Lai, S., Qian, X.: DBCFace: Towards PURE convolutional neural network face detection. IEEE Trans. Circuits Syst. Video Technol. (2021)
9. Lin, T.-Y., et al.: Microsoft COCO: common objects in context. In: Fleet, David, Pajdla, Tomas, Schiele, Bernt, Tuytelaars, Tinne (eds.) ECCV 2014. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48)
10. Lu, Z., Liao, J., Lv, J., Chen, F.: Relocation with coverage and intersection over union loss for target matching. In: VISIGRAPP (4: VISAPP), pp. 253–260 (2021)
11. Wang, L., et al.: Optimized detection method for Siberian crane (*Grus leucogeranus*) based on YOLOv5. Technical Report, EasyChair (2021)
12. Yu, J., Jiang, Y., Wang, Z., Cao, Z., Huang, T.: UnitBox: an advanced object detection network. In: Proceedings of the 24th ACM International Conference on Multimedia, pp. 516–520 (2016)
13. Zhai, H., Cheng, J., Wang, M.: Rethink the IoU-based loss functions for bounding box regression. In: 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), vol. 9, pp. 1522–1528. IEEE (2020)
14. Zhang, S., Chi, C., Yao, Y., Lei, Z., Li, S.Z.: Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9759–9768 (2020)
15. Zhang, S., Wen, L., Bian, X., Lei, Z., Li, S.Z.: Single-shot refinement neural network for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4203–4212 (2018)
16. Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D.: Distance-IoU loss: faster and better learning for bounding box regression. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 12993–13000 (2020)
17. Zheng, Z., et al.: Enhancing geometric factors in model learning and inference for object detection and instance segmentation. arXiv preprint [arXiv:2005.03572](https://arxiv.org/abs/2005.03572) (2020)



# Occlusion Robust Face Recognition Based on Mask Learning with Attention Mechanism

Quan Nguyen Minh<sup>1,2</sup>, Bang Le Van<sup>2(✉)</sup>, Can Nguyen Ngoc<sup>2</sup>,  
and Viet Dung Nguyen<sup>1(✉)</sup>

<sup>1</sup> Hanoi University of Science and Technology, Hanoi, Vietnam  
[dung.nguyenviet1@hust.edu.vn](mailto:dung.nguyenviet1@hust.edu.vn)

<sup>2</sup> Viettel High Technology Industries Corporation, Hanoi, Vietnam  
[{quannm23,banglv1,cannn1}@viettel.com.vn](mailto:{quannm23,banglv1,cannn1}@viettel.com.vn)

**Abstract.** Wearing masks in public places is an efficient strategy to prevent infection and slow the spread of COVID-19. However, masked face recognition is challenging due to a lack of information about facial features in the masked area. We present a unique Switching Replacement Attention Network (SRAN) for robust face identification based on attention mechanism, which is inspired by the human visual system, which exclusively focuses on non-occluded facial areas. Firstly, a replacement module is established by training a segmentation network to segment the location of the occlusion item. To exclude the corrupted feature components from recognition, we multiply the occluded object's segmentation mask with the original image features. To determine when the replacement module is applied, we use a lightweight switch module that is both fast and accurate. The proposed technique outperforms state-of-the-art systems on a variety of occluded and non-occluded facial recognition datasets, according to test results.

**Keywords:** Face recognition · Occlusion · Attention

## 1 Introduction and Related Works

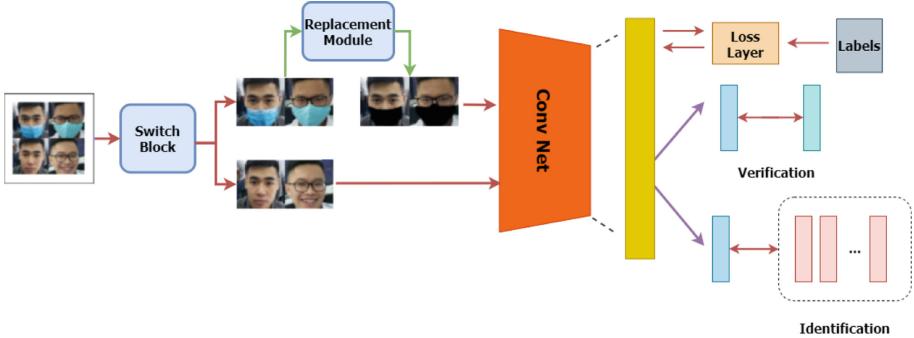
Face recognition has been the subject of extensive research for decades. Face recognition can be divided into two categories: face identification, which involves assigning a particular face to a certain identity, and face verification, which involves determining whether two face photos belong to the same person. Face recognition (FR) in realistic conditions has achieved great success thanks to the recent evolution of deep convolutional neural networks [1]. Approaches based on deep CNN intuitively learn the discriminative features to represent face images. For example, [2–4] use advanced network architecture to enhance the feature extraction capability. Another approach is to design new loss functions that allows for better discriminative power. In order to maximize face class separability, SphereFace [5], Cosface [6] and Arcface [7] combine angular margins

into well-established loss functions. Compared to the Euclidean margin, angular margin is preferred due to it's intrinsic consistency with softmax.

While facial feature extraction is well-resolved in constrained scenarios, there are uncontrollable variations in real-life images such as extreme facial expression, pose, illumination, and occlusion. Face occlusion is one of the most challenging problems to overcome among them. [8–12]. The core problem of occluded face recognition (OFR) is the lack of discriminative facial characteristics in the occluded areas. Occlusion items are also extremely diverse, such as hand covering face, sunglasses, scarf, and most notably, medical masks. The presence of occlusion items reduce the effectiveness of FR algorithms. Due to large intra-class variability and inter-class similarities induced by corrupted features from obstructed areas, Ghazi [13] demonstrated that CNN models which are trained using ordinary clean faces cannot correctly detect occluded faces.

Using state-of-the-art architectures and loss functions to train the network using a large number of occluded face photos is a simple technique to improve the efficiency of Convolutional networks under occlusion. Synthetic occluded faces are usually generated by substituting areas of the face with comparable occlusions objects like as masks or sunglasses due to the lack of masked face datasets. To improve the accuracy of generated masks, facial landmarks are first determined via traditional computer vision algorithms or deep learning [14]. These key points are then used to reasonably locate the location of occlusion items. However, this method does not consider various nuances such as face angle, mask fit, or lighting conditions. Recently, Anwar [15] proposed a new algorithm for masked face augmentation to generate more realistic synthetic occlusion items. However, generating synthetic masked data does not solve the problem fundamentally since there is still a discrepancy between the features of two faces with different occlusion elements. [16]. In addition to training the network with occluded faces, another approach to address OFR is partial face matching. We can delete feature elements that have been corrupted by occlusions in the same way as the human visual system pays attention to non-occluded facial areas for recognition (and ignores the occluded portions). The issue now is how to identify those feature elements that have been corrupted. When there is a clear correlation between image pixels and resulting feature components, traditional computer vision techniques like SIFT and HOG can be used, but not for deep convolutional features. Several approaches to resolving this issue have been offered, including partial face feature matching [17–19] and occlusion face recovery [20].

The core problem of OFR is to tell the network where to pay attention and where to ignore. We propose an unique Switching Replacement Attention Network (SRAN) to overcome this. In SRAN, the replacement module has the job of finding the occluded facial areas and replace them. This module answers the question “Where does occlusion item locate?”. However, not all faces contain occlusion items, therefore in those situations the replacement module is redundant. The switching module address this problem: “When is replacement module



**Fig. 1.** Our SRAN framework’s overall architecture. First, the input face is fed into a Switch Block based on Mobilenet-V2 [21]. If the face is occluded, the Replacement Module will replace the corrupted regions with zero vector. The discriminative face features are learned via a Loss Layer during the training phase (Large Margin Cosine Loss). The cosine similarity score is computed using extracted features during the testing phase.

applied?”. This module reduces the model’s overall processing time. The paper’s key contributions are twofold:

- We propose a novel SRAN framework that accurately locate occlusion items and remove them from the feature map. This makes our network focus more on the feature-rich areas of the face.
- We build a robust face-recognition system based on SRAN that performs well with both occluded and non-occluded face datasets.

## 2 Proposed Approach

Figure 1 depicts the overall flow of our approach. Our approach can be divided into 3 blocks:

- Switch block  $\mathcal{SW}$
- Replacement module  $\mathcal{R}$
- Backbone module  $\mathcal{B}$

In this paper, we leverage Resnet-50 [22] and state-of-the-art Large Margin Cosine Loss [6] as the backbone module.

### 2.1 Backbone Module

First, backbone model  $\mathcal{B}$  is used to extract high-level feature  $x_i$  from aligned face  $t_i$ . The feature extractor in this study is Resnet-50. The extracted feature is then passed through a fully-connected layer  $\mathcal{FC}$  with weight vector  $W_j$  and bias  $B_j$  (we fix  $B_j = 0$  for simplicity). The activation of  $\mathcal{FC}$  (denoted as  $f_j$ ) is given by the equation:

$$f_j = W_j^T x_i = \|W_j\| \|x_i\| \cos \theta_j \quad (1)$$

where  $\theta_j$  is angle between  $W_j$  and  $x_i$ . To make the norm of  $W$  invariable,  $W_j$  is normalized to 1 by  $L_2$  normalization. Since the recognition score is unaffected by the norm of the feature vector  $x$ ,  $\|x_i\|$  is l2-normalized and multiplied by a constant  $s$  during the training stage. Equation 1 now merely depends on the cosine of angle.

We rethink the softmax loss from a cosine perspective to create the Large Margin Cosine Loss (LCML). The function for softmax loss can be written as:

$$L_s = \frac{1}{N} \sum_{i=1}^N -\log p_i = \frac{1}{N} \sum_{i=1}^N -\log \frac{e^{f_{y_i}}}{\sum_{j=1}^C e^{f_j}} \quad (2)$$

where  $y_i$  is the groundtruth label and  $p_i$  is the posterior probability that  $x_i$  is correctly classified. By introducing Eq. 1 to the softmax function, we achieve the normalized softmax loss:

$$L_{nsl} = \frac{1}{N} \sum_i -\log \frac{e^{s \cos(\theta_{y_i, i})}}{\sum_j e^{s \cos(\theta_{j, i})}} \quad (3)$$

But NSL loss is not sufficient enough as it only emphasizes on correct classification. To address the issue cosine margin is introduced to the loss function. Considering an example of two classes  $C_1$  and  $C_2$ , and  $\theta$  as the angle between the feature and weight vector. NSL requires  $\cos(\theta_1) > \cos(\theta_2)$  when  $C_1$  is the ground truth class. We further add margin  $m$  so that the requirement now becomes  $\cos(\theta_1) - m > \cos(\theta_2)$ , where  $m$  is used to regulate the cosine margin. LCML is formally defined as:

$$L_{lmc} = \frac{1}{N} \sum_i -\log \frac{e^{s(\cos(\theta_{y_i, i}) - m)}}{e^{s(\cos(\theta_{y_i, i}) - m)} + \sum_{j \neq y_i} e^{s \cos(\theta_{j, i})}}, \quad (4)$$

## 2.2 Replacement Module

The backbone module works well on non-occluded face recognition, but it's effectiveness is limited when facial features are compromised. The extracted feature map contains information of both the occlusion item and the clean facial area, while ideally it should only focus on the information from the non-occluded regions. To address this, we integrate a replacement module  $\mathcal{R}$  to the baseline model to adjust the attention map so that it ignore the corrupted feature elements. This replacement module works as attention mechanism for the backbone module.

**The Choice of Replacement Module  $\mathcal{R}$ :** The task of module  $\mathcal{R}$  is to find the location the occlusion items via segmentation, and replace those regions with zero vector  $\mathcal{O}$ . The Replacement module is defined as.:

$$R_{i,j}(x) = \begin{cases} 0 & \text{if } x_{i,j} \text{ is occlusion pixel} \\ 1 & \text{otherwise} \end{cases} \quad (5)$$

where  $x$  is the normalized input face,  $x^s$  is replaced image, and  $\mathcal{R}$  is the replacement module. At it's core,  $\mathcal{R}$  can be seen as a semantic segmentation module. First, a segmentation map of occluded item is generated from the input face  $x$ . Next, the segmentation result is mapped to the original normalized feature via pixel-wise multiplication:

$$x^s = \mathcal{R}(x) \otimes x \quad (6)$$

In this study, we used Deeplab-V3 [23], a state-of-the-art segmentation network as the replacement module for best speed to accuracy trade-off.

### 2.3 Switch Module

While the replacement module works well on occlusion faces, it might introduce redundant computation since occlusion items are not always present. Therefore, we propose a switch module  $\mathcal{SW}$  that makes our system run faster and more precise. In Subsect. 2.2, probe face  $x$  is passed through Replacement module  $\mathcal{R}$  to achieve  $x^s$ . This approach obviously only works well when the face is occluded. As a result, we need a method to determine when the Replacement Module is used. To address this, the probe face is passed through a Switch Module  $\mathcal{SW}$  that classify it into two classes: occluded and non-occluded faces. The Replacement Module is then only activated when  $\mathcal{SW}$  determine that  $x$  is occluded. The following formula describes the Switch Module:

$$\mathcal{SW}(x) = \begin{cases} 0 (x \in X^{no}) & \text{then } x \rightarrow \mathcal{SW} \rightarrow \mathcal{B} \\ 1 (x \in X^\circ) & \text{then } x \rightarrow \mathcal{SW} \rightarrow \mathcal{R} \rightarrow \mathcal{B} \end{cases} \quad (7)$$

Since we want  $\mathcal{SW}$  to be accurate while not significantly impact the framework's overall speed, we use a customized version of MobilenetV2 [21] to construct the Switch Module. The classification network is trained with Focal Loss [24] to mitigate the class imbalance problems:

$$\mathcal{L}_{\mathcal{SW}} = -y_{\mathcal{SW}}(1-p)^\gamma \log p - (1-y_{\mathcal{SW}})p^\gamma \log(1-p) \quad (8)$$

### 3 Experimental Results

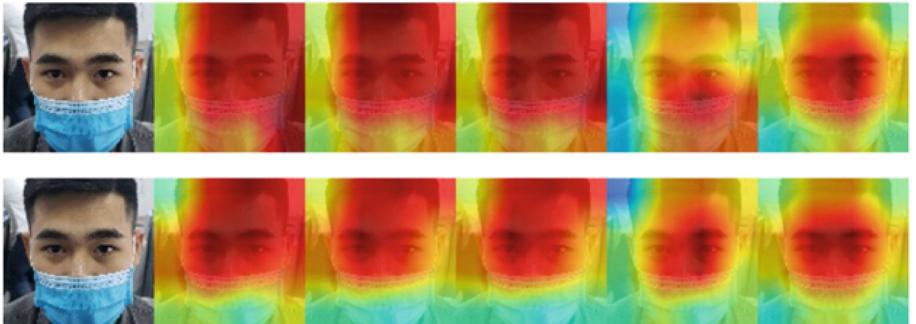
#### 3.1 Implementation Details

**Preparation of Data and Training:** The training process contains three stages: training the backbone module  $\mathcal{B}$ , training the Switch Module  $\mathcal{SW}$  and training the Replacement Module  $\mathcal{R}$ . These modules are trained separately. Similar to Cosface [6] we use 5.8 million face images of 85000 identities from MSCeleb-1M dataset [25] to train the backbone module. We further use face occlusion augmentation method from [15]. Occlusion items include medical mask, scarf, KN95, N95 and phone.

To generate the dataset used for training the Replacement module, we simply create segmentation masks of occlusion items from the augmented masked face. Attention-block training dataset consist of 1.6 million pairs of face - mask segmentation images.

Finally, to train the Switch Block, we use the MAsked FAces dataset (MAFA) [26] dataset. MAFA contains 30K images with 35K faces both with and without occlusion items, which make it suitable for training our robust  $\mathcal{R}$  block.

**Performance of Auxiliary Modules:** We train a Deeplab-V3 [23] segmentation network to determine the position of the occlusion. We fine-tune the model trained on COCO to provide good initialization and reduce the number of classes to one (occlusion object). Finally, our occlusion segmentation algorithm performs well on synthetic occluded Face-scrub dataset [27], with a mean IU of 99.14%. In Fig. 2, we visualize the impact of our Replacement module on the high-level feature map. Our  $\mathcal{SW}$  block also achieve 99.89% accuracy on MAFA test set.



**Fig. 2.** The first row illustrate the output attention map of six top layers without Replacement module  $\mathcal{R}$ . The second row illustrate the output attention map of six top layers with Replacement module  $\mathcal{R}$ . The network uses the attention module to focus on the clean face portions while ignoring the occluded areas.

### 3.2 Performance on LFW Benchmark

LFW (Labeled Face in the Wild) is a dataset of 13,233 images of faces in the wild. There are 5749 identities in this dataset, with 1680 persons having two or more images. The verification accuracies are reported on 6000 face pairings using the conventional LFW evaluation technique. We also benchmarked our method using a masked version of LFW using augmentation methods from [15], which we call occluded LFW (OLFW). We only use this augmentation method on half of the number of faces, the remaining are used for testing as normal. Results are shown in Table 1. Because the majority of the face images in LFW aren't covered, our method without the Switch block suffers slight performance drop compared to the baseline model. This is in line with [16], which found that a model that works well for occluded objects does not necessarily work well for non-occluded objects. With the addition of the Switch Block, our method is able to outperform the baseline model.

**Table 1.** Face verification (%) on LFW and synthesized OLFW dataset

Method	Training data	LWF	OLFW
Facenet [2]	200M	99.63	–
Centerface [3]	0.7M	99.28	–
SphereFace [5]	0.49M	99.42	–
ArcFace [7]	0.49M	99.53	–
CosFace (Baseline) [6]	5M	99.73	88.02
Ours (w.o Switch Block)	5M	98.64	89.34
<b>Ours</b>	5M	<b>99.8</b>	<b>92.89</b>

### 3.3 Performance on AR Dataset

We also test our methodology with real-life occlusions in the AR face dataset [28] with real-life occlusions. It comprises over 4,000 photos that correlate to the faces of 126 people. Frontal view faces with various facial expressions, lighting circumstances, and occlusions (sun glasses and scarf) are included in the images. There are two different testing procedures. To create the gallery set, Protocol 1 uses many images from each subject, but Protocol 2 only uses one image per subject. On all procedures and occlusion items, the experimental findings in Table 2 reveal that our technique greatly outperforms the baseline method.

**Table 2.** Rank-1 face identification accuracy (%) on the AR dataset with occlusions

Method	Protocol	Sunglass	Scarf
SRC [8]	1	87.00	59.50
NMR [9]	1	96.90	73.50
RPSM [10]	1	96.00	97.66
MaskNet [18]	1	90.90	96.7
CosFace (Baseline)	1	99.48	88.02
<b>Ours</b>	1	<b>99.68</b>	<b>100</b>
RPSM [10]	2	84.84	90.16
Stringface [11]	2	82.00	92.00
LMA [12]	2	96.30	93.70
CosFace (Baseline) [6]	2	96.57	96.28
<b>Ours</b>	2	<b>98.37</b>	<b>99.17</b>

## 4 Conclusion

In this paper, We present a robust occlusion recognition approach based on the attention mechanism that works on both occluded and non-occluded faces. In addition to the Attention block that explicitly helps the model omits corrupted information, we also introduce a switch block that helps improve the overall speed and accuracy of our algorithm. On both synthetic and natural face datasets, experimental results show that the proposed method is superior.

## References

- Masi, I., Wu, Y., Hassner, T., Natarajan, P.: Deep face recognition: a survey. In: 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), pp. 471–478. IEEE (2018)
- Schroff, F., Kalenichenko, D., Philbin, J.: Facenet. A unified embedding for face recognition and clustering (2015)
- Xu, Y., Yan, W., Yang, G., Luo, J., Li, T., He, J.: CenterFace: joint face detection and alignment using face as point. Sci. Program. **2020** (2020)
- Deng, J., Guo, J., Zhou, Y., Yu, J., Kotsia, I., Zafeiriou, S.: RetinaFace: single-stage dense face localisation in the wild. arXiv preprint [arXiv:1905.00641](https://arxiv.org/abs/1905.00641) (2019)
- Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., Song, L.: SphereFace: deep hypersphere embedding for face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 212–220 (2017)
- Wang, H., et al.: CosFace: large margin cosine loss for deep face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5265–5274 (2018)
- Deng, J., Guo, J., Xue, N., Zafeiriou, S.: ArcFace: additive angular margin loss for deep face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4690–4699 (2019)

8. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(2), 210–227 (2008)
9. Yang, J., Luo, L., Qian, J., Tai, Y., Zhang, F., Xu, Y.: Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(1), 156–171 (2016)
10. Wan, W., Chen, J.: Occlusion robust face recognition based on mask learning. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 3795–3799. IEEE (2017)
11. Chen, W., Gao, Y.: Recognizing partially occluded faces from a single sample per class using string-based matching. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6313, pp. 496–509. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-15558-1\\_36](https://doi.org/10.1007/978-3-642-15558-1_36)
12. McLaughlin, N., Ming, J., Crookes, D.: Largest matching areas for illumination and occlusion robust face recognition. *IEEE Trans. Cybern.* **47**(3), 796–808 (2016)
13. Mehdiipour Ghazi, M., Kemal Ekenel, H.: A comprehensive analysis of deep learning based representation for face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 34–41 (2016)
14. Wang, X., Wang, K., Lian, S.: A survey on face data augmentation for the training of deep neural networks. *Neural Comput. Appl.* **32**(19), 15503–15531 (2020). <https://doi.org/10.1007/s00521-020-04748-3>
15. Anwar, A., Raychowdhury, A.: Masked face recognition for secure authentication. arXiv preprint [arXiv:2008.11104](https://arxiv.org/abs/2008.11104) (2020)
16. Osherov, E., Lindenbaum, M.: Increasing CNN robustness to occlusions by reducing filter support. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 550–561 (2017)
17. Duan, Y., Lu, J., Feng, J., Zhou, J.: Topology preserving structural matching for automatic partial face recognition. *IEEE Trans. Inf. Forensics Secur.* **13**(7), 1823–1837 (2018)
18. Weng, R., Lu, J., Tan, Y.-P.: Robust point set matching for partial face recognition. *IEEE Trans. Image Process.* **25**(3), 1163–1176 (2016)
19. Song, L., Gong, D., Li, Z., Liu, C., Liu, W.: Occlusion robust face recognition based on mask learning with pairwise differential siamese network. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 773–782 (2019)
20. Wang, M., Hu, Z., Sun, Z., Zhao, S., Sun, M.: Varying face occlusion detection and iterative recovery for face recognition. *J. Electron. Imaging* **26**(3), 033009 (2017)
21. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C.: MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520 (2018)
22. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
23. Chen, L.-C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. arXiv preprint [arXiv:1706.05587](https://arxiv.org/abs/1706.05587) (2017)
24. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
25. Guo, Y., Zhang, L., Hu, Y., He, X., Gao, J.: MS-Celeb-1M: a dataset and benchmark for large-scale face recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 87–102. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46487-9\\_6](https://doi.org/10.1007/978-3-319-46487-9_6)

26. Ge, S., Li, J., Ye, Q., Luo, Z.: Detecting masked faces in the wild with LLE-CNNs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2682–2690 (2017)
27. Ng, H.-W., Winkler, S.: A data-driven approach to cleaning large face datasets. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 343–347. IEEE (2014)
28. Martinez, A., Benavente, R.: The AR face database: CVC technical report, 24 (1998)



# A Proposed Business Intelligence Framework for Autonomous and Non-public Higher Education Institutions in Vietnam

Han Minh Phuong<sup>1(✉)</sup>, Pham Minh Hoan<sup>2</sup>, Nguyen Trung Tuan<sup>2</sup>, and Doan Trung Tung<sup>3</sup>

<sup>1</sup> Thuongmai University, Hanoi, Vietnam

[hmphuong@gmail.com](mailto:hmphuong@gmail.com)

<sup>2</sup> National Economics University, Hanoi, Vietnam

<sup>3</sup> University of Greenwich (Vietnam), Hanoi, Vietnam

**Abstract.** Currently, the application of Business Intelligence systems in enterprises has become popular. However, research on applying Business Intelligence in the education field is still incomplete. Theoretical and experimental studies for applying the BI system in higher education institutions (HEIs) in a developing country like Vietnam are scarce. In this paper, we aim to study the application of Business Intelligence in Vietnamese HEIs. Based on previous framework studies and the current situation in VietNam, we proposed BI framework customizations for HEIs to create a suitable common framework for the autonomous and non-public HEIs in Vietnam. Besides that, the data warehouse will be stored on a cloud service, and inbuilt data science tools are utilized for analytical and forecasting reports.

**Keywords:** Business intelligence · Data science · Education · Framework · Data warehouse

## 1 Introduction

BI system is a system that allows the analysis and visualization of information and knowledge extracted from many data sources. BI helps business managers and leaders gain business-critical insights from data, thereby pointing out business problems or potential issues that need to be concerned. Furthermore, BI helps businesses identify market trends from which to make the right operating decisions, improve the operational efficiency of the business. BI also strengthens and creates a strong relationship between businesses and customers, increasing the competitiveness of businesses [1]. HEIs, primarily autonomous and non-public HEIs, can be considered a specific type of enterprise. The services that HEIs provide are training services. The more learner-oriented the training service is, the better quality it will be. The more it meet social requirements, the more

prestige it will bring to the educational institution. And the more it attracts learners, the more revenue it will generate for educational institutions. Therefore, the application of BI to educational institutions has become increasingly popular.

Studies on the application of BI systems in HEIs have been carried out for a long time [2]. These studies show that the BI system has a significant role in HEIs. For example, the role of BI is demonstrated in managing enormous data resources, improving the admissions process, improving student retention, improving student satisfaction, generating financial reports, etc., meaning reducing management costs. The researches focused on proposing general frameworks [3], technologies [4,5], and some aspects of learning data mining in the BI system [6,7]. Most of these researches were carried out in HEIs of advanced countries, in institutions with a high or very high level of information technology application in university administration. However, research on the BI framework that can be applied to an HEI in developing countries is not much. For example, in Vietnam, although the digital transformation promotes the application of technology in university administration, studies on BI for educational institutions are still rare [8]. Therefore, the purpose of this study is to propose an appropriate BI framework for HEIs in Vietnam, especially with autonomous and non-public HEIs. We make this proposal thanks to the inheritance of the published BI system frameworks, the study of the IT application characteristics in higher education administration, and current requirements in the management of Vietnamese HEIs.

The article consists of 5 parts. Section 1 introduces an overview of the article. Section 2 summarizes the published studies on BI and BI frameworks in HEIs, showing the main results corresponding to this aspect. Section 3 talks about the current application of IT in the university administration of Vietnamese HEIs. Section 4 proposes a framework solution for the BI system, and the last one includes conclusions related to the research.

## 2 Literature Review

In this section, we will review the current status of how BI is applying in HEIs in the world.

### 2.1 BI Systems

The term Business Intelligence was introduced by Hans Peter Luhn in 1958 [9]. From the 1990s, BI has been broadly understood as “terms and methods to improve business decisions by making use of support system based on business data.”

Today BI system is said to have been developed from the decision support systems (DSS) of the 1960s–1980s. DSS uses models to support decision-making and planning. Hans Peter Luhn also defines a BI System as an automated system developed to disseminate information to different parts of any industrial,

scientific or governmental organization. Also, he highlights the importance of effectively creating and distributing information to the right users.

According to Carlo's pyramid [10] that describes a general BI framework, a BI system performs two main activities: bringing data from discrete sources into an aggregated data warehouse. Through technological processes and application softwares, a BI system provides results in the form of reports and information for decision support purposes. BI's necessary techniques are extracting, transforming, and loading data combined with data visualization, forecasting, and data mining through machine learning. The ability of information visualization from multiple data sources helps managers make decisions based on relevant and complete data evidence. The predictive analytics capabilities of BI systems make businesses more intelligent. Instead of reacting to crises and opportunities, organizations anticipate them. The organization sees and capitalizes on opportunities in a competitive landscape, identifies and resolves problems before they escalate into crises, improves internal processes, products, and services to enhance customer satisfaction and loyalty. BI is a weapon for a 21st-century enterprise organization that can survive and thrive in a harsh and tough globalized environment [11].

## 2.2 BI Systems in HEIs

Recognizing the role of the BI system for organizations, HEIs have also begun to study BI and apply BI in their organizations. Although previous studies of BI from other disciplines may provide valuable insights, it is believed that studies of BI in the higher education sector still have their idiosyncrasies. Concepts such as academic analytics, data mining in higher education, and course management system audits are linked in research by Paul Baeppler and Cynthia James Murdoch, University of Minnesota, to create an initial idea of the BI system in HEIs [12].

When applying the BI system, the benefits that it brings to HEIs focus on the main goals that are to improve training quality, improve retention rate, dropout rate, increase student satisfaction, thereby increasing the school's enrollment rate [13–15]. In addition, the BI system also helps to cut management costs and rationally allocate the school's resources [16]. BI provides HEIs with a better outlook by providing valuable information relevant to university operations that can be used by top university management [17]. BI provides efficient capabilities in creating and delivering various reports, such as the annual university performance reports [18]. This report can make it easier for the university's senior manager to monitor how well it is doing. BI can support strategic vision by analyzing strategic planning and comparing educational quality outcomes between different units within the university and across universities [19, 20].

Although there are quite a few studies on applying BI for a specific HEI, research on the BI framework in general and the BI framework for HEIs, in particular, is still limited. The research of Muntean et al. presented a framework including eight components to guide the implementation of BI solutions in HEIs for university governance [19]. These eight components are university strategy,

financial and operational objectives, key business processes, university performance model, key performance indicators, people, process, BI technologies and information infrastructure.

These components are considered the main reference components for BI system research for HEIs. In addition, many other authors have based on these eight components to conduct other studies in applying BI to HEIs. Recent studies have focused heavily on the data mining layer to deepen understanding and increase predictability on the increasing volume of learning data in HEIs. However, it can be seen that the general framework studies are not based on the specific context of HEIs in a country because countries have very different levels of development and contexts. For developing countries, there is not much research.

In Vietnam, the problem of researching and applying BI systems at HEIs with limited funding, information infrastructure, and human resources has not yet been solved thoughtful. Educational organizations face too many problems when studying system implementation, such as the fit between the system's goals and strategy, readiness for the system, availability in terms of quantity and quality of data. The lack of practice also means creating a gap of limited knowledge related to the development of BI systems in the academic field.

### **3 Current Status in Vietnam**

In Vietnam, HEIs are categorized based on ownership into public and non-public HEIs. A public HEI is a state-invested institution in terms of funds and facilities and operates mainly with funding from public financial sources or non-profit contributions. A non-public HEI is established and invested by individuals or organizations, operating with funds contributed by students, customers, and donations. Non-public HEIs include non-profit and for-profit non-public HEIs, where the organizational and operational regulations are the same as ones for multi-member limited liability companies specified in the Enterprise Law of Vietnam.

In the period 2014–2020, the number of HEIs has a growth rate of 10% (aggregated data from <http://csdldh.moet.gov.vn/>), while the number of enrollments does not increase, even tends to decrease.

The public HEIs have started to implement university autonomy since 2015. After issuing Resolution 77/NQ-CP, Law 34/2018/QH14 (the law amending and supplementing some articles of the Higher Education Law), and Decree 99/2019/NĐ-CP, more and more public HEIs are transitioning to autonomous model. The autonomy model associated with the disclosure of education quality, the disclosure of ensuring education quality, and the disclosure of financial revenues and expenditures has brought many opportunities and challenges for public HEIs. Public HEIs promote their inherent strengths and potentials in creativity, dissemination, and knowledge transfer, when transitioning to autonomy, can more easily use capital to invest in infrastructure and improve the quality of teaching and learning.

However, the funding not gradually compensated by the state budget also makes the public HEIs encounter many difficulties. With 70% of revenue based

on tuition fees, they have to solve the problem of a sudden tuition fee increase but still have to maintain (or even increase) a large number of students. In addition, when university autonomy expands, special attention is paid to education quality accreditation, oriented to regional and international standards. Achieving quality accreditation is a prerequisite for any HEI to continue enrolling students, and autonomous public HEIs are allowed to increase the tuition fees 2–2.5 times non-autonomous public institutions or autonomous public but not yet qualified. Therefore, achieving quality accreditation has been a trend to affirm the position and HEI's quality and is carried out regularly (once every five years according to Circular 12/2017/TT-BGDDT of The Ministry of Education and Training).

For non-public HEIs, to attract more students and increase their competitiveness with public HEIs, non-public HEI, on the one hand, have invested in modern infrastructure and equipment, and on the other hand, constantly strives to change the curriculum in a more realistic direction. The alien training program helps students get international degrees right after graduation. However, as university autonomy expands, these programs always have to compete with an increasing number of alien training programs at public HEIs. Advertising in public HEIs also promoted is reducing the competitive advantage of non-public HEIs significantly.

Thus, within five years, the context of changing mechanisms has created a significant impact on the autonomous and non-public HEIs. That matter of fact pushes these HEIs to make changes in their operating strategies, apply advanced governance models, and increase technology in administration to create significant changes and breakthroughs. This change is also mentioned in the orientation of the National Digital Transformation Program to 2025 and the orientation to 2030 of the Prime Minister. However, a significant difficulty for current HEIs in applying a new and advanced governance model is the lack of synchronization and the failure to thoroughly apply the technology system to support all aspects of administration. Today's popular technology systems in HEIs are often discrete and involve some domains of training, admission, finance, and human resources, leading to overlapping data storage. HEIs cannot access and extract the hidden information in the large amount of data they have. In addition, there have not been computerization in some domains and activities, leading to data not being stored systematically, making complex administration and a lot of data loss.

The recent Covid-19 pandemic has created a big push and an opportunity for HEIs to demonstrate their ability to use information technology in teaching and learning. Thus, HEIs not only responded to the epidemic situation but also had strategic orientations, preparation, and substantial investment in developing online training, distance learning, and digital transformation. That is also the reason and motivation for us to research an appropriate framework to guide Vietnamese HEIs in applying the BI system in HEIs governance, decision-making, and development strategy support.

## 4 Proposed Framework

From the studies on the application of BI systems to HEIs, we found that when building the framework for the BI system, previous studies will stratify the components of the system. The layers /components of the system can be approached from one or more aspects: technology, organization, management. Although the components have been fully described, the level of detail and relevance, especially in technology, has not been mentioned when applied to Vietnamese HEIs.

Therefore, inherited the framework of the BI system applied to HEIs in the world, based on the current situation of information technology application in university administration activities in Vietnam and the urgent requirements for management information to response the higher education accreditation standards (TT 04-2016 TT-BGDDT), the study proposes a common technological framework for the BI system for Vietnamese HEIs.

The framework inherited from the research of Brenda Scholtz and colleagues [16] for HEIs, in general, includes eight layers and six steps specified by the author group to suit the requirements at Vietnamese HEIs. The following subsections will describe five customized layers to adapt to current situation of Vietnamese HEIs.

### 4.1 Data Source Layer

The data source layer is obtained from the school's Management Information Systems. University governance will be effective if three pillars: training, scientific research, and human resource are well-governed. Therefore, an HEI that applies information technology to university administration activities needs to focus on application groups serving these three domains, including:

**University Administration Application:** administrate admissions; manage records of staffs, lecturers, and students; manage training program; manage learning materials; manage grades of students; manage resources in the school, including physical resources, human resources and other resources; manage the work quality of staffs, lecturers in HEIs.

**Teaching Management Application:** manage teaching activities before, during, and after class lectures; manage the content of the lectures and the electronic lesson plans of the lecturers; evaluate the teaching quality of lecturers through students.

**Application to Manage Learning Activities of Students:** Learners use software on digital devices or computers to support self-study activities, online learning, or exploiting information on the internet. Lecturers can assign homework through online systems and evaluate self-study works submitted by students.

**Application in Research and Transfer of Science and Technology:** manage registration and grant of scientific and technological research projects; manage scientific research activities; manage the output products of scientific researches; manage and protect intellectual property rights.

**Application in Human Resources Management:** manage and evaluate the quality of the outputs of the training process; manage, assist students in finding jobs in the labor market manage the labor cooperation in the market, create links between HEIs, businesses, and students, bring benefits to all parties.

In the current Vietnamese HEIs, the main applications are Academic Information Portal, Course Registration System, Content/Learning Management System, Library Management System, Human Resource Management System, Customer Relationship Management System. Online learning platforms have especially exploded during the recent Covid period. However, not all of these applications are being used in HEIs.

In addition, data on online learning platforms such as Zoom, Trans, or Teams are restricted from use because the applications are exploited through the services of local or international service providers. Therefore, data can only be extracted from existing systems and updated when the HEI digitizes necessary operations.

## 4.2 ETL Layer

This layer includes tools and data warehouses that integrate data from multiple systems. We propose to use cloud computing technology to extract, transform and load data. ETL is a built-in service among the cloud computing services from Cloud Providers such as Google, Microsoft, or Amazon. Services like Google's Big Query apply technology to store big data in the cloud, support building data warehouses that store data extracted from many sources, and store data of many types. These services also support tools for data analysis, predictive machine learning algorithms on data. Google Big Query also provides APIs to embed tools to help users query and display data to support the automatic extraction and display of data.

## 4.3 Reporting Layer

The reporting layer is responsible for extracting data by ad-hoc queries to extract customized information according to the needs of management. In an enterprise BI system, the reports give information to management levels such as chief executive officer (CEO), chief financial officer (CFO), chief operating officer (COO), chief sales officer, and heads of departments of purchasing, production, logistics, marketing, customer care, warranty, and maintenance, etc. In the BI system of HEIs, reports are also reserved for managers from the heads of faculties and affiliated departments, which focus on functional departments in critical domains such as admission, training, science and technology, and human resource.

The image of the organizational hierarchy of bureaucracy organization (joint-stock company) is most similar to public HEIs or non-public HEIs with profits in Vietnam. Thus, at this layer, the system embeds visualization tools such as Tableau, PowerBI to provide information from ad hoc queries to managers up from the departmental level. The managers use reports according to the level of authority. For example, the Department of Training uses study plans, score management, academic complaint, and review reports. The Department of Legal Inspection or Testing and Quality Assurance access reports related to inspections, exams, and course evaluations. The Department of Student Affairs uses reports related to reward and learning support policies. The Science Department uses reports related to scientific research. Finally, the Human Resources Department uses reports related to the lecturers and staff.

#### **4.4 Analytical Layer**

Systems use advanced statistics and machine learning methods to dig deeper into data. Take advantage of large and diverse data sources obtained from many observations over a long period, such as student learning data and enrollment data over the years, to make forecasts in line with the goals of HEIs. Widespread forecasting problems are usually based on learners' data, such as forecasting students' course results based on mid-term assessments or online learning activities, predicting students who may pass or fail, and predicting student learning outcomes very early in the first or second year. The forecasting methods can be optional for each system, may include Logistic Regression, Decision Tree, Neural Network, Naive Bayes, KNN, Support Vector Machine or ensemble methods like Bagged trees, Adaptive boosting trees, Random forests. Machine learning methods have varying degrees of accuracy and are bound to HEI-specific data. Therefore, the establishment of studies is highly dependent on both the availability and accessibility of the institution's internal information systems.

#### **4.5 Monitoring Layer**

The information generated in the analysis layer is usually present to management in the monitor layer. In the enterprise BI system, the information generated in the monitoring layer will be dashboards or balanced scorecards presented to the Board of Directors, focusing on the four aspects of financial, customer, internal business process, learning/growth. For the BI system of HEIs, because there are differences with enterprises in general goals and output products, the information required in the monitoring layer will also follow different standards. The common goal of educational institutions is to train human resources to meet socio-economic development requirements and international integration.

Education is a service that includes both features, which are an exclusion in use because learners have to take entrance exams, pay tuition fees, and compete due to restrictions on the number of enrollments and school entry standards. The product of the educational institution is the knowledge and skills that learners gain after the study program. Therefore, the monitoring layer in the BI system

for HEIs should include a dashboard and a balanced scorecard, allowing the organization to better manage and monitor the indicators of HEIs in terms of enrollment, training, scientific research, and organization personnel. Monitoring layer reports present information to senior management (School Administrators) to help identify trends, patterns, and anomalies.

## 5 Conclusion

As organizations rely more and more on data, the effectiveness of leaders' and managers' decisions will depend on tapping into the organization's available potential data. Therefore, the need for a BI system will continue to grow and become increasingly essential. The application of the BI system is not only limited to enterprises but will spread more to organizations such as HEIs. In this article, the authors have carried out initial research on applying the BI system in HEIs in Vietnam. With the inheritance of framework studies and current situation studies, we have proposed customizations for the BI framework for HEIs to create a common framework suitable for the Vietnamese HEIs. This framework ensures data coverage from core business processes that impact the performance of HEIs and, at the same time, solves the problem of big data processing for HEIs with the infrastructure of technology that is unsolid and incomplete. Indicators to measure and evaluate the performance of HEIs that meet accreditation standards according to the Ministry of Education and Training are also proposed. It is expected that in the near future, the research team will continue to clarify the framework by empirical studies on several HEIs in Vietnam and evaluate the results of the model. From there, the research results can be widely applied in a rather specific context such as Vietnam's HEIs.

## References

1. Dmitri Apraxine, E.S.: Business intelligence in a higher educational institution. The case of University of Nicosia. In: IEEE Global Engineering Education Conference, Athens, Greece (2017)
2. Piedade, M.B., Santos, M.Y.: Business intelligence in higher education: enhancing the teaching-learning process with a SRM system. In: 5th Iberian Conference on Information Systems and Technologies (2010)
3. Zulkefli, N.A., et al.: A Business Intelligence framework for higher education institutions. J. Eng. Appl. Sci. (2016)
4. Hamed, M., Mahmoud, T., Gómez, J.M., Kfouri, G.: Using data mining and business intelligence to develop decision support systems in Arabic higher education institutions. In: Marx Gómez, J., Aboujaoude, M.K., Feghali, K., Mahmoud, T. (eds.) Modernizing Academic Teaching and Research in Business and Economics. SPBE, pp. 71–84. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-54419-9\\_4](https://doi.org/10.1007/978-3-319-54419-9_4)
5. Boulila, W., Al-kmali, M., Farid, M., Mugahed, H.: A business intelligence based solution to support academic affairs: case of Taibah University. Wirel. Netw. 1–8 (2018). <https://doi.org/10.1007/s11276-018-1880-3>

6. Villegas, W., et al.: Towards the integration of business intelligence tools applied to educational data mining. In: 2018 IEEE World Engineering Education Conference (2018)
7. Valdiviezo-Diaz, P., Cordero, J., Reategui, R., Aguilar, J.: A business intelligence model for online tutoring process. In: IEEE Frontiers in Education Conference (FIE), El Paso, TX, USA (2015)
8. Agustiono, W.: Academic business intelligence: can a small and medium-sized University Afford to build and deploy it within limited resources. *J. Inf. Syst. Eng. Bus. Intell.* **5**(1) (2019)
9. Luhn, H.: A Business Intelligence System (1958). <http://altaplana.com/ibmrd0204H.pdf>
10. Vercellis, C.: *Business Intelligence: Data Mining and Optimization for Decision Making*. Wiley, Hoboken (2009)
11. Eckerson, W.: *Smart Companies in the 21st Century: The Secrets of Creating Successful Business Intelligence Solutions*. The Data Warehousing Institute (2003)
12. Baeppler, P., Murdoch, C.J.: Academic analytics and data mining in higher education. *Int. J. Scholarship Teach. Learn.* **4**(2) (2010)
13. Falakmasir, M.H.: Business intelligence in eLearning (case study on the Iran University of Science and Technology dataset). In: The 2nd International Conference on Software Engineering and Data Mining (2010)
14. Villegas, W., Palacios-Pacheco, X., Lujan-Mora, S.: A business intelligence framework for analyzing educational data. *Sustainability* (2020)
15. Yellowfin (2013). <https://www.yellowfinbi.com>. Accessed Aug 2021
16. Scholtz, B., Calitz, A., Haupt, R.: Business intelligence framework for sustainability information management in higher education. *Int. J. Sustain. High. Educ.* (2018)
17. Mutanga, A.: A context-based business intelligence solution for South African higher education. *J. Ind. Intell. Inf.* **3**(2) (2015)
18. Chen, M.: Applying Business Intelligence in higher education sector: Conceptual models and users acceptance. University of Bedfordshire (2012)
19. Mihaela et al.: Business Intelligence Systems in Support of University Strategy. Recent Researches in Educational Technologies (2011)
20. Khatibi, V., Keramati, A., Shirazi, F.: Deployment of a business intelligence model to evaluate Iranian national higher education. *Soc. Sci. Humanit. Open* **2**, 100056 (2020)



# Fall Detection Using Omnidirectional Camera

Viet Dung Nguyen<sup>(✉)</sup>, Thi Men Tran, Xuan Bach Nguyen, and Minh Quan Nguyen

School of Electrical and Electronic Engineering, Hanoi University of Science and Technology,  
Hanoi, Vietnam  
dung.nguyenviet1@hust.edu.vn

**Abstract.** Falling is one of the crucial problems threatening health of the elderly. It is especially dangerous when the elderly lives alone. Using omnidirectional cameras instead of traditional pinhole cameras in surveillance systems has become an attractive choice recently, as they provide a wider field of view. However, images captured by omnidirectional camera are highly distorted, especially in peripheral areas. Therefore, it is required to modify surveillance approaches that are developed for conventional pinhole cameras to adapt with such omnidirectional cameras. In this paper, a fall detection method using omnidirectional camera is proposed. First, blobs corresponding to moving objects are detected directly from the omnidirectional video by background subtraction method. Then different blob's features based on its contour and minimum bounding rectangle are extracted to classify state of the blob as fall or not fall. Experiments on BOMNI omnidirectional video dataset show that the highest accuracy of 97.07% when using KNN as background subtraction method and classifier.

**Keywords:** Fall detection · Omnidirectional · Classification

## 1 Introduction

Growth in the number of older persons is a worldwide phenomenon. Living alone poses a great threat when emergency situations occur. Those emergency situations include natural disasters, health-related problems like stroke, heart attack, and unusual behaviors in daily life, which are mainly falls [1]. It is essential to detect and deal with these situations quickly and accurately to promptly notify to relatives or caregivers. Therefore, having video surveillance system is proper in such circumstances. In this paper, we propose an indoor fall detection method using omnidirectional camera. From the word “omni”, meaning all, an omnidirectional camera is a camera having field of view covers 360° in horizontal plane, or a hemisphere or (approximately) entire sphere [2]. In comparison to traditional pinhole camera, omnidirectional camera provides a larger field of vision. Thus it is more efficient for surveillance purposes.

In this work, the authors propose a method for fall detection using omnidirectional camera. The paper is organized as follow. Related works are briefly introduced in Sect. 2. In Sect. 3 is the proposed methodology. Experimental results and discussions are provided in Sect. 4 while conclusions are given in Sect. 5.

## 2 Related Works

Omnidirectional cameras (typically fisheye camera) have lot of potential for surveillance applications, since they give a large field of view without the need for multiple cameras or PTZ cameras. However, reduction in resolution of peripheral areas and structural integrity of the image is a major disadvantage of omnidirectional camera. That leads to difficulty for conventional object detection and action recognition to be effective. Several approaches have been proposed to overcome these shortcomings.

In [3], a solution using HOG + SVM detector and rotating people to make them appear upright was proposed. This method was computationally expensive as it requires the algorithm to rotate fisheye image many times.

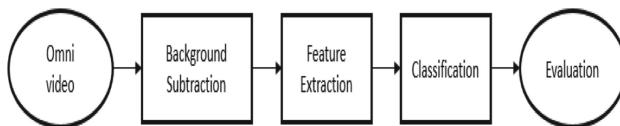
In [4], Convolutional Neural Network (CNN) was applied on fisheye image directly. But as human objects appear on fisheye image with different poses and angles, having not sufficient datasets for training CNN is the work's limitation.

In [5], the fisheye image was undistorted. In this approach, the peripheral pixels were removed to get an undistorted quadrilateral equiangular image. That made it not optimal in surveillance applications. In [6], similarly, the fisheye image was unwrapped. The difference is that the image was unwrapped into 360° not quadrilateral equirectangular image. However, it also suffered similar problem as in [5].

In [7], conventional features were modified to work on fisheye camera. However, this approach's complexity was increased, and the results was not comparable to modern standards.

## 3 Methodology

Depicted in Fig. 1 is the proposed method's block diagram. To detect moving objects on original omnidirectional video, background subtraction method is applied. Several features are extracted to represent each of the detected object or blob. The blob is then classified as fall or not fall based on those extracted features.

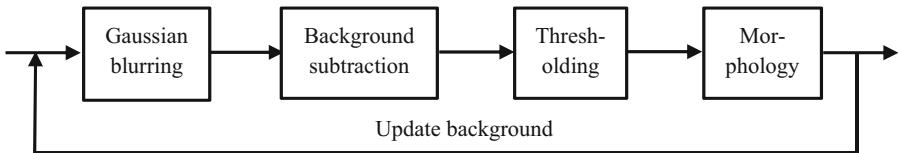


**Fig. 1.** Block diagram of the proposed method.

### 3.1 Background Subtraction

Among methods to detect moving or foreground objects in video sequences, the most popular one is background subtraction method. Given appropriated estimation of background pixel distribution, the objects can be detected by subtracting the current frame with background model. Most famous background subtraction methods are based on Gaussians Mixture Models (GMM) including GMG, MOG, MOG2 and KNN [8].

However, those algorithms mentioned above are not without flaws. Noise occurs in all these algorithms. Extracted objects sometimes are inseparable from their shadows. To overcome these problems, we did modification on traditional GMM background subtraction algorithm which is shown in Fig. 2.

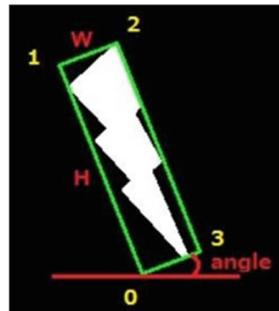
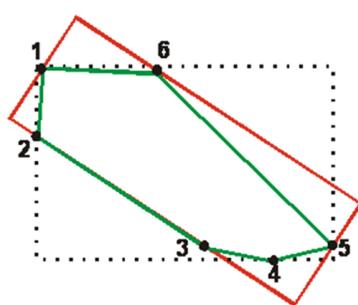


**Fig. 2.** Background subtraction improvement

Each frame is firstly blurred with Gaussian function [9] before GMM background subtraction is applied to extract moving objects. To remove object's shadow, thresholding method is applied [9]. Due to light changes or change in background, some small objects may appear. To eliminate these small objects, morphological operator is used [9].

### 3.2 Feature Extraction

As objects in omnidirectional images may appear in different directions therefore, instead of representing an object by its original bounding rectangle or box (like in perspective image), we use minimum bounding box as in Fig. 3.



**Fig. 3.** Object contour (green), original bounding box (dotted), minimum bounding box (red)

**Fig. 4.** Determination of angle of the minimum bounding box

Different features are used to describe each detected object. They are all extracted based the object's contour and minimal bounding box. They include.

- The contour's area
- The bounding box's area
- Ratio of the areas of the contour and the bounding box
- Angle of line fitted to the contour
- Ratio W/H, width and height of the minimum bounding box (see Fig. 4)
- Angle of the minimum bounding box (determined as in Fig. 4)

### 3.3 Classification

Detected objects are recognized as falling or not falling objects based on extracted features. In this work, Support Vector Machine (SVM) [10], K Nearest Neighbors (KNN) [11], Naïve Bayes [11] and Decision Tree (DT) [11] are used as classifier.

## 4 Results and Discussions

### 4.1 Dataset

All experiments are evaluated on BOMNI dataset [12]. This dataset contains two scenarios that are recorded by 2 omnidirectional cameras. One camera is mounted on the ceiling, and another camera is mounted on the wall.

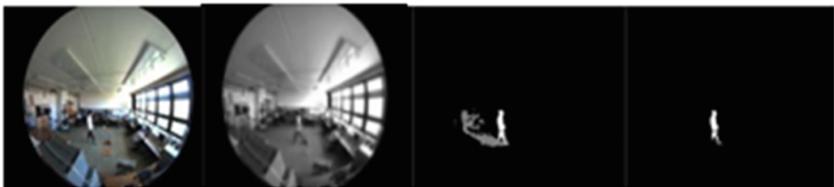
Annotations for the scenarios are as follow:

- Scenario #1: one person performs walking, sitting, drinking water, washing hand, opening-closing door, fainting/falling actions.
- Scenario #2: two or three persons perform walking, shaking hands, interested in object, sitting.

Since we focus on fall detection, only videos of scenario #1 are used in this work. Besides that, all actions are counted as “not fall” except fainting action (falling).

### 4.2 Background Subtraction

Through intensive trials with different GMM background subtraction methods, we found out that KNN method gave the highest accuracy. Optimal values of parameters are listed in Table 1. Results of our modified background subtraction method are depicted in Fig. 5.



**Fig. 5.** (From left to right). Original omni-directional; Gaussian blurred; background subtracted, and shadow removed images.

**Table 1.** Background subtraction parameter

Parameter	Value	
Shadow threshold	200	
History frame	500	
Opening	Kernel shape	Ellipse
	Kernel size	$3 \times 3$
	Iteration	1
Closing	Kernel shape	Ellipse
	Kernel size	$5 \times 5$
	Iteration	3

### 4.3 Feature Extraction

Illustrated in Fig. 6 are enlarged contour, minimum bounding box and fitted line to contour of detected object. Examples of extracted features for “fall” object and “not fall” object are given in Table 2.



**Fig. 6.** Contour of fall object (left). Minimum bounding box of fall object (middle). Line fitting to contour of not fall object (right)

**Table 2.** Extracted features

State	Ratio w_h	Area of contour	Area of bounding box	Ratio of areas	Angle of bounding box	Angle of fitting line
Fall	1.67	558	1728.48	0.3228	67.166	87.944
Not fall	4.46	1042	1776.11	0.5866	86.185	86.150

### 4.4 Classification

As mentioned in Sect. 3.3, SVM [10], KNN [11], NB [11] and DT [11] are used as classifier.

To evaluate performance of the classifiers, 10-fold cross-validation is utilized. The dataset is randomly divided into 10 folds while ratio between fall and not fall objects is kept same for all folds.

Hyper parameters of the classifiers are also fine-tuned. This is important because the entire model's performance is relied on the specified hyper parameter values.

**Table 3.** Classification performance

		DT	KNN	SVM	NB
Side view	Average	92.95%	<b>93.32%</b>	93.05%	91.43%
	Standard deviation	5.78%	4.65%	<b>4.37%</b>	6.16%
Top view	Average	96.64%	<b>97.07%</b>	95.36%	95.90%
	Standard deviation	<b>1.58%</b>	1.65%	2.34%	2.58%

Results from Table 3 prove that average and standard deviation of fall detection accuracy in top view omni-directional videos is better than that in side view videos. It is obvious that fall objects in top view videos are not obstructed by other furniture in the room.

For top view videos, among 4 considered classifiers, KNN provides highest performance of 97.07% and NB gives the worst 95.09%. Meanwhile, DT shows almost similar average and standard deviation accuracy of 96.64% and 1.58% respectively in comparison with corresponding 97.07% and 1.65% of KNN.

It should note that we recognize that fall false identifications happen mostly in following case (Fig. 7).

- Opening and/or closing door: when a person opens or closes the door, the detected bounding box will include both the person and the door area. It leads to false fall classification
- Stand nears by too bright places: when a person stands near too bright place, the brightness makes identifying the bounding box of the object difficult.



**Fig. 7.** False fall classification when opening/closing door (left) and standing near to bright place (left).

## 5 Conclusions

In this paper, one method to detect falling directionally in omnidirectional video is proposed. It composes of an improved background subtraction and a classifier. The highest accuracy is 97.07% with KNN as background subtraction as well as classifier with 6 features extracted from object contour and its minimum bounding box. This result can be used for future reference for further studies on omnidirectional camera-based elderly activity monitoring.

However, our approach cannot cope with occlusion or sudden change in illumination. It will be our future work.

## References

1. <http://www.un.org/en/sections/issues-depth/ageing/>
2. Scaramuzza, D.: Omnidirectional camera. In: Ikeuchi, K. (ed.) Computer vision: a reference guide. Springer, Berlin (2014). [https://doi.org/10.1007/978-0-387-31439-6\\_488](https://doi.org/10.1007/978-0-387-31439-6_488)
3. Chiang, A.T., Wang, Y.: Human detection in fisheye images using HOG-based detectors over rotated windows. In: 2014 IEEE International Conference on Multimedia and Expo Workshops ICMEW (2014)
4. Dupuis, Y., et al.: A direct approach for face detection on omnidirectional images. In: 2011 IEEE International Symposium on Robotic and Sensors Environments ROSE (2011)
5. Scaramuzza, D., Martinelli, A., Siegwart, R.: A flexible technique for accurate omnidirectional camera calibration and structure from motion. In: 4th IEEE International Conference on Computer Vision Systems ICVS (2006)
6. Kim, H., Jung, J., Paik, J.: Fisheye lens camera-based surveillance system for wide field of view monitoring. *Optik* **127**(14), 5636–5646 (2016)
7. Cinaroglu, I., Bastanlar, Y.: A direct approach for human detection with catadioptric omnidirectional cameras. In: 22nd Signal Processing and Communications Applications Conference SIU (2014)
8. Goyal, K., Singh, J.: Review of background subtraction methods using Gaussian mixture model for video surveillance systems. *Artif. Intell. Rev.* **50**(2), 241–259 (2017). <https://doi.org/10.1007/s10462-017-9542-x>
9. Gonzalez, R.C., Woods, R.E.: Digital image processing, 4th edn. Pearson, London (2018)
10. Boser, B., Guyon, I., Vapnik, V.: A training algorithm for optimal margin classifiers. In: 5th Annual Workshop on Computational Learning Theory (1992)
11. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern classification, 2nd edn. Wiley, New Jersey (2001)
12. Demiröz, D.E., et al.: Feature-based tracking on a multi-omnidirectional camera dataset. In: 5th International Symposium on Communications, Control, and Signal Processing ISCCSP (2012)



# Speed Up Communication for IoT Devices Using Small Form Factor in FPGA Platform

Trung Dang Tien<sup>1</sup>, Tuan Nguyen Ngoc<sup>2</sup>, Dat Tran Thanh<sup>3</sup>(✉) and Tinh Tran Xuan<sup>4</sup>

<sup>1</sup> Electric Power University, Ha Noi, Vietnam

<sup>2</sup> Military Technical Academy of Vietnam, Hanoi, Vietnam

<sup>3</sup> Hanoi University of Science and Technology, Ha Noi, Vietnam  
tranthanhdatt1506@gmail.com

<sup>4</sup> Air Defence - Air Force Academy of Vietnam, Hanoi, Vietnam

**Abstract.** IoT devices are being used more and more widely in today's life. With such great demand, IoT (Internet of Things) devices need to exchange more information. In this paper, a method using Small Form Factor (SFP module) on the FPGA platform instead of Ethernet ports is proposed to increase speed for IoT devices in the large SCADA system that requires real-time high-speed data exchange.

**Keywords:** Small Form Factor · FPGA · IoT device

## 1 Introduction

With IoT devices being used more and more widely in life leading to attracting more research from researchers on this issue such as: price reduction of IoT equipment [1]; wireless communication for IoT device [2]; energy saving for IoT device [3]. However, according to [4], the increasing number of IoT devices leads to more risks of data loss, information theft, ... Many security measures have been introduced such as encryption of transmission lines [5]; provide specific protocol standards for IoT devices [6]; ... However, there is an inherent weakness in IoT devices that is the low data transmission-receiving speed that leads to frequent IoT devices subject to denial of service (DDoS) attacks [7]. There have been many remedial measures by increasing the bandwidth of IoT devices such as using wireless communications such as wifi, zigbee, ... but only transmitting over short distances. In this paper, a method using the Small Form Factor (SFP module) on FPGA platform is introduced to increase speed for IoT devices.

## 2 Small Form Factor

With the goal of transmitting and receiving data of IoT devices greater than 1 Gbps, the team used the transmission line as fiber optic cable. The concept of fiber optic cable was mentioned in [8].

In this paper, the authors have presented the advantages of optical cable compared to copper cables for transmitting electrical signals such as: less interference, higher speed, especially can be used to transmit signals over very long distances.

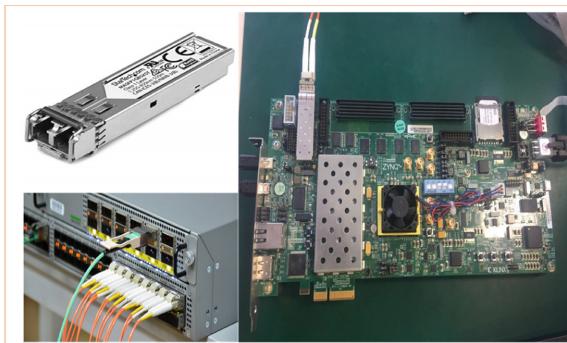
Fiber optic cable consists of a central conductor of glass fiber or refined plastic to allow maximum transmission of light signals. Optical fibers are coated with a lining to reflect the signals well.

Optical cable includes the following parts:

- Core: The reflective center of the fiber where the light goes.
- Cladding: External optical material encapsulates the core that reflects light back into the core.
- Buffer coating: The flexible outer coating protects the fibers from damage and moisture.
- Jacket: Hundreds or thousands of fibers are placed inside called fiber bundle. These bundles are protected by a cable coating called a jacket.

To ensure high-speed transmission from the physical downlink, devices need SFP Modules to process and control the transmitted information.

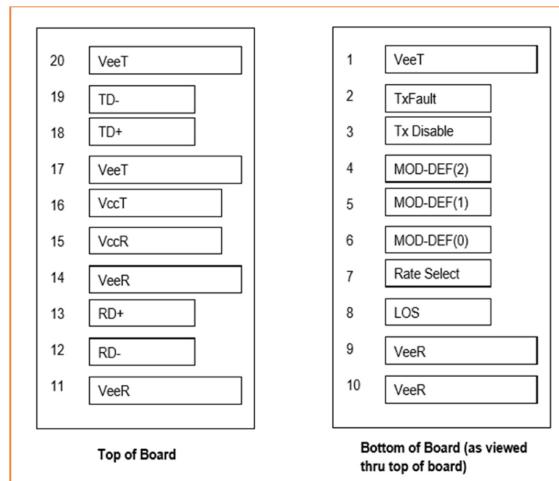
Small Form Factor (SFP) is a compact transceiver device, used for telecommunications and data communication. One end of the SFP attaches to devices such as switches, routers, media converters or similar devices; the other end is used to attach optical or copper cables. It is a popular standard protocol supported by telecommunications and networking equipment manufacturers (Fig. 1).



**Fig. 1.** Small Form Factor connect board FPGA

IOT and switches will connect to each other via Small Form Factor, the optical cable will act as a physical connection while the SFP Module acts as the speed control, handshake devices together and most importantly convert data. data from numbers to electricity [8]. With highly configurable customization, the SFP Module will have registers for the user to set the required parameters for data transmission. For software, when configuring on PuTTY devices such as switches, servers, ... will automatically install the programs for the Small Form Factor register, but for hardware devices such as FPGA board, 1 core is required to establish values to the registers.

### 3 Install Registers in the Driver to Control the Transmit and Receive Modes in Small Form Factor



**Fig. 2.** Small Form Factor Pins

Figure 2 is the pinout of the SFP Module, in which:

- + TD-, TD + are data transfer pins to optical ports.
- + RD-, RD + are pins that receive data to the optical port.
- LOS is the signal loss on the transmission line.
- + VeeR is the source pin of the receiving mode.
- + VccR is the ground pin of transmission mode.
- + VeeT ground pin of transmission mode.
- + VeeT source pin of receiver mode.
- + Tx Fault is the data pin of transmission error.
- + Tx Disable is the pin that allows data to be transmitted or not.
- + Rate Select is to adjust the bandwidth of data acquisition.
- + MOD-DEF (1), (2), (3) are data transmission pins according to standard into the setting registers.

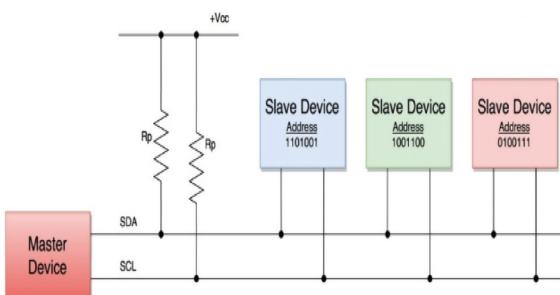
The Table 1 shows the values of the registers that need to be entered for the SFP Module to understand and control the transmission of data on the optical port. The user needs to know what the product is for to set the appropriate values. Here the team used a baud rate of 1000 BASE-T, the distance was long distance (L), the transmission mode was single mode, etc. [9].

**Table 1.** Registers in Small Form Factor

Data Address	Bit	Description	
<b>Identifier</b>			
0	7-0	SFP Transceiver (03h)	Fibre Channel transmitter technology
<b>Connector</b>			
2	7-0	SC (01h)	8 Shortwave laser w/o OFC (SN)
<b>Transceiver</b>			
SONET Compliance Codes			
5	4	OC 12 multi-mode short reach	Multi-mode, 50 m (M5)
Gigabit Ethernet Compliance Codes			
6	3	1000BASE-T	Fibre Channel speed
Fibre Channel link length			
7	4	long distance (L)	200 MBytes/Sec
<b>Encoding</b>			
11	7-0	8B10B(01h)	Encoding
<b>Options</b>			
64-65	7-0	Indicates which optional SFP signals are implemented	Options
<b>Date Code</b>			
84-91	7-0	Vendor's manufacturing date code	Date Code

## 4 Communication Standard in Small Form Factor

To configure the registers in the driver, the FPGA board is required to communicate with the SFP module according to the communication standard  $I^2C$  (Inter-Integrated Circuit).  $I^2C$  is a communication standard in a very popular embedded system. The communication standard  $I^2C$  connect between master (or multimaster) with single slave (or multi slave). Here the FPGA board acts as the Master and the SFP Module is the Slave. One of the outstanding advantages of the communication standard is that it only has 2 wires SCL (serial clock) and SDA (serial data) to transmit signals [10] (Fig. 3).

**Fig. 3.** Device model according to protocol

The working principle of the communication standard  $I^2C$  is as follows:

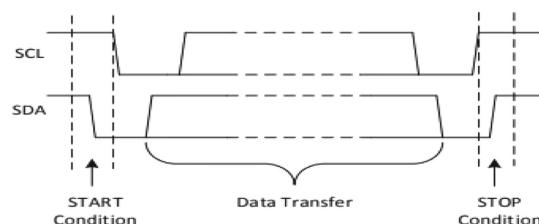
+ The Master device sends the data to the Slave device: Master is the transmitter. The START bit must be transmitted first. Second, the address of the Slave is sent. Finally, the Master transmits the STOP bit to terminate the sending.

+ The Master device sends the data to the Slave device: Master is the receiver (receiver IC). The START bit must be transmitted first. Second, the address of the Slave (transmitter). Then Master (receiver) sends the requested register to reads from the transmitter and receives data from the transmitted slave, and finally the master sends the STOP bit to terminate the sending.

Here the START bit and the STOP bit are defined as follows:

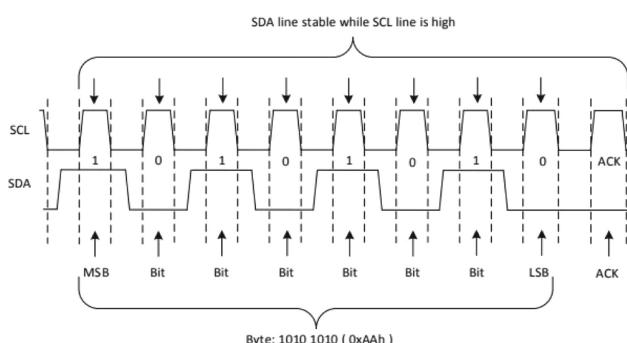
+ The START bit is set to equal 1 when the SDA wire is transferred from the active-high to the active-low, while the SCL bit is the active-high.

+ The STOP bit is set to equal 1 when the SDA wire is transferred from the active-low to the active-high, while the SCL bit is the active-high (Fig. 4).



**Fig. 4.** Condition to logic level “1” of the bit Start, Stop

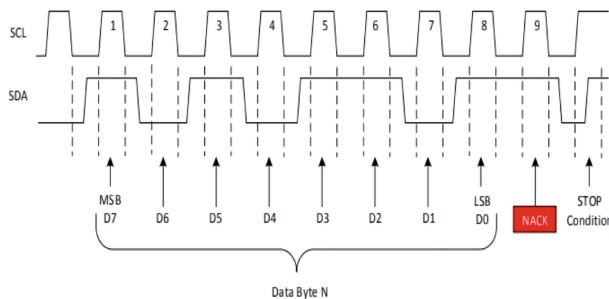
During transmission, each SCL pulse will transmit 1 bit of data correspondingly. Each transmission on the SDA will transmit 1 byte. One byte here can be IC’s address, register address, transmit and receive data. Data is transmitted with the First Most Significant Bit (MSB) (Fig. 5).



**Fig. 5.** The process of transmitting 1 byte of data according to standard

An ACK bit is sent immediately after each data byte from the re-triever. When the ACK bit is transmitted, it means that the byte has been received completely and that the next byte can be sent.

Before sending the ACK bit, the sender must release the SDA pin. The ACK bit is transmitted when the receiver must pull the SDA pin low while the SCL is high. If the case the SDA pin is still high, the NACK bit will appear (Fig. 6).



**Fig. 6.** The NACK bit indicates a transmission failure

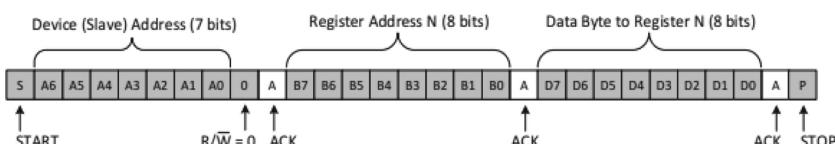
The NACK bit appears when:

- + The receiver is not ready to communicate with the Master.
- + When the data transmission is being taken, the receiver receives data or commands that it does not understand.
- + When the data transmission is being taken, the receiver cannot receive or transmit any more data.
- + One Master terminates the process of receiving data and notifies the slave via the NACK bit.

To write data via the bus, the Master must send the START bit over the bus and the address of the Slave and the last bit ( $R/W = 0$ ) denotes the writing. The Slave then sends the ACK to confirm the connection between the Master to the Slave. Then the Master will send the register address (Register Adress) and data to the register (Data byte to Register). To end data transmission Master sends STOP signal (Fig. 7).



#### Write to One Register in a Device



**Fig. 7.** Cycle to write data to Small Form Factor.

Reading data from a Slave is quite like writing data. After performing the same steps as reading the data above, the Master sends one more START bit (Repeated Start) followed by the address of the Slave and the bit ( $R/W = 1$ ) denotes the reading. The Master immediately released the SDA pin but continued to pulse the SCL pin, while the Master became the receiver, and the Slave became the transmitter.

Each time the Master receives 1 byte of data, it sends the ACK bit to the Slave to indicate that it wants to receive more data. Once the Master receives the required amount of data, it sends the NACK bit to stop the communication. Then comes the STOP bit (Fig. 8).

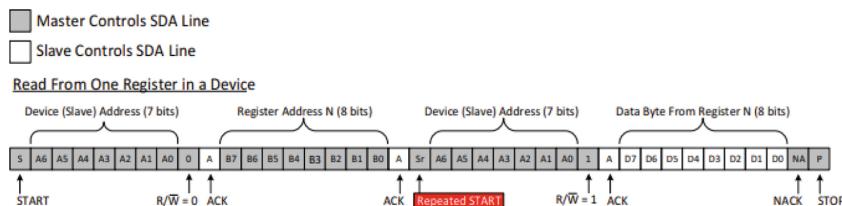


Fig. 8. Read cycle of register data for SFP Module

## 5 Simulation Results and Discussion

The figure is the simulation result of communication between FPGA Board and Small Form Factor; Small Form Factor txp-txn pins, Small Form Factor rxp-rxn pins in the FPGA board (the large SCADA system that require real-time high-speed data exchange) and the Small Form Factor interface according to  $I^2C$  communication standard. After the reset signal (sys\_rst) drops to “0”, the FPGA board will write the value 07 h to the 06 h register of the SFP Module whose address is “1100101”. That is, the group will select 1000 BASE-T mode. The bits START, WRITE is high up during data transmission. Due to the simulation on Vivado software, the ack signal sent from the SFP module will not be determined (red line) (Figs. 9, 10, 11 and 12).

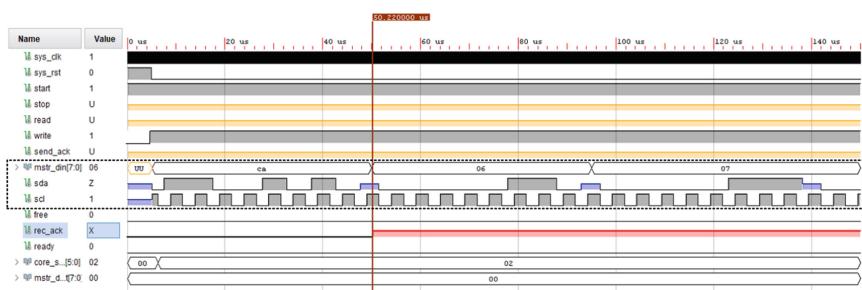
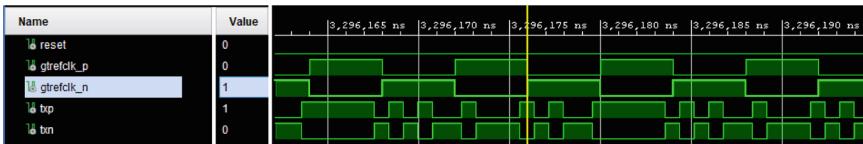


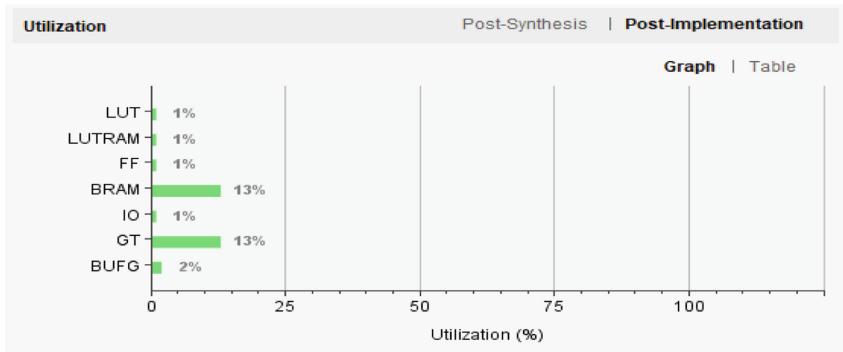
Fig. 9. Results of simulation of communication between FPGA Board and Small Form Factor.



**Fig. 10.** Results of simulation of Small Form Factor txp-txn pins in FPGA Board.



**Fig. 11.** Results of simulation of Small Form Factor rxp-rxn pins in FPGA Board.



**Fig. 12.** Resources for communicating with Small Form Factor

## 6 Conclusions

By configuring Small Form Factor according to the industrial communication standard on the FPGA board, the data transmission and reception speed of IoT devices has been significantly improved ( $>1000$  Mbs). The above study can be widely applied to IoT devices that need to exchange large data, especially in the large SCADA system that require real-time high-speed data exchange.

## References

1. Miguez, M., Marioni, M., Ortiz, M., Vogel, G., Arnaud, A.: An IoT-based electronic price-tag for food retail. In: 2019 26th IEEE International Conference on Electronics, Circuits and Systems (ICECS) (2019)
2. Shafiq, M., Tian, Z., Bashir, A.K., Du, X., Guizani, M.: CorrAUC: a malicious Bot-IoT traffic detection method in IoT network using machine-learning techniques. IEEE Internet Things J. **8**(5), 3242–3254 (2021)

3. Ku, T.-Y., Park, W.-K., Choi, H.: IoT energy management platform for microgrid. In: 2017 IEEE 7th International Conference on Power and Energy Systems (ICPES) (2017)
4. Garg, H., Dave, M.: Securing IoT devices and securely connecting the dots using REST API and middleware. In: 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU) (2019)
5. Bhayo, J., Hameed, S., Shah, S.A.: An efficient counter-based DDoS attack detection framework leveraging software defined IoT (SD-IoT). *IEEE Access* **8**, 221612–221631 (2020)
6. Tao, M., Hong, X., Qu, C., Zhang, J., Wei, W.: Fast access for ZigBee-enabled IoT devices using raspberry Pi. In: 2018 Chinese Control and Decision Conference (CCDC) (2018)
7. Kolias, C., Kambourakis, G., Stavrou, A., Voas, J.: DDoS in the IoT: Mirai and other botnets. *Computer* **50**(7), 80–84 (2017)
8. Charbonneau-Lefort, M., Yadlowsky, M.J.: Optical cables for consumer applications. *J. Lightwave Technol.* **33**(4), 872–877 (2015)
9. Cooperation Agreement for Small Form-Factor Pluggable Transceivers. <http://www.schelto.com/sfp/sfp%20msa%20091400.htm>
10. Son, H.M.: Industrial communication network. Science and Technology Publishing House, Hanoi (2004)



# The High-Speed Data Transmission System on Fiber Optic Cable for IoT Equipment Using FPGA Platform

Tuan Nguyen Ngoc<sup>1</sup>, Dat Tran Thanh<sup>2</sup> , Thang Le Dinh<sup>2</sup>, Tinh Tran Xuan<sup>3</sup>,  
and Tuyen Tran Van<sup>1</sup>

<sup>1</sup> Military Technical Academy of Vietnam, Hanoi, Vietnam

<sup>2</sup> Hanoi University of Science and Technology, Hanoi, Vietnam  
[tranhanhdat1506@gmail.com](mailto:tranhanhdat1506@gmail.com)

<sup>3</sup> Air Defence - Air Force Academy of Vietnam, Hanoi, Vietnam

**Abstract.** With the development of IoT devices there is a need to exchange data in large numbers on SCADA systems (SCADA - Supervisory Control And Data Acquisition). However, with the real-time response of the SCADA system, communication equipment is very expensive. This paper proposes a method to use FPGA (Field Programmable Gate Array) technology to transmit high-speed real-time data on optical fibers to help SCADA system exchange data at high speed, reducing costs. .

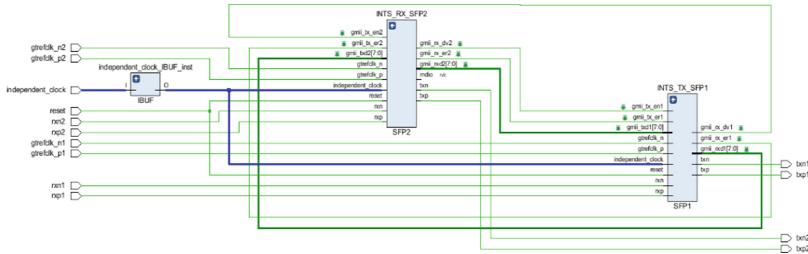
**Keywords:** FPGA · IoT device · SCADA system · Fiber optic cable

## 1 Introduction

With IoT devices being used more and more widely in life, leading to the attraction of many researchers on this issue such as: lowering the price of IoT equipment [1]; wireless communication for IoT devices [2]; saving energy for IoT device [3]. However, according to [4], the increasing number of IoT devices leads to more risk of data loss, information theft, ... Many security measures have been introduced such as encrypted transmission [5]; provides protocol standards specific to IoT devices [6]; However, there is an inherent weakness in IoT devices that is the low data transmission and reception speed, which leads to frequent denial of service (DDoS) attacks by the IoT device [7]. There have been many remedial measures by increasing the bandwidth of IoT devices such as using wireless media such as wifi, zigbee, etc. but only transmitting over short distances. In addition, IoT-type devices with high data transmission-receiving rates are often very high due to the need to satisfy real-time requirements [8]. In this paper, we propose a High-speed data transmission System on fiber optic cable for IoT equipment using FPGA platform.

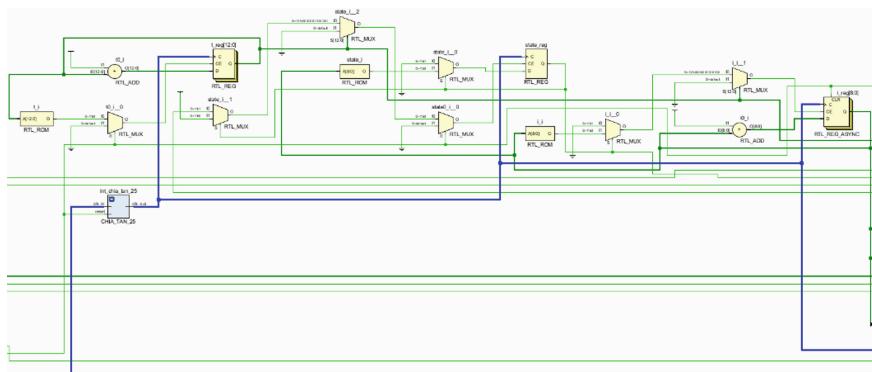
## 2 Clock Synchronization

With the high-speed data transmission system on fiber optic cable for IoT equipment uses FPGA platform, clock synchronization is a prerequisite for functional cores to understand each other because each core uses a separate clock. Running synchronous will help the system to be able to shake hands with other devices in the computer system such as Switches, computers, ... and help reduce delay time, response time of packets, and increase speed line [9]. The system needs to synchronize 250 MHz, 125 MHz, 25 MHz.



**Fig. 1.** Independent clock in high-speed data transmission system

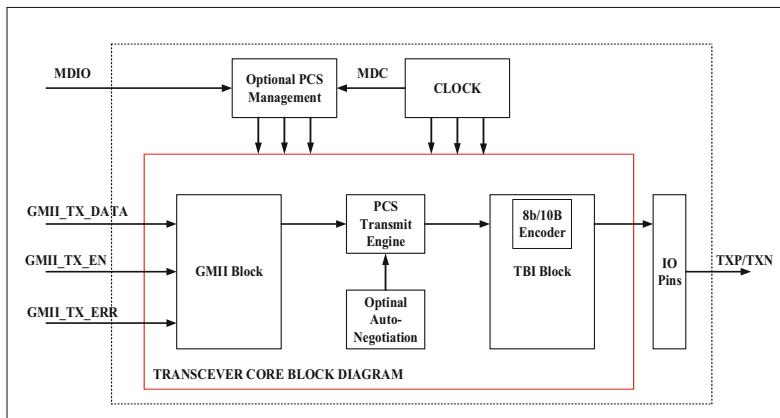
Figures 1, 2 shows a clock synchronization circuit of the design. By using an independent clock as the model clock and one block as the common clock for the whole design, I have successfully performed clock synchronization, avoiding clock deviation, which leads to delay as well as system is not synchronized. Next, the design team will focus mainly on data processing and data propagation between registers.



**Fig. 2.** Common clock in high-speed data transmission system

### 3 Core Transmission-Method of Transmission

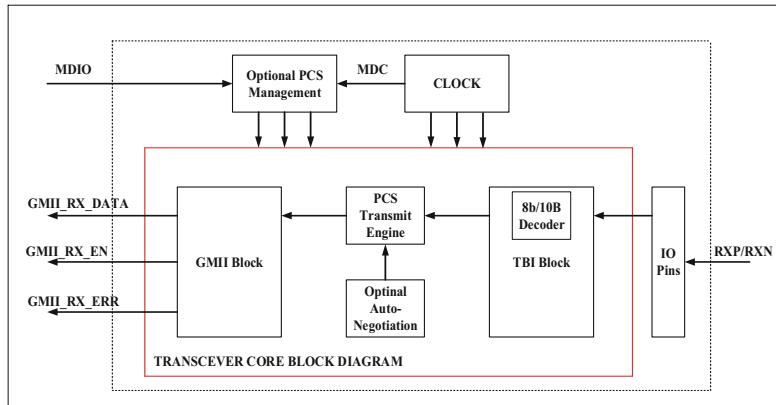
Figure 3 is a diagram of the core block that transmits data with the input data (GMII\_TX\_DATA), the transmission enable signal (GMII\_TX\_EN), the error signal (GMII\_TX\_ERR). At the GMII Block, the data will go through preprocessing, check to ensure that the transmitted packet is full of a data frame structure and correct with protocol standards. Then the data will be transmitted by the internal data transmission block to the encryption block (TBI (ten-bit interface) block). Here data will be encoded as 8 bits into 10 bits to give the transmission line (TXP/TXN). In addition, the transmitting Core also has a Clock block to synchronize signals and a common control block to manage the component cores as well as install registers on Fiber Optic Cable [10].



**Fig. 3.** Core block diagram of data transmission

### 4 Core Receive-Receive Method

In essence, the receiving core will be the opposite of the core that transmits the data. Here the input signal is data transmitted from the fiber optic cable into the on Fiber Optic Cable. The signal is then decoded from 10 bits to 8 bits at the TBI block. Then the signal will go through the preprocessing step before entering the GMII Block to be transported in three ways: the data line (GMII\_RX\_DATA), the allowable signal line (GMII\_RX\_EN) and the error signal line (GMII\_RX\_ERR). In addition, the Clock block is used to synchronize the signals and the common control unit to manage the component cores and install the registers on Fiber Optic Cable (Fig. 4).

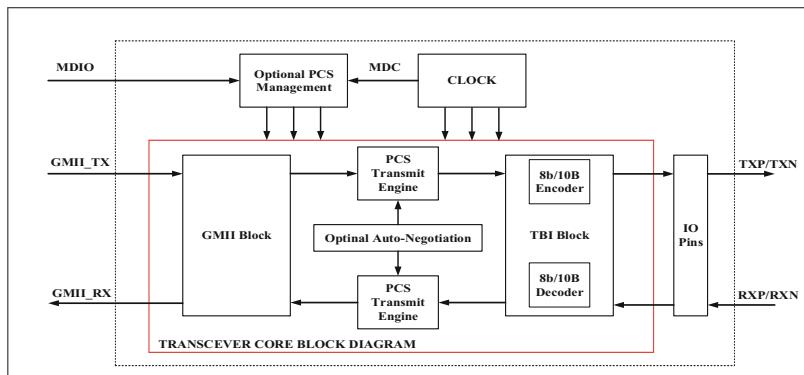


**Fig. 4.** Core block diagram receives data

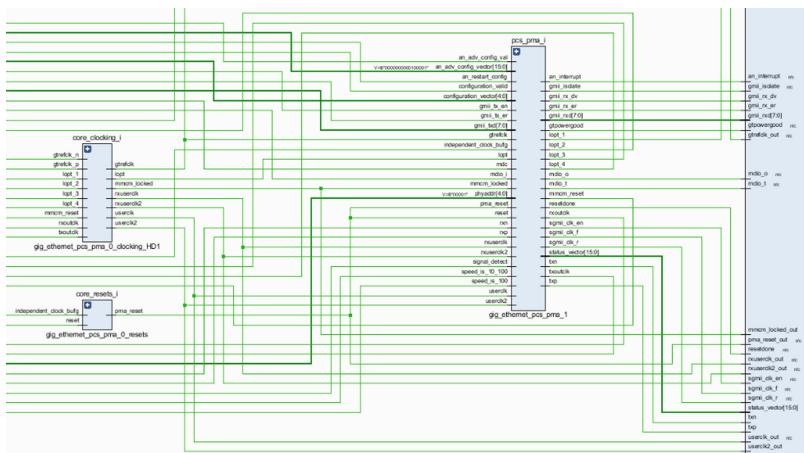
Here, the main function of the TBI unit is to encode the data moving from 8 bits to 10 bits in the sending direction while the data encoding moves from 10 bits to 8 bits in the corresponding receiving direction. This is aimed at complying with the rules of data transmission on optical cable.

## 5 Simulation Results and Discussion

The Figs. 5, 6 below shows the block diagram and principal diagram of Core Trasmit Data on Vivado software. In addition to the previously designed GMII and TBI blocks, the team added configuration emulators for Fiber Optic Cable, the block that creates data frames for transmission and retrieval, and the TBI (ten-bit interface) unit that decodes and encodes the data received and transmitted respectively. (PCS-physical coding sublayer)



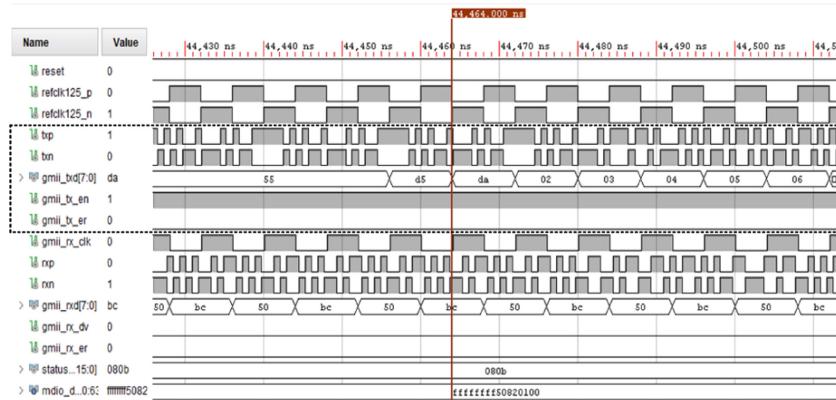
**Fig. 5.** Block diagram of high-speed data transmission system



**Fig. 6.** Principle of high-speed data transmission system on Vivado

With the reset logic low, the data transmission unit operates. With the device speed set to 1Gbits, the clock input will be 125 MHz. The picture is the simulation result of the data transmission-receiving unit on Vivado.

At the transmission core, the inputs include transmission data (gmii\_txd), transmit enable signal (gmii\_tx\_en), error signal (gmii\_tx\_er). Output the encoded data (txp(txn)). During transmission the allow signal (gmii\_tx\_en) is always 1, each clock edge sends one byte of data, which corresponds to 10 bits encoded at the output (txp(txn)) (Fig. 7).



**Fig. 7.** Results of simulating data transmission on Vivado

At the receiving core, the input is the encoded data (txp/rxn) transmitted on Fiber Optic Cable. Outputs include data price (gmii\_rx\_d), transmit enable signal (gmii\_rx\_dv), error signal (gmii\_rx\_er). During receiving the allow signal (gmii\_tx\_en) always equal to 1, each clock deceleration will receive 10 encoded bits (txp/txn), which will be one

byte of received data (gmii\_rxd). When the process is over, the process of receiving the enable signal falls to 0 and returns to the standby state (Fig. 8).

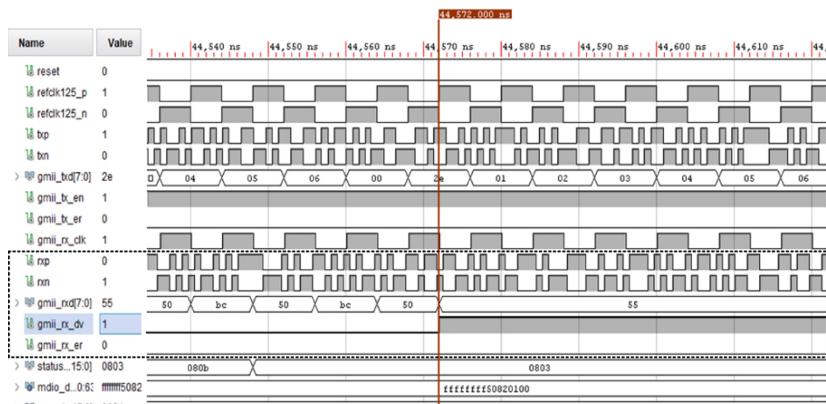


Fig. 8. Results simulating the process of receiving data on Vivado

The authors tested on the Xcku040-fbva676-1-c FPGA Board with 2 modes: 1000 Mbs and 2500 Mbs. Figure 9 is the test result of the high-speed data transmission system in the 1000 Mbs mode with the Sprient flow injection device (Fig. 10).

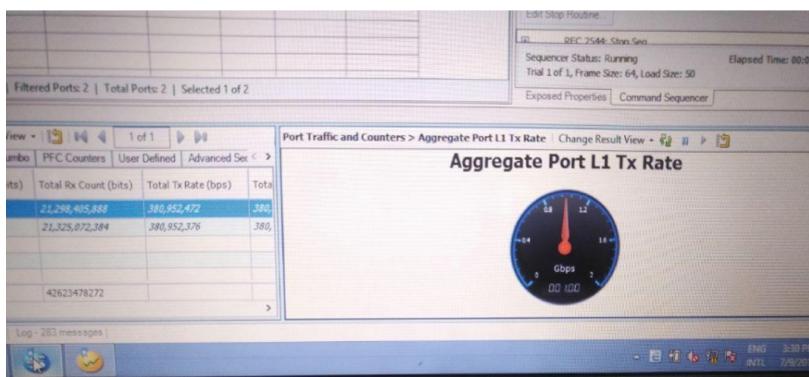
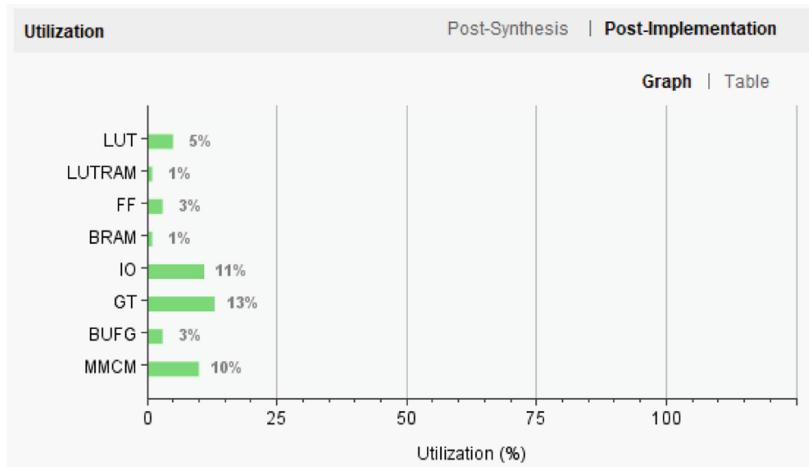


Fig. 9. Results with the 1 Gbs speed of the high-speed data transmission system tested with the Sprient flow injection device



**Fig. 10.** Resources for communicating of the high-speed data transmission system

## 6 Simulation Results and Discussion

With the High-speed Data Transmission System on Fiber Optic Cable for IoT Equipment Using FPGA platform, the data transmission and reception speed of IoT devices has been significantly improved (1000 Mbs or 2500 Mbs). The above study can be widely applied to IoT devices that need to exchange large data, especially in the large SCADA system that require real-time high-speed data exchange. By applying a clock-based stream processing algorithm, the system can achieve a speed of 1000 Mbs or 2500 Mbs. Therefore, the system can completely replace the high-speed Switches in the SCADA system, minimizing equipment investment costs.

## References

1. Miguez, M., Marioni, M., Ortiz, M., Vogel, G., Arnaud, A.: An IoT-based electronic price-tag for food retail. In: 2019 26th IEEE International Conference on Electronics, Circuits and Systems (ICECS) (2019)
2. Ohori, F., Itaya, S., Maruhashi, K., Kojima, F.: Subdividing one channel of 5GHz wireless LAN into narrow channels for factory IoT. In: 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU) (2019)
3. Ku, T.-Y.; Park, W.-K., Choi, H.: IoT energy management platform for microgrid. In: 2017 IEEE 7th International Conference on Power and Energy Systems (ICPES) (2017)
4. Garg, H., Dave, M.: Securing IoT devices and securely connecting the dots using REST API and middleware. In: 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU) (2019)
5. Shen, M., Tang, X., Zhu, L., Du, X., Guizani, M.: Privacy-preserving support vector machine training over blockchain-based encrypted IoT data in smart cities. *IEEE Internet Things J.* **6**(5), 7702–7712 (2019)
6. Tao, M., Hong, X., Qu, C., Zhang, J., Wei, W.: Fast access for ZigBee-enabled IoT devices using raspberry Pi. In: 2018 Chinese Control and Decision Conference (CCDC) (2018)

7. Kolias, C., Kambourakis, G., Stavrou, A., Voas, J.: DDoS in the IoT: mirai and other botnets. *Computer* **50**(7), 80–84 (2017)
8. Charbonneau-Lefort, M., Yadlowsky, M.J.: Optical cables for consumer applications. *J. Light. Technol.* **33**(4), 872–877 (2015)
9. XILINX: Clocking wizard v6.0. LogiCORE IP product guide, Vivado design suite, PG065, 5 Feb 2020
10. XILINX: Tri-mode ethernetMAC v9.0. LogiCORE IP product guide, Vivado design suite, PG051, 8 July 2020



# Design LMI Sliding Mode Control for the Anti-aircraft Gun Speed Control Loop

Tuan Nguyen Ngoc<sup>1</sup>, Trung Dang Tien<sup>2</sup>, Tinh Tran Xuan<sup>3</sup>✉, and Tuyen Tran Van<sup>1</sup>

<sup>1</sup> Military Technical Academy of Vietnam, Hanoi, Vietnam

<sup>2</sup> Electric Power University, Hanoi, Vietnam

<sup>3</sup> Air Defence - Air Force Academy of Vietnam, Hanoi, Vietnam  
tinhpk79@gmail.com

**Abstract.** This paper presents the results of designing a finite-time sliding controller for the anti-aircraft gun transmission speed control loop. The test results by simulation and experiment show that this controller ensures the tracking accuracy; stabilize cannon speed; can reduce oscillations under system conditions affected by nonlinear factors due to mechanical structure and impulse when firing.

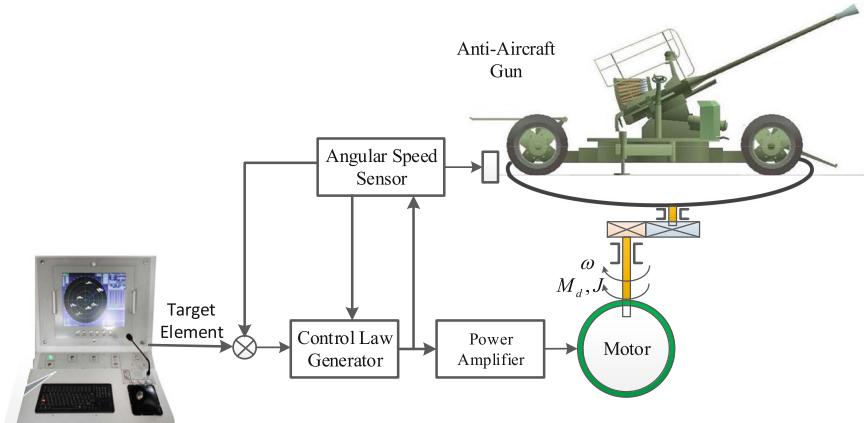
**Keywords:** Anti-aircraft · Sliding mode control · Nonlinear systems

## 1 Introduction

Currently, for modern warfare, anti-aircraft artillery is still an indispensable part of creating an effective low-altitude anti-aircraft fire network. The main research direction of the paper is to focus on overcoming the disadvantages of the existing anti-aircraft artillery controller, that is, fast acting, stability of rotational speed in the presence of uncertain noise. To meet that need a controller with high persistence. The survey shows that the sliding controller gives good results even when there are nonlinear noises [2–4]. Sliding controllers have gained attention in recent years for their ability to deal with noise and uncertainty. It can be seen that the sliding control is stable with the ability to reduce the influence of noise through the selection of the sliding surface as well as the corresponding controller. However, the disadvantage of sliding control is chattering. To reduce the effect of Chattering effect, recently higher order sliding controller has been used [5–8]. The advantages obtained of the terminal sliding mode control (TSMC) considered in [9] guarantee finite time stability, reduced steady-state error, and dynamic response. fast. The linear matrix inequality (LMI) sliding mode controller proposed in [10, 11] ensures convergence with finite time not only for the sliding variable but also for the tracking error.

## 2 Dynamic Model and Problem Statement

According to [1] there is a functional block diagram of the anti-aircraft artillery drive system (Fig. 1).



**Fig. 1.** Functional block diagram of anti-aircraft artillery angle drive system

The equation describing the motion of the control object has the form:

$$T = J_{eq} \frac{d\omega}{dt} + T_L \quad (1)$$

where  $\omega$  is the rotor angular speed;  $T$  is the moment of Motors (N.m);  $T_L$  is load torque (N.m);  $J_{eq}$  are the inertia moment of Motors coefficient ( $\text{Kg} \cdot \text{m}^2$ ).

The friction force on the bearing and the rotating part can be converted to the resisting moment acting on the load-side rotating shaft:

$$T_f = (T_{f0} \text{sign}\phi + T_{f1} \frac{d\phi}{dt}) \quad (2)$$

where

$T_{f0}$  - The coefficient of static friction, the sign of which depends on the direction of motion;

$T_{f1}$  - The coefficient of dynamic friction depends on the speed of motion;

$\phi$ - Motor shaft rotation angle.

If the drive system is applied on weapon systems, then when firing will generate impulses (reactions) acting on the rotating part. This component is random, mutable, and has a limited value, denoted  $T_i$ . Thus, the components impeding the motion of the angular mechanism are:

$$T_L = T_f + T_i \quad (3)$$

The electromagnetic torque equation generated by an independently excitation motor has the form:

$$T = K_i i_r \quad (4)$$

With voltage  $u$  applied to both ends of armature winding, we have rotor voltage balance equation:

$$L_r \frac{di_r}{dt} = -R_r i_r - K_b \omega + u \quad (5)$$

The state variable:

$$\underline{x} = [x_1 \ x_2]^T = [\omega \ i_r]^T$$

According to [2], there is a system of equations of state of the transmission system:

$$\begin{cases} \dot{x}_1 = -\frac{1}{J_{eq}} T_f x_1 + \frac{K_i}{J_{eq}} x_2 - \frac{1}{J_{eq}} (T_f 0 sign\phi - T_i) \\ \dot{x}_2 = -\frac{K_b}{L_r} x_1 - \frac{R_r}{L_r} x_2 + \frac{1}{L_r} u \end{cases} \quad (6)$$

### 3 LMI Sliding Mode Control Design

The system of Eqs. (6) returns to the equation of state of the general form:

$$\dot{\underline{x}} = A \cdot \underline{x} + B \cdot u + F(\underline{x}, t) \quad (7)$$

With

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} -\frac{1}{J_{eq}} T_f 1 & \frac{K_i}{J_{eq}} \\ -\frac{K_b}{L_r} & -\frac{R_r}{L_r} \end{bmatrix}; B = [B_{11} \ B_{12}]^T = \left[ 0 \ \frac{1}{L_r} \right]^T;$$

$$F(\underline{x}, t) = \left[ -\frac{1}{J_{eq}} (T_f 0 sign\phi - T_i) \ 0 \right]$$

Select the surface:

$$s(e) = \Lambda e \quad (8)$$

where

$$\Lambda = [F, g]; e = [e_1 \ e_2]^T; \underline{e} = \underline{x}^{ref} - \underline{x} \quad (9)$$

$\underline{x}^{ref}$  is desired value

$F$  is the gain vector,  $g$  is a constant.

When the system satisfies the sliding condition  $s(e) = 0$  we have

$$e_2 = -g^{-1} F e_1 \quad (10)$$

According to (7), (9), (10) we have the error of the system

$$\dot{e}_1 = (A_{11} - A_{12} g^{-1} F) e_1 + A_{11} x_{m1} + A_{12} x_{m2} - \dot{x}_{m1} \quad (11)$$

Assume that the control law  $u_m$  guarantees for  $\dot{x}_{m1} = A_{11} x_{m1} + A_{12} x_{m2}$ .

From (7) there is the error of the closed system:

$$\dot{e}_1 = (A_{11} - A_{12} g^{-1} F) e_1 \quad (12)$$

Use the LMI linear matrix inequality to design the sliding surface. According to [2, 10] with  $g > 0$  and matrices  $X, Y, W > 0$ , if the LMI inequality is satisfied, the system will be stable:

$$\begin{pmatrix} A_{11}X - A_{12}Y + XA_{11} - Y^T A_{12} & X \\ X & -W \end{pmatrix} < 0 \quad (13)$$

With  $P$  being a positive definite matrix  $P = X^{-1}$ ;  $F = gYX^{-1}$ , the error of the system will approach close to zero.

**Proof:** Based on the Lyapunov candidate function of the form:

$$V_1(e_1) = e_1^T P e_1 \quad (14)$$

According to formula (11):

$$\begin{aligned} \dot{V}_1(e_1) &= e_1^T P \dot{e}_1 + \dot{e}_1^T P e_1 \\ &= e_1^T P \left( A_{11} - A_{12}g^{-1}F \right) e_1 + e_1^T \left( A_{11} - A_{12}g^{-1}F \right)^T P e_1 \end{aligned} \quad (15)$$

Assume the following inequality is satisfied:

$$\begin{aligned} P \left( A_{11} - A_{12}g^{-1}F \right) + \left( A_{11} - A_{12}g^{-1}F \right)^T P &\leq -W^{-1} \\ \Rightarrow \dot{V}_1(e_1) &\leq -e_1^T W^{-1} e_1 \leq -\lambda_{\min}(W^{-1}) \|e_1\|^2 \Rightarrow \dot{V}_1(e_1) \leq -\alpha_1 V_1(e_1) \end{aligned}$$

where  $\alpha_1 = \frac{\lambda_{\min}(W^{-1})}{\lambda_{\max}(P)}$ .

With  $X = P^{-1}$  we have:

$$A_{11}X - A_{12}g^{-1}FX + XA_{11} - \left( A_{12}g^{-1}FX \right)^T \leq -XW^{-1}X$$

When  $Y = g^{-1}FX$  the condition of the LMI inequality is proved.

From (10) it can be concluded that when the sliding mode  $s(e) = 0$  is satisfied, then  $e_2(t)$  will also converge to the asymptote to 0. Choose a control law so that the state error approaches the face slide in finite time.

$$u = -(gB)^{-1}(\Lambda Ax - \Lambda \dot{x}_m + Q \operatorname{sgn}(s) + \gamma s + \sigma \operatorname{sgn}(s)|s|^\eta) \quad (16)$$

where  $\sigma$  and  $\gamma$  are positive constants.  $Q$  is the vector of upper bounds of  $f(\underline{x}, t)$ .

**Proof:** Choose a positive definite Lyapunov function.

$$V_2(s) = \frac{1}{2} s^T s \quad (17)$$

Derivative (17), through the transformation we have:

$$\dot{V}_2(s) = s^T \dot{s} = s^T (\Lambda Ax + gBu + gf(x, t) - \Lambda \dot{x}_m) \quad (18)$$

Replace (16) into (18)

$$\begin{aligned}\dot{V}_2(s) &= s^T \dot{s} = -s^T \sigma \operatorname{sgn}(s) |s|^\eta - s^T \gamma s - s^T Q \operatorname{sgn}(s) + s^T g f(x, t) \\ \Rightarrow \dot{V}_2(s) &\leq -\lambda_{\min}(\gamma)(\|s\|)^2 - \lambda_{\min}(\sigma)(\|s\|)^{\eta+1} = -\alpha V_2(s) - \beta (V_2(s))^{\eta_2} < 0\end{aligned}$$

where  $\alpha = 2\lambda_{\min}(\gamma) > 0$ ,  $\beta = 2^{(\eta+1)/2}\lambda_{\min}(\sigma) > 0$ .

According to [1, 5] the state of the system will advance towards the sliding surface in finite time:

$$t_r = \frac{1}{\alpha(1-\eta_2)} \ln \left( \frac{\alpha V_2(s(e(t_0)))^{1-\eta_2} + \beta}{\beta} \right) \quad (19)$$

So according to standard Lyapunov stable system.

## 4 Simulation, Experiment and Results Assessment

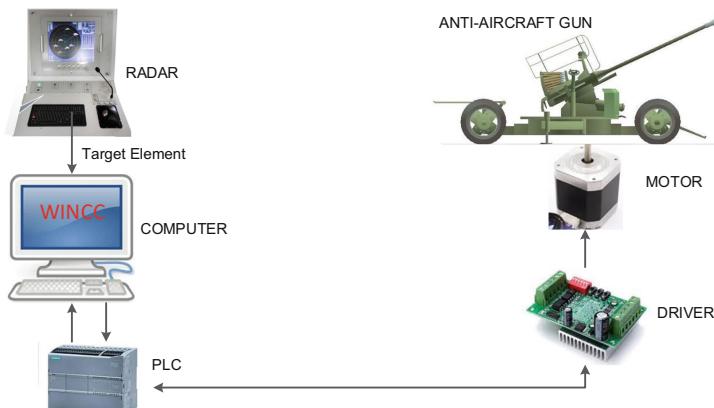
In this section, we consider several simulation results to demonstrate the effectiveness of the proposed LMI sliding mode control law based on the anti-aircraft system. Simulation parameters are as follows:  $R_r = 5 \Omega$ ,  $L_r = 0.2H$ ,  $K_b = 0.1 \text{ V/rad/s}$ ,  $K_i = 0.1 \text{ Nm/A}$ ,  $J_{eq} = 2.10^{-3} \text{ Kg.m}^2$ , gear ratio of the gear box  $\frac{n_1}{n_2} = 1/1076$ . Through calculation, we have the coefficient matrix

$$A = \begin{bmatrix} -6.10^5 & 50 \\ -0.5 & 25 \end{bmatrix}; B = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$$

Solving condition (13) using Matlab's LMI toolbox determined:

$$P = X^{-1} = \begin{bmatrix} 2.1 & -0.17 \\ -1.35 & 2.1 \end{bmatrix}; F = gYX^{-1} = \begin{bmatrix} 10.5 & -0.85 \\ -6.75 & 10.5 \end{bmatrix}$$

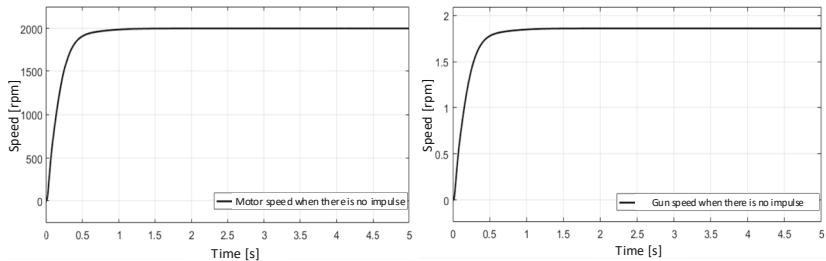
Then determine the control law  $u$ .



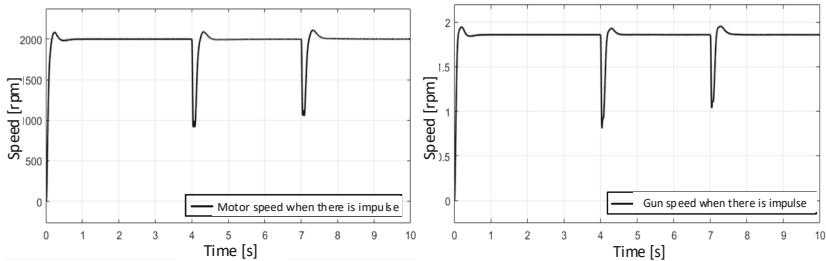
**Fig. 2.** Experimental model of anti-aircraft artillery drive system

Design experimental model of anti-aircraft artillery on the scale of 1/20. The control program is loaded on PLC S7-1200, the control interface is implemented by WinCC software. Parameters of the target are created on the computer, passed through the software to calculate the angle of fire. The firing angle will be passed through the controller at the PLC through the Driver to control the speed and rotation of the actuator motor. The angular speed sensor signal will be fed back to the Encoder (Fig. 2).

Simulation results in Fig. 3 and Fig. 4 describe the speed response of the drive system in the cases where there is no impulse and there is an impulse  $T_i$  when firing.



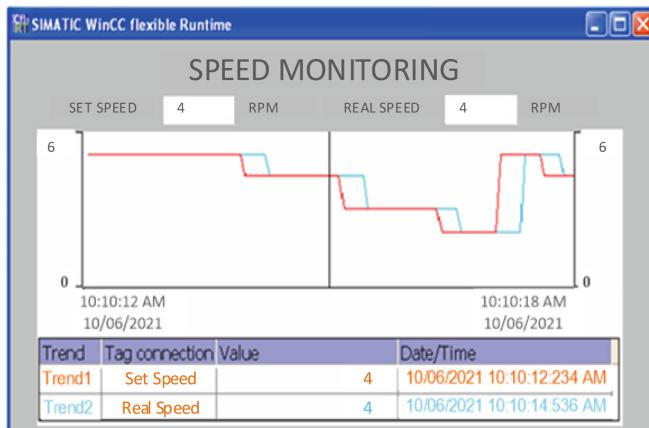
**Fig. 3.** Motor speed and Gun speed when there is not impulse when firing



**Fig. 4.** Motor speed and Gun speed when there is impulse when firing

When there is no impulse when firing, the speed is stable, the transient time is short, the overshoot is zero. In case of impact when firing: at 4 s and 7 s for  $T_i = 1000$  Nm and 900 Nm, the output response of engine speed and turret remains stable after time 0.4 s.

Experimental results when changing the speed set for anti-aircraft guns, the actual speed always reaches the set value with zero error, the transition time is 2 s (Fig. 5).



**Fig. 5.** Set speed and actual speed of anti-aircraft artillery

From the simulation and experimental results, it is found that the LMI slide controller for anti-aircraft artillery drive system under the influence of impulse when firing has good quality, meeting the stability requirements of the system. Under the influence of nonlinear factors such as: frictional moment, moment of inertia on load, impulse when firing, the system still ensures accuracy in both dynamic mode and static mode. Static error always goes to 0 in different cases. Vibrations in the control process are very small, the transient time is short, helping to improve the ability to destroy the target.

## 5 Conclusions

This paper presents the design results of the finite-time sliding controller for the anti-aircraft artillery electromechanical system. the test results show:

- Firstly, the sliding controller increases the quick response of the system, ensuring the overshoot within the allowable threshold.
- Second, reduce errors and oscillations when the system is affected by nonlinear factors.
- The controller can be applied to real equipment to improve the control quality of the anti-aircraft artillery drive system.

## References

1. Duc, T.N., Manh, T.P.: Anti-Aircraft artillery Equipment and Structure, Air Defence - Air Force Academy (2007)
2. Ngoc, T.N., Xuan, T.T.: Synthesis of sliding controller on the basis linear matrix inequality for anti-aircraft artillery drive system. Sci. Technol. J. Hanoi Univ. Ind. **56**(6) (2020)
3. Zhang, L., Wang, Y., Hou, Y.: Fixed-time sliding mode control for uncertain robot manipulators. IEEE Trans. Ind. Inform. **07**, 149750–149763 (2019)

4. Nasirin, A., Nguang, S.K., Swain, A.: Adaptive sliding mode control for a class of MIMO nonlinear systems with uncertainties. *J. Franklin Inst.* **351**, 2048–2061 (2013)
5. Feng, Y., Han, F., Xinghuo, Y.: Chattering free full-order sliding-mode control. *Automatica* **50**, 1310–1314 (2014)
6. El-Sousy, F.F.M.: Adaptive dynamic sliding mode control system using recurrent RBFN for high-performance induction motor servo drive. *IEEE Trans. Ind. Inform.* **09**(4), 1922–1936 (2013)
7. Koliass, C., Kambourakis, G., Stavrou, A., Voas, J.: DDoS in the IoT: Mirai and other botnets. *Computer* **50**(7), 80–84 (2017)
8. Boukattaya, M., Mezghani, N., Danak, T.: Adaptive nonsingular fast terminal siding-mode control for the tracking problem of uncertain dynamic systems. *ISA Trans.* **77**, 1–19 (2018)
9. Corradini, C.: Nonsingular terminal sliding-mode control of nonlinear planar systems with global fixed-time stability guarantees. *Automatica* **95**, 561–565 (2018)
10. Yin, C., Chen, Y., Zhong, S.-m.: LMI based design of a sliding mode controller for a class of uncertain fractional-order nonlinear systems. In: 2013 American Control Conference (ACC), Washington, DC, USA, 17–19 June (2013)
11. Leu, V.Q., Choi, H.H., Jung, J.-W.: LMI-based sliding mode speed tracking control design for surface-mounted permanent magnet synchronous motors. *J. Electr. Eng. Technol.* **7**(4), 513–523 (2012)



# Extraction of Characteristic Breast Dimensions of North Vietnamese Girl-Students Using Random Forest Algorithm

Nguyen Thi Le<sup>1(✉)</sup>, Luu Thi Hong Nhung<sup>2(✉)</sup>, and Nguyen Nhat Trinh<sup>3</sup>

<sup>1</sup> Hanoi University of Industry, Hanoi, Vietnam

le.nguyenth@haui.edu.vn

<sup>2</sup> Hung Yen University of Technology and Education, Hung Yen, Vietnam

luuhongnhung228@gmail.com

<sup>3</sup> Ha Noi University of Science and Technology, Hanoi, Vietnam

trinh.nguyennhat@hust.edu.vn

**Abstract.** Women's breasts are measured with many different dimensions, which are quite diverse and complicated depending on the purpose of use. The aim of this study is to select the typical dimensions of girl-student breasts to obtain important dimensions for breast classification, thereby reducing noise, reducing costs, and contributing to building a basis for improving the sizing system, designing, and choosing the right woman's bras. 460 girl students in Northern Vietnam between the ages of 18 and 24 with a body mass index of 14.5–24.3 kg/m<sup>2</sup> are selected in this study. Their breasts were 3D scanned to identify 3 parameters of the body (height, weight and body mass index) and 18 breast parameters. The Principal Component Analysis method, Random Forest, and Learning Vector Quantization algorithms are applied to extract the characteristic breast dimensions. The results show that the accuracy reaches 0.9615, the sensitivity reaches 0.989 and the specificity reaches 0.977 with the Random Forest method. The most important dimensions for breast classification are breast volume, the difference between bust and underbust circumference, outer breast, bust, upper bust, inner breast, and bust circumference. Height, weight, body mass index, prolapse, and thoracic arch are the least important parameters when classifying girl-student breasts.

**Keywords:** Breast dimensions · Breast classification · Random forest algorithm · Features extraction

## 1 Introduction

The extraction of characteristic breast dimensions is an important basis for classifying, building a sizes system, designing and selecting women's bras. The studies contribute to collecting information to improve the suitability of the bra dimension system with the wearer's body. Many woman's breast dimensions have been measured in anthropometric, categorical, and bra fit studies. Seolyoung Oh et al. studied to determine the breast dimension of 32 Korean women and proposed to determine the bra dimension based

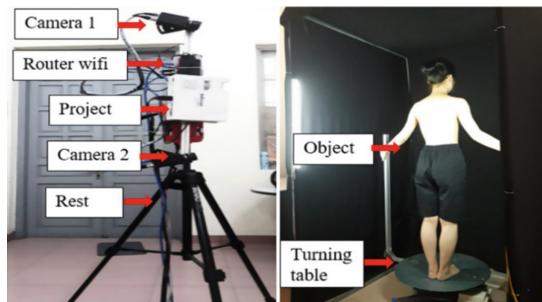
on the breast arc length [1]. Chang-Lixia et al. divided the breast shapes of 270 young women in Western China into 9 clusters [2]. Rong Zheng classified breast shape based on 8 parameters: weight, breast volume, internal breast arch, external breast arch, height, nipple orientation, thoracic slope, and breast shape [3]. Deirdre E. McGhee et al. measured breast volume and assessed the fit of bra dimensions in 104 women based on breast volume [4]. 19 woman's breast dimensions have been studied for breast reconstruction plastic surgery [5]. 75 Shanghai schoolgirls' breasts were 3D scanned and studied to improve breast aesthetics with a proper bra [6]. The change in breast shape of women aged 18–45 years old ( $BMI < 30$ ) was studied based on 66 identified breast measurements from scanning 478 North American women's breasts [7]. Breast classification studies based on dimensions were also performed to improve the fit of women's bras [8]. Jie Zhou et al. also selected the breast dimensions of young women in Northwest China using the Random Forest algorithm [9]. The results showed that with fewer dimensions, the women's breast recognition models are more effective. The contour and bust dimensions are more important than the straight cross-sectional dimensions. The breast positioning dimensions are the least important. Five parameters such as bust, underbust circumference, and bust circumference, and two breast angles of 200 Vietnamese women were measured to determine the body shape index [10].

The above studies were conducted on women from regions with different anthropometric characteristics. There are many features that reflect breast shape characteristics. However, if too many features are selected, the classification process will be disturbed and expensive. Extraction of features to identify important breast dimensions, reduce costs and improve accuracy in breast classification. This study aims to select the featured dimensions of the breasts of North Vietnamese girl-students to contribute to building a basis for the classification and determination of breast morphology, breast plastic surgery, improvement of the sizing system, the designing and selecting women's bras that are suitable for the wearer's body.

## 2 Methods

**Measurement Object:** 460 girl students from 3 universities in Hanoi (Vietnam), aged 18–24, are in normal health,  $BMI$  from  $14.5\text{--}24.3 \text{ kg/m}^2$  were 3D scanned the breast to determine the dimensions. These students came from the provinces of North Vietnam and volunteered to participate in this study. Subjects were currently non-pregnant or non-lactating and had never undergone any breast surgery [11].

**Experiment to Determine Girl Student Breast Dimensions:** The girl students' breasts were scanned with the designed Scan3D MB2019 device using structured light with 2 cameras (Fig. 1) [12]. The girl students only wore pants and leave their breasts bare during the chest scanning. Take measurements in the measuring chamber, standing posture according to ISO 20685.



**Fig. 1.** Designed MB2019 3D scanning measurement system [12]

21 body parameters of girl students including height  $h$  (cm) measured in standing position according to ISO 20685, weight  $w$  (kg), mass index  $BMI$  ( $\text{kg}/\text{m}^2$ ), and 18 breast dimensions were determined by Geomagic Design X software based on scanned data in the study [11]. These dimensions are presented in Table 1.

**Selection of Features of Breast:** Feature extraction is very important in data discovery. This process shows the importance of the breast features to select the best set of features for classifying or predicting models, increasing performance, and shortening training time. Caret and Random Forest Packages in R were applied to the feature extraction on the obtained data of girl-student breast dimensions. The data were firstly probed by Principal Component Analysis (PCA). Then, the process of extracting the characteristic breast dimensions was done through 3 stages: removing the redundant dimensions in the data set, ranking the breast dimensions by importance, and finally selecting the breast dimensions in the data set.

PCA is the technique used when analyzing multidimensional data. PCA reduces the large dimensionality of the data but retains the necessary information of the original dataset. The goal of PCA is to bring data to a new space with fewer dimensions without affecting the accuracy of the clustering and forecasting models that are subsequently established on that dataset. In which, the coordinate axes in the new space are built so that on each axis, the variability of the data is explained as the largest possible (maximize the variability) [13]. The essence of PCA is to construct new principal components by linearly combining the original variables whose principal components are not linearly correlated with each other.

The removal of excess dimensions in the female bust dimension dataset was based on the pairwise linear correlation coefficient of the dimension pairs. A pairwise linear correlation coefficient between pairs of dimensions in the dataset was determined. If the data pair has  $r \geq 0.95$ , it were considered to be closely correlated and can remove one of the two dimensions when building classification or forecasting models later [13].

The ranking of breast dimensions according to their importance for data clustering was done by the Random Forest (RF) machine learning algorithm because of its objectivity and has been successfully applied in many fields. From the original data set, a sample  $L$  of  $n$  objects was extracted to calculate the parameters of the RF model. In the next steps, iterative  $b$  times create a sample  $L_b$  that also includes  $n$  objects by resampling with the replacement of objects in the original sample, then computing the parameters

**Table 1.** Breast dimensions determined based on 3D scanning data

Dimension	Description	Dimension	Description
Upper-bust girth (vnt)	The circumference of the upper breast passes through the point under the right armpit and under the left armpit, lying on the plane that crosses the armpit of the body	Bust girth (vn)	Breast circumference is the line around the body lying on a plane perpendicular to the body axis and passing through the widest point of the bust
Underbust circumference (vcn)	Cross-sectional circumference by a plane perpendicular to the body axis through the lowest point of the thorax	Distance between 2 nipples (cn)	Is the sum of the distances from the right nipple and the left nipple to the midline of the torso
Outer arch of the right thorax (cnnp)	From the outermost point of the right breast to the tip of the right breast, following the curve of the right breast surface	Outer arch of the left breast (cnnt)	From the outermost point of the left breast to the tip of the left breast following the curve of the right breast surface
Inner arch of the right breast (cntp)	From the innermost point of the right breast to the tip of the right breast, following the curve of the surface of the right breast	The inner bow of the left breast (cntt)	From the innermost point of the breast to the tip of the left breast following the curve of the left breast surface
Under right breast arc length (ccnp)	The curved arc from the innermost point of the right breast through the lowest point of the right breast to the outermost point of the right breast	Under left breast arc length (ccnp)	The curved arc from the innermost point of the left breast through the lowest point of the left breast to the outermost point of the left breast
Upper breast height (snt)	The distance from the tip of the breast to the plane that cuts through the top edge of the breast	Lower right breast height (sndp)	Distance from the top of the right breast to the plane that intersects the lowest point of the right breast

(continued)

**Table 1.** (continued)

Dimension	Description	Dimension	Description
Lower left breast height (sndt)	The distance from the left breast tip to the plane that cuts through the lowest point of the left breast	Difference between vn and vcn (cl)	Dimension difference between bust and bust
Sternum to nipple left breast (xut)	Distance from the sternum to the left nipple left breast	Sternum to nipple right breast (xup)	Distance from the sternum to the left nipple right breast
Volume of left breast (ttt)	Volume of left breast	Volume of right breast (ttp)	Volume of right breast

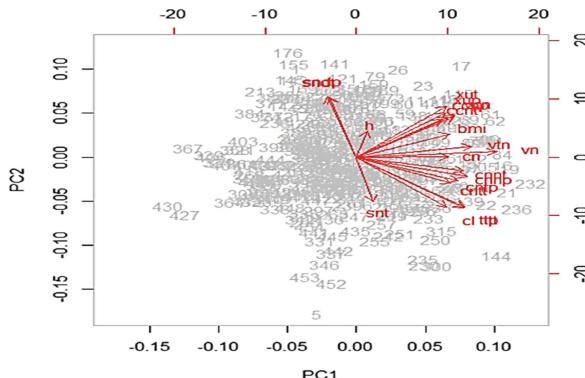
[14]. The estimation of important variables is considered according to two indicators, namely MDA (Mean Decrease Accuracy) or MDG (Mean Decrease Gini) for each variable [15, 16]. Ranking the importance of the dimensions in the dataset was performed using the Learning Vector Quantization (LVQ) method. Extracting the dimensions in the breast data was set by this method and using the Recursive Feature Elimination algorithm including six steps as follows [13]:

- (1) Import and initialize the dataset F.
- (2) Select classifier C.
- (3) Weighting of each breast dimension in dataset F based on the accuracy of categorical prediction.
- (4) Remove the minimum weighted bust dimension of  $f_j$  and update F.
- (5) Repeat steps 3 and 4 until F is only one bust dimension.
- (6) Rank the importance of bust dimensions.

Important bust dimensions for breast clustering are chosen based on this ranking.

### 3 Results and Discussion

**Principal Component Analysis Results:** The results of PCA are shown in Fig. 2 with 2 main components PC1 and PC2. The vectors corresponding to the dimensions show the relationship between the original variables and the principal components. The length of the vector indicates the strength of the correlation between the original variable and the principal component. With longer vectors representing more important features. Vectors showing breast volume ( $ttt$ ,  $ttt$ ), difference  $cl$ , bust  $vn$  have the strongest impact on the principal components. Body height  $h$  has the weakest effect on principal components.



**Fig. 2.** Importance of variables represented with 2 principal components

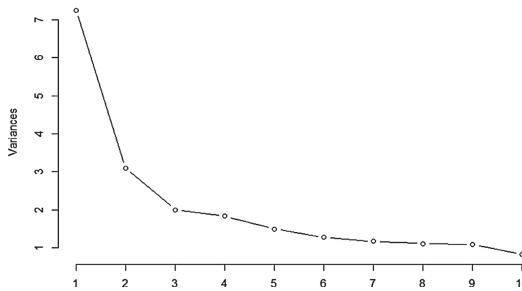
The cumulative proportion of the explanatory power between subjects, the proportion of variance, and standard deviation according to the principal components is shown in Table 2.

**Table 2.** Importance of components

Principal Component	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard Deviation	2,69	1,76	1,40	1,23	1,12	1,06	1,05
Proportion of Variance	0,34	0,15	0,09	0,07	0,06	0,05	0,05
Cumulative Proportion	0,34	0,49	0,59	0,66	0,72	0,77	0,82
Principal Component	PC8	PC9	PC10	PC11	PC12	PC13	PC14
Standard Deviation	0,93	0,84	0,79	0,73	0,52	0,41	0,40
Proportion of Variance	0,04	0,03	0,03	0,02	0,01	0,008	0,007
Cumulative Proportion	0,86	0,90	0,93	0,95	0,97	0,974	0,981
Principal Component	PC15	PC16	PC17	PC18	PC19	PC20	PC21
Standard Deviation	0,35	0,34	0,27	0,21	0,18	0,04	1,63e <sup>-15</sup>
Proportion of Variance	0,006	0,006	0,004	0,002	0,016	0,000	0,000
Cumulative Proportion	0,987	0,993	0,996	0,998	0,999	1,000	1,000

This result shows that using the first 8 principal components explained 86% of the differences between subjects; The first 11 principal components explain 95%, the first 15 principal components explain 99% of the differences between subjects. The proportion of variance did not differ much after the 10th principal component (Fig. 3).

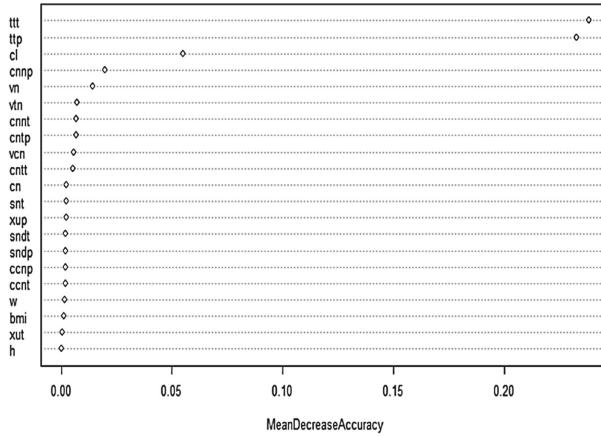
**Linear Correlation Between Girl-student Breast Dimensions:** The pairwise linear correlation coefficient  $r$  between breast dimensions was determined using the *cor* function in the *caret package on R*. One of the two dimensions with a linear correlation is



**Fig. 3.** The proportion of variance according to the number of principal components

removed when  $r \geq 0.95$ . The results show that the 18<sup>th</sup> dimension is *sndp* (right lower breast prolapse) which is excluded from the data set because it has a strong linear relationship with left lower breast prolapse *sndt* ( $r = 0.9633$ ). The remaining dimensions all have  $r < 0.95$  when considering the linear relationship one by one, so they are used continuously in the algorithms to extract the breast features.

**Selection of Breast Features:** Random Forest algorithm is used on each iteration to evaluate the model. RF is set up to evaluate all possible subsets of the feature set in the dataset. The results of determining the importance of dimensions for breast classification are shown in Fig. 4.



**Fig. 4.** Importances of breast dimensions determined by RF

Figure 4 shows that the first group of dimensions can be clearly defined: *ttt* (left breast volume) and *ttp* (right breast volume) are the most important, followed by a group of eight variables difference *cl*, *cnnp* (right outer breast), *vn* (breast circumference), *vtn* (upper breast circumference), *cntp* (right inner breast arc), *cntt* (left outer breast arch), *vcn* (breast circumference) and *cntt* (inner breast arch) left) is less intuitive. The least important variables are height *h*, distance from sternum to left cephalothorax,

mass index BMI, body weight  $w$ ,  $ccnt$  (left thoracic arch),  $ccnp$  (right thoracic arch) and  $sndp$  (lower right breast prolapse), respectively. Dimensions  $cn$  (breast distance),  $snt$  (thoracic prolapse above),  $sup$  (breastbone to right cephalothorax),  $sndt$  (left lower thoracic prolapse) are intermediate.

The Accuracy and Kappa coefficient of the RF model with the training results are shown in Table 3. The model has good clustering ability, the accuracy reaches 97.5%. The test results shown in Table 4 show that the model has an accuracy of 0.9615. The sensitivity is up to 0.989 and the specificity is up to 0.977.

**Table 3.** Accuracy and Kappa of RF model

mtry	Accuracy	Kappa
2	0.9497253	0.9179362
11	0.9749909	0.9596642
21	0.9745930	0.9587882

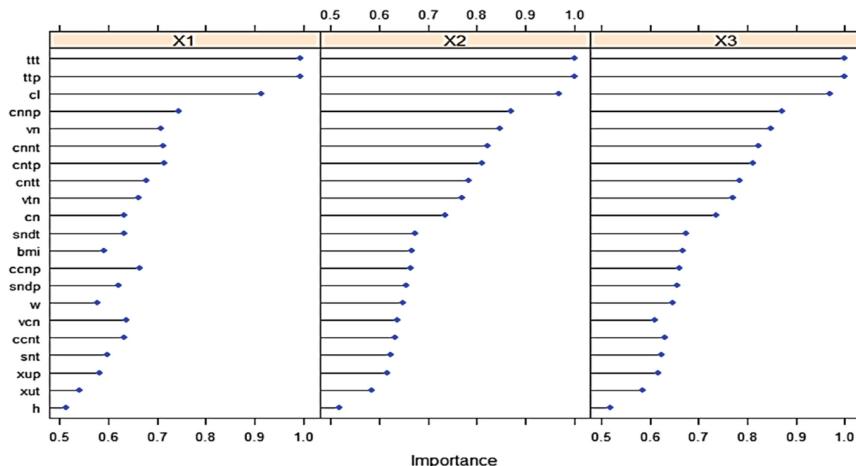
**Table 4.** Testing results of RF model

Accuracy:	0.9615		
95% CI:	(0.9224, 0.9844)		
P-value:	<2.2e-16		
Kappa:	0.9377		
Statistics by class	Class 1	Class 2	Class 3
Sensitivity	0.9216	0.9487	0.9891
Specificity	0.9771	0.9930	0.9667

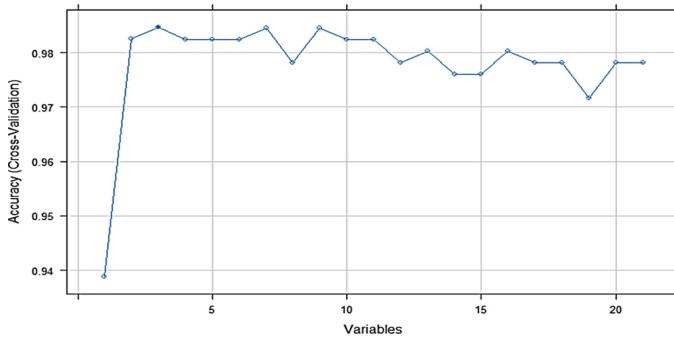
The importance of the variables determined by the LVQ method shown in Fig. 5 also shows that the most important dimensions for breast classification include  $ttt$  (left breast volume),  $ttp$  (right breast volume), difference  $cl$ , followed by the dimensions  $ccnp$  (right outer breast),  $vn$  (breast circumference),  $cntp$  (right inner breast),  $cnnt$  (left outer breast arch),  $cint$  (left inner breast arch),  $vtn$  (breast circumference) and  $vcn$  (breast circumference). This result is essentially consistent with the RF determination results above. The importance of breast dimensions when distinguishing individuals in clusters may not be uniform (Fig. 5).

The Accuracy and Kappa for each variable obtained in Table 5 show that the model is accurate and reliable. Increasing the number of dimensions measured after the 10th variable did not increase the accuracy of the breast classification model (Fig. 6). Thus, it is possible to apply RF to select important characteristic dimensions for better breast classification.

Thus, the results show that the most important dimensions for classifying North Vietnamese girl students' breasts are their volume of the breasts, followed by  $cl$  (the difference between bust and underbust circumference),  $ccnp$  (arc), right outer breast),

**Fig. 5.** Importances of breast dimensions determined by LVQ**Table 5.** Testing results of LVQ model

Variables	Accuracy	Kappa	Accuracy SD	Kappa SD
1	0.9389	0.900	0.03119	0.05186
2	0.9825	0.9716	0.02248	0.03636
3	0.9846	0.9748	0.02092	0.03444
4	0.9824	0.9710	0.02492	0.04132
5	0.9824	0.9710	0.02492	0.04132
6	0.9825	0.9712	0.02026	0.03336
7	0.9846	0.9747	0.02093	0.03448
8	0.9781	0.9642	0.01803	0.02972
9	0.9846	0.9746	0.02093	0.03448
10	0.9824	0.9712	0.02027	0.03337
11	0.9825	0.9712	0.02026	0.03338
12	0.9782	0.9643	0.01803	0.02972
13	0.9803	0.9677	0.01931	0.03181
14	0.9760	0.9607	0.01633	0.02698
15	0.9760	0.9608	0.01633	0.02698
16	0.9803	0.9676	0.01931	0.03182
17	0.9782	0.9643	0.01803	0.02972
18	0.9782	0.9643	0.01803	0.02972
19	0.9716	0.9538	0.01496	0.02453
20	0.9782	0.9643	0.01803	0.02972
21	0.9782	0.9643	0.01803	0.02972



**Fig. 6.** Accuracy by number of variables

*vn* (breast circumference), *cntp* (right outer breast arch), *cnnt* (left outer breast arc), *cint* (left inner breast arch), *vtm* (top breast circumference) and *vcn* (breast circumference). The parameters of height *h*, *xut* (distance from sternum to left cephalothorax), *ccnp* (right thoracic arch), *BMI*, *snt* (breast prolapse), *ccnt* (left thoracic arch) and body weight *w* are least important in the corresponding order. Dimensions *xup* (mid sternum to right thoracic tip), *sndt* (left lower thorax), *cn* (distance between 2 nipples), and *sndp* (right lower thoracic prolapse) were average between the above two groups. This may contribute to the phenomenon of many women wearing ill-fitting bras when bra dimensions are determined based on bust and underbust circumference dimensions. The important dimensions for the chest classification of Northern Vietnamese girl students have been extracted and shown in more detail and completeness than the study [10] that only mentioned a few breast parameters. On the other hand, this result also eliminates the measurement of breast parameters that are not significant for the classification of girl-students breasts. This result is somewhat consistent with the conclusion of Jie Zhou et al. in the study [9]. That is an important basis for the study to build a model to classify women's breasts based on the importance of the identified dimensions.

## 4 Conclusions

Principal Component Analysis was applied to consider the parameters and North Vietnamese girl students' breast dimensions. Random Forest and Learning Vector Quantization algorithms are applied to extract the characteristic breast dimensions. The applied models all have significant accuracy and specificity results (accuracy reaches 0.9615, sensitivity reaches 0.989 and specificity reaches 0.9771 in the RF model).

The most important features for classifying the North Vietnamese girl students' breasts are the volume of the breasts, the difference between the bust and the underbust circumference, followed by the outer breast, breast circumference, upper breast circumference, and inner breast arc, and breast circumference. Height, body weight, body mass index, thoracic prolapse, and thoracic arch are the least important parameters for breast classification. This result should be verified by the girl student breast classification.

Therefore, the study of classifying girl students' breasts using K-means clustering, Machine Learning, and artificial neural network models will be performed next to evaluate the classification efficiency with selected features of the breast.

## References

1. Oh, S., Chun, J.: New breast measurement technique and bra sizing system based on 3D body scan data. *J. Ergon. Soc. Korea* **33**(4), 299–311 (2014). <https://doi.org/10.5143/JESK.2014.33.4.299>
2. Li-xia, C., Xin, Z., Jing, Q.: Research on subdividing of female breast shapes based on 3-D body measurement. *J. Text. Res.* **27**(12), 21–24 (2006)
3. Zheng, R., Winnie, Y., Fan, J.: Development of a new Chinese bra sizing system based on breast anthropometric measurements. *Int. J. Ind. Ergon.* **37**, 697–705 (2007). <https://doi.org/10.1016/j.ergon.2007.05.008>
4. McGhee, D.E., Steele, J.R.: Breast volume and bra dimension. *Int. J. Cloth. Sci. Technol.* **23**(5), 351–360 (2011). <https://doi.org/10.1108/0955622111166284>
5. Movassaghi, K.: Shaping the breast: a comprehensive approach in augmentation, revision, and reconstruction. Springer, Berlin (2021). <https://doi.org/10.1007/978-3-030-59777-1>
6. Sheng Nan, C.: Analysis of female breast shape based on 3D human body scan in Shanghai. *Int. J. Biomed. Sci. Eng.* **4**(4), 34 (2016). <https://doi.org/10.11648/IJBSE.20160404.11>
7. Pei, J., Park, H., Ashdown, S.P.: Female breast shape categorization based on analysis of CAESAR 3D body scan data. *Text. Res. J.* **89**(1), 590–611 (2019). <https://doi.org/10.1177/0040517517753633>
8. Pei, J., Fan, J., Ashdown, S.P.: A novel optimization approach to minimize aggregate-fit-loss for improved breast sizing. *Text. Res. J.* **90**(15–16), 1823–1836 (2020). <https://doi.org/10.1177/0040517519901318>
9. Zhou, J., Mao, Q., Zhang, J., Lau, N.M.L., Chen, J.: Selection of breast features for young women in northwestern China based on the random forest algorithm. *Text. Res. J.* (2021). <https://doi.org/10.1177/00405175211040869>
10. Phạm Thị Thắm: Nghiên cứu đặc điểm hình dáng phần ngực phụ nữ Việt nam trưởng thành dựa trên kỹ thuật quét 3D cơ thể người. *Tạp chí KHCN trường ĐH CNHN*, số 17 (2013)
11. Nhung, L.T.H., Tu, N.N., Trinh, N.N., Hieu, P.M., Linh, N.T.D., Le, N.T.: Development of 3D breast measurement system using structured light for bra design. In: Tran, D.-T., Jeon, G., Nguyen, T.D.L., Lu, J., Xuan, T.-D. (eds.) ICISN 2021. LNNS, vol. 243, pp. 373–385. Springer, Singapore (2021). [https://doi.org/10.1007/978-981-16-2094-2\\_46](https://doi.org/10.1007/978-981-16-2094-2_46)
12. Lưu Thị Hồng Nhung, Nguyễn Nhật Trinh, Hà Công Minh Hoàng, Nguyễn Thị Lê: Nghiên cứu đặc điểm nhân trắc phần ngực nữ sinh miền Bắc Việt Nam. *Kỷ yếu Hội nghị KHCN HaUI lần 5*, trang 827–838, 6 (2021)
13. Kenett, R., Zacks, S., Amberti, D.: Modern Industrial Statistics with applications in R MINITAB and JMP. Wiley (2021). <https://doi.org/10.1002/9781118763667.ch9>
14. Genuer, R., Poggi, J.-M.: Random forests with R. Springer (2020). <https://doi.org/10.1007/978-3-030-56485-8>
15. Vũ Văn Luân: Rừng ngẫu nhiên cải tiến cho lựa chọn thuộc tính và phân loại dữ liệu gen, luận văn thạc sĩ, trường Đại học Công nghệ, đại học Quốc gia Hà nội (2017)
16. Đồng Thị Ngọc Lan, Nghiên cứu, xây dựng phương pháp trích chọn thuộc tính nhằm làm tăng hiệu quả phân lớp đối với dữ liệu đa chiều, luận văn thạc sĩ, trường Đại học Công nghệ, đại học Quốc gia Hà nội (2011)



# Vacant Parking Car Detection and Notification by Using Multi-camera System and Deep Learning

Dinh Cong Nguyen<sup>1(✉)</sup>, Viet Nam Le<sup>1</sup>, Hoang Long Nguyen<sup>1</sup>,  
Doan Minh Tran<sup>1</sup>, and Van Luyen Cao<sup>2</sup>

<sup>1</sup> Hong Duc University, Thanh Hoa, Viet Nam

{nguyendinhcong,levietnam,trandoanminh}@hdu.edu.vn, longlcf0311@gmail.com

<sup>2</sup> Department of Information and Communications, Thanh Hoa, Viet Nam

luyencv.stttt@thanhhoa.gov.vn

**Abstract.** A system of parking car assistance is a necessary implementation in the real environment. To obtain a good parking space in a narrow and limited area is sometimes inconvenient, especially in administrative agencies, judicial authorities. This depends on whether the empty place is available or not. In this paper, a new parking assist system is given by first discussing the parking space detection based on camera surveillance systems and then proposing a new approach to send notifications to users. The proposed system works well in terms of car detection accuracy up to 100% and the maximum fault rate of notifications to users is about 0.46%.

**Keywords:** Parking assist system · Camera surveillance · Car detection · Free-space detection

## 1 Introduction

Along with the significant economic development in Vietnam recently, we could witness a sharp increase in the number of cars. However, the infrastructure of car parking stations in particular is somewhere quite low, and this raises a lot of difficulties for the drivers or owners. Moreover, parking spaces in each official department are usually small and narrow, Fig. 1. Let citizens sometimes waste time to find a good place. With the growth of technology, parking car detection systems have been simplified in different ways.

From literature, it could be divided into two kinds of approach as visual and non-visual detection methods [1]. Non-visual detection approach is based on ultrasonic sensors and laser scanners [2,3]. This approach fits well the requirements of parking detection in term of accuracy. The usages of sensors become more popular as a clue many of the shopping centers equips their own systems in which the red light means the place is occupied while the blue one is available. However, the implementation is quite expensive when each parking location



**Fig. 1.** Image acquisition from camera with different official locations.

requires at least one sensor. And it cannot apply in the places where there are no physical boundaries for each parking space, as shown in Fig. 1.

In this paper, we pay our attention on the visual approach where camera surveillance systems are available. In addition, artificial intelligence (AI) and computer vision (CV) being recently more popular with their efficiency are a promising approach to solve this problem [1]. Therefore, our contributions are followed:

- We propose 10 new datasets corresponding with 10 official places in Thanh Hoa, Viet Nam where there are no physical boundaries for each parking car place.
- We present an end-to-end solution in order to notify users whether any parking car place is occupied or not.

The remainder of our paper is given as follows. Section 2 discusses the related works. Section 3 describes our protocols and how to obtain materials. Section 4 presents our solution while Sect. 5 shows the experimental results. Finally, Sect. 6 concludes the paper and proposes some perspectives.

## 2 Related Works

As mentioned above, there are two main approaches for this topic. The first one is to use a non-visual strategy while the second is to follow a visual strategy. In the scope of our paper, we follow the second approach. With a specific scenario that the parking locations (mostly in official places) do not contain physical boundary for each parking lot, and those places need to follow instructions from the guards or staffs.

For the visual approach, authors in [6] gave car parking model applying smart camera system and Deep learning model. This model could use on the low-hardware devices such as Raspberry PI by used mLetNet and mAlexNet models. The obtained results was near 98% with one camera, and about 91% for many cameras. [7] proposed a new approach to detect vacant parking spaces from drone. The model applied HoG feature descriptor to extract the region of interest (ROI). This approach worked well with the accuracy up to 96%. However, it have a little sense in the real application.

Other work in [8] presented a pipeline used the Tiny YOLOv1 model [9]. The method tended to track license plate number of each car by edge camera when cars enter or leave the parking lot. However, it is undeniable that the YOLOv1 model has several disadvantages in the literature. Recently, authors in [10] proposed a new dataset namely CNRPark Ext and then applied different models such as LeNet and AlexNet on it. It proved as an efficient and affordable approach through IP camera.

However, our approach is quite different from others. We try to propose a new end-to-end system in which could notify to users about vacant parking locations through application. At the best our knowledge, the works considering this system are little discussed in the literature.

### 3 Materials

#### Data Collection

In this paper, parking locations were collected from different official places through camera surveillance systems. It is worth noting here that those places do not have any physical boundaries for each parking slot.

For multiple cameras, we have employed the camera configuration as presented in [12] with IP HIK vision camera. Moreover, the data collection process has obtained with WIFI connection through Kurento Protocol<sup>1</sup>.

Videos are continuously stored during the days and then extracted into images whenever the cars or the parking places have changed their status. To create our datasets, each car is cropped and labeled by LabelImg annotation tool<sup>2</sup>. In addition, we use Matplotlib library<sup>3</sup> to draw soft boundaries for each parking slot. To verify them, we ask managers from official places for the accurate organizations of ground truths. Some visual examples are presented in Fig. 2.

### 4 Proposed Solution

The overall framework is illustrated in Fig. 3. To detect the cars from video streaming, we have used YOLO (You Only Look Once) model. To detail, YOLOv1 model was first proposed in [9]. In early 2020, YOLOv4 model [4]

<sup>1</sup> <https://doc-kurento.readthedocs.io/>.

<sup>2</sup> <https://github.com/tzutalin/labelImg>.

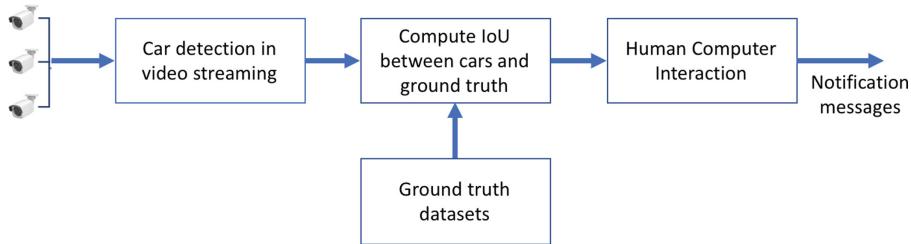
<sup>3</sup> <https://matplotlib.org/>.



**Fig. 2.** The obtained ground truths from several camera views at diverse locations and times.

was published with a lot of improvements. Recently, YOLOv5 model was also published at the model level without a paper. Even that this version do not have many improvements in compared with YOLOV4 model, it reduced model size, faster speed and had the same performance as YOLOv4 model.

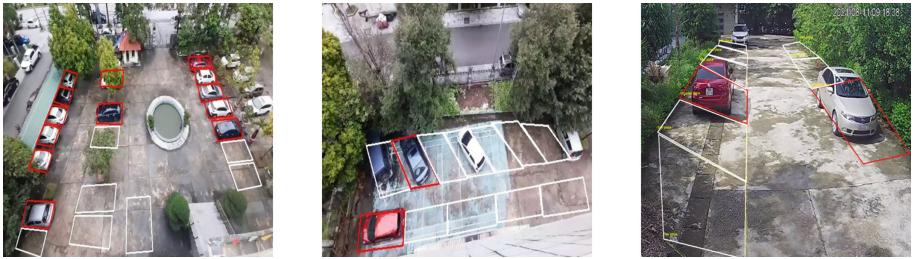
To detect cars in video streams and to fit with real-time constraint in our context, we inherit the pre-trained model of YOLOv5<sup>4</sup>. This means that car detection part will find cars in video streams and to store the bounding boxes of cars.



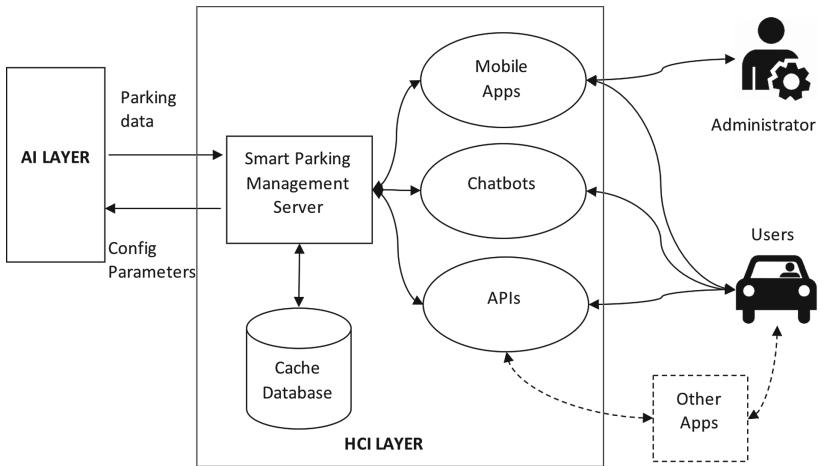
**Fig. 3.** Our framework includes of different main parts.

To remind here, ground truth bounding boxes are defined for each parking lot by using Matplotlib library in Python to draw polygon as a parking lot and label it with an ID, which is used for detect the free parking lot. Therefore, in the next step, using the ground truth (already stored in the datasets) to calculate IoU (Intersection over Union) [5] between the detected and the ground truth boxes. If the IoU is larger than 50% with a ground truth and a detected box, the parking lot is occupied, otherwise the parking-lot is free. For multiple cameras, the detected IDs will be saved in a set, the aim is to remove the duplicate IDs between cameras.

<sup>4</sup> <https://github.com/ultralytics/yolov5>.



**Fig. 4.** Results of detected cars, the red parking lots mean that lots were occupied while white lots have been free.



**Fig. 5.** The general pipeline in at the HCI layer.

The parking lot situation, as results of object detection layer, will be informed to the human-computer interaction (HCI) layer, Fig. 5. The HCI layer in a smart parking system is built to work over the internet, so we deployed it on a cloud virtual private server (VPS). This layer, on the one hand, is directly connected to the AI module to get near-real-time parking space data for storage on a caching database. On the other hand, it communicates with users in several ways, either directly via Mobile apps, Telegram, Chatbots, or indirectly by providing application programming interface (API) to other software developers. We can divide the HCI layer into several divisions such as:

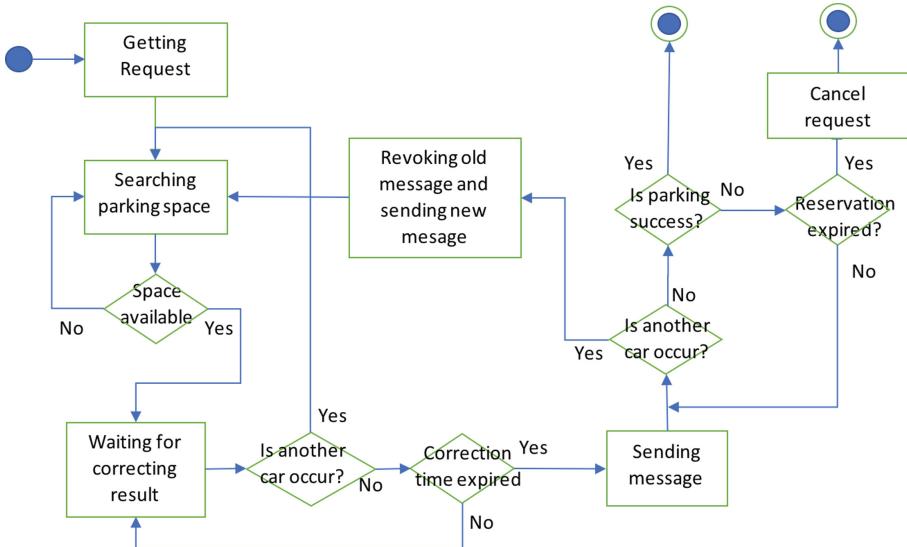
- **Communicate AI-HCI:** The AI module takes the working parameters at initialization or when there is some updates by the system administrator. These parameters contain reporting interval (seconds), the parking status and some AI configuration values. Periodically, mostly within a few seconds, the AI module will report the status of the parking status with the proper

images to the server. The HCI module will store these data in the database for querying purposes.

- **Communicate HCI-users:** The HCI layer allows the user to submit a parking request. After searching the database, the system will send a response whether there is a parking space, along with information about the parking spaces. When this information becomes obsolete, the system will revoke it and send a new notification to the user. There are many ways in which users can communicate with the system:

- Mobile applications (app): Through a mobile app, users can find their packing spaces via smartphones, both Android and iOS. For the best experience, these apps should be allowed to send notifications and access the user's location, those are likely optional.
- Chatbots: If the user does not install the app on the phone, they can use instant messaging to query data. To do that, the user has to add the chatbot's account and communicate with the chatbot using a few simple syntaxes. We built a chatbot based on Telegram because of the convenience of the APIs they provided.
- APIs: We provide APIs for developers to create their widgets or integrate to their system. That makes the system more flexible with a variety of user demands. To use these APIs, users must have an access token that is provided by the administrator. For each request, the system will check token authorization and return the result in JSON format.

Our diagram of the HCI layer could be precise as Fig. 6.



**Fig. 6.** The diagram of the HCI layer.

## 5 Experiments

### Experimental Materials

The pre-trained model of the YOLO5 model was downloaded from Ultralytics<sup>5</sup> and configured with class = 1 at the training process. It is common to apply augmentation techniques such as dilation, erosion to enhance the dataset.

For the metrics, we used the regular metric as accuracy (%) to measure the performance of the car detection model. With the HCI layer, a typical error that can occur in message-related systems is inconsistency. For example, information changes after as soon as the user gets a message. To avoid that, we use two different techniques:

- **Adding delay time for sending messages:** When it recognizes a parking space that matches the user requests, the system immediately locks this space and waits for a period called “Correction-time” to see if any other vehicles are approaching the parking space. Obviously, the longer the delay system sends the message, the less fault message it will send later. However, if this period is too long, the user will have to wait a lot, even if that will cause him to give up the intention to park. That is why we set it for less than one minute.
- **Sending invocation messages:** When a user receives a message that there is a parking space that matches his request. It cannot be ruled out that another driver occupies this empty parking space shortly after. If that happens, we will immediately retrieve the old message and send a new repair message.

Moreover, there are several parameters taken part into our experiments such as:

- Reservation time (minus): The time that the driver holds the parking space. During this time, the parking space is in a reserved state.
- Correction time (seconds): The time the system waits to see if the status of the parking space is changed.
- Number of requests(times): The number of requests that the user sends to the system.
- Number of responses(times): Number of times the system returns the result to the users.
- Revocation (times): The number of times that the system has to revoke a notification because an empty parking space is occupied by another vehicle.
- Fault (times): The number of times the user arrives at the parking space designated by the system but this parking space has been occupied by another vehicle.
- Fault rate (%) = Fault/Number of request.

---

<sup>5</sup> <https://github.com/ultralytics/yolov5>.

## Experimental Results

The performance of the detection model was completely estimated on the test set. The experiments have been applied on the computer with CPU Intel Core i7-10750H 6 cores 12 threads 2.6 GHz (up to 4.9 GHz), RAM 16 GB dual channel (1463.2 MHz).

The accuracy on the test sets was 100% for car detection task among 10 our datasets with the frame rate per second (FPS) was 15. It could be explained that the YOLOv5 works well with our data and our model was converged and robustness under different conditions.

For performance of the HCI layer, we show the results of our system in Table 1. For our context, we select Chatbot in Telegram for our evaluation.

**Table 1.** Performance of the proposed HCI system recorded during three weeks.

Week	Occupy time (minus)	Correction-time	Number of request (times)	Number of responses (times)	Revocation (times)	Fault (times)	Fault rate (%)
1	5	20	2150	2150	15	10	0.467071462
2	5	40	2225	2225	6	6	0.270392068
3	5	60	2041	2041	7	5	0.196656834

From the experimental aspects, we need to choose an acceptable range of the correction time. With several surveys for user experiences, we decide to employ the correction time was 20 s in the first week, the revocation messages and fault times were much larger than when the correction time was 40 s and 60 s in the second and third week, respectively. This could be explained that 20 s are not enough time to wait some cases of cars moving to the vacant parking places.

In contrast, when correction time was 40 s in the second week, revocation times, fault times and fault rates reduced a half. In the third week, correction time set to 60 s, the fault times and revocation messages remained the same as in the second week. This means that 40 s could be suitable for our system to see any change with the parking status before sending notification to users. 60 s sometime are quite long with some users while the performance does not improve much.

One of the weaknesses of our approach is dependent on the WIFI connection. The system will stop when this connection fails. We could think about the cable solution for camera connection as a redundant approach.

## 6 Conclusions and Perspectives

In this paper, the end-to-end system for vacant parking cars was proposed. The system applied the YOLOv5 model to detect parking places with 15 FPS. It could be used in video streaming from multiple camera surveillance systems. Then, we proposed the HCI system to interact with users who would like to reserve the free-space parking lots in the place without physical boundaries for

each parking lot. The HCI system could notify to user after 20, 40 or 60 s to inspect the parking status.

As perspective, we could perform our system with larger places such as supermarkets or airports. We also would like to embed users' locations into our system to estimate the their distances. By doing that, we expect to give the better instruction with the lower correction time.

## References

1. Ma, Y., Liu, Y., Zhang, L., Cao, Y., Guo, S., Li, H.: Research review on parking space detection method. *Symmetry* **13**(1), 128 (2021). **17**(12), 2738 (2017)
2. Pohl, J., Sethsson, M., Degerman, P., Larsson, J.: A semi-automated parallel parking system for passenger cars. *Proc. Inst. Mech. Eng. J. Auto Eng.* **220**, 53–65 (2006)
3. Jeong, S.H., et al.: Low cost design of parallel parking assist system based on an ultrasonic sensor. *Int. J. Automot. Technol.* **11**, 409–416 (2010)
4. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: YOLOv4: optimal speed and accuracy of object detection. arXiv preprint [arXiv:2004.10934](https://arxiv.org/abs/2004.10934) (2020)
5. Nguyen Dinh, C., Delalandre, M., Conte, D., Pham, T.A.: Fast RT-LoG operator for scene text detection. *Journal of Real-Time Image Processing* **18**(1), 19–36 (2020). <https://doi.org/10.1007/s11554-020-00942-7>
6. Amato, G., Carrara, F., Falchi, F., Gennaro, C., Vairo, C.: Car parking occupancy detection using smart camera networks and Deep Learning. In: (SCC), pp. 1212–1217 (2016)
7. Peng, C.F., Hsieh, J.W., Leu, S.W., Chuang, C.H.: Drone-based vacant parking space detection. In: WAINA, pp. 618–622 (2018)
8. Bura, H., Lin, N., Kumar, N., Malekar, S.: An edge based smart parking solution using camera networks and deep learning. In: ICCC, pp. 17–24 (2018)
9. Redmon, J., Divvala, S., Girshick, R.: You only look once: unified, real-time object detection. In: CVPR, pp. 779–788 (2016)
10. Farley, A., Ham, H.: Real time IP camera parking occupancy detection using deep learning. *Procedia Comput. Sci.* **179**, 606–614 (2021)
11. Redmon, J., Farhadi, A.: Yolov3: an incremental improvement. arXiv preprint [arXiv:1804.02767](https://arxiv.org/abs/1804.02767) (2015)
12. Hoang, V.Q., Nguyen, D.C.: Detecting leftover food and the shrimp for estimating of the shrimp body length based on CNN. In: Tran, D.-T., Jeon, G., Nguyen, T.D.L., Lu, J., Xuan, T.-D. (eds.) ICISN 2021. LNNS, vol. 243, pp. 239–247. Springer, Singapore (2021). [https://doi.org/10.1007/978-981-16-2094-2\\_30](https://doi.org/10.1007/978-981-16-2094-2_30)



# Implementation of Real-Time Human Tracking System Based on Deep Learning Using Kinect Camera

Van-Truong Nguyen<sup>(✉)</sup> and Duc-Tuan Chu

Department of Mechatronics Engineering, Hanoi University of Industry, Hanoi, Vietnam  
nguyenvantruong@hau.edu.vn

**Abstract.** In this study, an effective real-time tracking target human method using Kinect camera is developed in different condition environments using deep learning. The proposed algorithm combines the YOLOv3 and the deep SORT method. To be more detailed, the YOLOv3 algorithm is adopted to detect the target human. To trace the identified human, the deep SORT is employed. The efficiency of the proposed system is tested with the complex environments with four cases: low light, normal light, outdoor, and rain environments. The experiment results indicate that the suggested solution performed effectively in different complex environments. Besides, the developed algorithm can detect the human target in the situation of partial occlusion and loss of the human target.

**Keywords:** YOLOv3 · Deep SORT · Human tracking · Kinect camera

## 1 Introduction

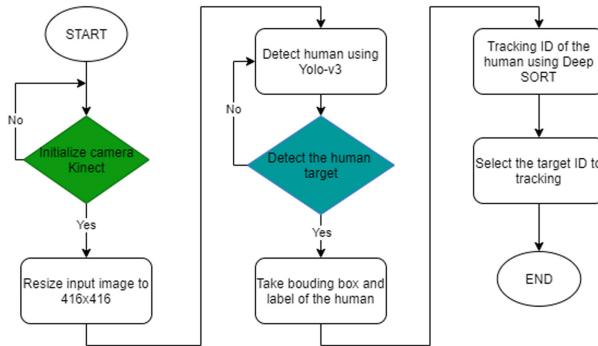
The human tracking motion is among the most important parts of computer vision with a range of applications including robot vision, medicine or traffic monitoring, etc. In the 4.0 technology era, object tracking algorithms have been developed based on machine learning and deep learning. In the process of practical tracking humans, the system requires working in real-time, appearance change, complex environments when the target human is partial occlusion or object losing [1, 2]. Hence, the human tracking problem has remained a challenging and fascinating issue for many scientists.

In recent times, several deep learning-based approaches have been developed to increase the performance of human tracking algorithms. For instance, a hierarchical convolutional features (HCF) approach is proposed based on three convolutional layers in the VGG network and correlation filters technique [3, 4]. The HCF method can obtain the object position, but the anti-interference ability is not good enough. In [5], the real-time human tracking approach is developed as an effective human tracking algorithm. However, this algorithm neglects the rain environment. On the other hand, various human tracking methods are proposed based on computer vision in [6, 7]. Those methods can be implemented for real-time applications when the noise environment is distributed normally.

In this research, the deep learning method build on the combination of YOLOv3 and deep SORT is proposed to track the target human under complex environments. The YOLOv3 is used to detect the object target, and deep SORT is adopted to track the human target. The proposed algorithm is used in most complex scenes, real-time speed (10–12 fps), partial occlusion, and loss of human target. Besides, the experiment in a Kinect camera shows that the system adapts to different conditions environment such as low light, normal light, outdoor, and rain environments.

## 2 Real-Time Human Tracking System

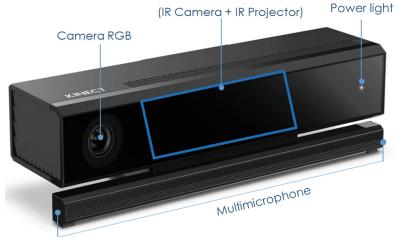
### 2.1 System Architecture



**Fig. 1.** System's flowchart.

The suggested algorithm is presented in Fig. 1. Firstly, the system starts with the initialization Kinect camera. In the next stage, the input images are scaled to size  $416 \times 416$ . Next, the human target is detected based on the YOLOv3. If there is a human target, YOLOv3 takes bounding boxes and labels of the human target, or else the system continues to detect humans. Then, the proposed approach tracks the ID human using Deep SORT. Each person is mounted an ID number on the bounding box and displayed results. Finally, we select the target ID to follow.

## 2.2 Camera Kinect



**Fig. 2.** Kinect camera

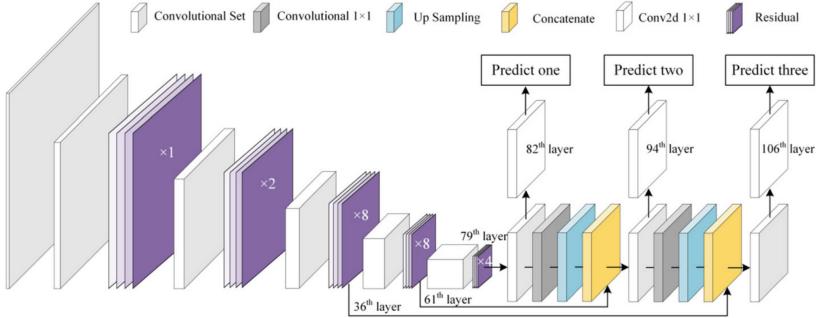
Nowadays, the 3D camera draws more and more attention, due to its applications in computer vision. One of them is Kinect camera (Fig. 2) which is invented by Microsoft. It provides RGB color and depth images, a multi-microphone, and the ability to rotate using a motor. The Kinect camera is a 3D depth sensor and is usually used as an in-depth estimation algorithm [8]. The system includes an infrared camera and an infrared projector. In this study, the Kinect camera is used to track the human system. All RGB images and depth information are obtained from the Kinect camera.

## 2.3 YOLOv3 Algorithm

YOLOv3 [9] is a CNN model with 53 convolution layers as shown in Fig. 3. Each layer has a normalization layer (batch normalization) and an activation (Leaky Relu). The input images of the YOLOv3 have input the size of  $416 \times 416$  or  $608 \times 608$ . Both sizes  $416 \times 416$  and  $608 \times 608$  have 3 types of feature maps ( $13 \times 13$ ,  $26 \times 26$ , and  $52 \times 52$ ) and ( $19 \times 19$ ,  $38 \times 38$ , and  $72 \times 72$ ), respectively. The outputs of the YOLOv3 have a lot of the bounding boxes and the classes. There are six predictions in each bounding box:  $p_c$ ,  $b_x$ ,  $b_y$ ,  $b_h$ ,  $b_w$  and  $c$ . The  $(b_x, b_y)$ ,  $b_h$  and  $b_w$  are the center point, the height, and the width of the bounding box, respectively. The  $p_c$  is confidence prediction which signifies the IOU between the objects after detection and ground truth area. The  $c$  is prediction class probabilities, that are conditioned on the grid cell containing an object.

YOLOv3 predicts three types of feature maps. For each feature map, the grids predict three bounding boxes by three anchors. To reduce the bounding box of outputs, the YOLOv3 filters bounding boxes by confidence scores. The low score bounding boxes are deleted. The overlap bounding boxes are eliminated based on Non-max suppression. Yolov3 has two kinds of loss as localization loss  $L_{loc}$  and confidence loss  $L_{cls}$ . By adding them together, the overall loss is calculated:

$$\begin{aligned}
 L_{loc} = & \rho_{coord} \sum_{i=0}^{m^2} \sum_{j=0}^K 1_{ij}^{obj} \left[ (a_i - \hat{a}_i)^2 + (b_i - \hat{b}_i)^2 \right] + \rho_{coord} \\
 & \sum_{i=0}^{m^2} \sum_{j=0}^K 1_{ij}^{obj} \left[ (\sqrt{c_i} - \sqrt{\hat{c}_i})^2 + (\sqrt{d_i} - \sqrt{\hat{d}_i})^2 \right]
 \end{aligned} \quad (1)$$



**Fig. 3.** YOLOv3 architecture.

$$\begin{aligned}
 L_{cls} = & \sum_{i=0}^{m^2} \sum_{j=0}^K 1_{ij}^{obj} \left( E_i - \hat{E}_i \right)^2 + \rho_{noobj} \sum_{i=0}^{m^2} \sum_{j=0}^K 1_{ij}^{obj} \left( E_i - \hat{E}_i \right)^2 \\
 & + \sum_{i=0}^{s^2} 1_i^{obj} \sum_{e \in E} \left[ g_i(e) - \hat{g}_i(e) \right]^2
 \end{aligned} \quad (2)$$

$$Loss = L_{loc} + L_{cls} \quad (3)$$

where  $[(a_i, b_i), (c_i, d_i)]$  is the center point of the anchor box, the width, and the height of the anchor box;  $\left[ (\hat{a}_i, \hat{b}_i), (\hat{c}_i, \hat{d}_i) \right]$  is the prediction of the center point of the anchor box, the width, and the height;  $1_i^{obj}$  indicates whether or not the object is present in the cell  $i$ ;  $1_{ij}^{obj}$  denotes  $j^{\text{th}}$  bounding box responsible for the prediction of the object in the cell  $i$ ;  $E_{ij}$  is confidence score in the cell  $i$ ;  $\hat{E}_{ij}$  is prediction confidence score in the cell  $i$ ;  $E$  is the number of classes;  $g_i(e)$  is a conditional probability, whether or not the cell contains an object of the class  $e \in E$ ;  $\hat{g}_i(e)$  is the prediction of conditional probability;  $\rho_{coord}$ , and  $\rho_{noobj}$  are regularisation parameters required to balance the loss function.

## 2.4 The Deep SORT Tracking Algorithm

In this paper, Deep SORT is developed based on Wide Residual Network (WRN) [10] to follow the human targets in the current frame. Kalman filter is a crucial component in Deep SORT which predicts new track humans. Each track contains eight variables as  $x = [h, l, y, k, h', l', y', k']$ , where  $[h, l]$  is the center point of the bounding box,  $[y]$  is the aspect ratio,  $[k]$  is the height of the bounding box,  $[h', l', y', k']$  are the respective velocities. For each detection, Deep SORT creates a track that has all state information of the target. With each track, Deep SORT counts the number of frames since the last successful measurement association. During Kalman filter prediction, this counter is incremented. If the track is associated with measurement, the counter is reset to zero. On the contrary, the first three frames are removed.

### 3 Experimental Results

#### 3.1 YOLOv3 Training Model

To train the YOLOv3, we collected 6000 images of humans from various sources, including 2400 images from Google and 3600 images from our camera. On both the COCO and our datasets, the transfer learning method is combined with the pre-trained model to reduce training model time.

#### 3.2 Running System and Evaluating Results

The proposed algorithm is executed on a computer with 8 GB RAM, INTEL Core i7 7500U CPU 2.9 GHz (2 CPUs), and a Kinect v3 camera sensor. The algorithm is implemented in Python 3.7 and uses the OpenCV 4.4 Library and Numpy Library. In all experiments, a set of parameters is as follows: cosine\_distance is 0.5; nms\_max\_overlap is 0.3;  $\rho_{coord}$  is 5; and learning rate is 0.001.

The system's efficiency is evaluated in a variety of environments as low light, normal light, outdoor, and rain. Figure 4 illustrates the test result. The results show that the proposed method is effective as follows.

Case 1: Normal light condition (500 lx).

In this paper, we test the proposed algorithm indoors with light intensity 500 lx. The YOLOv3 detects and mounts ID for each object. Next, the humans are selected object targets to track. The system works sufficiently stable when the humans are overlapped or go past each other.

Case 2: Outdoor condition (700 lx).

In this case, the developed algorithm is tested ranging from outdoor with strong light. We used light sensor LUTRON LM-81LX to evaluate light intensity. The system works well in the case of partial occlusion and loss of human targets.

Case 3: Low light (250lx).

Like case 1 and case 2, we used light sensor LUTRON LM-81LX to evaluate light intensity. The conditional test is indoor with some fluorescent lamps off. The result is shown that the detection speed is lower than case 1 and case 2. However, the system still detects and tracks the human targets.

Case 4: Rain environment.

To evaluate the ability to respond to different environments, the proposed method is tested in a rainy environment. The experiment shows that the system works efficiently and tracks object targets in the case of partial occlusion and loss of human targets.



**Fig. 4.** Test algorithm with different condition environment a, Normal light b, Outdoor c, Low light d, Rain environment

**Table 1.** Test algorithm with different condition environments.

Test condition	Light intensity	Total of the frames	Result of the CAM-shift [6]		Results of the proposed method	
			Accuracy (%)	FPS	Accuracy (%)	FPS
Normal light	500	900	92.6	14	94.5	10
Outdoor	700	990	89.9	16	92.3	11
Low light	250	900	88.3	13	90.1	10
Rain	—	1300	90.39	15	91.5	12

Table 1 shows the comparison results of the proposed algorithm and the CAM-shift [6]. The proposed system has high precision with maximum accuracy of up to 94.5% in the case of normal light, partial occlusion, and loss of human targets. In addition, FPS is ranged from 10 to 12. Therefore, the system works well in real-time applications.

## 4 Conclusion

In this study, the effective real-time tracking target human algorithm is proposed in different condition environments based on the YOLOv3 and the deep SORT. The proposed approach detects the human targets with high accuracy and tracks through longer periods of occlusion. The experimental results show that our approach achieves great performance under a variety of conditions. Even in bad conditions such as low light and rain, the detection rate is still high. In near future, the algorithm will be applied in a real system in the industrial environment.

## References

- Chen, Y., Sheng, R.: Single-object tracking algorithm based on two-step spatiotemporal deep feature fusion in a complex surveillance scenario. *Math. Probl. Eng.* **2021**, 1–11 (2021)
- Tran, Q.V., et al.: Pyramidal Lucas-Kanade-based noncontact breath motion detection. *IEEE Trans. Syst. Man Cybern. Syst.* **50**(7), 2659–2670 (2020)
- Li, F., et al.: Learning spatial-temporal regularized correlation filters for visual tracking. In: CVPR, pp. 4904–4913 (2018)
- Mueller, M., et al.: Context-aware correlation filter tracking. In: CVPR, pp. 1396–1404 (2017)
- Nguyen, V.-T., Nguyen, A.-T., Nguyen, V.-T., Bui, H.-A.: A real-time human tracking system using convolutional neural network and particle filter. In: Tran, D.-T., Jeon, G., Nguyen, T.D.L., Lu, J., Xuan, T.-D. (eds.) ICISN 2021. LNNS, vol. 243, pp. 411–417. Springer, Singapore (2021). [https://doi.org/10.1007/978-981-16-2094-2\\_50](https://doi.org/10.1007/978-981-16-2094-2_50)
- Nguyen, N.Q., et al.: Real time human tracking using improved CAM-shift. In: IFS-A-SCIS, pp. 1–5 (2017)
- Tran, Q. V., et al.: Real-time non-contact breath detection from video using adaboost and Lucas-Kanade algorithm. In: IFS-A-SCIS, pp. 1–4 (2017)
- Han, J., et al.: Enhanced computer vision with Microsoft Kinect sensor: a review. *IEEE Trans. Cybern.* **43**(5), 1318–1334 (2013)
- Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint [arXiv:1804.02767](https://arxiv.org/abs/1804.02767) (2018)
- Wojke, N., et al.: Simple online and realtime tracking with a deep association metric. In: ICIP, pp. 3645–3649 (2017)



# An Approach to Evaluate the Reliability of the Face Recognition Process Using Adversarial Samples Generated by Deep Neural Networks

Dinh Cong Nguyen<sup>1,2(✉)</sup>, Nhan Dong Le<sup>1</sup>, The Cuong Nguyen<sup>1,2</sup>, Tuong Quan Nguyen<sup>3</sup>, and Van Quan Nguyen<sup>3</sup>

<sup>1</sup> ThinkLABs JSC, Thanh Hoa, Viet Nam

{cong,dongln,cuong}@thinklabs.com.vn

<sup>2</sup> Hong Duc University, Thanh Hoa, Viet Nam

{nguyendinhcong,nguyenthecuong}@hdu.edu.vn

<sup>3</sup> Ministry of Public Security, Hanoi, Vietnam

quannsc@gmail.com, quanhvktqs85@gmail.com

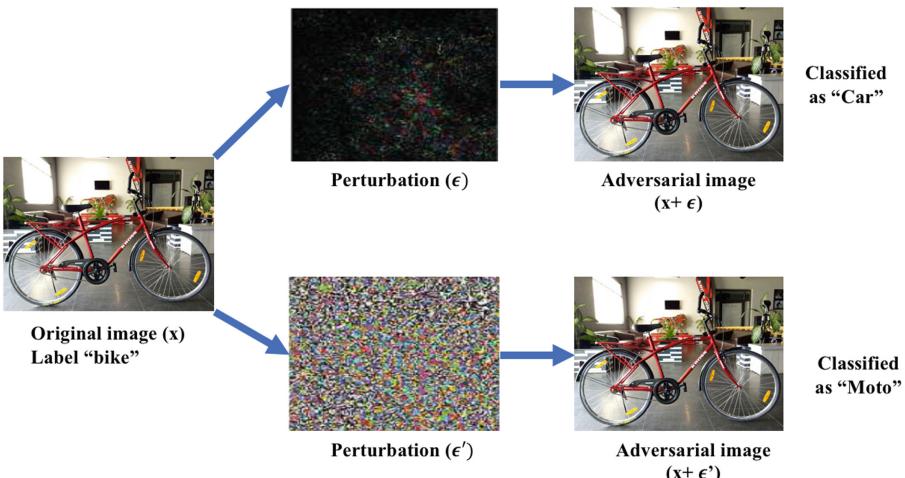
**Abstract.** Many studies have shown that Deep Neural Networks (DNNs) can be fooled by deceptive inputs, known as adversarial samples. Adversarial samples are produced by putting a small amount of perturbations into input images. These samples might mislead the object detection/recognition phases, thus it leads to the decreasing in the overall reliability of the object recognition process. Therefore, understanding the characteristic of adversarial samples is a very important step to propose the efficient attacking algorithms. In the paper, we present a new approach to particularly attack the face recognition process based on Fast Gradient Sign Attack (FGSA) and Jacobian-based Saliency Map Attack (JSMA) algorithms for the white-box attacks. By this way, we can evaluate the reliability of the face recognition process which are using MobileFaceNets and Caffe models.

**Keywords:** Adversarial samples · Face recognition · FGSA · JSMA

## 1 Introduction

Deep neural network (DNN) is a narrow branch of machine learning (ML) known as extremely changing the world. It is true that DNN obtains incredible results on most tasks which could not be defined by hand-crafted features and/or rule-sets. Its application can be listed as object recognition [1], machine translation [2], or speed recognition [3]. Since new DNNs have been deployed in commercial projects, their ability to keep the same performance even facing with data from slightly dissimilar properties than trained samples are required. This requirement is explained as the model's reliability to generalize the prediction outside the training data. A robust DNN system could not only achieve high accuracy

but it also performs well in a certain level of data. However, [4] proved that this reliability could be destroyed with a small and intentional level of noise in the input images. The term “adversarial samples” was first applied to define the status of input images added to these kinds of noises Fig. 1. From this exploration, many publications were conducted to investigate the aspects. Their summarization could be found in the comprehensive survey [5].



**Fig. 1.** An example of adversarial samples by adding different amounts of perturbations  $\epsilon$  into the original image. The first row is classified as “Car” and the second row is classified as “Moto”.

In this paper, we continue to discover the extent to which face recognition processes are in danger of attacks. Nowadays, these processes are extensively used in different sensitive aims such as access control<sup>1</sup>, or surveillance<sup>2</sup>. Therefore, attackers who fool them could result in serious consequences. To the best of our knowledge, there are usually two different kinds of attacks to classifiers considering a white-box attack. The first is to play with the pixels of images while the second is to create new physical items such as stickers adding into real scenes of images. In the scope of the paper, we pay a particular attention on the first approach with the contributions as follows.

- We propose a new approach to attack face recognition processes using two the FGSA and the JSMA methods.
- We discuss performance evaluations of two attack methods.

The following parts of the paper could be listed. Section 2 shows the related works with a particular attention on the FGSA and the JSMA methods. Next,

<sup>1</sup> <https://www.nec.com/en>.

<sup>2</sup> <https://www.neurotechnology.com/index.html>.

Sect. 3 presents our approach to attack the face recognition processes. Section 4 shows and discusses the results. At last, Sect. 5 will conclude and discuss some perspectives.

## 2 Related Works

It is true that DNNs are quite easy to attack by adversarial samples. This point was investigated from [6–8]. FGSM was proposed in [9] to produce adversarial examples. This method processed a gradient update with the direction of the gradient sign at every pixel. The added noises into images could not be recognized by manual inspection of humans. However, [6] proved that this method was easy to defend. Based on FGSM, there are several modified versions in order to improve the method. It can be listed as I-FGSM [6], RAND-FGSM [10]. I-FGSM used momentum to create iteratively adversarial samples while RAND-FGSM proposed Ensemble Adversarial Training which added small amounts of noise so as to defeat adversarial training.

The JSMA method [11] is another approach for this direction. This method focuses on a constant number of features from images and then maximizes the saliency map at each iteration step. Compared to above methods, the JSMA method treats with fewer pixels. Deepfool [12] provided a new method to efficiently estimate perturbations by computing and finding the closest projection from the input images to the adversarial samples on large-scale datasets.

From early works for face detection in [13], face detection processes are widely applied as basic object detection and then used in real applications<sup>3</sup>. From that, [14] proposed a new model to camouflage the face detection system in [13] by making up and designing new hair styles. With a similar motivation, [15] took a weakness of camera sensors with infra-red light to design adding a special kind of noise into the captured images. Although these methods are effective for evading systems, they are conspicuous and unreal.

Other work [16] attacked on the face recognition processes using either Eigenfaces or Fisherfaces [17]. However, this method proved recently that it does not perform well with the advanced systems of the face recognition [18]. [19] proposed a new approach in which applied adversarial generative nets (AGNs) to design eyeglasses to fool face recognition processes with the better performance compared to previous methods. However, this method is out of the scope of our paper.

In this work, we are interested in methods adding slight perturbations into the images. These methods are not easy to distinguish from human eyes. Moreover, we prove that the volume of images still guarantees for time-processing of the face recognition processes.

## 3 Our Proposed System

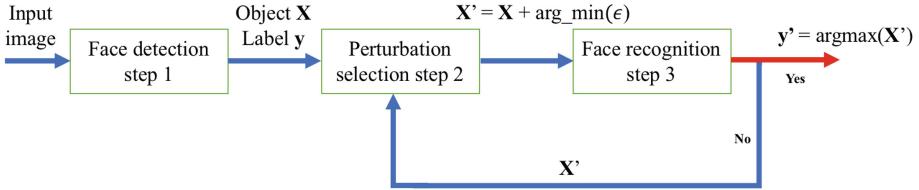
In our context, we assume that an attacker has already accessed into a trained DNN. The opponent cannot change the parameters of the DNN. This means

---

<sup>3</sup> <https://github.com/mobilesec/authentication-framework-module-face>.

that he/she only modifies the input images in order to fool the DNN for an untargeted attack.

To detail the assumption, we note  $f(\cdot)$  as the DNN classifier with  $X$  as the input image corresponding to its label  $y$ . With every  $X$ , it is assumed that  $X'$  is an adversarial sample. For our purpose, it is to force the DNN to classify  $X'$  into other classes  $y'$  with  $y' \neq y$  ( $y' = \text{argmax}(f(X'))$ ). Our proposed system could be presented as Fig. 2.



**Fig. 2.** The proposed approach is given. It includes three main steps: Face detection, noise creation, and face recognition.

For the convenience, we detail here three main steps in the proposed approach.

- Step 1: For the face detection task, we inherit TinaFace model [20] which the advanced model for face detection in the challenging face detection benchmark WIDER FACE<sup>4</sup>. Our purpose is to detect efficiently all faces from images.
- Step 2: As above suggestions, we employ two methods for perturbation selection as the FGSM method [9] and the JSMA method [11]. These methods could easily be controlled by the amount of adding noises. We aim to estimate the limitation of any face recognition after. Our approach differs from other works when we only add perturbation in face parts returned by TinaFace without considering whole input images.
- Step 3: Normally, this step results in true labels of detected faces corresponding to the trained face images of people. However, due to a untargeted attack, we aim to fool this DNN model applied in this step. To do that, we use MobileFaceNets [21], Caffe [22] as models for face recognition step.

## 4 Experiments

In this section, the performance evaluation of our approach is illustrated on our own dataset. Section 4 introduces our dataset. Section 4 details protocols used in the experiment. At last, Sect. 4 presents the results.

<sup>4</sup> <http://shuoyang1213.me/WIDERFACE/>.

## Dataset

Images were collected from the camera verification system of our company in front of the main door. To enrich data for training, videos were recorded at the different times during a day with illumination/rotation changes. 30 staffs volunteered in the data collection process with more than 1000 face images for each volunteer. Each face was carefully annotated with a label. Next, we divided respectively into ratios of 60%, 10%, 30% for train, val, test sets. Figure 3 shows the faces extracted from the dataset.



**Fig. 3.** The faces were extracted from our dataset. To prepare data for the training step, we extracted 1000 face images for each person under different conditions.

## Protocol

In this paper, we only pay our attention on the face recognition step. As known, with a given set of input images  $X$  with the corresponding output labels  $Y$ . Basically, a DNN algorithm tries to find a mapping  $f : X \mapsto Y$  in order to minimize the amount of mis-classified images.

In our scenario, for each clean image  $x$  which gets accurately classified with label  $y$  by function  $f(\cdot)$  we will create an adversarial image  $x'$  by using the minimal noise  $\epsilon$  to input image  $x$ . By doing that, our purpose is to let  $x'$  get classified with another label  $y'$  by  $\operatorname{argmin}_{\epsilon} f(x + \epsilon) = y'$  with two methods the FGMS and JSMA on our dataset.

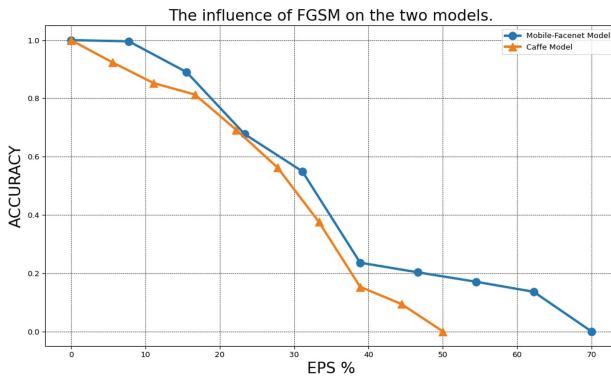
As usual, the models for face detection and face recognition have re-trained with the dataset before an inference process. In our work, noise addition will be processed at each location of faces from images. This period is done iteratively to discover the limitation of face recognition models: MobileFaceNets and Caffe models

## Results

We present here first the influence of the FGSM method on the two models MobileFaceNets and Caffe for face recognition. For the convenience, the results of the system before attacks as given in Table 1. After controlling the amount of perturbations into the detected face images, we produce the Fig. 4. It is clearly seen that the performance evaluation of the system falls down when we put more noises into the input images. Caffe model completely fooled at the percentage of perturbations 50% while this percentage with MobileFaceNets is 70%. This could explain that the MobileFaceNets model is stable and harder to attack than the Caffe model.

**Table 1.** Performance evaluation of our system before attacks.

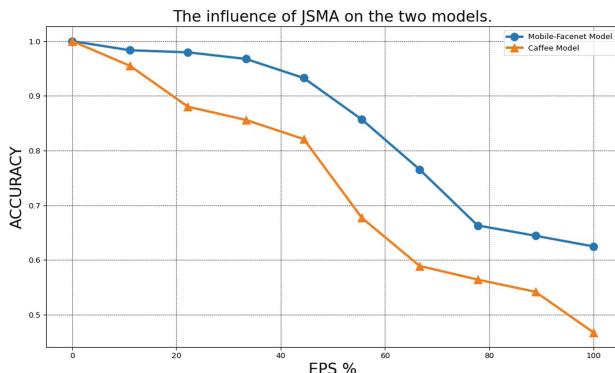
	Accuracy of face detection		Accuracy of Face recognition
TinaFace	100%	MobileFaceNets Caffe	100% 100%



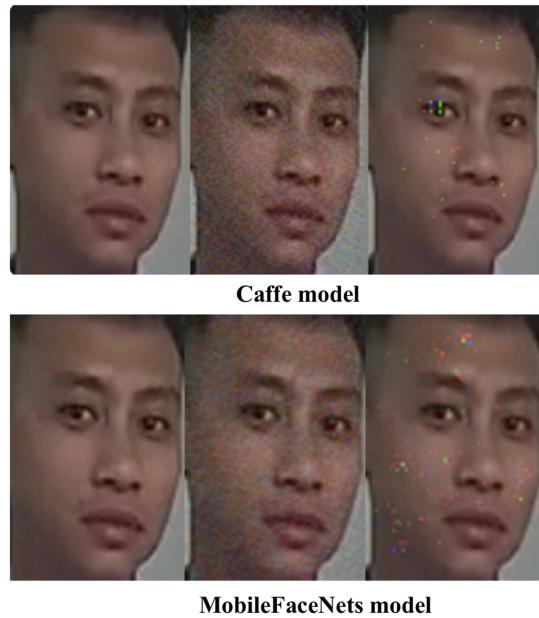
**Fig. 4.** The results of the influence of the FGSM method on two models with our dataset.

With the JSMA method, we present in Fig. 5 the influence of the amount of perturbations into the face recognition with two models. We can witness the same trend when we add noises into the input images. The overall accuracy of the two models decreases with the increases in the amount of adding noises. MobileFaceNets model still performs better than Caffe models.

To be clear about two kinds of methods, we present the visualization results about perturbations given by the FGSM and JSMA methods Fig. 6. Compared to the FGSM method, the JSMA method gave less intensity of perturbation. However, the overall accuracy of system attacked by the JSMA method never fell down into 0% due to a small number of perturbed features.



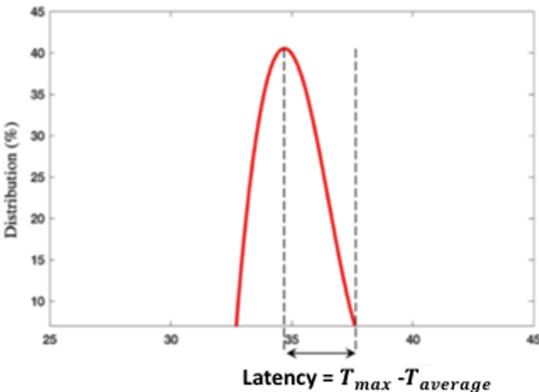
**Fig. 5.** The results of the influence of the JSMA method on two models with our dataset.



**Fig. 6.** An visualization example of perturbations was produced by two models Caffe and MobileFaceNets model. In the first column is the input image, the middle column is the adversarial image by the FGSM method and the last column is the adversarial image by the JSMA method.

In addition, we report here our architecture for testing the physical changes from input images when adding noise: CPU Intel® Core i7-10750H 6 cores 12 threads 4.9 GHz. With 30 videos at the HD resolution ( $1080 \times 810$ ), we increase the amount of adding noises into the face regions, then measure the changes of input image volume. The biggest gap between the original image and the

added noise image is less than 2%. It could be explained that our approach is only to put noise into the face region from the whole image. Moreover, with the time-processing aspect we measure the latency of added noise videos as shown in Fig. 7. We can see that it is less than 15% of the latency. This still guarantees that the system could process in the real-time [23].



**Fig. 7.** The latency of the added noise videos in the face recognition system with  $T$  is processing time in millisecond (ms),  $T_{max}$  is the maximal processing time while  $T_{average}$  is the average processing time.

## 5 Conclusions and Perspectives

In the paper, we conducted a performance evaluation of the face detection and recognition processes using the well-known models: the TinaFace for the face detection step and the Caffe and MobileFaceNets models for the face recognition step. Then, we measure how perturbation given by the FGSM and JSMA methods affects the overall accuracy of two face recognition models.

We can witness the overall accuracy significantly reduced when more noises were added. Although the JSMA method created a slight amount of noise and looked like sophisticated, the effect of the FGSM method is still better in our proposed dataset.

As a perspective, the strategy to reinforce the accuracy of the recognition models could be taken into account. Adversarial samples, which are produced by the JSMA and FGSM methods, are likely retrained the recognition models. Moreover, we would consider the Generative Adversarial Networks (GAN) to create different kinds of perturbations to attack on the face detection and recognition processes and to estimate the reliability of these processes.

## References

- He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)

2. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: NeurIPS, pp. 3104–3112 (2014)
3. Vaswani, A., Shazeer, N., Parmar, N.: Attention is all you need. In: NeurIPS, pp. 5998–6008 (2017)
4. Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J.: Intriguing properties of neural networks. preprint [arXiv:1312.6199](https://arxiv.org/abs/1312.6199) (2013)
5. Serban, A., Poll, E., Visser, J.: Adversarial examples on object recognition: a comprehensive survey. CSUR **53**(3), 1–38 (2020)
6. Kurakin, A., Goodfellow, I., Bengio, S.: Adversarial examples in the physical world. arXiv preprint [arXiv:1607.02533](https://arxiv.org/abs/1607.02533) (2016)
7. Carlini, N., Wagner, D.: Towards evaluating the robustness of neural networks. In: SP, pp. 39–57 (2017)
8. Eykholt, K., Evtimov, I., Fernandes, E.: Robust physical-world attacks on deep learning visual classification. In: CVPR, pp. 1625–1634 (2018)
9. Goodfellow, I.J., Shlens, J., Szegedy, C.: Explaining and harnessing adversarial examples. [arXiv:1412.6572](https://arxiv.org/abs/1412.6572) (2014)
10. Tramèr, F., Kurakin, A., Papernot, N., Goodfellow, I.: Ensemble adversarial training: attacks and defenses. preprint [arXiv:1705.07204](https://arxiv.org/abs/1705.07204) (2017)
11. Papernot, N., McDaniel, P., Jha, S.: The limitations of deep learning in adversarial settings. In: EuroSP, pp. 372–387 (2016)
12. Moosavi-Dezfooli, S.M., Fawzi, A.: Deepfool: a simple and accurate method to fool deep neural networks. In: CVPR, pp. 2574–2582 (2016)
13. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: CVPR 2001, vol. 1 (2001)
14. Harvey, A., Dazzle, C.V.: Camouflage from face detection. Master's thesis, New York University (2010)
15. Yamada, T., Gohshi, S., Echizen, I.: Privacy visor: method based on light absorbing and reflecting properties for preventing face image detection. In: ICSMC, pp. 1572–1577 (2013)
16. Feng, R., Prabhakaran, B.: Facilitating fashion camouflage art, pp. 793–802. ACM (2013)
17. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. PAMI **19**(7), 711–720 (1997)
18. Sharif, M., Bhagavatula, S., Bauer, L.: Accessorize to a crime: real and stealthy attacks on state-of-the-art face recognition, pp. 1528–1540. ACM (2016)
19. Sharif, M., Bhagavatula, S., Bauer, L.: A general framework for adversarial examples with objectives **22**(3), 1–30 (2019)
20. Zhu, Y., Cai, H., Zhang, S., Wang, C., Xiong, Y.: Tinaface: strong but simple baseline for face detection. [arXiv:2011.13183](https://arxiv.org/abs/2011.13183) (2020)
21. Chen, S., Liu, Y., Gao, X., Han, Z.: MobileFaceNets: efficient CNNs for accurate real-time face verification on mobile devices. In: Zhou, J., et al. (eds.) CCBR 2018. LNCS, vol. 10996, pp. 428–438. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-97909-0\\_46](https://doi.org/10.1007/978-3-319-97909-0_46)
22. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S.: Caffe: convolutional architecture for fast feature embedding. preprint [arXiv:1408.5093](https://arxiv.org/abs/1408.5093) (2014)
23. Nguyen Dinh, C., Delalandre, M., Conte, D., Pham, T.A.: Fast RT-LoG operator for scene text detection. Journal of Real-Time Image Processing **18**(1), 19–36 (2020). <https://doi.org/10.1007/s11554-020-00942-7>



# Multi Deep Learning Model for Building Footprint Extraction from High Resolution Remote Sensing Image

Ho Trong Anh, Tran Anh Tuan, Hoang Phi Long, Le Hai Ha,  
and Tran Ngoc Thang<sup>(✉)</sup>

School of Applied Mathematics and Informatics, Hanoi University of Science  
and Technology, No. 1 Dai Co Viet, Ha Noi, Viet Nam

Anh.HT211306M@sis.hust.edu.vn, {ha.lehai,thang.tranngoc}@hust.edu.vn

**Abstract.** 3D city modeling is a new development trend in cartography that has a lot of practical and scientific value. The project necessitates the extraction of a building footprint using remote sensing images. This research examined how to solve the Building Footprint problem using automatic segmentation methods. We reviewed popular segmentation models as Mask-RCNN, U-net, and U2-net, and developed two multi-models that generated more stable and good results than the single models.

**Keywords:** Segmentation · Convolution neural networks · Building footprint · Remote sensing

## 1 Introduction

Building footprint extraction from huge remote sensing images is a challenging endeavor. In recent years, high-tech businesses have made significant investments in surveillance equipment with high resolution, providing a rich supply of image data for remote sensing. Automatically detecting the footprint of a structure has never been easier thanks to recent developments in computer vision.

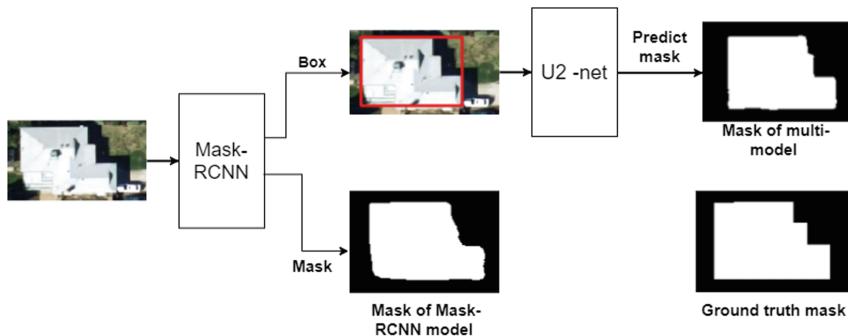
In recent years, many powerful segmentation models, such as U-net, U2-net, and Mask-RCNN, have been created. Through experiments [1], U-net has demonstrated its potential in the Building Footprint problem. Wei et al. [2] advocated using the U2-net model to create an accurate position and exact building outline. To tackle the problem, Mask-RCNN has been combined with post-processing [3] or preprocessing [4] stages. These studies are helping to solve the problem of the Building Footprint, but more research is needed to improve and validate their accuracy and usefulness.

In remote sensing images, a building's pixels are substantially larger than the  $28 \times 28$  pixels of a mask produced by Mask-RCNN, hence the sampling results are less detailed. When we replaced the Mask-RCNN branch with U-net and

---

T. A. Tuan and H. P. Long—These authors contributed equally to this work.

U2-net, we were able to generate two multi-models and compare the results to single segmentation models (Fig. 1).



**Fig. 1.** Mask-R CNN and U2-net combined to form a generalized multi-model. The similarity between the Multi-model mask and the ground truth mask was obvious, which confirms the effectiveness of the Multi-model.

## 2 Method

### 2.1 U-net

The U-net model developed by Olaf Ronneberger et al. aims to partition cells from biomedical imaging. The model architecture consists of two main branches: the downsampling path and the upsampling path in the same way as the encoding and decoding process. The downsampling path is responsible for learning the features from the input image like a conventional CNN network, with the key trait that each time it reduces, the number of filters rises correspondingly such as Resnet, VGG16, etc. As for the upsampling path, its job is to restore the size after the downsampling path to its original size. When it comes to the upsampling approach, its primary function is to restore the size to its original size after downsampling. The model's downsampling path may help reduce information loss by reusing the data that was previously used in the downsampling process. We chose available architectures such as Resnet50 from the pre-trained model ImageNet when choosing the initialization architecture for the research process with a change in activation function, that was ReLU.

### 2.2 Mask R-CNN

Recently, many deep learning models have been presented, mainly for object classification. These models have many different approaches, such as SSD bounding box approach, box-based object search on different aspect ratios, and scale per image location. Or U-net approaches in the direction of finding masks of objects

in the image. These methods each have their advantages and disadvantages that seem to complement each other. There is a model that approaches object classification in both ways and is called Mask R-CNN. This is an extension of the Fast R-CNN model by adding a branch to predict the object's mask based on parallelism with bounding box detection. This model is quite simple to train, and the cost is not too much greater than the Fast R-CNN model.

### 2.3 Multi-model

The U-net model's identification of dense objects is quite difficult, but it has high accuracy when detecting sparse objects. In contrast, with the Mask R-CNN model, we get the ability to identify dense areas accurately. Therefore, when combining the two models, that will give us an expectation of the model's efficacy for the Building Footprint problem. To accomplish this, we combined two models: the U-net and the Mask R-CNN. During the training phase, we merged the Mask-RCNN model outputs into the identified object branch, which were utilized as inputs to the U-net model. Although we expected a multi-model between the Mask-RCNN model and the U-net model, which was made effective, the evaluation metrics are contrary. So, we replaced the U-net model with the U2-net model, which improved the disadvantages that the two original models (the U-net model and the Mask-RCNN model) have. When compared to applying each model independently, the results on the test set demonstrated that the two integrating models considerably modify the AP value at the IoU thresholds, as shown in Fig. 3. The difference between the mask of the Mask-RCNN model and the multi-model is shown in Fig. 2.



**Fig. 2.** Top left: Input image. Top right: Ground truth mask. Lower left: Mask was predicted by Mask-RCNN model. Lower right: Mask was predicted by multi-model (Mask-RCNN model combined with U2-net model).

## 3 Experiments

### 3.1 Datasets

This research was conducted on three different datasets: SUNNY VALE, USA UAV, and VN UAV. The SUNNY VALE dataset [5] contains High-Resolution Orthoimagery images with a resolution of 30 cm. For the training/test/validation set, we picked 16 images with a size of  $5000 \times 5000$ , split them into images with sizes of  $256 \times 256$ , and randomly divided them with a scale of  $6 : 2 : 2$ . Furthermore, the data label was obtained from the building dataset on the website OSM (Open Street Map) [6]. The following dataset, USA UAV, is a collection of UAV images from several locations in the United States, including Santa Ana, Visalia, California, and Orem, Utah. They were labeled by the Skymap Global company's label and were utilized for research purposes only. These images were separated into  $512 \times 512$  images that included 3600 images for the training set, 1200 images for the validation set, and 1200 images for the test set, with resolution quality ranging from 7.5 cm to 15 cm. Finally, the VN UAV (Viet Nam UAV) dataset was also provided by the Skymap Global company. They were split into images with dimensions of  $512 \times 512$ , which included 3000 images for the training set, 1000 images for the validation set, and 1000 images for the test set, with a resolution quality of 10 cm.

### 3.2 Evaluation Metrics

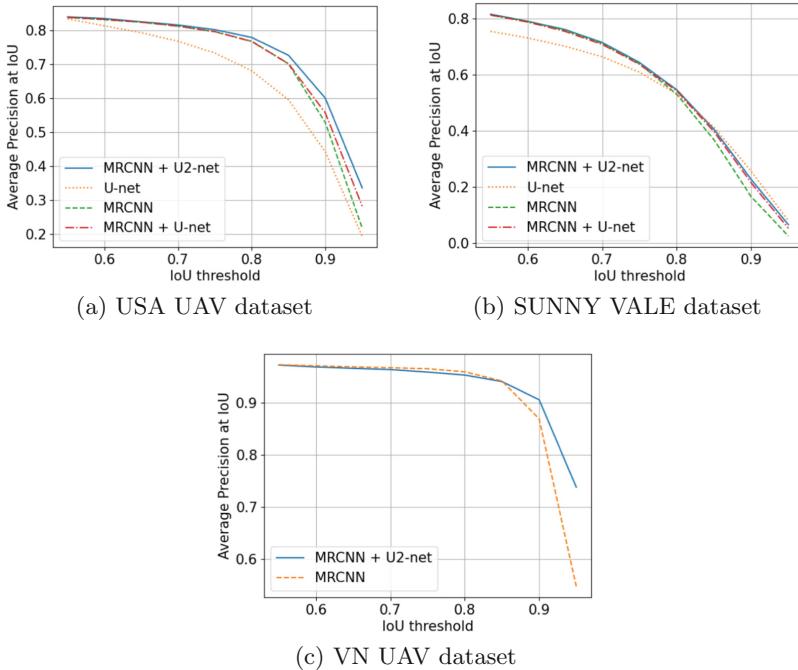
Before calculating evaluation metrics, we excluded objects that were located on the contour of the image and were recognized more than once by the non-maximum-suppression method. We used Kaggle's contest-based evaluation metrics [7] including mAP and mAR. Both mAP and mAR were averaged by AP values at IoU thresholds, which were in  $[0.55 : 0.05 : 0.95]$ . At the same AP's IoU criteria, the accuracy and recall were likewise calculated with a batch size of 4.

## 4 Results

Table 1 show evaluated results on HP-Z800 workstation with configuration: 02 CPU Intel Xeon Processor X5650, 32 GB RAM, GPU Nvidia GTX 1080 Ti 11 GB.

**Table 1.** Results of models on 3 datasets with mAP, mAR in IoU range [0.55 : 0.05 : 0.95]. The entries are blank because of the non-converging model.

Model	USA UAV		SUNNY VALE		VN UAV	
	mAP	mAR	mAP	mAR	mAP	mAR
U-net	0.65	0.615	0.526	0.52	–	–
MRCNN	0.701	0.787	0.533	0.54	0.906	0.903
MRCNN+U-net	0.712	0.795	0.544	0.55	–	–
MRCNN+U2-net	<b>0.728</b>	<b>0.813</b>	<b>0.551</b>	<b>0.56</b>	<b>0.930</b>	<b>0.927</b>



**Fig. 3.** Visualizing the AP value at the IoU thresholds on 3 datasets.

Construction works are typically spaced and moderate in countries with strong construction planning, such as the United States, allowing the U-net model to function well on this data. In contrast, in countries with poor building planning, such as Vietnam, construction activities are near and dense, making the combination of the Mask R-CNN and the U-net ineffective. The adoption of the U2-net model to replace the U-net model, later on, improves identification in instances when buildings are hidden by trees or are close together. The predicted mask of the Multi-model, as shown in Fig. 4, exhibits accuracy when compared to the mask of Mask R-CNN model, which substantially assists further work when using geometry correction algorithms to obtain the best linear outlines for objects.



**Fig. 4.** The first top-left image is the input image, second and third is the Mask R-CNN predictions and geometry correction algorithm predictions respectively. The first lower-left image and second are the Multi-model predictions and corresponding geometry correction algorithm predictions, and the third image is the expected output image.

## 5 Conclusion

Except for U-net, all of the empirical models perform well on various data sets. Multi-models provided significant improvements, according to the findings. As a result, we have proposed an effective solution to the Building Footprint problem, which may be used as a foundation for post-processing procedures such as geometry correction and proving its multi-model potential application for other tasks.

## References

1. Emek, R.A., Demir, N.: Building detection from SAR images using UNET deep learning method, pp. 215–218 (2020). <https://doi.org/10.5194/isprs-archives-XLIV-4-W3-2020-215-2020>
2. We, X., et al.: Building outline extraction directly using the u2-net semantic segmentation model from high-resolution aerial images and a comparison study. *Remote. Sens.* **13**, 3187 (2021)
3. Zhao, K., Kang, J., Jung, J., Sohn, G.: Building extraction from satellite images using mask R-CNN with building boundary regularization. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 242–2424 (2018). <https://doi.org/10.1109/CVPRW.2018.00045>
4. Qinzhe, H., Yin, Q., Zheng, X., Chen, Z.: Remote sensing image building detection method based on mask r-cnn. *Complex Intell. Syst.* (2021). <https://doi.org/10.1007/s40747-021-00322-z>

5. USGS: Sunnyvale uav images. <https://earthexplorer.usgs.gov/>
6. OSM: Sunnyvale uav labels. <https://www.openstreetmap.org/>
7. Kaggle: 2018 data science bowl (2018). <https://www.kaggle.com/c/data-science-bowl-2018>



# Optimizing Energy Consumption in Smart Homes Based on OCF Standard Connectivity

Anam Nawaz Khan<sup>1,3,4</sup>, Atif Rizwan<sup>1,3,4</sup>, Wenquan Jin<sup>2,3,4</sup>, Rashid Ahmed<sup>2,3,4</sup>, Anh Ngoc Le<sup>2,3,4</sup>, and Dohyeun Kim<sup>2,3,4(✉)</sup>

<sup>1</sup> Department of Computer Engineering, Jeju National University, Jeju City, Republic of Korea  
[{anamnawaz,atifrizwan}@jejunu.ac.kr](mailto:{anamnawaz,atifrizwan}@jejunu.ac.kr)

<sup>2</sup> Big Data Research Center, Jeju National University, Jeju City, Republic of Korea  
[{wenquan.jin,kimdh}@jejunu.ac.kr}, rashid.ahmad@cuiatk.edu.pk](mailto:{wenquan.jin,kimdh}@jejunu.ac.kr)

<sup>3</sup> Swinburne Vietnam, FPT University, Hanoi, Vietnam  
[ngocla2@fe.edu.vn](mailto:ngocla2@fe.edu.vn)

<sup>4</sup> Department of Computer Engineering and Research Center of Advance Technology, Jeju National University, Jeju City, Republic of Korea

**Abstract.** In this paper, an OCF connectivity based smart home energy management controller has been proposed. The problem has been formulated using wind driven optimization algorithm that focuses the occupant comfort and energy savings. The proposed optimal control mechanism utilizes the energy consumption data, constraints involving the load limits and preference of the occupants and provisions optimal actuator set points and a suitable time frame for the house hold appliance to be operational. The proposed framework provides excellent results in terms of energy cost minimization and occupant comfort satisfaction by achieving energy saving percentage of 37.14%.

**Keywords:** Energy optimization · Energy management · Smart homes · OCF

## 1 Introduction

Optimizing the energy use in smart environments has evolved from technological fringe and became today's reality. Today's Smart home is equipped with smart appliances where smart grid, smart IoT networks and renewable energy sources converged together requiring energy optimization as a de-facto necessity for sustainable environment [1]. The current situation demands a smart home energy controller that has the ability to achieve energy efficiency, ultimately leading to resource wastage. Since energy is a scarce resource, and there exist no reliable means to store it [2]. Optimizing the use of energy can be immensely beneficial for economic development of the country. However, developing a smart home energy management controller according to the desires and need of the occupant based on grid response is a challenging task due to conflicting objectives.

Existing works of the optimal energy use in smart home mainly involve [3] where an integer liner programming model is devised to minimize the energy cost based on

event driven energy controller. Another study on energy optimization proposed a proactive energy control mechanism based on fuzzy learning rules and adaptive approaches [4]. Likewise, several studies such as [5] and [6] proposed mono-objective optimization approaches for optimal energy management in smart home. Another work developed an integer multi-objective model for analyzing the cost of planning and maintenance of power generators. Extensions of similar work has been proposed by [7] where several multi-objective optimization models are developed for managing the electrical loads in smart homes with varying constraints such as temperature, comfort and technological specifications. Other notable contributions are proposed by [8] and [9] based on various theoretical loads and price tariffs for real time energy demand optimization employing integer linear programming. Another study presented a load classification scheme with various comfort use cases. Article [10] developed a real time energy optimization mechanism based on occupant comfort.

The rest of the paper is organized as follows; Sect. 2 presents the problem formulation and mathematical model of the proposed optimal control framework. In Sect. 3 the proposed architecture is described in details, while Sect. 4 discusses the results of the proposed study.

## 2 Problem Formulation

Finding an optimal dispatch time for smart home appliances and electrical loads is complicated and challenging activity involving multi-parameters objectives and constraints. the main goal of the solution must focus how the electrical loads must be scheduled in the presence of several constraint, decision variables, occupant comfort and discomfort schemes thus tackling the said multi-objective in an efficient manner. The proposed study employs time of use tariffs defined by national energy agency republic of Korea. The occupant comfort is quantified as the minimum distance between the occupant desired operational time of the appliance against the actual start time. The smart home occupant preferences are acquired through smart phone based on the variable weighted aggregation function (Table 1).

$$O_1 = \sum_{E_c=1}^{E_c} \sum_{j=ST_k}^{ET_k-D_k} \left( \sum_{l=j}^{j+(D_k+1)} M_p[l-j] E_c[l] \right) \lambda_{kl} \quad (1)$$

$$O_2 = \sum_{k=1}^k \sum_{l=ST_k}^{ET_k-D_k} \left( OC_k \sqrt{l - O_{st_k}^2} \right) \lambda_{kl} \quad (2)$$

Equation (1) is formulated for energy cost while Eq. (2) is formulated for managing occupant comfort. Both these equations can be summarized in a complimentary combination as follows:

$$O_{\cos t} = \gamma(O_1) + (1 - \beta) \times O_2 \quad (3)$$

**Table 1.** Nomenclature.

Nomenclature	
$E_c$	Energy cost for a certain time period
$ST_k$	start time (Minimum) of kth appliance
$ET_k$	End time (maximum) of kth appliance
$D_k$	Time duration of the kth appliance
$M_p$	Mean power of kth appliance
$\lambda_{kl}$	Decision variable for kth appliance
$OC_k$	Occupant comfort for kth appliance
$O_{st_k}^2$	Optimal start time for kth appliance
$NL$	No. of loads
$P_z$	Peak limit of zth time-slot

Subject to the following constraints:

$$\sum_{j=ST_k}^{ET_k-D_k} \lambda_{kl} = 1 \quad (4)$$

$$\sum_{k \in NL_z} \left( \sum_{l=z-(D_k-1)}^{z-(z-ET_k+D_k)} M_p[z-l] \lambda_{kl} \right) \leq P_z \quad (5)$$

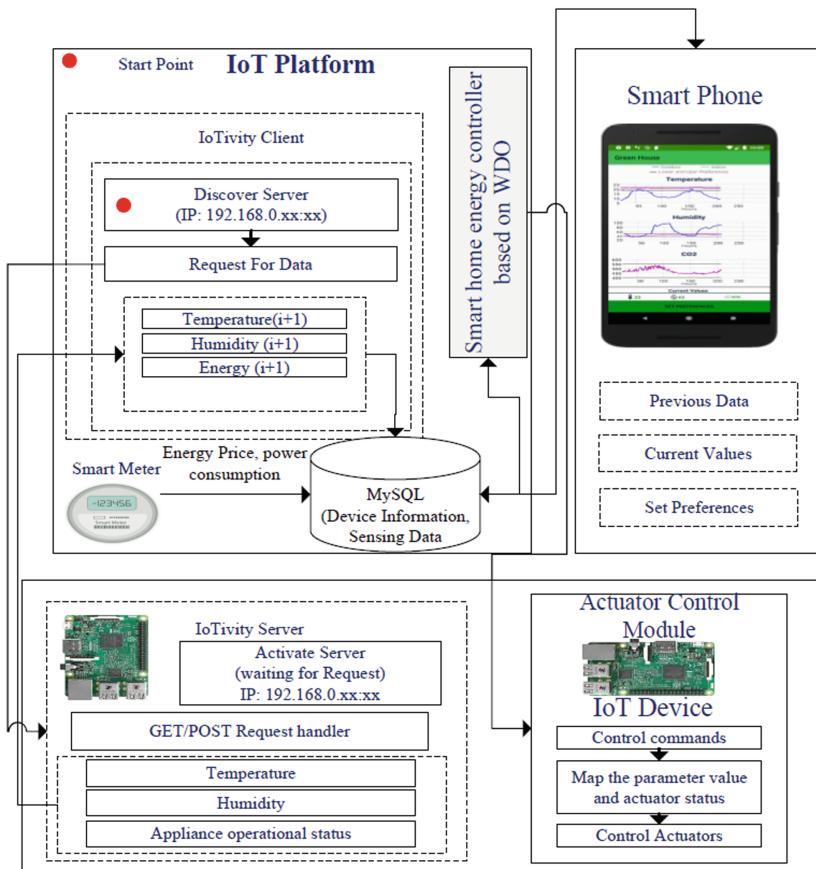
$$\lambda_{kl} \in [0, 1]$$

The energy cost is computed through Eq. (1) as a minimization objective function while the comfort is quantified in terms of distance from expected time to start and actual operational time. The objective function is formulated as minimization of cost function formulated as association of Eq. (1) and (2). The assignments of weights have been done using Eq. (3) while constraints are specified in Eq. (4) and (5) that impose limits for each appliance to possess only one start time. The proposed constraints limit the loads to not cross the minimum and maximum start and end time respectively. Constraints are set to ensure the appliance to be operational such that the smart home loads should not exceed the peak limit. Moreover, it imposes limits the overall appliance loads to not exceed the peak limit load beginning from zth start time till last possible end time of the appliance operation and time duration considering all decision variable belonging to binary classes.

### 3 Smart Home Energy Management Controller with OCF Standard Connectivity

The architecture of proposed optimal energy management mechanism based on OCF standard connectivity is presented in Fig. 1. The proposed framework manages the

energy consumption of controllable appliances that include heaters and chillers that can be turned on and off for a specific time period while other loads are specified as detectable loads having a specific power consumption these including smart home lighting. The smart home energy controlling mechanism defines when a certain appliance can be operational based on the information from their respective actuators that is stored in the central resource management information database. The proposed framework provides the actuator control commands based on the objective function implemented through optimization model based on wind driven optimization algorithm (WDO). The communication functions among the entities of smart home are based on OCF standard connectivity using COAP protocol. The working overview of smart home energy controller is as follows.



**Fig. 1.** Smart home controller architecture

Communication with smart home actuating units based on OCF standard for acquiring the power consumed by the appliances and send actuating commands to operate the actuators. Inspecting and analyzing the operational status of the appliances and their

respective loads. Moreover, receive the occupant preferences through smartphone and check the expectancy of future loads. The optimal control mechanism sends control commands to controllable appliance with optimal set points.

## 4 Proposed Architecture and System Definition

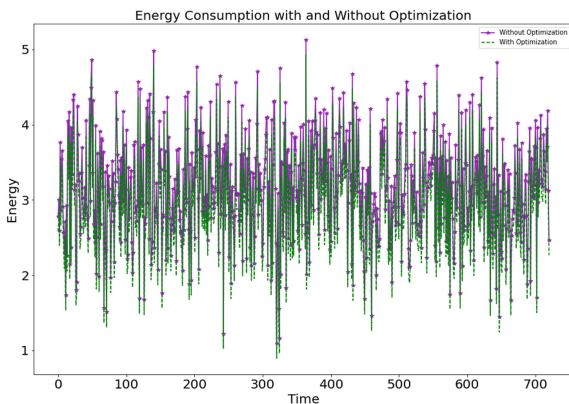
Firstly, to develop the proposed optimal control framework we begin by deploying the IoT device in the smart home networks to provide monitoring, control and environment update services. The employed IoT device is a raspberry pi model b with Ubuntu server. The IoT device host the environmental resources acquired through sensors and sends the requested data to OCF client. Iotivity version 2.2.2 java version is configured on both IoT device as well as IoT platform for providing OCF services. The environmental data is stored in the MySQL database which reside in the AIoT Platform to be retrieved by the smart phone client. IoT device is an OCF entity comprising of OCF server that communicates to OCF Client residing in IoT platform. Various scenarios have been tested based on appliances, stages, cycles and price tariffs. stages can be defined as appliances performing specific functions while cycle involve an all stages of operations of an appliance. The appliances considered for energy optimization in smart home are listed in Table 2 with their detailed characteristics including power consumption, time zone and expected hours of operations recorded through smart meters.

**Table 2.** Appliance description.

Appliance type	Appliance Name	Mean Power	Peak power	Up-start time	Schedule start time	Schedule End time	$\Delta T$	Appliance type
Controllable	Heater	1.20	1.30	21:00	16:00	24:00	250	Controllable
Controllable	Chiller	1.00	1.20	20:00	16:00	24:00	250	Controllable
Detectable	Smart lights	0.16	0.35	18:00	17:30	23:30	270	Detectable

The optimization model restricts the smart home occupants to not exceed the load limits of 3.5 kW by imposing constrains on the user. The load limit threshold is set by the grid operator depicting the maximum allowed energy demand by the user. The multi objective optimization aims to reduce the cost and maximize the user comfort based on wind driven optimization. The wind driven optimization take user preferences, current energy consumption, environmental parameters and operational status of the appliances as input and provides the optimal actuator set point and optimal operational schedule for the appliances. The energy pricing has been considered on the basis of peak and off-peak loads. Firstly, we considered time of use tariff that involve hourly basis pricing policy. The environmental data is acquired from sensors attached to the device in smart home using client server communication.

The energy consumption data, operational status and load characteristics are collected using smart-meters. Figure 2 presents the comparison optimization results of energy consumption with and without optimization.



**Fig. 2.** Comparison of energy consumption optimized vs unoptimized.

## 5 Conclusions

This paper proposed an optimal control mechanism for managing occupant comfort and energy consumption based on OCF standard connectivity. The occupant preferences are acquired through smartphone. Smart meter is used to record the energy consumption, operational status, time of use pricing, load characteristics and time duration of appliance usage. The recorded information along with other input parameters is fed to optimization model which act as smart home energy controller. The objected function is implemented by the wind driven optimization algorithm based on constrains. The experimental results show improved performance after applying the optimization techniques. The energy saving percentage of the proposed system is 37.14% that proves the effectiveness of the proposed study.

**Acknowledgement.** This research was supported by Energy Cloud R&D Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (2019M3F2A1073387), and this work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2021-0-00188, Open-source development and standardization for AI enabled IoT platforms and interworking)). Any correspondence related to this paper should be addressed to Dohyeun Kim.

## References

1. de Souza Dutra, M.D., Anjos, M.F., Le Digabel, S.: A realistic energy optimization model for smart-home appliances. *Int. J. Energy Res.* **43**(8), 3237–3262 (2019). <https://doi.org/10.1002/er.4454>
2. Frontiers | Investigation of Smart Home Energy Management System for Demand Response Application | Energy Research. <https://www.frontiersin.org/articles/10.3389/fenrg.2021.772027/full>. Accessed 04 Feb 2022
3. Di Giorgio, A.: An event driven Smart Home Controller enabling consumer economic saving and automated Demand Side Management. *Appl. Energy* **96**, 92–103 (2012). <https://doi.org/10.1016/j.apenergy.2012.02.024>

4. Abbas, S., Fatima, A., Asif, M., Saleem, M.: Energy optimization in smart homes by using fuzzy inference system, January 2019
5. Yu, L., Xie, D., Huang, C., Jiang, T., Zou, Y.: Energy optimization of HVAC systems in commercial buildings considering indoor air quality management. *IEEE Trans. Smart Grid* **10**(5), 5103–5113 (2019). <https://doi.org/10.1109/TSG.2018.2875727>
6. Asare-Bediako, B., Ribeiro, P.F., Kling, W.L.: Integrated energy optimization with smart home energy management systems. In: 2012 3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe), October 2012, pp. 1–8 (2012). <https://doi.org/10.1109/ISGTEurope.2012.6465696>
7. Antunes, C., Martins, A., Brito, I.: A multiple objective mixed integer linear programming model for power generation expansion planning. *Energy* **29**, 613–627 (2004). <https://doi.org/10.1016/j.energy.2003.10.012>
8. Cho, S.S.: Energy-efficient smart home system: optimization of residential electricity load management system, p. 45
9. Casella, I.R.S., Sanches, B.C.S., Filho, A.J.S., Capovilla, C.E.: A dynamic residential load model based on a non-homogeneous poisson process. *J. Control, Autom. Electr. Syst.* **27**(6), 670–679 (2016). <https://doi.org/10.1007/s40313-016-0269-8>
10. Rasheed, M.B., et al.: Energy optimization in smart homes using customer preference and dynamic pricing. *Energies* **9**(8) (2016). Art. no. 8. <https://doi.org/10.3390/en9080593>



# Physical Therapy Exercise Design for Lower Limb Rehabilitation Robot Using Character's Animation Clips

Minh-Duc Duong<sup>1</sup>, Ngoc-Thuan Ly<sup>1</sup>, Anh-Tu Hoang<sup>1</sup>, Van-Vuong Dinh<sup>1,2</sup>,  
and Quy-Thinh Dao<sup>1(✉)</sup>

<sup>1</sup> Hanoi University of Science and Technology, Hanoi, Vietnam  
[{duc.duongminh,thinh.daoquy}@hust.edu.vn,](mailto:{duc.duongminh,thinh.daoquy}@hust.edu.vn)  
[{thuan.ln174249,tu.ha174316}@sis.hust.edu.vn](mailto:{thuan.ln174249,tu.ha174316}@sis.hust.edu.vn)

<sup>2</sup> Hanoi Vocational College of High Technology, Hanoi, Vietnam  
[vuongdv@hht.edu.vn](mailto:vuongdv@hht.edu.vn)

**Abstract.** In recent years, rehabilitation robots have been widely applied in the medical field. Rehabilitation robots help promote the reorganization of functions and regenerate the nervous system, create reflexes to the brain, and limit muscle atrophy due to sedentary. This paper uses Unity software to design sample physical therapy exercises for a lower limb rehabilitation robot with the goal: doctors perform a few movements of the exercise and this exercise data are recorded with a 3D camera, through specific processing and calculation on Unity software will obtain this exercise in the form of the final character's animation clip. The results show that the experimental physical therapy exercises can be built according to different patients' body conditions.

**Keywords:** Rehabilitation robot · Blender · Unity3D · Animation & Rigging

## 1 Introduction

Physical therapy is a method of treatment for patients without drugs, with many outstanding advantages, especially for those with occupational accidents and congenital disabilities [1]. After various injuries, the muscles can be weakened or restricted, leaving you with pain and impaired function. Physical therapy will help you manage pain, restore fitness and strength, and facilitate overall recovery.

In the coming years, the rehabilitation robot market is forecast to grow rapidly [2], prompting technology companies to develop more advanced robots that serve humans. Post-stroke patients will suffer severe sequel that take a long time to recover their mobility. Patients use recovery robots that help them to recover faster, be proactive, and be independent.

---

M.-D. Duong, N.-T. Ly, A.-T. Hoang, V.-V. Dinh and Q.-T. Dao—These authors contributed equally to this work.

Following this forecast, there have been various research in robotic field to develop the rehabilitation robot system, especially for lower limb rehabilitation [3,4]. The main work is to build a rehabilitation robot that can force/assist patients to move following preset trajectories, or move freely. In these system, the robot is controlled to track the reference trajectory precisely, with safe and appropriate interaction between the patient and robot. For the safety and comfort interaction, the first problem that must consider is the actuators of rehabilitation robot. In the commercial market, rehabilitation robots are usually driven by electric motors. However, these systems have some major concerns related to high costs and the low power/weight ratio of motorized actuators. Due to this fact, in the medical field, there are more and more studies on the application of Pneumatic Artificial Muscles (PAMs) to rehabilitation robots [5,6]. PAMs possess many advantages of an actuator that is light in weight, cheap in price, and easy to machine.

The second problem with the system is exercises for the patients who use this rehabilitation robot system. Normally, there are two main types of exercises including passive and assisted exercises. In the assisted exercises, the patients are moved freely or follow pre-defined trajectories and the robot will assist the patients when it recognize the requirements from the patients. In the passive exercises, the robot forces patients to track pre-defined trajectories. To guarantee the recovery of motor functions, the moving trajectories that are very important and they are created by experienced physicians. One of the efficient way to make the trajectory is to record the standard the trajectory and re-generate this trajectory to the rehabilitation robot. In order to do this process, the human pose need to be detected.

In recent years, the development of computer vision and AI allow us to estimate 3D human pose using RGB-D cameras [7,8]. However, 3D human pose detection for assisted and rehabilitation robotic environments needs further improvements to achieve real-time tracking for human motion analysis with enough accuracy [9].

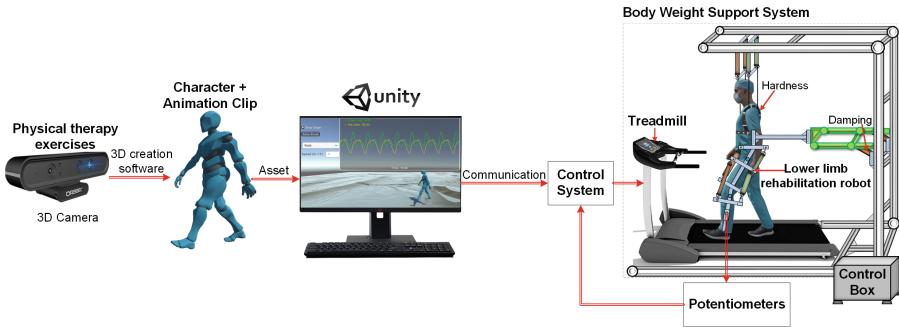
In this paper, in order to create the exercise trajectory for lower limb rehabilitation, Unity3D, a video game engine with the advantages of vivid, intuitive images and rich resources [10] is used. Unity allows creating various rehabilitation exercises with customization levels from complex to easy will help improve the quality of training and create excitement for patients. The software as an intelligent solution makes the output trajectory data quickly processed without the need for digital filters, helping to provide high-quality exercise and muscle recovery performance. The generated trajectories then is applied to a PAMs based rehabilitation robot system [11], to verify the effectiveness of the proposed trajectory generation approach. In summarize, this paper has the following major contributions:

- Apply the character's animation clips to build different physical therapy exercises.
- Using the calculations coded in Unity software calculate joint angle trajectories to serve as sample trajectories for the lower limb rehabilitation robot.

## 2 Prototype Rehabilitation System

The experimental procedure with prototype rehabilitation system is detailed in Fig. 1, specifically as follows:

- **Step 1:** The physician performs physical therapy exercises within the recording range of the 3D camera (such as Orbbec, Kinect, etc.) for a short time - about 1 min. Exercise data will be recorded through this camera.
- **Step 2:** Recorded data of physical therapy exercises collected from step 1 is sent to the 3D creation software for further processing. The character design and animation clips corresponding to each exercise are done in this step.
- **Step 3:** The character's animation clips are imported into Unity software as assets. At this step, hip and knee's joint angle trajectories are calculated.
- **Step 4:** Final data collected in step 3 are packaged and sent to robot's trajectory controller. This controller will control trajectories of the lower limb rehabilitation robot in the body weight support system. The patient is guided or assisted by robot to follow the designed trajectories.



**Fig. 1.** Prototype rehabilitation system.

## 3 Character and Animation Clips Design

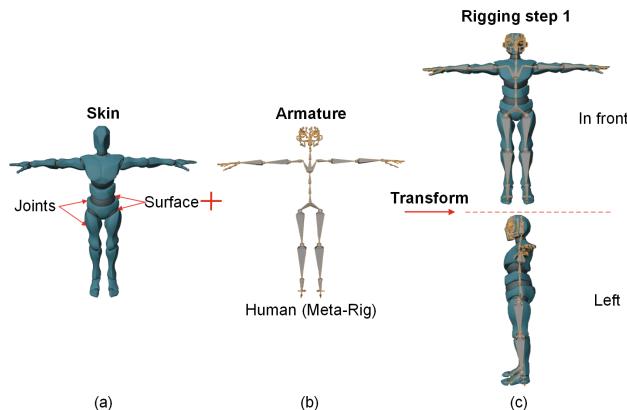
### 3.1 Design in 3D Creation Software

To design 3D characters with animation, this paper uses Blender, a free and open-source 3D creation suite software [12]. Blender supports the entirety of the 3D pipeline modeling, rigging, animation, simulation, rendering, compositing and motion tracking, video editing, and 2D animation pipeline. The process of designing a character with animations is shown in detail in Fig. 2 and Fig. 3.

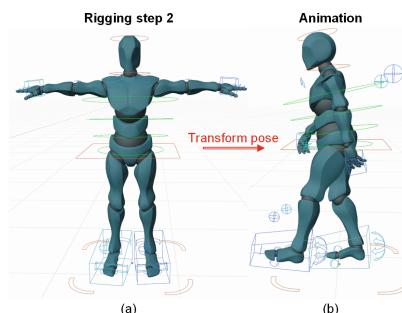
A complete character is designed, including two main components: skin and bones. First, we create a simple skin in the shape of a robot, which consists of joints and surface [13], linked together as shown in Fig. 2a. After final processing and smoothing, the skin obtained will apply rigging tool to create connective

bones and its drivers. Rigging is a general term used for adding controls to objects, typically for animation. Rigging consists of two steps: bone-skin bonding and driver creation, specifically as follows:

- **Step 1:** Create a human armature - meta rig (skeleton) as shown in Fig. 2b. Then, adjust this skeleton and the skin to connect, and this skeleton fits in the skin. The result is a simple character consisting of skin and bones linked together, as shown in Fig. 2c.
- **Step 2:** Create drivers from the skeleton obtained from step 1; the result is as shown in Fig. 3a. By using drivers, the rig can control many different values at the same time, as well as make some properties automatically updated used on changes elsewhere. For this change to apply to the character's skin, it is necessary to set the parent using **set the object's parenting: with automatic weights**. The final obtained character will be a parent object containing parameters of the child objects: the skin and skeleton with drivers attached. Through which it is easy to create a keyframe animation using pose transformations (move, rotation, scale), detail in Fig. 3b.



**Fig. 2.** Rigging step 1 with components: (a) skin; (b) bones and (c) bone-skin bonding.



**Fig. 3.** Rigging step 2 with: (a) drivers creation and (b) animation design.

The complete design of an animation clip is based on data about poses over time and the length of a physical therapy exercise movement obtained from the 3D camera. From which it is possible to determine the number of keyframes to generate, and the recording function in Blender software will help create a complete animation clip; it includes the attached character and is packaged as an asset. Since the animation clip is a single linear recording and with the use of rigging tool in Blender software, it is smooth, without noise during the playback.

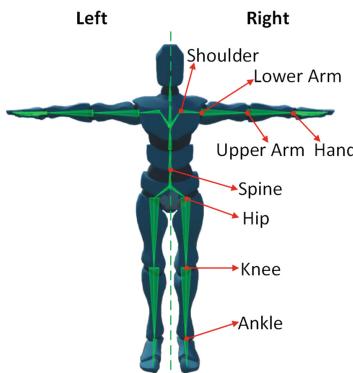
### 3.2 Unity Software Application

Data obtained from Sect. 3.1 is imported as assets, including character information and animation clips into Unity. Animation clips contain information about the character's position, rotation, and scale over time. Each clip can be thought of as a single linear recording, created by external tools such as Blender, Max, Maya, or from motion capture studio systems. However, the movement can be more complicated by combining multiple clips or simply changing each clip, clip timing, and repetition. So the Animator is used to accomplish this. In this paper, physical therapy exercises that require movement are pretty simple, so the Animator has the following tasks: change clips, time a clip, create continuity for clips, and combine clips (if needed). In designing restorative exercises for the lower limbs, the skeleton must follow a standard hierarchy: Hips - Knees - Ankles as shown in Fig. 4 [13].

As the character performs the exercises, the hip and knee joints will be monitored and joints' positions are collected to calculate joint angles.

### 3.3 Joint Angle Calculation

The lower limb rehabilitation robot consists of two joints: knee and hip, the control input is the joint angle values. Through physical therapy exercises in the form of the character's animation clips, based on the position of joints obtained



**Fig. 4.** Body skeleton structure used in Unity software.

from the character's skeleton in 3D, it is possible to calculate the knee and hip joint angle values. These values will be used as the model trajectory for the robot.

Assuming there are two vectors  $\vec{a} = (x_1, y_1, z_1)$  and  $\vec{b} = (x_2, y_2, z_2)$ , the angle between these two vectors is calculated as follows:

$$\cos(\vec{a}, \vec{b}) = \cos(\theta) = \frac{x_1x_2 + y_1y_2 + z_1z_2}{\sqrt{x_1^2 + y_1^2 + z_1^2}\sqrt{x_2^2 + y_2^2 + z_2^2}} \quad (1)$$

From (1), it is possible to calculate the lower limb's joint angles. The knee angle is calculated as the angle between knee-hip vector and knee-ankle vector. The hip angle is calculated as the angle between hip-knee vector and the vertical vector  $\vec{u} = (0, 1, 0)$ .

## 4 Physical Therapy Exercises Design

Physical therapy provides polio and stroke patients with treatments aimed at restoring mobility and function impairments. The primary goal is to restore the patient's inherent resiliency to near-normal activities.

The treatment method presented in this paper is to use the sample physical therapy exercises built from the character's animation clips to control the lower limb rehabilitation robot, maximally supporting the patient to perform the movements of exercises.

**Table 1.** Parameters of physical therapy exercises

Exercises	Exercise cycle	Exercise time	Other parameters
Standing	4–6 (second)	4–5 (minute)	Standing time: 0.5–1 (second) Standing posture: upright or stoop with a stooped angle 10–20 (degree)
Walking	1–2.5 (km/h)	3–4 (minute)	Foot kick angle when walking: 35–70 (degree)
Kneeing	1–2 (second)	2–3 (minute)	Leg kick angle: 70–110 (degree)

Physical therapy exercises are mainly applied to the patient's legs, including basic to advanced exercises that can flexibly change the exercise cycle and exercise time to best suit the patient's condition. This paper refer to four physical therapy exercises from essential to advanced: standing, walking, and kneeing. The details of these exercises are presented below, and their parameters are listed in detail in Table 1.

#### 4.1 Standing Exercise

The first essential exercise in physical therapy exercises is standing. Polio and stroke patients have very weak leg muscles, unable to stand up, so this exercise aims to restore the patient's ability to stand.

First, the patient is arranged to sit in a wheelchair or soft chair, and then the body weight support system will assist him in standing up with an upright posture for a specific time. He then sits back down to the starting position.

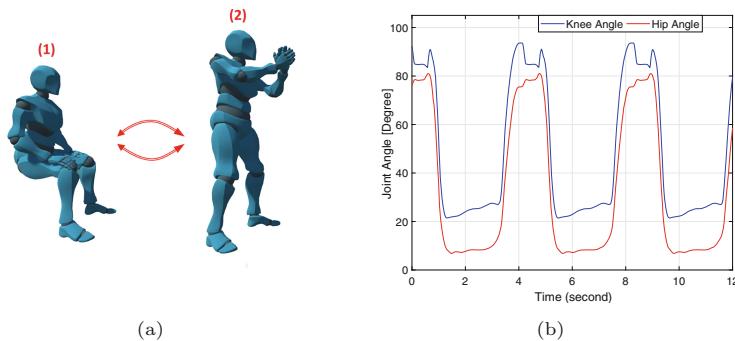
Figure 5 shows state steps of standing exercise and the result of standing exercise with the following parameters: standing posture is upright, and standing time is approximately 1 s.

#### 4.2 Walking Exercise

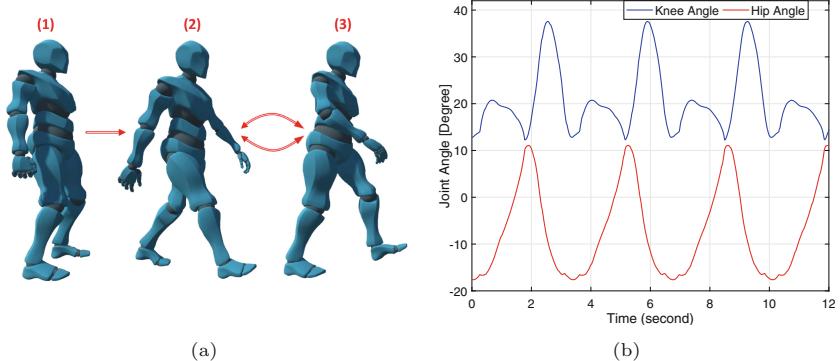
The next essential exercise is walking after the patient has completed the standing exercise. He will perform this exercise on a treadmill.

The patient will be suspended in the air by the body weight support system, about 0.5 m from the treadmill. First, the rehabilitation robot will assist him in performing a walk in the air at speed synchronous with the treadmill's set speed. Next, the body weight support system will help lower his body gradually down the treadmill's conveyor. After landing, the robot assists him to continue walking on the conveyor.

The walking cycle is depicted as shown in Fig. 6a. State 1 of this exercise is standing when the patient is in the air. Then, he will switch to walking mode: right swing phase, then left swing phase, do it sequentially and continuously from state 2 to state 3. Figure 6b shows the result of walking exercise with parameters: treadmill set speed is 1.5 km/h and foot kick angle when walking is about 38°.



**Fig. 5.** Standing exercise: (a) state steps in Unity software and (b) trajectory of joint angles.

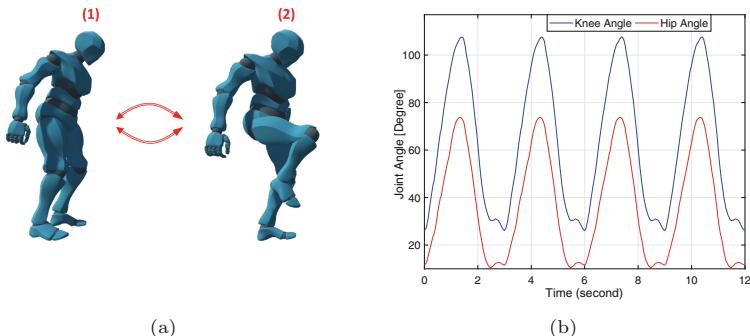


**Fig. 6.** Walking exercise: (a) state steps in Unity software and (b) trajectory of joint angles.

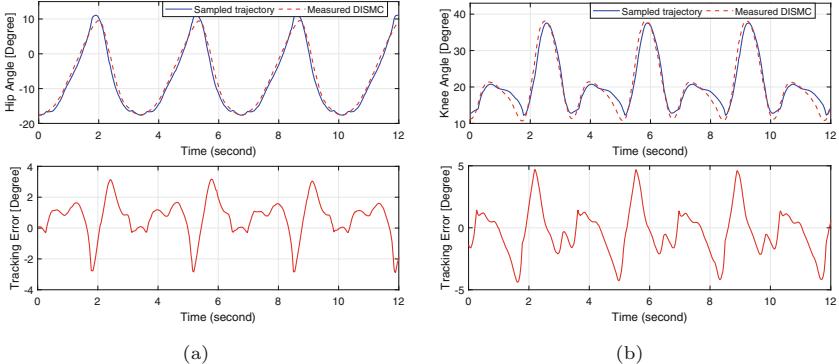
#### 4.3 Kneeling Exercise

After completing the essential exercises, the patient will have access to advanced physical therapy exercises. This exercise helps to improve the patient's leg muscles, avoiding muscle atrophy due to inactivity for a long time.

Figure 7a shows state steps of kneeing exercise: initially, the patient will stand stooped as shown in state 1, then the rehabilitation robot will assist in kicking his leg up, as state 2. After that, he retracts the leg back to standing position (state 1) and ends an exercise cycle. Figure 7b shows an experimental result of kneeing exercise with leg kick angle of about  $110^\circ$ .



**Fig. 7.** Kneeling exercise: (a) state steps in Unity software and (b) trajectory of joint angles.



**Fig. 8.** Experiment results when tracking walking exercise at 1.5 km/h of the treadmill speed: **(a)** hip joint **(b)** knee joint.

## 5 Experimental Results

To verify the efficiency of the proposed exercise generation method, physical therapy exercises with different scenarios were performed. Experiments are carried out with all the exercises presented in Sect. 4 using the robot system in [11]. Through the experimental process, these physical therapy exercises are evaluated as suitable for the lower limb rehabilitation robot. For example, when the robot tracks the walking exercises with 1.5 km/h of the treadmill speed (Fig. 8), the root mean square tracking error is less than  $2.1^\circ$  and the maximum tracking error does not exceed  $5.0^\circ$ . The measured trajectory results of proposed DISMC controller (dash red line) could well track the sample trajectories (blue line). This result is accordance to the LOKOMAT a well known gait training system [14].

## 6 Conclusion

This paper applies animation clips of characters to build model physical therapy exercises suitable for different patient conditions. The proposed approach can create a smooth and flexible trajectories from recorded physician's motion. These exercises are used as reference trajectories to control the lower limb rehabilitation robot. In the future, these approach is applied to create more rehabilitation trajectories and these trajectories are evaluated in practice.

**Acknowledgments.** This research was partially supported by the Asahi Glass Foundation.

## References

1. MedicineNet Inc. [https://www.medicinenet.com/physical\\_therapy/de-finition.htm](https://www.medicinenet.com/physical_therapy/de-finition.htm)

2. Global Industry Analysts Inc. <https://www.strategyr.com/market-report-rehabilitation-robots-forecasts-global-industry-analysts-inc.asp>
3. Diaz, I., Gil, J.J., Sanchez, E.: Lower-limb robotic rehabilitation: Literature review and challenges. *J. Robot.* (2011). <https://doi.org/10.1155/2011/759764>
4. Bradley Hobbs and Panagiotis Artermiadis: A review of Robot-assisted lower limb stroke therapy: unexplored paths and future directions in gait rehabilitation. *Front. Neurorobot.* **14**, 19 (2020)
5. Yeh, T.-J., Wu, M.-J., Lu, T.-J., Wu, F.-K., Huang, C.-R.: Control of McKibben pneumatic muscles for a power-assist, lower-limb orthosis. *Mechatronics* **20**, 686–697 (2010). <https://doi.org/10.1016/j.mechatronics.2010.07.004>
6. Dao, Q.-T., Yamamoto, S.: Assist-as-needed control of a robotic orthosis actuated by pneumatic artificial muscle for gait rehabilitation. *Appl. Sci.* **8**(4), 499 (2018)
7. 3DiVi Inc. <https://nuitrack.com>
8. DeepMotion Inc. <https://www.deepmotion.com>
9. Hernández, Ó.G., Morell, V., Ramon, J.L., Jara, C.A.: Human pose detection for robotic-assisted and rehabilitation environments. *Appl. Sci.* **11**, 4183 (2021). <https://doi.org/10.3390/app11094183>
10. Unity Technologies. <https://assetstore.unity.com>
11. Trinh, M.-C., Do, T.-H., Dao, Q.-T.: Development of a rehabilitation robot: modeling and trajectory tracking control. *ASEAN Eng. J.* (2021)
12. Blender's GNU General Public License (GPL). <https://www.blender.org/>
13. Adobe. <https://www.mixamo.com>
14. Riener, R., Lunenburger, L., Jezernik, S., Anderschitz, M., Colombo, G., Dietz, V.: Patient-cooperative strategies for robot-aided treadmill training: first experimental results. *IEEE Trans. Neural Syst. Rehabil. Eng.* **13**, 380–394 (2005)



# Real-Time Object Following Based on Visual Information and Online Video Stabilization for Unmanned Aerial Vehicles

Quoc-Sang Tran<sup>(✉)</sup>

Department of Automatic Control,  
Ho Chi Minh City University of Technology, Ho Chi Minh 70000, Vietnam  
1870059@hcmut.edu.vn

**Abstract.** In recent, Unmanned Aerial Vehicles (UAV) have been becoming widely used for many applications such as documentary, agriculture, rescue, surveillance, etc. This research addresses two problems of UAVs. The first is designing the solution for real-time object following using visual information to extract the target object on a current frame defined by the user in the initial state and estimate the position control parameter for the drone system based on the ArduPilot opensource flight controller. In complex systems and wind-disturbance can lead to vibration of video sensor, and the research also focuses on designing the online video stabilization based Kalman multi-rate to improve the navigation performance. The architecture is validated using the ArduPilot flight controller and Jetson TX2 for the object tracking-based OpenTLD tracker method. The experiment shows the system's capabilities can track a great variety of things present in outdoor areas and video stable robustness against wind perturbations and target occlusion in outdoor environments.

**Keywords:** Object following · Video stabilization · UAV object tracking

## 1 Introduction

The control algorithm of UAVs to handle the task of detecting and tracking moving objects has always been a topic of research focus in recent times. Technology forum, opensource community grow fast to offer products that apply to real life. Many commercial products are also launched and packaged in both hardware and software for each specific application. The first approach to detecting the object is to calculate the object-containing area from the image data and then estimate the object's boundary extracted from the original image. One widely used method is contrast analysis to determine the difference between the subject and the background. The area-based object detection method is also strongly developed with superpixel algorithms [1]. Except for contrast and area separation methods, there are many other approaches by exploiting the correlation of the boundaries of the objects [2]. Author Wei [4] proposed an object detection method based on the contrast of the object and the background. Zhu [3] provides a link between the superpixel algorithm and the background image estimation.

Object tracking aims to extract the object from which to estimate the object's trajectory in each frame. Designing an efficient and stable tracking control system is always a complex issue, especially in real-world conditions that are heavily influenced by light changes, background noise, environmental influences, and machine influences flying, image is shaken, blurred, sampling frequency changes (FPS), ... In the study of author YuanWei Wu et al. [5], an algorithm based on Kalman filter was built to estimate the state of the image. Object after each frame, the research relies on a recursive algorithm for the processes of prediction, object detection, and correlation to calculate the object position to determine the error for the aircraft controller. In addition, the study of Can Baykara [6] presents the architecture of using AI to design a real-time system for classifying and tracking moving objects using DJI's UAV model.

Besides, research on video stabilization algorithms for drone applications is also interested in development, including research by Andrew Litvin [7] about image stabilization with Kalman filter and automatic filling of gaps using the mosaic algorithm. Alexandre Karpenko [8] researched digital image stabilization based on Gyroscopes sensor.

This paper proposes the approach for designing the solution for real-time object following based IBVS (Image-Based Visual Servoing) and also the algorithm for video stabilization using Kalman multi-rate for fusing both data from video sensor with state estimated by the Ardupilot flight controller. The object tracking utilizes OpenTLD, which allows tracking a large number of targets accurately compared to prior research such as balloon tracking [9], targets marked with blobs of different sizes [10].

## 2 Video Stabilization Based Kalman Multirate

Accurate estimation of an object's state of motion is at the heart of autonomous robotic applications, and the system controls the trajectory and moves automatically. When analyzing the kinematic model of the object, the relationship between the motion state and the sensor sources, the extended Kalman and Kalman filters are helpful tools for sensor fusion and estimate the motion of an object precisely. When the assumed conditions are satisfied for the noise parameters of the sensor and the kinematic equation of the thing, the Kalman filter is an algorithm for estimating the optimal state of the object. Therefore, Kalman filters are often used in navigation systems requiring high accuracies, such as UAVs, missiles, and robots. In this research, Kalman Multirate will be used for fusing two information sources: the frameshift information from Optical Flow extracting from the camera and the parameters from the ArduPilot state estimator.

Since the sources of measurement information from the sensors are temporally discontinuous, 20 Hz for the Optical Flowmeter and 200 Hz for the estimated parameters extracted from Ardupilot for the combination the Kalman filter consists of two main steps:

- The kinematic model to predict the next state of the system before the measurement from low-frequency sensor output are available (Predict/Time update)
- Use the measurement output of Optical Flow to adjust the state predicted in step 1 optimally.

On the other hand, Kalman is a recursive algorithm, so it can be run online and is optimal in terms of the average error of the system when the assumptions about the noise variance of the system are met. The system diagram has shown in Fig. 1.

Consider a discrete domain linear system:

$$x(k+1) = A_d x(k) + B_d u(k) + w(k)$$

$$z(k) = Hx(k) + v(k) \quad (1)$$

In the formula,  $A_d$  is state-transition model;  $B_d$  is control-input model;  $H$  is observation model;  $w(k)$  is the process noise is zero-mean Gaussian noise with covariance  $Q$ ;  $v(k)$  is the measurement noise is zero-mean Gaussian noise with covariance  $R$ ; With  $w(k) \sim N(0, Q)$ ,  $v(k) \sim N(0, R)$ .

Let the estimated state be an object's position in the 3D field, then vector  $x$  as follows:

$$\hat{x} = \begin{bmatrix} S_x \\ S_y \\ S_z \end{bmatrix} \quad (2)$$

Otherwise, choose the observed variable as the frame-shift from the Optical Flow system:

$$Z_k = \begin{bmatrix} OF_{posx} \\ OF_{posy} \\ OF_{posz} \end{bmatrix} \quad (3)$$

The kinematic equation of the system:

$$\begin{aligned} S_x(k+1) &= S_x(k) + V_x(k) * T \\ S_y(k+1) &= S_y(k) + V_y(k) * T \\ S_z(k+1) &= S_z(k) + V_z(k) * T \end{aligned} \quad (4)$$

where  $S_x$ ,  $S_y$ ,  $S_z$  are object's position in space;  $V_x$ ,  $V_y$ ,  $V_z$  are velocity toward each axis of drone;  $T$  is update time. So that,

$$A_d = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, B_d = \begin{bmatrix} T & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T \end{bmatrix}, H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

When there are only velocity measurement results from the drone but no Optical Flow frame-shift signal, EKF only performs state prediction over time and does not update the observed variable:

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) = A_d \hat{x}(k|k) + B_d u(k)$$

$$P(k+1|k+1) = P(k+1|k) = A_d P(k|k) A_d^T + Q_d \quad (6)$$

When the Optical Flow system signal is available, both Kalman filter steps are performed: estimate new state and update the observation variable.

$$\hat{x}(k+1|k) = A_d \hat{x}(k|k) + B_d u(k)$$

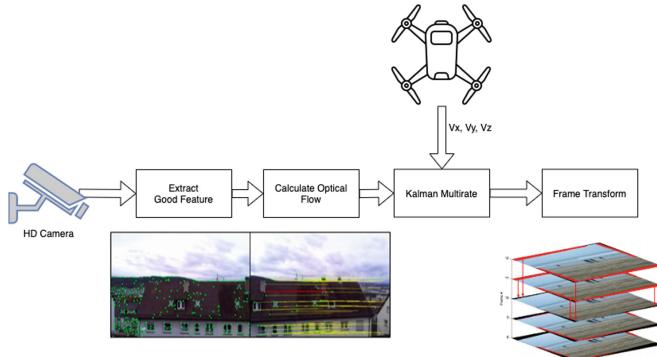
$$P(k+1|k) = A_d P(k|k) A_d^T + Q_d$$

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + K(k+1)[z(k+1) - H\hat{x}(k+1|k)] \quad (7)$$

$$P(k+1|k+1) = [I - K(k+1)H]P(k+1|k)$$

$$K(k+1) = P(k+1|k)H^T \left[ H P(k+1|k) H^T + R_d \right]^{-1}$$

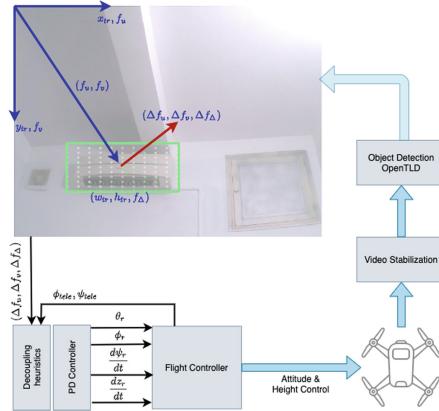
After the Kalman Multirate, the system will use the results to transform the corresponding image frame and update the new structure into the next processing system as object tracking or communication with the graphical interface can easily observe the output of the video stabilizer.



**Fig. 1.** Diagram of the video stabilization. The transformed frame will be an input of object detection and tracking algorithm.

### 3 Real-Time Object Following Controller

This research uses the image-based visual servoing (IBVS) control method, one of the most commonly used methods in robotic position control, with the input of images taken directly on the video device attached to the robot, which is similar to the quadrotor model in this paper. The control system described in Fig. 2, object tracker with image input from the video stabilization processed with the algorithm presented in Sect. 2. The smooth image is put into the OpenTLD tracker, and the algorithm can accurately track the object by simply selecting the original thing. OpenTLD is capable of learning changes and



**Fig. 2.** Image-based Visual Servoing controller architecture.

improving detection when running online. The whole algorithm design onboard using Jetson TX2 equipped with GPU to accelerate processing capacity while reducing latency time and data loss due to data transmission when transmitting over Wi-Fi.

The error is extracted by comparing the position and size of the object with the corresponding desired parameter configured in the initial step and as input to the IBVS control method. From the results of the OpenTLD object tracking algorithm, we can determine the position and size of the target, thereby determining the parameters for the controller:

$$f_u = \frac{x_{tr} + (\frac{w_{tr}}{2})}{w_{img}}; f_v = \frac{y_{tr} + (\frac{h_{tr}}{2})}{h_{img}}; f_\Delta = \sqrt{\frac{A_{img}}{A_t}} \quad (8)$$

where  $f_u, f_v, f_\Delta$  are the feedback parameters for the controller. The object tracking method provides the information  $x_{tr}, y_{tr}$  are the horizontal and vertical location of upper right corner and  $w_{tr}, h_{tr}$  denote width and height of target.  $A_{img}$  is the entire image area and  $A_t$  is area of target at the current frame, the ratio between two approximate  $A_{img}, A_t$  is the depth information of the target from which the distance to the tracked object can be estimated.

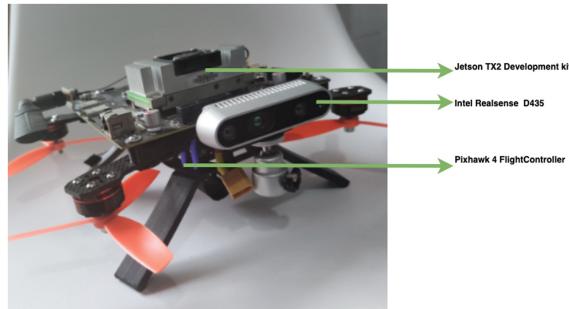
In a visual servoing system for a quadrotor with a fixed camera, the parameters extracted from the image plane will naturally couple the control variables. Therefore, we need a method of decoupling control heuristics for the altitude and lateral movement degrees of freedom calculated through the expression:

$$\Delta f_{u\psi} = \Delta f_u; \Delta f_{uy} = \Delta f_u + \frac{\psi_{tele}}{FOV_u}; \Delta f_{uz} = \Delta f_v + \frac{\theta_{tele}}{FOV_v}; \Delta f_{\Delta x} = \Delta f_\Delta \quad (9)$$

In the formula,  $\psi_{tele}$  and  $\theta_{tele}$  are yaw and pitch angle of the vehicle which obtained from drone's telemetry data.  $FOV_u$  and  $FOV_v$  are horizontal and vertical camera's field of views can obtain from formula  $FOV_u = 2\tan^{-1}\frac{w}{f}$ ,  $FOV_v = 2\tan^{-1}\frac{h}{f}$ .

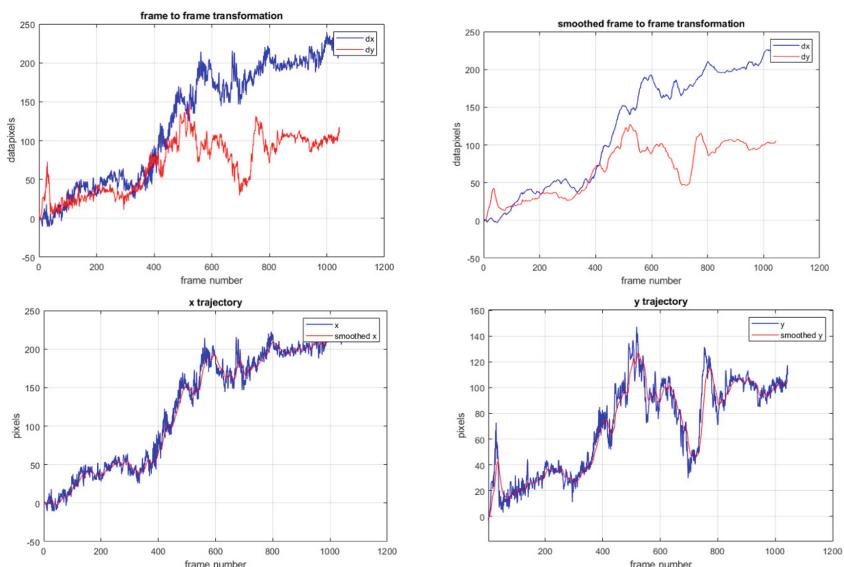
The altitude and lateral are finally sent to the PD controller to decompose and output the speed commands.  $\left\{ \theta_r, \phi_r, \frac{d\psi_r}{dt}, \frac{dz_r}{dt} \right\}$  which is pitch, roll, heading rate and altitude rate and also the final reference command sent to the ArduPilot flight controller.

## 4 Experiments and Results



**Fig. 3.** Quadrotor model build for testing and experiment

The experiment was performed on the quadrotor model with a flight controller using ArduPilot opensource. Also, the Jetson TX2 development kit and Intel real-sense camera for running video stabilization and object-tracking based OpenTLD method, the actual design described in Fig. 3.



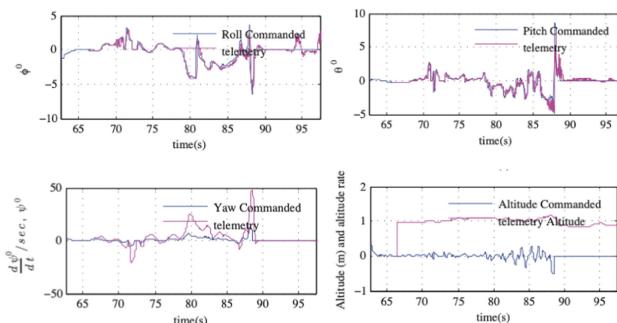
**Fig. 4.** The output of video stabilization show video smoothed in x and y-axis.

The research has conducted many test cases for the video stabilization algorithm with changes in system and measurement noise parameters. It shows clearly in Fig. 4.  $dx$ ,  $dy$  axes indicate the initial data before processing are a lot bumpy due to the vibrations from the camera equipment. The second graph shows the  $x$  and  $y$  axes after the image stabilizer; also, the noise is compensated, and the output becomes smoother. Some other experiments show that changes in system noise parameters do not have much effect on the results of the stabilizer; instead, the measured noise parameters affect more when  $R$  is larger, the smoother the frame will be. However, active camera movements are also known as vibration, and compensation causes distortion of the frame.



**Fig. 5.** The OpenTLD IBVS tracker in Gazebo Simulation

The simulation system is performed on Gazebo as shown in Fig. 5, including a drone model simulated based on the AR. Drone 2.0 source code with full control parameters and physical effects like natural conditions. The turtle bot machine act as the tracked object, which moves in a repeating trajectory. The blue square is the result obtained from the OpenTLD method. Figure 6 shows the roll, pitch, yaw rate, and altitude rate command from the IBVS controller and the corresponding values from the telemetry, which show the accuracy and excellent tracking performance demonstrated in the simulation.



**Fig. 6.** Roll, pitch, yaw rate, and altitude rate commands and corresponding values obtained from telemetry on simulation test.

## 5 Conclusions

This research achieved two significant problems for today's intelligent UAV applications. First, the video stabilization system combined two inputs, such as the camera's image optical flow and the flight controller's estimation parameters, using the EKF filter. The results show that images remain stable when affected by vibrations; changes in the environment, especially for camera systems capable of zooming up to 100x, lead to the tiny vibrations will significantly affect image quality and also affect subsequent detection and tracking systems. The approach will help resolve these problems for large UAV applications.

Second, the research demonstrates an overall algorithm for an image-based visual servoing system using the OpenTLD tracking method, which allows accurate detection of objects. Auto-learning ability also helps with good tracking results provides a reliable input to the IBVS control. The IBVS control method allows decoupling the information from the image plane into corresponding control variables that can be directly sent to the flight controller to control the position quadrotor model to follow the object, from which the system is capable of tracking multiple targets with unknown objects of different sizes and a range of distances.

The future work is to perform more realistic flight tests in various indoor and outdoor environments to find more accurate parameters to improve the IBVS controller. The image stabilization system will be integrated with cameras capable of large zoom or combined with gimbal equipment to provide a comprehensive video stabilization solution for both hardware and software.

## References

1. Borji, A., Cheng, M., Jiang, H., Li, J.: Salient object detection: a survey (2014). <https://arxiv.org/abs/1411.5878>
2. Zhang, J., Sclaroff, S., Lin, Z., Shen, X., Price, B., Mech, R.: Minimum barrier salient object detection at 80 FPS. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1404–1412 (2015)
3. Zhu, W., Liang, S., Wei, Y., Sun, J.: Saliency optimization from robust background detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2814–2821 (2014)
4. Wei, Y., Wen, F., Zhu, W., Sun, J.: Geodesic saliency using background priors. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7574, pp. 29–42. Springer, Heidelberg (2012). [https://doi.org/10.1007/978-3-642-33712-3\\_3](https://doi.org/10.1007/978-3-642-33712-3_3)
5. Wu, Y., Sui, Y., Wang, G.: Vision-based real-time aerial object localization and tracking for UAV sensing system. *IEEE Access* **5**, 23969–23978 (2017)
6. FahdRafi, K.S., Shah, M.: Autonomous target following by unmanned aerial vehicles. In: Proceedings of the SPIE 6230, Unmanned Systems Technology VIII, 623010 (2006)
7. Litvin, A., Konrad, J., Karl, W.C.: Probabilistic video stabilization using Kalman filtering and mosaicking. In: Proceedings of the SPIE 5022, Image and Video Communications and Processing (2003). <https://doi.org/10.1117/12.476436>
8. Karpenko, A., Jacobs, D., Baek, J., Levoy, M.: Digital video stabilization and rolling shutter correction using gyroscopes. Stanford University Computer Science Tech Report CSTR 2011-03 (2011)

9. Mondragon, I., Campoy, P., Olivares-Mendez, M., Martinez, C.: 3D object following based on visual information for unmanned aerial vehicles. In: Robotics Symposium, 2011 IEEE IX Latin American and IEEE Colombian Conference on Automatic Control and Industry Applications (LARC), pp. 1–7 (2011)
10. Zhang, H., Ostrowski, J.: Visual servoing with dynamics: control of an unmanned blimp. In: Robotics and Automation, 1999. Proceedings, 1999 IEEE International Conference, vol. 1, pp. 618–623 (1999)



# Mobility Prediction on a Location-Based Social Network Using K Latest Movements of Friends

Hai Thanh Nguyen<sup>1</sup>(✉), Chi Le Hoang Tran<sup>2</sup>, and Huong Hoang Luong<sup>3</sup>

<sup>1</sup> Can Tho University, Can Tho, Vietnam

[nthai.cit@ctu.edu.vn](mailto:nthai.cit@ctu.edu.vn)

<sup>2</sup> FPT Polytechnic, Can Tho, Vietnam

<sup>3</sup> FPT University, Can Tho, Vietnam

**Abstract.** With recent emerging events such as the pandemic of coronavirus disease (COVID) 2019, human mobility has caused significant concern in the spread of this dangerous pandemic, so mobility prediction is considered as one of the crucial factors to prevent the pandemic. Therefore, there have been many proposed and highly functional studies. Applications of social networks have stored vast data of user movements and brought a vast of interesting research on human mobility. Friendship on social networks has also revealed some effects on the movement. In this study, we have attempted to explore the influence of friendships in location-based social networks on human mobility. We conduct the movement based on the K latest check-ins of friends of the user to predict mobility. We have deployed Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm (using Haversine distance) to cluster original check-in points and filtered the K latest friends' check-ins of the user to predict the user's next movement with the Random Forest algorithm. The prediction conducted from movement history of friends has obtained better performances compared to the prediction without considering the Friendship. The highest accuracy is 0.3176 (with a radius of 400 m and four latest check-ins of friends). Besides, we compare and evaluate the results of the proposed method with the clustered dataset with the original dataset. As observed from the experiments, clusters generated by DBSCAN with wider radii can reveal that their friends' movements can influence users' mobility on a location-based social network (LBSN).

**Keywords:** Location-based social network · Human mobility prediction · Clustering · Haversine · DBSCAN · Random Forest · Friendships · Check-ins

## 1 Introduction

Location-based social networks (LBSNs) allow people to share their location with others to expose their emotions/recommendations and so on. Researchers

can collect data and research in analyzing and predicting human mobility for various purposes. Predicting the path of people has many vital applications and dramatically affects everyone's life. For example, it is applied in analyzing behavior and predicting the direction of criminals to help police make more accurate decisions in the investigation process. In addition, the information that users share on LBSNs can improve customer satisfaction with travel services by making recommendations on places that customers might like and want to visit. Besides, LBSNs also help advertising services make accurate recommendations about restaurants, hotels, and beautiful places that users find interesting. Especially in the context of the current COVID 19 epidemic, this is very important in tracing the epidemiology, and at the same time, locating those who are likely to spread the infection more quickly.

Human behavior and movement can depend on both subjective and objective factors. One of the significant factors contributing to their behavior is their friendships. We have found that people tend to move around where their friends have been. For instance, when a person goes to a particular place and feels like they share it on social networks, the user's friends will see it, and they will likely go there. Human movement history data is collected through LBSNs that provide information about the user's time and location. Based on the data we have obtained, we can use it in conjunction with data on friends' relationships to predict their movements in the near future. Therefore, this study proposes a method to predict human movement in location-based social networks based on the movement history of friends.

In this study, we use the Gowalla dataset<sup>1</sup> consisting of users' check-ins data and their friendship network. First, we adjust the dataset by linking the check-ins points of a person and that person's friends; on each data line, we add the check-ins information of five friends sorted by the closest time compared to the time of that line. Next, we refine and remove confounding data lines. We then used the Haversine formula [1] to determine the distance between the longitude and latitude of the data. Finally, we use the DBSCAN algorithm [2] to cluster and apply the coordinate classification Random Forest (RF) model to make predictions. In addition, we compare and evaluate the results of the proposed method with the clustered dataset and the RF model with the original dataset.

## 2 Related Work

The appearance of the LBSNs has brought about a remarkable development in human behavior and movement. Therefore, many studies and approaches have been presented in [3–5], especially in predicting the next location that a person will visit in the future based on his or her travel history data. Using Deep Learning (DL) approaches such as Recurrent Neural Networks (RNN), Long Short-term Memory (LSTM), Convolutional Neural Networks (CNN), etc., has overcome many problems compared to the traditional approach. In LBSNs studies, the most commonly used datasets are Gowalla and Brightkite<sup>2</sup>. The work

---

<sup>1</sup> <http://snap.stanford.edu/data/loc-Gowalla.html>.

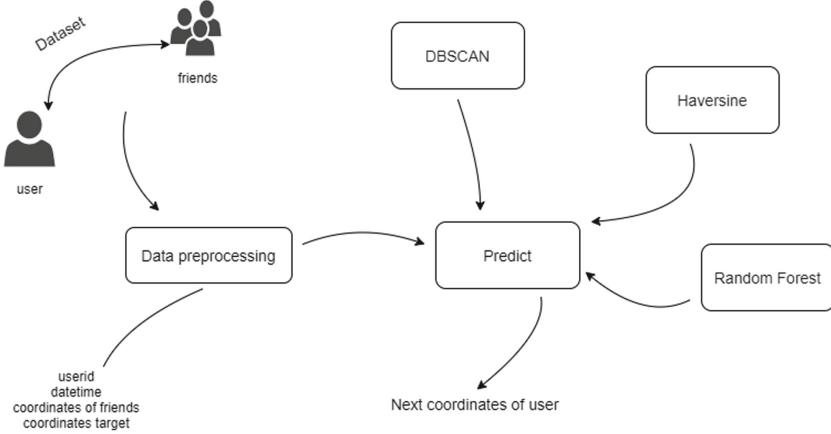
<sup>2</sup> <https://snap.stanford.edu/data/loc-brightkite.html>.

in [6] proposed the Predicting Mobility in the Near Future (PMNF) model that can be used in predicting the mobility of people. The results showed that the proposed model significantly improved the accuracy of the prediction. Thereby, the author also emphasizes the importance of approaching movement prediction using LBSNs and can improve future prediction accuracy based on your group's influence on users. In another study in [7], the authors showed that periodic time travel (e.g., commuting to work, going home) was not affected by social network structure, while long-distance travel is affected quite a lot by relationships on social networks. In addition, they also developed a model based on combining cyclical daily movement with influences from friendships. Moreover, Zhao et al. [8] studied to improve the accuracy in predicting the next Point-Of-Interest (POI) based on the original LSTM model. Thereby, they proposed the Spatio-temporal LSTM, in which they used additional time and distance gates to control the influence of recently visited POIs on the next POI recommendation. The study in [9] considered two human behavioral tendencies, including exploratory and cautious, to investigate mobility data. The authors in [10] proposed a time-variant Markov model based on Gibbs sampling to predict the movement of an app-collected dataset of 100,000 individuals. Another work in [11] also deployed a hybrid Markov-based model and obtained an accuracy of more than 56% accuracy in movement prediction. The authors in [12] presented the k-Latest Check-ins model to predict the movement of people by using an algorithm to get the check-in data of the nearest users. They also use the Haversine formula to calculate the distance between two points with latitude and longitude. They concluded that the accuracy increased proportionally with the radius value considered, and their model gave better results when compared with the Spatio-temporal LSTM method presented in [8].

### 3 Method

This work explores the influence of friends' movement on user mobility on location-based social networks. The proposed method is exhibited in Fig. 1. First, we extract the friends' movement history of users and order them by the time of checked-in to consider and evaluate their effects on predicting user movement using the Random Forest algorithm for the prediction tasks. Then, the locations are clustered with DBSCAN with various radii (using Haversine distance) to compare the original locations.

For a given raw check-in  $u(userid, time, lat, lon, locationid)$  where  $userid$  denotes the User ID,  $time$  indicates when the user checked in at that original check-in point ( $locationid$ ) with Latitude ( $lat$ ) and Longitude ( $lon$ ), we added the  $K$  latest  $locationid$  locations of friends. Let say, from the original check-in  $u(userid, time, locationid)$  (with  $K = 0$ ) which will be fetched into training phase. If  $K = 1$ , we obtain a modified check-in  $u_1(userid, time, locationid_1, locationid)$  where  $locationid_1$  is the latest location where some friend of the user has checked-in at the time being the nearest to  $time$ . If  $K = 2$ , we have a changed check-in



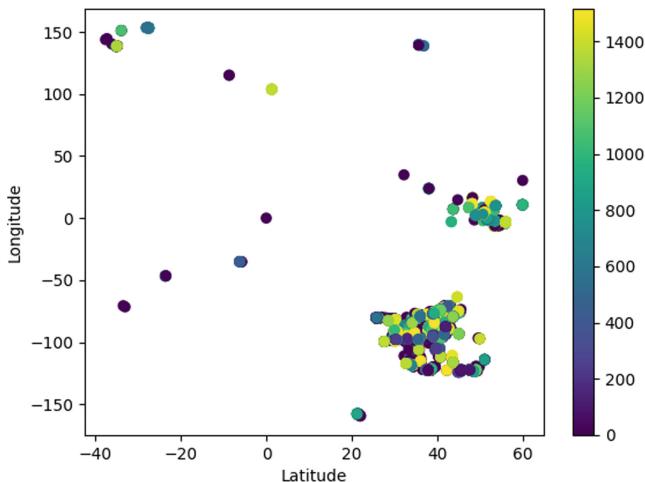
**Fig. 1.** The architecture of the K latest locations model.

**Table 1.** The dataset description has been preprocessed for prediction.

	Users	Check-ins	Ratio
Total	772	18429	100%
Train	741	17000	92%
Test	515	1429	8%

history  $u_2(userid, time, locationid_1, locationid_2, locationid)$  where  $locationid_1, locationid_2$  are 2 check-ins made with the time being the nearest ones to  $time$ . We use clustered check-in points as locations,  $locationid$  is cluster ID. We removed confounding data such as incomplete check-in rows of the five closest friends and users with less than ten check-ins. We use about 20,000 rows to run our experiments by sorting and dividing them into training and test sets by time. Next, to overcome the confounding data in the testing set, we remove the check-ins that do not exist in the training set. Finally, we use that training and testing set to cluster and predict, detailed in Table 1. In the next step, we use the Haversine formula to calculate distance D as illustrated in Formula 1 and the DBSCAN algorithm to cluster the coordinates of the checked points with  $min\_samples=2$  (Fig. 2). The color bar displayed is based on the number of coordinates of one cluster. The reason we chose the Haversine formula to determine the distance is that many studies have shown that this formula is more suitable for the coordinates of the earth than the Euclidean algorithm [13, 14]. Finally, we train the model and predict the next destination using the RF algorithm with 16 trees.

$$D = 2r \times \sin^{-1} \left( \sqrt{\sin^2 \left( \frac{\varphi_2 - \varphi_1}{2} \right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2 \left( \frac{\lambda_2 - \lambda_1}{2} \right)} \right) [1] \quad (1)$$



**Fig. 2.** An illustration of clusters with radius  $R = 400$  m.

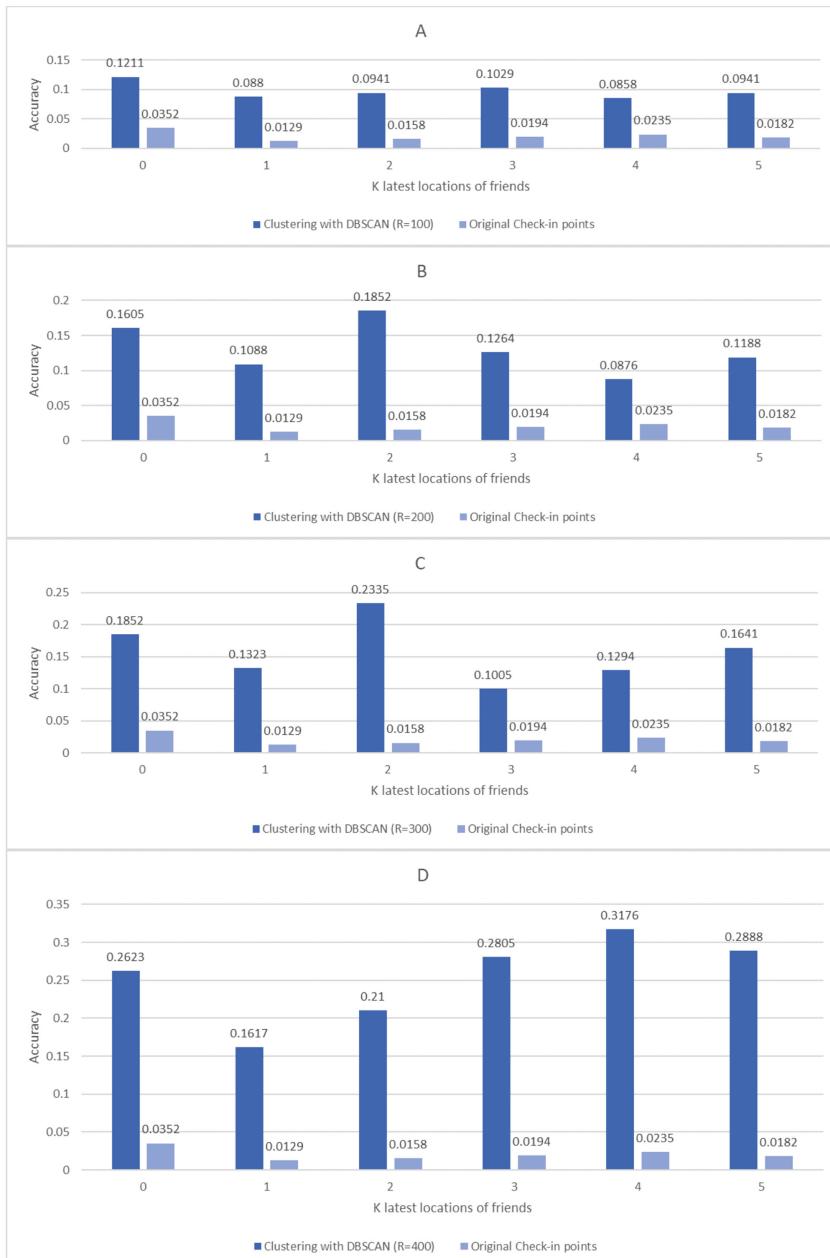
## 4 Results

We evaluate the performance of the K latest of friends method using Acc@K (Formula 2). K is the latest location of user relationships considered.  $K = 0$  means data without adding friends. In this article, we choose  $K = \{0, 1, 2, 3, 4, 5\}$ . At the same time, we also consider the radius between coordinates when clustering them with  $R = \{100, 200, 300, 400\}$  (unit: meters (m)).

$$ACC = \frac{\text{The number of rows is predicted exactly}}{\text{The total number of rows is predicted}} \quad (2)$$

**Table 2.** Comparison of human mobility prediction performance of various K latest movements of friends and different radii in use of clustering algorithms for check-ins points. The **bold** results reveal that the performances with  $K \geq 2$  give better results.

K	Clustering with DBSCAN (R)				Original check-in points
	100 m	200 m	300 m	400 m	
<b>0</b>	0.1211	0.1605	0.1852	0.2623	0.0352
<b>1</b>	0.0880	0.1088	0.1323	0.1617	0.0129
<b>2</b>	0.0941	<b>0.1852</b>	<b>0.2335</b>	0.2100	0.0158
<b>3</b>	0.1029	0.1264	0.1005	<b>0.2805</b>	0.0194
<b>4</b>	0.0858	0.0876	0.1294	<b>0.3176</b>	0.0235
<b>5</b>	0.0941	0.1188	0.1641	<b>0.2888</b>	0.0182



**Fig. 3.** Accuracy comparison between using clusters generated by DBSCAN with various radii (100 m (A), 200 m (B), 300 m (C), 400 m (D)) and original check-in points as destinations for human mobility prediction.

We compare the accuracy between the original data and the data after fine-tuning with K latest check-ins of friends. We also experiment with different values of K and radius R, and the results are shown in Table 2. The table shows that the accuracy has improved after applying the K latest check-ins of friends method. For example, within a radius of 400 m, when not considering friendship, the accuracy is 0.2623; if considered based on the number of friends, the accuracy is 0.2805 (with K = 3), 0.3176 (with K = 4), and 0.2888 (with K = 5) (Fig. 3). Besides, with K = 5, the accuracy from 0.0941 (with R = 100) has increased to 0.2888 (with R = 400). Thereby, we find that the accuracy increases proportionally with the latest locations of friends (K) of the user and the radius (R) when considering the distance of the coordinates.

Figure 3 illustrates improved results in human mobility prediction using friendships when increasing the radius with cluster coordinates. The figure reveals the results with various radii. The charts of A, B, C, D in Fig. 3 present the results with radius R = 100 m, 200 m, 300 m, 400 m, respectively. The highest accuracy is obtained with 0.3176 (using K = 4 and R = 400 m). The graphs illustrate that increasing the radius and based on an appropriate number of friends' check-ins can increase the accuracy. For example, with R = 100 m, then K = 0 gives the highest accuracy (0.1211), but with R = 200 m and R = 300 m, K = 2 gives the highest accuracy. Similar to R = 400 m, when K >= 3, the result is higher than K = 0 (without considering friends' movements). These results show that the user's movement is more related in a clearer manner to the friends (K) when we increase the considered radius.

## 5 Conclusion

This work presented an approach to predict human mobility based on K's latest check-ins of friends in a location-based social network. We investigated check-ins of users and their friends to show the influence of friends on human mobility. Experimental results on clusters generated by DBSCAN algorithm show that the accuracy has been improved (0.3176) compared to the original check-ins points (0.2623). For original points of check-ins, we cannot see the influence of friendships on users' movement. However, when we increase the radius to consider more expansive areas, we can observe more evidently the influence of friends' movement on their mobility.

Further research can combine the K latest check-ins of friends with other methods to give better results. Besides, increasing the size of the dataset can bring better results. Future work can investigate more on behavioral analysis, which can be collected from other social networks. In addition, further experiments can also focus on finding which K is the best for human mobility prediction. It can be a big challenge due to significant differences in human movement patterns. Especially in the context of the current COVID 19 epidemic, this is very important in tracing the epidemiology and at the same time locating those who are likely to spread the infection more quickly.

## References

1. Robusto, C.C.: The cosine-haversine formula. *Am. Math. Monthly* **64**(1), 38–40 (1957). <http://www.jstor.org/stable/2309088>
2. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, pp. 226–231. KDD 1996. AAAI Press (1996)
3. Luca, M., Barlacchi, G., Lepri, B., Pappalardo, L.: A survey on deep learning for human mobility. *ACM Comput. Surv.* **55**(1), 1–44 (2023). <https://doi.org/10.1145/3485125>
4. Islam, M.A., Mohammad, M.M., Das, S.S.S., Ali, M.E.: A survey on deep learning based point-of-interest (POI) recommendations (2020). <https://arxiv.org/abs/2011.10187v1>
5. Wei, X., Qian, Y., Sun, C., Sun, J., Liu, Y.: A survey of location-based social networks: problems, methods, and future research directions. *GeoInformatica* **26**(1), 159–199 (2021). <https://doi.org/10.1007/s10707-021-00450-1>
6. Hai, N.T., Nguyen, H.H., Thai-Nghe, N.: A mobility prediction model for location-based social networks. In: Intelligent Information and Database Systems, pp. 106–115. Springer, Berlin Heidelberg (2016). [https://doi.org/10.1007/978-3-662-49381-6\\_11](https://doi.org/10.1007/978-3-662-49381-6_11)
7. Cho, E., Myers, S.A., Leskovec, J.: Friendship and mobility. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 2011. ACM Press (2011). <https://doi.org/10.1145/2020408.2020579>
8. Zhao, P., Zhu, H., Liu, Y., Li, Z., Xu, J., Sheng, V.S.: Where to go next: a spatio-temporal LSTM model for next POI recommendation (2018). <https://arxiv.org/abs/1806.06671v1>
9. Liu, E.J., Yan, X.Y.: A universal opportunity model for human mobility. *Sci. Rep.* **10**(1) (2020). <https://doi.org/10.1038/s41598-020-61613-y>
10. Wang, H., Zeng, S., Li, Y., Jin, D.: Predictability and prediction of human mobility based on application-collected location data. *IEEE Trans. Mob. Comput.* **20**(7), 2457–2472 (2021). <https://doi.org/10.1109/tmc.2020.2981441>
11. Qiao, Y., Si, Z., Zhang, Y., Abdesslem, F.B., Zhang, X., Yang, J.: A hybrid Markov-based model for human mobility prediction. *Neurocomputing* **278**, 99–109 (2018). <https://doi.org/10.1016/j.neucom.2017.05.101>
12. Dao, T.C., Nguyen, H.T.: Human mobility prediction using k-latest check-ins. In: Future Data and Security Engineering. Big Data, Security and Privacy, Smart City and Industry 4.0 Applications, pp. 36–49. Springer Singapore (2021). [https://doi.org/10.1007/978-981-16-8062-5\\_3](https://doi.org/10.1007/978-981-16-8062-5_3)
13. Wang, Q., Taylor, J.E.: Patterns and limitations of urban human mobility resilience under the influence of multiple types of natural disaster. *PLOS ONE* **11**(1), e0147299 (2016). <https://doi.org/10.1371/journal.pone.0147299>
14. Laylavi, F., Rajabifard, A., Kalantari, M.: A multi-element approach to location inference of twitter: a case for emergency response. *ISPRS Int. J. Geoinf.* **5**(5), 56 (2016). <https://doi.org/10.3390/ijgi5050056>



# A New System for License Plate Recognition in Traffic Violation Scenarios in Vietnam

Ha Nguyen-Xuan<sup>1</sup> , Dong Hoang-Nhu<sup>2</sup>, Kien Kim-Van<sup>2</sup>,  
and Tuan Dang-Minh<sup>2,3</sup>

<sup>1</sup> Hanoi University of Science and Technology, 1 Dai Co Viet, Hai Ba Trung, Hanoi, Vietnam  
ha.nguyễnxuan@hust.edu.vn

<sup>2</sup> CMC Institute of Science and Technology, 11 Duy Tan, Cau Giay, Hanoi, Vietnam

<sup>3</sup> Posts and Telecommunication Institute of Technology, KM 10 Nguyen Trai, Ha Dong, Hanoi,  
Vietnam

**Abstract.** License plate recognition in traffic violation scenarios is a key issue for any traffic management system. In this work, a new system that can detect traffic violations and output the license plate numbers of violating vehicles and evidence was introduced. The system was developed based on image processing techniques combined with deep neural networks. A dataset, which contains thousands of images of traffic vehicles from public cameras in Hue City, was collected. Based on this dataset, three deep learning networks for different tasks in the overall pipeline process were developed. A pretrained-model of the YOLOv5s6 neural network was used to create a model to detect four classes of vehicles, including cars, buses, trucks, and motorbikes with riders. The so-called “Deep SORT” model was exploited for vehicle tracking and traffic violation detection tasks. For license plate detection, a model based on the RetinaFace with backbone MobileNet was trained. Rectification techniques for the detected license plates were applied. Finally, an OCR model for license plate number recognition was finetuned from a pre-trained model with CRNN architecture. Experimental evaluations show that all developed models are very lightweight and have high accuracy. The YOLOv5s6 achieves an mAP of 87.8%; the RetinaFace has a maximal mean square error of only 1.6, and the OCR model attains a very high accuracy of 99.72%. The results of the implementation of the pipeline on embedded hardware, namely the Jetson Xavier Development Kit, show that our system is very computationally efficient. The total computing time is only 41 ms and the use of RAM is less than 3 GB. These results show high potential for practical applications.

**Keywords:** License plate recognition · Traffic violation · Deep neural network

## 1 Introduction

With the rapid development of image processing techniques based on the OpenCV library and deep neural networks, many key issues in intelligent transportation systems can be solved. The problem of automatic license plate detection and recognition has received much investigation in recent years [1, 2]. The research trend for this topic is focusing on

the exploitation of deep neural networks for key processing tasks, including problems of object detection, tracking, landmark detection, and optical character recognition (OCR). Many studies have tried to improve the accuracy, robustness, and reliability of developing deep neural network models [3]. However, only a few studies have been conducted on automatic license plate detection and recognition in traffic violation scenarios [4], [5], [6]. This issue becomes especially difficult in developing countries such as Vietnam, where traffic is always crowded and chaotic, and many motorbikes are subjected to poor traffic infrastructure and conditions [7].

In this work, based on a baseline introduced in [8], a new system for license plate detection and recognition in traffic violation scenarios in Vietnam was proposed. Our approach is the use of several state-of-the-art deep neural network architectures, including the YOLOv5s6 [9], Deep SORT [10], RetinaFace [11], and RCNN-based OCR [12], to develop models for the tasks of vehicle detection and tracking, license plate detection, and localization, and license plate number recognition. The dataset used in this work is self-collected from public-traffic camera streams in Hue City. All developed models are combined with the algorithms of the preprocessing tasks using the OpenCV library [12]. The complete system was deployed on embedded hardware, namely the Jetson Xavier Development Kit [14], for the computational evaluation. It is expected that the results of this work can be used efficiently for many intelligent-traffic applications in Vietnam.

## 2 Method

A general processing pipeline for traffic violation detection is proposed in Fig. 1. There are three main modules for processing: i) detection and tracking of vehicles, ii) detection of traffic violations and iii) detection and recognition of license plates. The overall processing pipeline is described as follows:

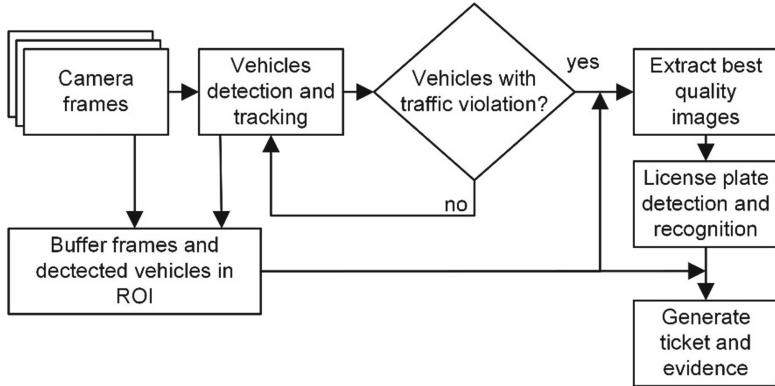
First, the sequences of consecutive frames from the camera are acquired and pushed to a model for detecting vehicles of interest. Each vehicle is assigned an ID and tracked through a tracking algorithm. In parallel to this task, all frames and images of detected vehicles in a region of interest (ROI) are buffered in local memory for further usage in the next step.

Second, an algorithm is applied to detect tracked vehicles with traffic violations. If there is no traffic violation in ROI, a loop for tracking is continued. Otherwise, the next step for license plate detection and recognition is performed. It is noticed that this next step is only used when a traffic violation is detected. This ensures that the required computational resources for this task are minimized.

License plate detection and recognition are implemented in the third step. In this step, a sequence of consecutive images of vehicles with traffic violations, which are stored in the buffer, are allocated. A best-quality image from this set of images is chosen to be processed. When a vehicle is detected to be violating the law, its position can be very far from the camera, or its license plate can be hidden by other vehicles. As a result, the license plate is either too small to recognize or hidden. Additionally, the image quality is also strongly influenced by the light conditions, which are believed to be very challenging in reality. This leads to a reduction in the accuracy of the system. The criteria for choosing the best quality image of a violated vehicle are: the license

plate is seen without or less distortion; the size of the license plate in the image is as large as possible; the image does not have motion blur, dazzling or glossy light. The buffer module is a significant improvement in this work.

In the last step, a ticket for the violating vehicle is generated. The evidence, which contains three time-based consecutive images describing the violation, is also created.



**Fig. 1.** Proposed overall processing pipeline for traffic violation detection

## 2.1 Vehicle Detection and Tracking

For the data preparation, a dataset was created. The dataset contains 5131 images with sizes of  $1920 \times 1080$ , which were collected from public traffic surveillance cameras in Hue City. There are 1986 images captured in conditions of light at night and rainy weather. Four classes of vehicles are considered, including the car, the truck, the bus, and the motorbike with the rider. Vehicle objects in each image are labeled with bounding boxes and class-ID. Vehicle objects whose bounding box dimensions are smaller than 150 pixels are eliminated. The reason for this is that when the vehicle dimensions are too small, the quality of their cropped image will not be good enough for the next processing. The dataset is divided into the training set with 4464 images and the testing set with 667 images. All images are resized to dimensions of  $1280 \times 1280$ , which are suitable for the input of the YOLOv5s6 network.

The YOLOv5s6 network was chosen for the vehicle detection task. In this work, a pre-trained model, which was pre-trained on the MS COCO dataset [15], was chosen to be finetuned. The transfer learning process on the pre-trained YOLOv5s6 model with the above-mentioned self-collected dataset was implemented. Four classes of objects are considered. The configuration parameters for the transfer learning task are set according to the values in Table 1.

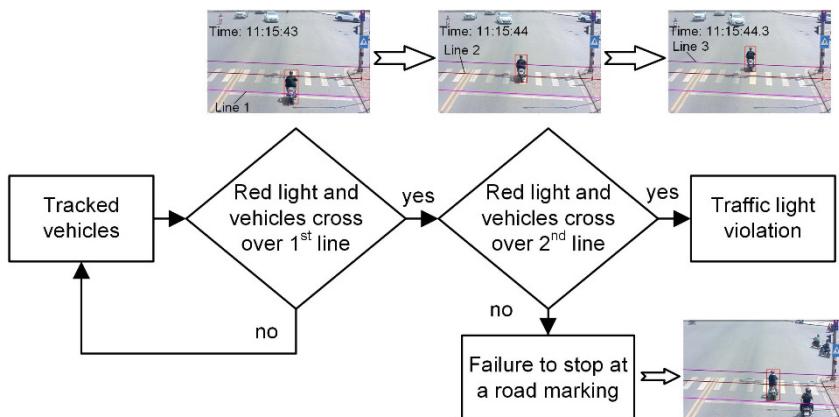
For vehicle tracking, a state-of-the-art method called DEEP SORT [10, 16] (Simple Online and Realtime Tracking with Deep Association metric) was applied. This method is an extension of the original SORT [16], in which the association metric is replaced by a more informed metric that combines motion and appearance information. By integration,

a convolution neural network is trained to discriminate moving objects. This method is very robust against misses and occlusions while keeping the system very easy and efficient to deploy. In our case, because the traffic situation in Vietnam is very crowded and chaotic, and vehicles move at high speeds, the use of Deep SORT will maximally reduce occlusions, misses, or ID swapping.

**Table 1.** Transfer learning configurations of YOLOv5s6 [9]

Parameter	Value
No. of images	5131 (4464 for the train and 667 for the test)
Input image size	1280 × 1280
Batch size	32
Epochs	200
Learning rate	0.0032
No. of classes	4 (motorbike with rider, car, bus, and truck)

## 2.2 Traffic Violation Detection



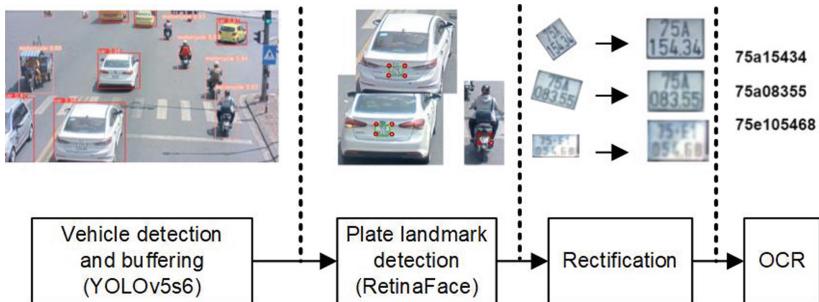
**Fig. 2.** Algorithm for red light violation detection

Algorithms for traffic violation detection were proposed. Several violations were considered, including no stopping, or standing, failure to stop or run across a road marking, failure to run according to a road marking and running through a red light. With these algorithms, the positions of the bounding boxes of tracked vehicles are compared with the road marking lines, which are combined with the traffic signs and traffic lights to find violations according to the traffic rules. Figure 2 shows exemplarily a situation where a motorbike with a rider crosses the road marking line during the red light, resulting in a

violation. As described in the algorithm, during the red light, if the motorbike runs over the first line, we will track its bounding box positions. If the motorbike continues running through the second line, we can conclude that the rider has a “red light” violation. If the position of the bounding box of the motorbike remains in the middle of the first and second lines, the rider has a “failure to stop at a road marking” violation. When the motorbike runs over the third line, it will not be tracked anymore.

### 2.3 License Plate Detection and Recognition

A processing pipeline for license plate detection and recognition is shown in Fig. 3. When a vehicle is detected to be a traffic violation, its cropped images, which are buffered in the temporary local memory, are acquired and the best quality one is selected to detect and recognize its license plate number. For license plate detection, a model that is based on the RetinaFace network [11] with a backbone of MobileNet0.25 [17] was developed. The developed model is used to detect four landmark points on the license plate of the violating vehicle. The landmarks of the license plate are the four corner points. The model was trained from scratch with our self-collected dataset. Parameters for the training process are shown in Table 2. The collected dataset contains 29225 images. Each image has only one vehicle, which is cropped from the resulting images of the vehicle detection model. The dataset is divided into 28757 images for the training set and 650 images for the testing dataset. The training dataset has 4257 images captured at night and 15104 images captured in rainy weather conditions, while the numbers for the testing dataset are 125 images and 300 images, respectively.



**Fig. 3.** Processing pipeline for license plate detection and recognition

Depending on the viewpoint of tracking cameras as well as the position and configuration of the license plate attached to vehicles, detected license plates can be distorted. Also, the size of the license plate can vary depending on the relative position between the tracking camera and the vehicles. Thus, the detected license plates must be rectified as illustrated in Fig. 3. The license plates are rotated, translated, and resized so that they have a front view of the tracking camera. For this task, two functions of the OpenCV library [13], namely `cv2.getPerspectiveTransform()` and `cv2.warpPerspective()`, were used. The first function outputs a transformation

matrix, which transforms the coordinates of the four landmarks of each detected plate into targeted points with a front-view perspective. This transformation matrix is used in the second function to warp the image of the detected license plate to a new one having a straight view from the front. This image is then resized to the input dimension of the license plate number recognition network.

For the task of license plate number recognition (Optical Character Recognition – OCR), a pre-trained model, which is based on the Convolution Recurrent Neural Network (CRNN) architecture [12], was used for the finetuning process. Parameters for the training process are shown in Table 3. A dataset that contains 14698 images with sizes  $32 \times 100$  was created for this model. The dataset contains 2200 images captured in bad light conditions. Since the characters on the license plate of Vietnam are alphabetical, a vocabulary of alphabets is configured for the OCR process.

**Table 2.** Parameter configurations of the license plate detection model based on RetinaFace [11]

Parameter	Value
No. of images	29225 (28571 for the train and 650 for the test)
Image size	$640 \times 640$
Batch size	32
Epochs	250
Learning rate	0.0032
No. of classes	4 (motorbike with rider, car, bus, and truck)

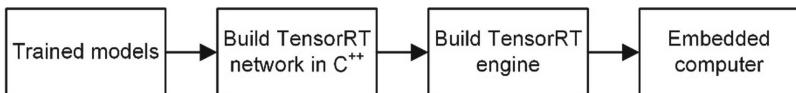
**Table 3.** Parameter configurations of the license plate number recognition model based on CRNN [12]

Parameter	Value
No. of images	14698 (11758 for the train, 1000 for the test, and 1940 for the validation)
Image size	$32 \times 100$
Batch size	32
Epochs	250
Learning rate	0.001
Vocabulary	Alphabet

## 2.4 Model and Algorithm Deployment on Embedded Hardware

Computational efficiency is of great importance for traffic processing problems because much of the computation must be implemented on the very limited resources of edge

computing devices. In this work, the Jetson Xavier Development Kit [14] (GPU: 512-core Volta GPU with Tensor Cores; CPU: 8-core ARM v8.2 64-bit CPU, 8 MB L2 + 4 MB L3) was chosen to evaluate the performance of our developed models and algorithms. The deployment pipeline is shown in Fig. 4. All fine-tuned models on the PyTorch framework are converted to the dictionary format of “.wts”, which contains weights and layers. These parameters are used to build a computational network in the TensorRT format in C++, which enables the models to run efficiently on the Jetson Xavier Kit. The weights are then loaded according to the dictionary onto the TensorRT network to build a TensorRT engine. This engine is run on the hardware for computing tasks. In our experiments, the environment installed on the kit was Ubuntu18.04/cuda10.2/tensorRT7.1/opencv4.5.0.



**Fig. 4.** Deployment pipeline of the proposed algorithms and models on Jetson Xavier [12]

### 3 Results and Discussions

#### 3.1 Evaluation of Vehicle Detection and Tracking Model

The metric Average Precision (AP) with an Intersection Over Union (IoU) threshold of AP<sub>50</sub> was used to evaluate the precision of the traffic vehicle detection model. The mean average precision for all four vehicle classes on the test dataset is 87.7%. Table 4 shows the AP<sub>50</sub> result for individual traffic vehicle classes, which is evaluated on the test dataset of 667 images. It is critically noticed that the first three classes reach a very high AP<sub>50</sub>. The AP<sub>50</sub> for the truck class is still low because the training dataset, which was collected from inner-city traffic cameras, includes very few trucks.

In the case of a motorbike with a rider, there is a big difference in the traffic situation in Vietnam in comparison to other developed countries. The capacity to detect a motorbike with a rider is of great importance. This will make the vehicle tracking process more reliable and robust. It is experienced from practical deployments that if the motorbikes are detected without accompanying riders, their bounding boxes are very small, and their size variation is very large. This makes the tracking process for the motorbike very difficult without riders. The ID of the tracking motorbike is regularly swapped or missed. Therefore, our developed model is a significant contribution to such a difficult situation.

The tracking model using the Deep SORT was evaluated in actual traffic situations in Hue City. Several scenarios were experimented with, including very crowded traffic, different weather, and light conditions (rainy, sunny, day-night). The robustness of the Deep SORT combined with the high-quality vehicle detection model results in a robust tracking module that makes the traffic violation detection very reliable.

**Table 4.** Obtained AP<sub>50</sub> of the vehicle detection model based on YOLOv5s6 (%)

ID	Class	Value
1	Car	96.4
2	Motorbike with rider	89.2
3	Bus	92.0
4	Truck	73.3
	<b>mAP<sub>50</sub></b>	<b>87.7</b>

### 3.2 Evaluation of License Plate Detection and Recognition Model

The quality of the license plate detection model was evaluated based on the Mean Square Error (MSE) between the predicted landmarks  $\hat{\mathbf{p}}$  and the ground truth  $\mathbf{p}$ , according to Eq. (1). The number of predictions  $N$  is set at 650. Table 5 shows the evaluation results. It is noticed that the MSE is very small for all four landmark points. The maximal mean square error is only 1.60, while the dimension of the vehicle images is comparatively large of  $640 \times 640$ . This result proves the efficiency of the fine-tuned model.

$$MSE(\mathbf{p}, \hat{\mathbf{p}}) = \frac{1}{N} \sum_{i=0}^N (p_i - \hat{p}_i)^2 \quad (1)$$

**Table 5.** Mean square error of the license plate detection model

	Top left	Top right	Bottom left	Bottom right
MSE	1.58	1.39	1.60	1.58

The quality of the developed license plate number recognition model (OCR) was evaluated on the test dataset of 1000 plate images. The evaluation standard is based on Eq. (2), where the accuracy is a ratio between the number of times of correct detection  $N_a$  and the total number of tests  $N(N = 1000)$ . When all the characters on the license plate exactly match the ground truth, the recognition is true. Otherwise, the recognition is false.

$$Acc = \frac{N_a}{N} \quad (2)$$

The quality of the model also strongly depends on the quality of the input image. As mentioned, at the beginning of Sect. 2, by introducing the buffering system, we can obtain the license plate image of the violating vehicles with the best quality. This will make the OCR model output more accurate results. Experiments show that without a buffering system, we can only achieve an accuracy of 95.7%. In the meantime, with the buffering system, an accuracy of 99.72% is achieved. This result is higher than

the results presented in [6] (78%) and [8] (96.3%). This is an impressive result, which satisfies regulations for traffic cameras issued by the Ministry of Public Security of Vietnam [18].

### 3.3 Computational Efficiency

The computational efficiency of the developed models and algorithms was evaluated based on the Jetson Xavier Development Kit from NVIDIA [14]. The results are shown in Table 6. It is obvious that the developed models are very lightweight, as indicated by the computing time and used RAM. The computing time of the overall process is less than 41 ms and the allocated RAM is only 2.89 Gb. Since the trained models are converted to a mode of FP16, which is suitable for the embedded edge computing hardware, the accuracy is slightly reduced in the range of 0.5–5%. With these results, our system can be deployed on edge computing hardware for AI-cameras.

**Table 6.** Results of the deploying models on Jetson Xavier Development Kit [14]

Parameter	YOLOv5s6	RetinaFace	OCR
Bath size	1	1	1
Mode	FP16	FP16	FP16
Computing time (ms)	27.70	8.53	4.32
RAM usage (GB)	0.83	1.16	0.9
Accuracy	82.5% (mAP)	1.62 (MSE)	99.65% (Acc)

## 4 Conclusions and Outlook

In this work, a new system for license plate recognition in traffic violation scenarios in Vietnam has been developed successfully. The system has several noticeable advantages, as follows:

- The system has high accuracy in satisfying the regulations of the Ministry of Public Security on traffic cameras (1914/2009/Q-BCA-E11).
- The system has several enhancements that are suitable for the traffic conditions of Vietnam, such as crowded traffic and many motorbikes. The introduction of the buffering module and the ability to detect motorcycles with riders has improved the quality of the system.
- The system is very lightweight so that it can be easily deployed on edge computing hardware like Jetson Xavier. This makes it possible to deploy the hardware of the AI camera for practical applications.

For future work, we will perform more experiments for other traffic violation scenarios. Deployment of the system on edge computing cameras is planned. Evaluations of the influence of vehicle speed on the performance of the system will also be considered.

**Acknowledgment.** This work was funded by the CMC Institute of Science and Technology, CMC Corporation, Hanoi, Vietnam.

We kindly acknowledge Hue City for allowing us to use the camera streams.

## References

1. Lubna, Mufti, N., Shah, S.A.A.: Automatic number plate recognition: a detailed survey of relevant algorithms. *Sensors* **21**, 3028 (2021). <https://doi.org/10.3390/s21093028>
2. Sanjana, S., Sanjana, S., Shriya, V.R., Vaishnavi, G., Ashwini, K.: A review on various methodologies used for vehicle classification, helmet detection and number plate recognition. *Evol. Intel.* **14**(2), 979–987 (2020). <https://doi.org/10.1007/s12065-020-00493-7>
3. Soroori, S., Tourani, A., Shahbahrami, A.: Employing deep learning approaches for automatic license plate recognition: a review. In: 3rd International Conference on Soft Computing at: Rudsar, Guilan, Iran
4. Wijetunge, A.W.G.C.D., Ratnaweera, D.A.A.C.: Real-time recognition of license plates of moving vehicles in Sri Lanka. In: 2011 6th International Conference on Industrial and Information Systems, pp. 82–87 (2011). <https://doi.org/10.1109/ICIINFS.2011.6038045>
5. Mukhija, P., Dahiya, P.K., Priyanka, P.: Challenges in automatic license plate recognition system: an Indian scenario. In: Fourth International Conference on Computational Intelligence and Communication Technologies (CCICT), pp. 255–259 (2021). <https://doi.org/10.1109/CCICT53244.2021.00055>
6. Hendry, H., Chen, R.C.: Automatic license plate recognition via sliding-window darknet-Yolo deep learning. *Image Vis. Comput.* **87** (2019). <https://doi.org/10.1016/j.imavis.2019.04.007>
7. Nguyen, P., Vu, Q.C.: Proposing WPOD-NET combining SVM system for detecting car number plate. *IAES Int. J. Artif. Intell. (IJ-AI)* **10** 657 (2021). <https://doi.org/10.11591/ijai.v10.i3.pp657-665>
8. Silva, S.M., Jung, C.R.: License plate detection and recognition in unconstrained scenarios. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) *ECCV 2018. LNCS*, vol. 11216, pp. 593–609. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-01258-8\\_36](https://doi.org/10.1007/978-3-030-01258-8_36)
9. YOLOv5 Homepage. <https://github.com/ultralytics/yolov5>. Accessed 09 Nov 2021
10. Deep SORT Homepage. [https://github.com/nwojke/deep\\_sort](https://github.com/nwojke/deep_sort). Accessed 09 Nov 2021
11. RetinaFace in PyTorch Homepage. [https://github.com/biubug6/Pytorch\\_Retinaface](https://github.com/biubug6/Pytorch_Retinaface). Accessed 09 Nov 2021
12. Shi, B., Bai, X., Yao, C.: An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition (2015). [arXiv:1507.05717v1](https://arxiv.org/abs/1507.05717v1)
13. OpenCV library Homepage. <https://opencv.org/>. Accessed 09 Nov 2021
14. Jetson AGX Xavier Developer Kit Homepage. <https://developer.nvidia.com/embedded/jetson-agx-xavier-developer-kit>. Accessed 09 Nov 2021
15. Lin, T.-Y., et al.: Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014. LNCS*, vol. 8693, pp. 740–755. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48)

16. Wojke, N., Bewley, A., Paulus, D.: Simple online and realtime tracking with a deep association metric [arXiv:1703.07402v1](https://arxiv.org/abs/1703.07402v1) (2017).
17. Andrew, G., et al.: MobileNets: efficient convolutional neural networks for mobile vision applications [arXiv:1704.04861](https://arxiv.org/abs/1704.04861) (2017).
18. Regulations for traffic cameras no. 1914/2009/Q-BCA-E11 issued by the Ministry of Public Security of Social Republic of Vietnam (2009)



# Development of a Real-Time Obstacle Detection System on Embedded Computer Based on Neural Network and Multi-camera Streaming

Quang Khoi Tran<sup>1</sup>✉), Khanh Hieu Ngo<sup>2</sup>, Anh Huy Le Dinh<sup>2</sup>, Lu Tien Truong<sup>2</sup>,  
Hai Quan Tran<sup>1</sup>, and Anh Tuan Trinh<sup>1</sup>

<sup>1</sup> Ho Chi Minh City University of Technology, VNU-HCM, Real-Time Robotics Vietnam Ltd.,  
Ho Chi Minh City, Vietnam

tqkhoi.sdh21@hcmut.edu.vn

<sup>2</sup> VNU-HCM Key Laboratory for Internal Combustion Engine, 268 Ly Thuong Kiet Street,  
District 10, Ho Chi Minh City, Vietnam  
ngokhanhhieu@hcmut.edu.vn

**Abstract.** The paper presents the study of implementing a real-time obstacle avoidance program based on the multi-camera streaming and a single neural network to process all of input frames come from the cameras. When a collision situation is detected, the collision mode forces all operations to be disabled. The neural network uses the PyTorch framework and based on the DroNet model. The optimization framework used to enhance the neural network real-time performance is TensorRT. Feeding the neural network are 4 camera streams which provide 30 FPS each without any problems. The testing on NVIDIA Jetson Nano's Jetbot with a low-cost Pi camera V2 equipped with 160-degree FOV lens shows that the obstacle detection system could obtain an accuracy of 80%. The results in this paper demonstrate the feasibility of this approach for further development of a real-time obstacle avoidance system with multi-camera streaming on drones to prevent collisions as well as the ability to improve the P-DroNet accuracy for its applications with low-cost cameras by adding data and network retraining.

**Keywords:** Autonomous navigation · Pi camera · DroNet · PyTorch · MAVLink · Obstacle avoidance

## 1 Introduction

Nowadays, UAV/drones are playing an important role in real-life applications. Complex flight missions that require coordinated work between control systems, navigation, and awareness of real-world situations sometimes take place out of the blue.

Studies in 2009, Grzonka et al. [1] used the SLAM (Simultaneous Localization and Mapping) algorithm on a quadcopter which is capable of determining the current position of the aircraft and reconstructing the map with small, low-weight, and low energy consumption sensory equipment [2]. Their purposes are to estimate the trajectory, reconstruct the environment, and measure the depth.

In 2013, the author trains a Deep Learning Network model capable of recognizing trails along cliffs [3]. By using RGB color images from a camera onboard the drone, the system can recognize the object in real-life situations. In 2018, A. Loquercio et al. created the DroNet, a light neural network model built to carry on obstacle avoidance task as well as driving the drone [4]. This network model become a perfect candidate for applications utilizing the embedded computer thanks to its light weight and short inference time. As example, a relevant application for fully-autonomous nano-size drone is based on DroNet approach with a low-power computing platform [5] was a success. Also, another approach in article [6] used unsupervised methods on a commercial drone. The network is trained from crashed data and be able to avoid collision in complex situations such as flight in forest [7], the prediction of obstacle avoidance is performed by a network called AlexNet. But it takes more than a day-time to retrain the network, so long training time is another difficult problem to implement neural networks in obstacle avoidance applications.

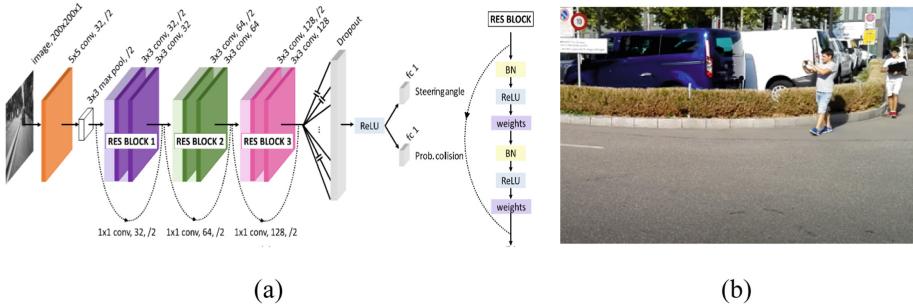
Tran Quang Khoi and Ngo Khanh Hieu in [8] have presented a research result in 2021 shown that implementing neural networks on Raspberry Pi embedded computer having a lot of challenges, mostly because of the poor real-time performance of neural networks. But this approach has great potential if future works could take advantages of GPU alongside CPU as well as using optimization algorithms to improve the real-time performance.

This article proposes an approach to develop a real-time obstacle detection system on embedded computer based on the DroNet network architecture [4], the PyTorch platform, and the multi-camera streaming. Thereby, the neural network would be optimized by using the TensorRT framework (named P-DroNetRT) with the 4 cameras streaming to observe all around the drone. Consequently, the P-DroNetRT real-time obstacle avoidance system could perform at 120 fps (frames per second) with 4 cameras streaming (30 fps for each camera).

The rests of this paper are presented in the followings: Sect. 2 represents the P-DroNetRT model; Sect. 3 presents the evaluation of the P-DroNetRT model; Sects. 4 and 5 present conclusions and future works.

## 2 The P-DroNet Model

DroNet is a CNN (Convolutional Neural Network) designed for small vehicles with limited computation strength onboard computers [4] in the Fig. 1a. Nevertheless, their deployment on the Bebop drone is still based on a PC computer as can be seen in the Fig. 1b. It's less flexible than the approach using a companion computer onboard the drone itself because of the transmission latency. Figure 2 shows the VIAN drone of the Real-time Robotics Vietnam Ltd. (RtR) using a NDIVIA Jetson Nano embedded computer onboard and can handle many computation-intensive tasks such as HD video streaming in real-time thanks to the GPU onboard of the NDIVIA Jetson Nano embedded system.

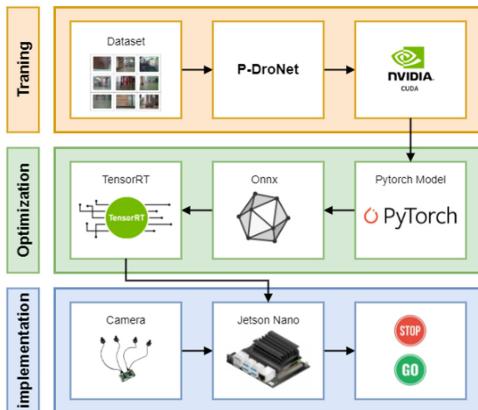


**Fig. 1.** a) DroNet CNN architecture; b) DroNet’s purpose is to reliably drive an autonomous drone using ground station computer.

Figure 3 represents our methodology to develop of a real-time obstacle detection system on embedded computer based on neural network and multi-camera streaming. The combined FOV (Field-of-View) from all 4 cameras provides 360-degree observation, so the object avoidance program can collision detection all around the drone.



**Fig. 2.** RtR’s VIAN Drone which has a Jetson Nano companion computer onboard and using Pixhawk Cube flight controller [9].



**Fig. 3.** P-DroNetRT methodology.

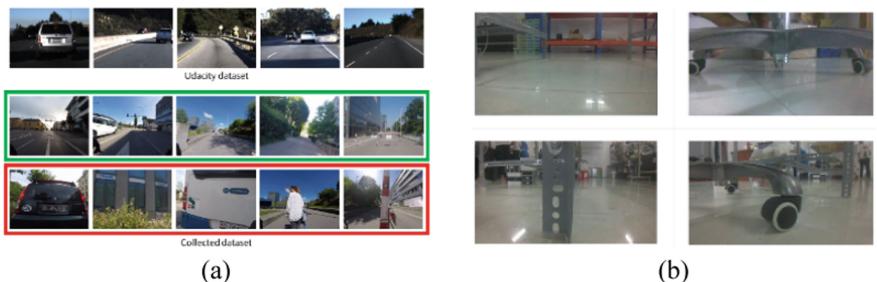
The processes in Fig. 3 are as following:

- In the re-training process, we have combined the datasets from UZH Zurich with the same driving scenarios in indoor environments as in Viet Nam.
- In the optimization process, the retrained P-DroNet model is converted to ONNX format using engine optimization of the TensorRT to speed up inference on Jetson Nano expected to increase significantly processing ability in real-time system.
- The implementation is to feed for the P-DroNetRT with frames from the 4 cameras streaming and processes them to predict collision possibility.

## 2.1 P-DroNet Building Process

### The PyTorch Conversion

PyTorch is supported by NVIDIA directly on Jetson Nano's Jetpack software suite [11]. Therefore, by successfully converting DroNet to PyTorch, the NVIDIA Jetson Nano embedded system is expected to perform the neural network faster than the one on Keras even without the optimization process. After the conversion, we compare the DroNet and P-DroNet side by side to make sure our network is rightly converted and have the same performance with the original. In the Table 1, we show quantitative and qualitative results of both DroNet and P-DroNet. The Avg. Accuracy and F-1 score of both networks are relatively the same with only 1–3% in difference, so we consider the DroNet is successfully converted into P-DroNet (Fig. 4).



**Fig. 4.** a) UZH Zurich's dataset [4]; b) HCMUT's dataset.

The authors then carry out the experiment of both network performance with the datasets at Ho Chi Minh City University of Technology (HCMUT) in Vietnam. During these experiments, both networks show a much lower accuracy as shown in Table 2. This could be caused by some reasons such as the quality of the camera and/or the uncalibrated camera parameters. It is also possible to retrain with the images of collision situations captured by the being used camera itself.

### Re-train the P-DroNet with Additional Data

The indoor datasets are called HCMUT's datasets and belong to the Ho Chi Minh City University of. So, the HCMUT datasets is combined with the UZH Zurich's datasets to form a new training dataset for P-DroNet.

The comparison results after retraining are presented in Table 3. The Avg. Accuracy indicates how accurately a collision is detected by the confusion matrix of the collision probability on testing images of the HCMUT’s dataset are shown in Fig. 5. At 80.8%, this is a good result consider the quality of input frames. The higher the F-1 score, the less imbalanced the network is. The P-DroNet has significantly lower accuracy performing on the HCMUT dataset than the UZH Zurich’s dataset, but it has superior processing speed thanks to build on the PyTorch platform expected to be suitable for applications that need fast response time.

**Table 1.** Training results of DroNet and P-DroNet on UZH Zurich’s dataset.

Model	EVA	RMSE	Avg. Accuracy	F-1 score
DroNet (Keras)	0.737	0.109	95.4%	0.901
P-DroNet (PyTorch)	1.043	0.111	92.1%	0.916

**Table 2.** Evaluating result of P-DroNet with dataset at HCMUT.

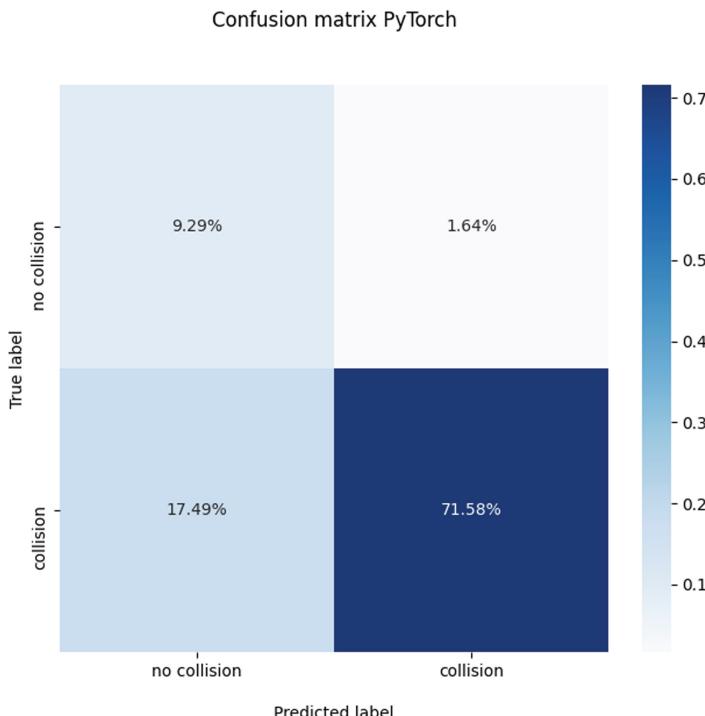
Model	Avg. Accuracy	F-1 score
P-DroNet (PyTorch)	77.3%	0.853

## 2.2 Network Optimization

The TensorRT [10] can help solve this problem by boosting the network efficient to a higher level by optimizing the P-DroNet structure for NVIDIA GPU using. There are two important factors in deploying the optimization process. For the first factor, we convert the PyTorch-based network to ONNX format that is a framework-agnostic model format (independency of any framework) and then build it into a TensorRT engine. For the second factor, we decided to use a standalone runtime and fed a batch of data into our engine then extract their predictions. The steps of this process are shown in Fig. 6 below.

## 3 Evaluation of the P-DroNet Model

In the Subsect. 3.1, the test cases were performed by feeding the net-works with video streaming at different resolutions. In the Subsects. 3.2, the test cases were conducted by feeding the networks with 4. The test cases in the Subsects. 3.3 were carried on by using a Jetbot (a Rover that would be explained more in the sub-Sect. 3.1.a) with a single-camera stream at the resolution of  $320 \times 240$ .



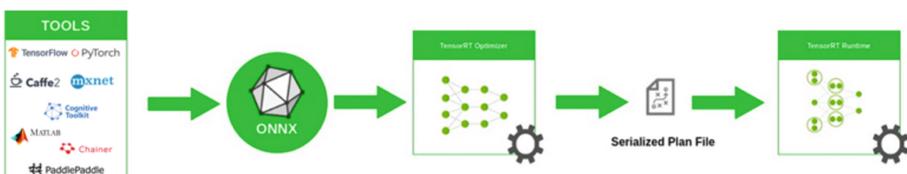
**Fig. 5.** Confusion matrix on the collision classification.

**Table 3.** Quantitative results on collision classification.

Model	Avg. Accuracy	F-1 score	Precision	Recall
P-DroNet (PyTorch)	80.8%	0.882	0.977	0.803

**Table 4.** Single-camera streaming performance on P-DroNet and P-DroNetRT.

Stream	PyTorch		TensorRT		
	FPS	CPU	GPU	CPU	GPU
1	70	50%	15%	150 FPS	60%



**Fig. 6.** TensorRT workflow [10].

### 3.1 Single-Camera Streaming Performance

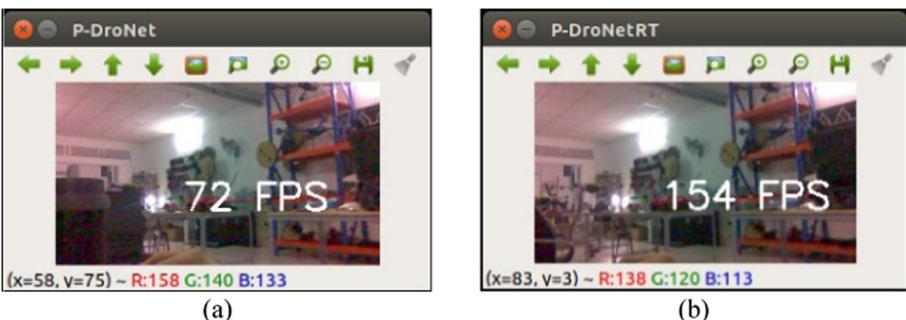
The P-DroNet achieves a processing speed above 40 fps at the resolution of  $320 \times 240$  pixels when combined with the control algorithm on the Jetbot. As the required processing speed for drones is at least 70 fps, the test cases have shown that it could be achieved at the resolution of  $256 \times 144$  pixels (see Fig. 7).

As the resolution of camera streaming reduces from  $320 \times 240$  pixels to  $256 \times 144$  pixels, the ability to detect collisions of the P-DroNet changes from 0.92 (true) to 0.34 (false) at the same distance. In case of a single-camera streaming, with higher resolution of camera streaming, and closer obstacles the collision detections are more accuracy. To maximize the real-time performance of the networks according to the suitable distance for the ability to detect collisions, the experiments in the next paragraphs were conducted at the resolution of  $256 \times 144$  pixels. By consequence, the P-DroNetRT network could achieve a processing speed above 150 fps. This result is almost twice the ability of the P-DroNet network (see Fig. 8).

The Table 4 shows the effect of the optimization process on the P-DroNet with single-camera streaming in comparison with the original one.



**Fig. 7.** a) Inference P-DroNet on Jetson Nano  $320 \times 240$  pixels at 43 fps; b) Inference P-DroNet on Jetson Nano  $256 \times 144$  pixels at 74 fps.



**Fig. 8.** a) Inference P-DroNet on Jetson Nano  $256 \times 144$  at 70 fps; b) Inference P-DroNetRT on Jetson Nano  $256 \times 144$  at 150 fps.

### 3.2 Multi-camera Streaming Performance

The development of using multi-camera streaming for P-DroNetRT network brought both good and bad results. It means that the P-DroNet is not able to process multi-camera streams due to the lack of resources. And the P-DroNetRT can reach a processing speed below 30 fps. In Fig. 9, the fps value rises to approximately 30 fps.

### 3.3 Real-World Experiments

Real-world experiments were conducted to test the performance of P-DroNetRT under real conditions. These experiments also aim to identify situations that confuse the neural network, thereby serving to improve the network in future.

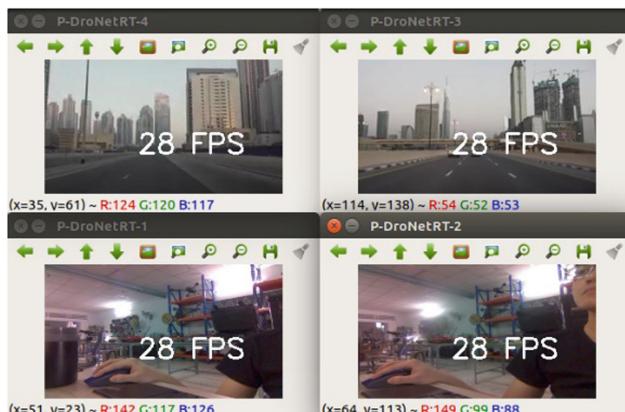


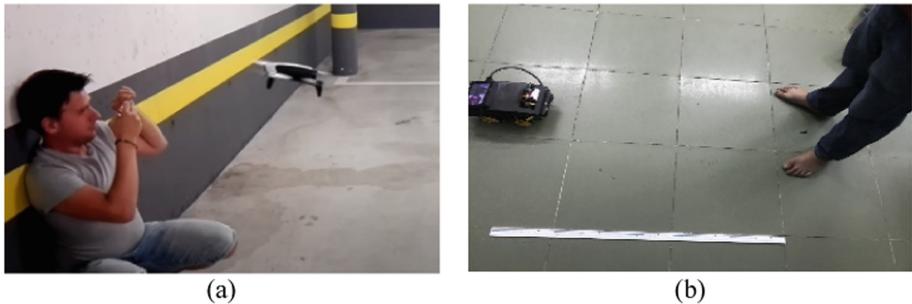
Fig. 9. P-DroNetRT multi-camera processing.

#### Fixed Obstacle Conditions

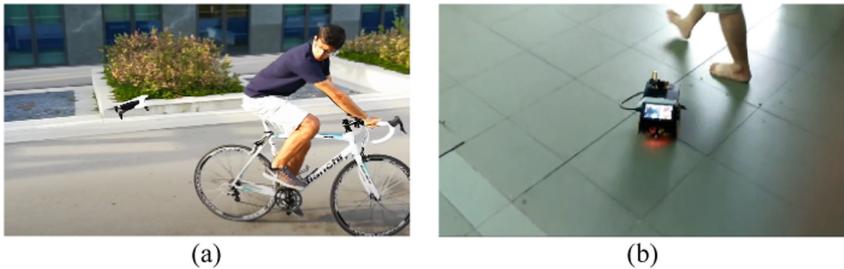
In this scenario, a person is placed at a distance of 200 cm from the Jetbot. The Jetbot is expected to stop when it encounters an obstacle within the designed recognition range. As a result, the P-DroNetRT model performs well within the range of 67 cm to 85 cm in both normal and low-light conditions. The experimental result of the P-DroNetRT detecting collision situation where the Jetbot moves to a person is like the DroNet in Fig. 10.

#### Moving Obstacles Conditions

In the unexpected collision situations, the person moves forward and suddenly hinders Jetbot's movement, thereby the collision distance is estimated since Jetbot encounters an obstacle and stops completely in Fig. 11. Once an impending collision situation is detected, the collision chance value continuously increases as the distance between the person and the Jetbot is shortened. In both of the above tests, Jetbot recognized a collision situation well at a distance of less than 40–50 cm.



**Fig. 10.** a) Avoiding a person in DroNet; b) Avoiding a person in P-DroNetRT.



**Fig. 11.** a) Person crosses over the Drone in DroNet; b) Person crosses over the Jetbot in P-DroNet.

### 3.4 Results Summary

Overall, the tests showed that the P-DroNetRT model was able to achieve the expected real-time performance at 30 fps for each camera stream. This performance satisfies most real-time applications. The resource consumption of the P-DroNetRT model is not quite high and could be optimized in the future if the 360-degree obstacle avoidance system is to be deployed on the drone. The overall program consumes about 70–80% of the NDIVIA Jetson Nano computer’s CPU usage. In real-life experiments, the P-DroNetRT model performs quite well in recognizing collision situations.

## 4 Conclusions

The study in this paper proposes an approach on how to convert, retrain and optimize the DroNet neural network model to develop on the NDIVIA Jetson Nano embedded computer. It is named the P-DroNetRT model. With 4-camera streaming, the P-DroNetRT model could recognize in collision detection at approximately 30 fps for each camera. This processing speed meets the input capacity of the camera system and has a good potential for deployment on the commercial drones of the Real-time Robotics Vietnam Ltd. These drones will be equipped with the ability to recognize obstacles around them

in real time to improve safety for the human pilot as well as the drone itself. The results of the P-DroNet model show that it uses a quite high re-source consumption and not able to process multi-camera streaming. While the P-DroNetRT model seem to be “better” and success in processing 4-camera streams. The obstacle avoidance program consumes approximately 70% of the NDIVIA Jet-son Nano computer’s CPU and 30% of the GPU. This is a quite high consumption rate, however, there is still space for other applications to be deployed on this computer onboard the drone.

## 5 Future Work

In this research, the obstacle avoidance system was only implemented on a Jetbot, which is an omni ground vehicle instead of a quadcopter. Subsequently, some further development for this application can be (1) Implement the obstacle avoidance system to quadcopter and examine its capability in real life. (2) Enhance the accuracy of the program in complex environment lightning condition by collect and retrain the network with more data of these environment.

**Acknowledgement.** We sincerely thank the Real-time Robotics Vietnam Ltd. and Ho Chi Minh City University of Technology (HCMUT), VNU-HCM. This research is funded by VNU-HCM under grant number C2021-20-11. We acknowledge the support of time and facilities from Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for this study.

## References

1. Grzonka, S., Grisetti, G., Burgard, W.: Towards a navigation system for autonomous indoor flying. In: IEEE International Conference on Robotics and Automation, ICRA’09, pp. 2878–2883. IEEE Press, New York (2009)
2. Von Stumberg, L., Usenko, V., Engel, J., Stuckler, J., Cremers, D.: From monocular SLAM to autonomous drone exploration. In: Proceedings of the 2017 European Conference on Mobile Robots, ECMR 2017, Paris, France, 6–8 September (2017)
3. Giusti, A., Guzzi, J., Cirean, D.C., He, F.L., Rodrguez, J.P., Fontana, F., Faessler, M., Forster, C., Schmidhuber, J., Caro, G.D., Scaramuzza, D., Gambardella, L.M.: A machine learning approach to visual perception of forest trails for mobile robots. In: IEEE Robotics and Automation Letters (2016)
4. Loquercio, A., Maqueda, A.I., Del-Blanco, C.R., Scaramuzza, D.: DroNet: Learning to fly by driving. *IEEE Robot. Autom. Lett.* **3**, 1088–1095 (2018)
5. Palossi, D., Loquercio, A., Conti, F., Flamand, E., Scaramuzza, D., Benini, L.: A 64-mW DNN-based visual navigation engine for autonomous nano-drones. *IEEE Internet Things J.* **6**, 8357–8371 (2019)
6. Gandhi, D., Pinto, L., Gupta, A.: Learning to fly by crashing. In: Proceedings of the IEEE International Conference on Intelligent Robots and Systems, Vancouver, BC, Canada, 24–28 September 2017; Vol 2017, pp. 3948–3955
7. Dionisio-Ortega, S., Rojas-Perez, L.O., Martinez-Carranza, J., Cruz-Vega, I.: A deep learning approach towards autonomous flight in forest environments. In: Proceedings of the 2018 International Conference on Electronics, Communications and Computers (CONIELECOMP), Cholula, Mexico, 21–23 February 2018, pp. 139–144

8. Tran, Q.K. et al.: Object detection for drones on Raspberry Pi: Potentials and challenges. In: IOP Conference Series: Materials Science and Engineering, vol. 1109, 012033 (2021)
9. Real-time Robotics Viet Nam Company Limited: <https://rrobotics.com/>
10. TensorRT NVIDIA: <https://docs.nvidia.com/deeplearning/tensorrt/developer-guide/index.html>
11. PyTorch 1.9.0 Documentation: <https://PyTorch.org/docs/stable/nn.html>



# Optimizing Penta-Band Vivaldi Antenna for 5G, WiFi 6, Radar and Satellite Applications

Thi Thanh Tu Duong<sup>(✉)</sup>, Minh Duc Hoang, and Thi Thu Nga Nguyen

Faculty of Telecommunication 1, Posts and Telecommunications Institute of Technology,  
Hanoi, Vietnam

{tudtt, ngavt}@ptit.edu.vn, duchm.B18VT109@stu.ptit.edu.vn

**Abstract.** The Vivaldi antenna is used to design for ultra-wideband applications due to its characteristic of broadband traveling waves. In this paper, we optimized a Vivaldi structure to become a pentaband antenna. Fabricated on FR4 substrate with a dielectric constant of 4.3, the size of antenna is rather compact with a patch dimension of 34.61 mm × 40 mm which is suitable for modern terminals. Operating at 3.8 GHz, 5.8 GHz, 8 GHz, 10.5 GHz, and 12.8 GHz with a large bandwidth of 769 MHz, 1959 MHz, 1965 MHz, 1906 MHz, and 2857 MHz respectively, the proposed antenna can apply for applications of 5G, WiFi 6, radar, and the satellite of C, X, Ku band. The antenna design dimensions are calculated and optimized. The S11 parameter, current density, radiation pattern, and resistance matching are analyzed by using CST-MW software. Besides, the reflection coefficient simulation and measurement results agree well.

**Keywords:** Vivaldi structure · Wideband · Pentaband

## 1 Introduction

At the present, wireless communication technology has a significant development that created a lot of improvement in transmission performance, speed. These make the great motivation for the development of novel antenna structures that achieve wider bandwidth as well as long transmission capability. In addition, terminals nowadays tend to integrate multi-technology and multi-application within their compact size. Therefore, low-configuration planar antennas that can be easy to integrate on the printed circuits often attract the attention of researchers [1].

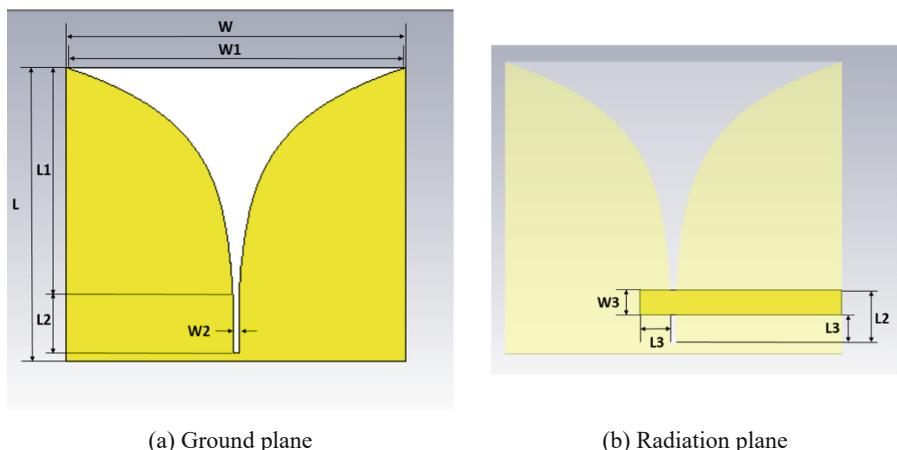
With the characteristic of broadband performance, the Vivaldi structure is usually applied for ultra-wideband (UWB) design [2]. The fabricated antenna gets a low cost and the radiation characteristic has many advantages such as low side lobes, constant beam width, and high directivity [2]. In the last few years, quite a lot of research on Vivaldi antennas for advanced communication systems has been published. Most of these publications focus on UWB antenna design [1, 3–7]. Because of its ultra-wide bandwidth characteristic (from a few GHz to tens of GHz), Vivaldi UWB antennas only work in short ranges (<100 m). Additionally, this ultra-wideband feature often interferes with the spectrum of many available bands of other advanced communications technologies

such as 802.11ac in the 5 GHz band, 802.16 m in the 3.5 GHz band, and 5G in the sub-6GHz band. Hence, UWB antennas often had additional slots or auxiliary parts to remove these bands, which makes the design more complex and difficult to optimize [8]. A new research trend in Vivaldi antenna is the multiband antenna that solutions for wideband applications in today's advanced communications such as 5G/6G, IoT/IIoT, radar, and satellite systems [9–11]. R. Sturdivant and E. Chong [9] proposed a wideband Vivaldi antenna for 5G/ IoT applications. However, this antenna only works in a single band from 0.7 GHz to 2.4 GHz. This is also a common feature of many Vivaldi antenna designs because there is only one center resonant frequency and crossover frequency that can be calculated based on geometric parameters of the original Vivaldi antenna structure [2]. Tale Saeidi [10] and A.K. Patnaik [11] and his colleagues proposed a Vivaldi multiband antenna for wideband application, but the proposed design only has three operating bands.

In our study, we propose a multiband antenna that can operate in five broad bands with good gain and radiation efficiency, which can be used for broadband applications for 5G, WiFi 6 communication, radar, and satellite communications in the C, X, and Ku bands. Besides, by using CST simulation software, the influence of parameters is also analyzed and evaluated to optimize the bandwidth and operating frequency of the antenna.

## 2 Vivaldi Antenna Design

The detailed design and all dimensions of pentaband antenna are shown in Fig. 1 and Table 1. This design is a coplanar Vivaldi structure that has a copper layer on both sides of the substrate.



**Fig. 1.** The structure of the pentaband antenna

The tapered slot is created by two exponential curves, which is determined by the following formula [2]:

$$y = \pm A \times e^{px} \quad (1)$$

where  $y$  is an exponential function with the variable  $x$ .  $A$  is half of the uniform slot's width ( $W_2$ ).  $p$  is the taper rate. This exponential curve is very important in determining the raw values of the design parameters such as  $L_1$ ,  $L_2$ ,  $W_1$ , and  $W_2$ . The crossover length ( $L_3$ ) is determined to be approximately one-quarter of the length of the wavelength at the lowest band. That wavelength is calculated according to the formula [2]:

$$\lambda(m) = \frac{3 \times 10^8}{f \times \sqrt{\epsilon}} \quad (2)$$

where  $f$  is the center frequency of the lowest band, which is the operating frequency that we determined from the beginning in the antenna design goal for 5G communication;  $\epsilon$  is the dielectric constant of the substrate. In this design, we use conventional material, FR4. The substrate thickness is 1.6 mm. The thickness of copper layer is 0.035 mm and the dielectric constant is 4.3. Because of specificity of the Vivaldi antenna and also to facilitate optimization, the substrate's width is equal to the printed antenna's width. All details of the proposed antenna's geometrical dimensions are calculated, analyzed and optimized on the commercialized CST simulation tool.

**Table 1.** Dimensions of the proposed antenna

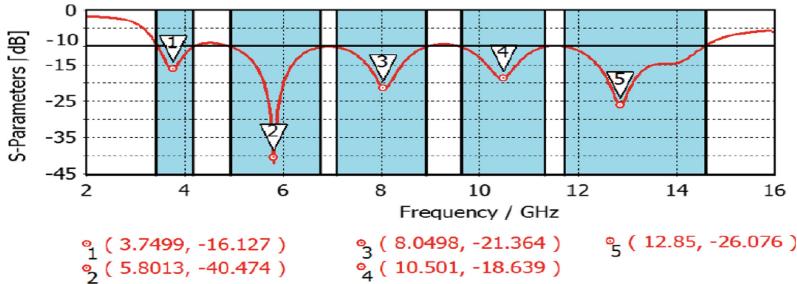
Dimensions	L	W	L1	W1	L2	W2	L3	W3
Size (mm)	34.61	40	27	39.8	6.61	0.7	3.61	3

### 3 Simulation Results

Basic antenna parameters for example bandwidth that is determined by the reflection coefficient ( $S_{11}$ ), resistance matching which is evaluated by the input impedance, directivity and gain which attained from 2D and 3D radiation patterns, and current density at the resonant frequencies are analyzed and evaluated based on CST electromagnetic field simulation software.

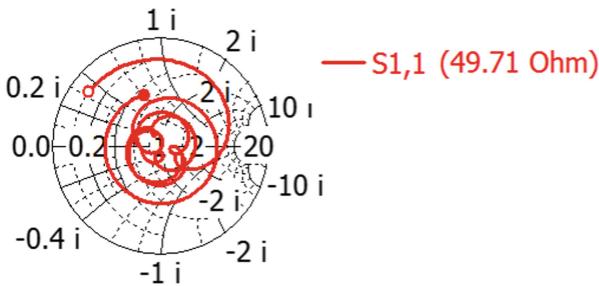
As shown in Fig. 2, the antenna resonates at 5 frequencies corresponding to 5 available bands of 5G, WiFi 6, radar, and satellite communications.

- Band 1: from 3.444 to 4.213 GHz (769 MHz wide); resonant at 3.75 GHz.
- Band 2: from 4.989 to 6.948 GHz (1.959 GHz wide); resonant at 5.8 GHz.
- Band 3: from 6.948 to 8.913 GHz (1.965 GHz wide); resonant at 8.05 GHz.
- Band 4: from 9.642 to 11.548 GHz (1.906 GHz wide); resonant at 10.5 GHz.
- Band 5: from 11.743 to 14.6 GHz (2.857 GHz wide); resonant at 12.85 GHz.

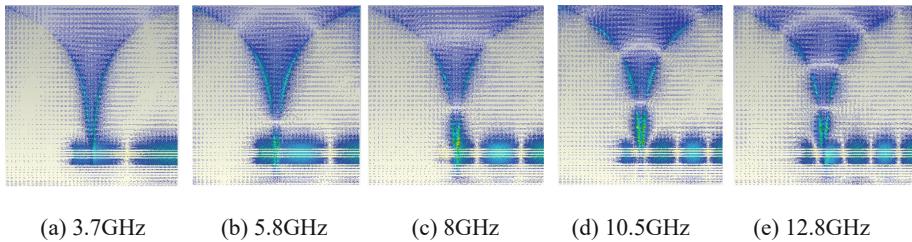


**Fig. 2.** The reflection coefficient

Figure 3 illustrates the Smith chart. We can see that at the resonant frequencies, the antenna's impedance reaches a value of approximately  $50\Omega$ , which ensures impedance matching conditions for power supply in real life. The electric field in the tapered slot is bent to create radiation. As shown in Fig. 4, we can easily notice that the higher the resonant frequency is, the more electric field line appear in the tapered slot.



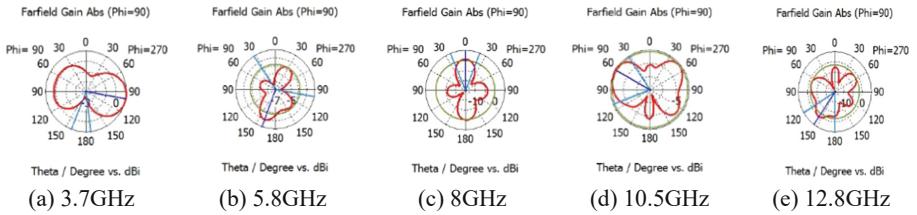
**Fig. 3.** Smith chart



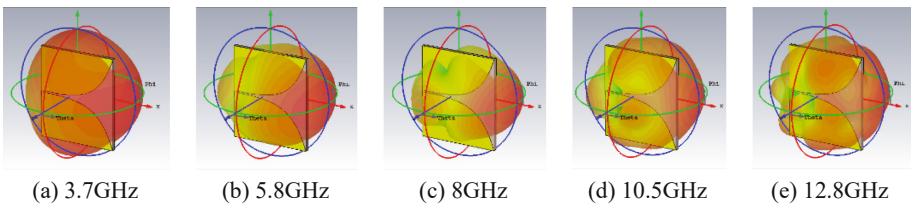
**Fig. 4.** The current density at different resonant frequencies

Figures 5 and 6 show the radiation patterns of the pentaband antenna in 2D and 3D mode. The gains of antenna are 2.2 dBi, 5.213 dBi, 6.77 dBi, 4.74 dBi, and 3.33 dBi at the resonant frequencies of 3.7 GHz, 5.8 GHz, 8 GHz, 10.5 GHz, and 12.8 GHz respectively. The directivity also are 3 dBi, 6.2 dBi, 8.08 dBi, 6.54 dBi and 6.17 dBi, respectively. Thus, the antenna is capable of operation in 5 bands with good gain and

radiation efficiency at the band for 5G, WiFi 6 applications or 8GHz band of radar communication, and acceptable ones at the remaining bands and applications.



**Fig. 5.** The 2D radiation pattern at different resonant frequencies



**Fig. 6.** The 3D radiation pattern at different resonant frequencies

From the statistics which are detailed in Table 2, we can see that the size of the pentaband antenna is smaller [11, 12] and only our study achieves operation at 5 bands, all bands are wider than ones in other works.

**Table 2.** Performance comparison of the pentaband antenna

Ref	Dimension (mm <sup>2</sup> )	Resonant frequency (GHz)	Bandwidth – 10 dB	Gain (dBi)	Number of bands
[10]	24.8 × 30	3.8	1.6 GHz	5.78	3
		5.2	1.2 GHz	4.73	
		8	1 GHz	5.29	
[11]	40 × 60	3.9	296 MHz	NA	3
		8.7	210 MHz		
		13.9	389 MHz		
[12]	40 × 58.5	2.4	189 MHz	3.76	3
		4	500 MHz	2.68	

(continued)

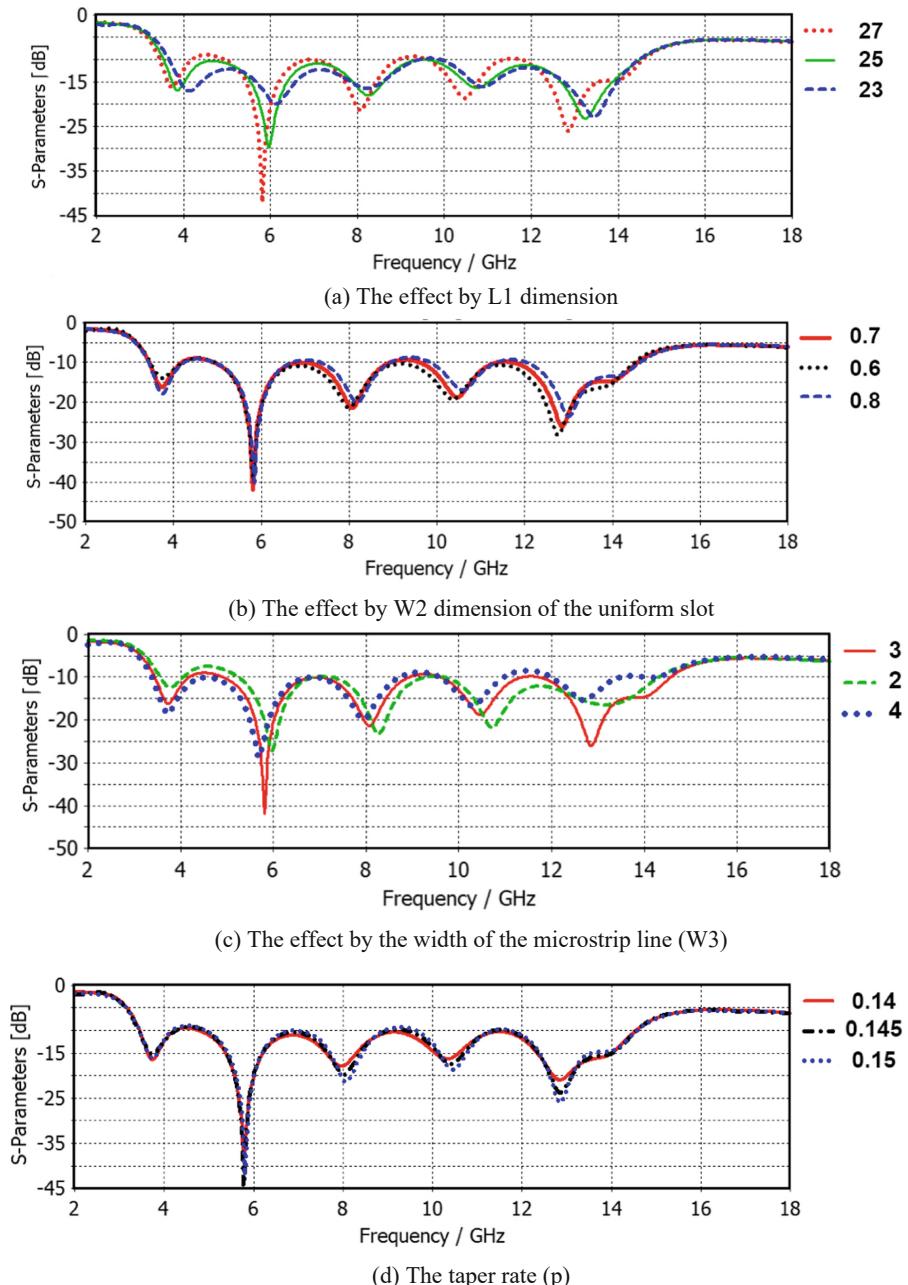
**Table 2.** (continued)

Ref	Dimension (mm <sup>2</sup> )	Resonant frequency (GHz)	Bandwidth – 10 dB	Gain (dBi)	Number of bands
This paper	34.58 × 40	5.6	1 GHz	4.5	5
		3.7	769 MHz	2.2	
		5.8	1.959 GHz	5.213	
		8	1.965 GHz	6.77	
		10.5	1.906 GHz	4.74	
		12.8	2.857 GHz	3.33	

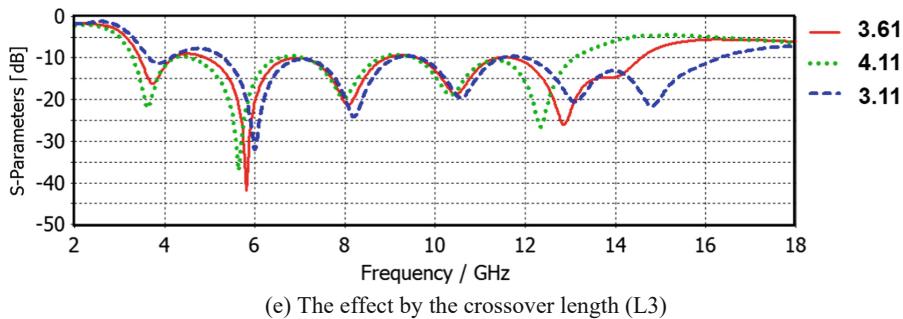
## 4 Effect of the Dimensional Parameters on the Antenna Performance

Based on the conventional Vivaldi structure, the dimension of each part has interferences to a certain extent with the simulation's results as well as the performance of the antenna and they will analyze in this section.

- **The tapered slot length (L1):** The tapered slot length affects greatly the resonant frequency of the antenna. As shown in Fig. 7(a), when the length reduces from 27 mm to 23 mm with a step of 2 mm, the resonance at all frequencies decreases. The reason is that reducing the length will lead to a decrease in the directionality of the antenna, and also cause the maximum width of the tapered slot to decrease, which makes the radiation weaker.
- **The width of the uniform slot (W2):** The influence of the uniform slot's width is clearly shown in Fig. 7(b). As can be seen, this parameter almost has no affection on low-frequency bands but it can change the resonance at high frequencies with the resonance inversely proportional to the resonant bandwidth. Nevertheless, we need to keep in mind the lower limit of the slot's width. If the width is too narrow, the radiation's energy will be lost. Furthermore, it is also difficult to fabricate printed circuit technology.
- **The width of the microstrip line (W3):** In Fig. 7(c), the influence of this line's width show that if the width increase, the resonance at low frequencies will increase and the resonance at high frequencies will greatly decrease. However, what we need to be concerned about is that this parameter is the only parameter that can adjust the impedance matching problem. Through Smith charts shown in Fig. 8 and the simulation result, the impedance of the antenna with the width of the microstrip line at 4 mm, 3 mm, and 2 mm are 42.28 Ω, 49.71 Ω, and 60.94 Ω. We can see easily that the wider the microstrip line, the larger the area connect to the source, the lower the input impedance.
- **The taper rate (p):** The taper rate of the antenna is determined by the parameter  $p$ . As shown in Fig. 7(d), the increase of the taper rate leads to an increase in the maximum



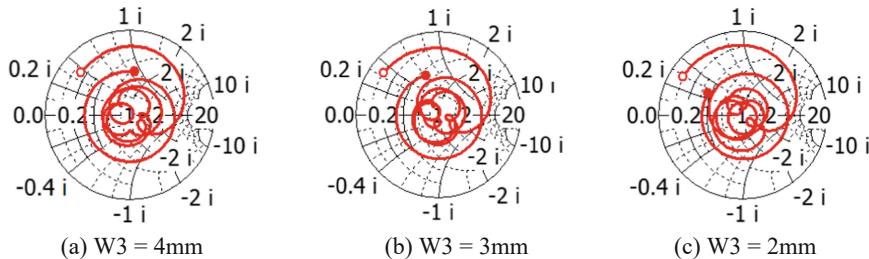
**Fig. 7.** Analysis of the size parameters affecting the reflection coefficient characteristics



(e) The effect by the crossover length (L3)

**Fig. 7.** continued

width of the tapered slot, which makes it easier for the antenna to radiate. As the antenna's aperture increases, we also see that low frequencies' resonance decreases and high frequencies' resonance increases. Moreover, when adjusting both the taper rate and the length of the tapered slot, we can successfully change the resonance's range as well as the number of resonant bands of the antenna.

**Fig. 8.** Smith charts with different widths of the microstrip line**Table 3.** The change of gain at the resonance frequencies

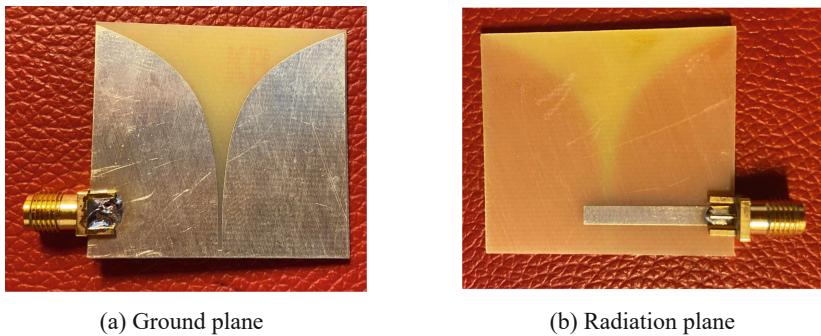
Frequencies	The crossover length (L3)		
	3.11 mm	3.61 mm	4.11 mm
3.7 GHz	1.96 dBi	2.2 dBi	2.25 dBi
5.8 GHz	5.152 dBi	5.213 dBi	5.18 dBi
8 GHz	6.699 dBi	6.771 dBi	6.817 dBi
10.5 GHz	4.78 dBi	4.74 dBi	4.63 dBi
12.8 GHz	4.03 dBi	3.33 dBi	2.43 dBi

- **The crossover length (L3):** As discussed in Sect. 2, this is the parameter that determines the center frequency's wavelength as well as the gain's distribution. As shown in Fig. 7(e), with the optimal limit range of 1/4, if the L3 is longer, the resonance

area will shrink, the resonant frequency at the center will decrease, and the resonance of a few bands will increase to maximum then decrease. Additionally, the crossover length also affects the distribution of gain in the resonance area. According to Table 3, when the crossover length increases, the gain at high frequencies reduces the low frequencies' gain increases. Therefore, when designing antennas, it is necessary to carefully calculate and adjust the balance between resonance and gain to meet the requirements.

## 5 Measured Result

To ensure the pentaband operation of the proposed antenna, we use FR4 substrate to fabricate as shown in Fig. 9. The antenna substrate thickness is 1.6 mm and the dielectric constant of FR4 material is 4.3.

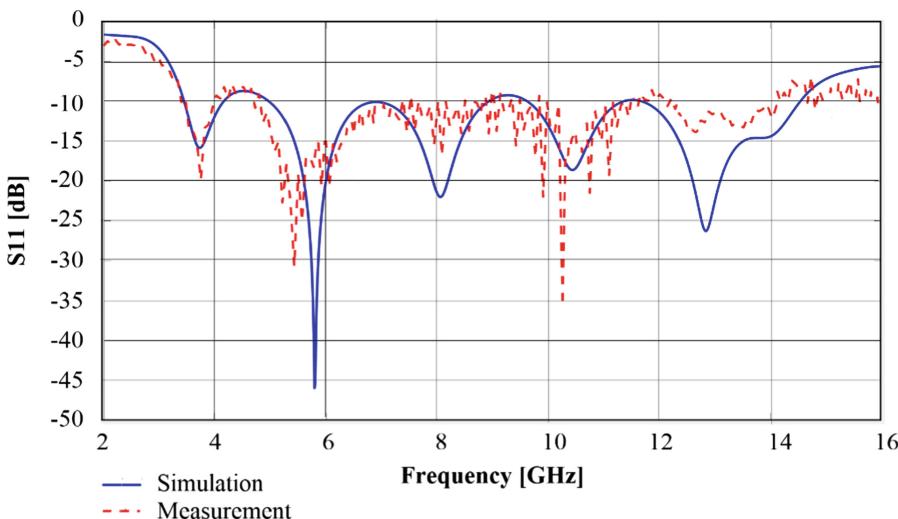


(a) Ground plane

(b) Radiation plane

**Fig. 9.** The pentaband Vivaldi antenna based on FR4 material

The comparison of measured and simulated S11 parameter results is shown in Fig. 10. It can be seen that they agree well. The fabricated antenna has five operating bands with wideband for modern communication systems such as 5G, WiFi 6, radar, and satellite.



**Fig. 10.** The comparison of measured and simulated pentaband Vivaldi antenna

## 6 Conclusions

This paper proposed a pentaband Vivaldi antenna for broadband applications in advanced communications. Besides, based on the analysis and the optimization of the size parameters, this study has found out that not only basic parameters as determined in theory such as the width of the substrate, the taper rate, the crossover length, but other parameters also have a significant influence on the resonant frequencies, bandwidth or gain of Vivaldi antenna. This analysis will help researchers to optimize and design multi-band Vivaldi antennas with desired resonant frequencies for different wireless applications.

## References

1. Madhav, P.V., Prasad, M.S.G.: Characterization of printed pedestal Vivaldi antenna (8–18 GHz) on RT duroid with single and double cavity. In: Advances in Array Optimization. IntechOpen (2020)
2. Balanis, C.A: Antenna Theory: Analysis and Design, 4th edn. Wiley, Hoboken (2016)
3. Bhattacharjee, A., Bhawal, A., Karmakar, A., Saha, A., Bhattacharya, D.: Vivaldi antennas: a historical review and current state of art. *Int. J. Microwave Wirel. Technol.* **13**(8), 833–850 (2020)
4. Tseng, V., Chang, C.Y.: Linear tapered slot antenna for ultra-wideband radar sensor: design consideration and recommendation. *Sensors* **19**(5), 1212 (2019)
5. Kumar, M., Agarwal, S., Sharma, A.: A multi-application compact ultra wideband Vivaldi antenna for IoT, 5G, ITS, and RFID. In: 2019 IEEE Indian Conference on Antennas and Propagation (InCAP) (2019)
6. Yin, Z., He, G., Yang, X.-X., Gao, S.: Miniaturized ultra wideband half-mode Vivaldi antenna based on mirror image theory. *IEEE Antenna Wirel. Propag. Lett.* **19**(4), 695–699 (2020)
7. Rittiplang, A., Phasukkit, P.: 1-Tx/5-Rx through-wall UWB switched-antenna-array radar for detecting stationary humans. *Sensor* **20**(23), 250 (2020)

8. Wu, W., Li, Y., Wu, R., Shi, C.B., Cui, T.: Band-notch UWB antenna with switchable and tunable performance. *Int. J. Antenna Propag.* **2016** (2016). Article no. 9612987
9. Sturdivant, R., Chong, E.K.P.: Dielectric notch radiator antennas with integrated filtering for 5G and IoT access. In: 2017 IEEE Radio and Wireless Symposium (RWS), pp. 197–200 (2017)
10. Saeidi, T., et al.: High gain triple-band metamaterial-based antipodal Vivaldi MIMO antenna for 5G communications. *Micromachines* **12**(3) (2021). 24 pages
11. Patnaik, A.K., Suresh, S., Swathi, A., Prasad, L.B., Kishore Babu, G., Koppisetti, A.: Analysis and design of balanced Vivaldi antenna for triple band applications. In: Satapathy, S.C., Bhateja, V., Chowdary, P.S.R., Chakravarthy, V.V.S.S.S., Anguera, J. (eds.) Proceedings of 2nd International Conference on Micro-Electronics, Electromagnetics and Telecommunications. LNEE, vol. 434, pp. 437–443. Springer, Singapore (2018). [https://doi.org/10.1007/978-981-10-4280-5\\_46](https://doi.org/10.1007/978-981-10-4280-5_46)
12. Tu, D.T.T., Duy, P.D., Do, V.D., Tan, N.V.: Triple band Vivaldi antenna for construction monitoring based on the IIoT systems. In: International Conference on Advanced Technologies for Communications (ATC 2021), HoChiMinh City, Vietnam, 14–16 October 2021, pp. 179–183 (2021)



# The Method Calculates of Radial Distance from the Human Body Surface to the Inner Surface of a Women's Shirt

Do Thi Thuy<sup>1,2(✉)</sup>, Nguyen Thi Thuy Ngoc<sup>2</sup>, and Ngo Chi Trung<sup>2</sup>

<sup>1</sup> Faculty of Garment Technology and Fashion Design, Hanoi University of Industry, Hanoi,  
Vietnam

thuy.dothi@hau.edu.vn

<sup>2</sup> School of Textile – Leather, and Fashion, Hanoi University of Science and Technology, Hanoi,  
Vietnam

**Abstract.** Designing garments requires visualization, fast and fitting to not only meet the requirements for production in the garment industry but also satisfy the needs of the wearer. There are many studies on 3-dimensional garment design to respond to these requirements, however, it has not been fully addressed. A method to calculate the radial distance from the human body surface to the inner face of a shirt for 3-D shirt design is presented in this paper. Rapidform software is used to model and measure the distance from the center to the cross-section contour of the human body and shirt. R and Excel software are used to analyze and calculate the radial distance from the surface of the human body to the inner face of the shirt based on a multivariable linear regression model for each region, testing and evaluating the reliability of the model generated. As result, the multivariable linear regression model of 192 radial distance regions has been determined. Based on this model, the study has shown that the method of calculating the radial distance from the surface of the body to the inside of the shirt with a custom width, calculates the ease of the bust, and gives the same result as ease on the real shirts. The study also showed that the variation of the radial distance at the bust region of the side and back is mainly influenced by the ease of bust circumference, while the radial distance near the center front is mainly influenced by the position of the cross-section and position of the point.

**Keywords:** Radial distance · Ease · Multivariable linear regression model

## 1 Introduction

With the advancement of 3D scanning techniques, individual bodies and shirts can be easily obtained, then modeled and shirts designed in some different ways. S.Thomassey et al. have designed a shirt with ease based on the reverse 3D method [1]. The authors calculated the distance from the body to the clothes by taking measurements and marking the intersection points of the shirt with the body, then scanning the human body in 2 cases of wearing a shirt and not wearing a shirt; Acquisition of human body surfaces

across predefined boundary parameters; Stack images on top of each other; Divide the space between the clothes and the body; Calculate the distance between the body and the shirt; Acquisition of convergent data for each cross-section; Get the final parts out in the CAD platform and check the results by simulating the fabric on the reference body. This study provides a method to identify, quantify and control allowable ease in 3D for mass or custom production. However, with this design, only 1 shirt can be designed with the same reference body, the same distance, and the same garment material. To study the static and dynamic ease distribution of bust and waistlines on a shirt, Zheng et al. [2] used a three-dimensional motion capture system (3DMCS). Experiments to test the ease changes in distribution with body movement. The study showed the distribution of static ease on different samples at different body landmarks. In the chest, ease tends to be more concentrated in the left and right regions, while the waist ease seems to be more concentrated in the anterior and posterior regions. Using 3D scan data, Jihong Xu et al. [3] investigated the ease relationships between the standard body and a particular shirt style with different sizes. The normal distance between the body surface and the shirt curves is used to calculate the ease relations between the body and the shirt. Ease distribution analyzed the curves of different sections of the body and shirt along with the variation of body angles. Jihong Xu et al. [4] analyzed the change of the ease distance with the change of body angles. The results show that the ease distance is abnormally changed with the increase of shirt size. Ease distances are affected by the allowable ease and the mechanical properties of the material. Fangfang Zhang and Trevor John Little, the 3D ease will be evenly added into the garment during the 3D garment pattern design process, but the ease amount varies in terms of different parts of the garment [5]. Garment visual appearance and distance ease between straight fit dress and mannequin in respect to fabrics properties were investigated by Agn   LAG   et al. [6]. It was defined that distance ease values at bust and at waist girth of real and virtual dresses differed because of wrinkles in this area. Yuxiang Zhu et al. present a dual adaptive garment slice adjustment technique for automatic resizing of apparel products with variant body shapes, and this technique can quickly generate clothed characters [7]. Their contribution is to propose a novel fit evaluation method. Their method enables fully automatic adjustment, preserving the original pattern style of garments between characters with vastly body shapes. Yongjoon Lee et al. proposed a method that can transfer garment models between human models without any user guidance even when the reference and target human models have different poses [8]. Their goal is not to resize or deform the original garment model according to the target human model but to yield realistic fitting results of the given garment on the target human models. Using pose-independent segmentation and cloth simulation, they achieve realistic and automatic fitting results in reasonable running time.

It can be seen that the studies have had different methods to support the design of 3D clothing, but they are still incomplete. A method for calculating the radial distance from the surface of the human body to the inner surface of a shirt for 3-D shirt design is shown below.

## 2 Method and Objective

### 2.1 Research Subjects

Research subjects selected according to the national standard TCVN 5782:2009 [9], Vietnamese women from 18 to 25 age, with a standing height, is 158 (155–160) cm, a bust circumference is 88 (87–90 cm) and a bottom circumference is 92 (90–93) cm. Making patterns and sewing 3 women's shirts with 10, 35, and 50 mm ease of half bust circumference, corresponding to A1, A2, and A3 shirts. All 3 shirts are without collars and sleeves.

### 2.2 Using Sampling Survey Method to Determine the Sample Size for the Research

According to the preliminary survey of the 5 human body measurements as required in the research subject item and analyzing the variance of the data gives the results the probability  $p_1$  of the human bodies with the same height, bust circumference, and bottom circumference is 60%. The probability  $p_2$  of the shirt has the same length, bust circumference, and hem circumference is 100% (because the shirt is ready-made according to the original measurements, so it does not change during the experiment). Therefore, the effect is  $d = p_2 - p_1$ , so,  $d = 0,4$ . Choose power for the study to be 0,9.

The standard deviation of  $d$  is  $SD = \sqrt{p_1(1 - p_1) + p_2(1 - p_2)} = 0,5$ . Effect size  $ES = \frac{d}{SD} = 0,8$ . Therefore, the number of sample sizes needed for the research is  $N$ .

$$N = \frac{C(\alpha, \beta)}{ES^2} = 16,4 \sim 17 \text{ (samples)}$$

In which,  $C$  is a constant depending on the error  $\alpha$  and  $\beta$ ,  $C$  is taken according to the normal distribution,  $C = 10,5$ ;  $\alpha$  is type 1 error and  $\alpha = 0,05$ ;  $\beta$  is the error of type 2 and  $\beta = 0,1$ . Thus, it is necessary to have at least 17 human body samples to scan and measure for this research.

### 2.3 Using Interpolation Method to 3D Model the Human Body and Shirt

Based on the research "Mannequin model from 3D scanning data" [10] to model human bodies and shirts to ensure the accuracy required for this research.

#### Identify Anthropometric Points on the Human Body and 3D Scan

The identification of anthropometric markers and human body landmarks is very important, affecting the accuracy and consistency of the data obtained between scans. Use the pen to trace the landmarks on the body and the corresponding landmarks on the shirt. The landmarks taken according to the international standard ISO 8559-1:2017(E) [11] and to support and identify in the process of data processing are the 7th vertebrae, pharynx, 2 points of shoulder, 2 points of the breast, 2 points of chest root line, 2 points

of underarm, 2 points of the waist at the side, 2 points of protruding hip. 3D scanning of the human body while wearing a T-Shirt bra and shirts with the Artec Eva handheld scanner under normal conditions, then processing the scanned data on Rapidform XOR3 software.

### **Model of the Human Body when Wearing a T-Shirt Bra and when Wearing Shirts**

In order to fully model an object from 3D scan data, the research follows these steps: Identify major cross-section of the human body and shirt passing through anthropometric points on the body; Identify sub-cross-section on the human body and shirt; Create a surface that passes through the contours of the cross-section using boundary curve interpolation. Assuming the human body model surface, the shirt model is symmetrical on the right and left sides. Create a smooth surface passing through the left half of the cross-section contour. Then, take the symmetry to the right half and attach the two halves of the model surface together to form the complete model.

#### **2.4 Use Rapidform XOR3 Software to Measure the Radial Distance Between the Body Surface and the Inner Surface of Each Shirt**

Choose the coordinate origin point for the human body model and the shirt model is the intersection of the protruding hip cross-section, the middle front vertical slice, and the vertical slice passing through two points under the armpits.

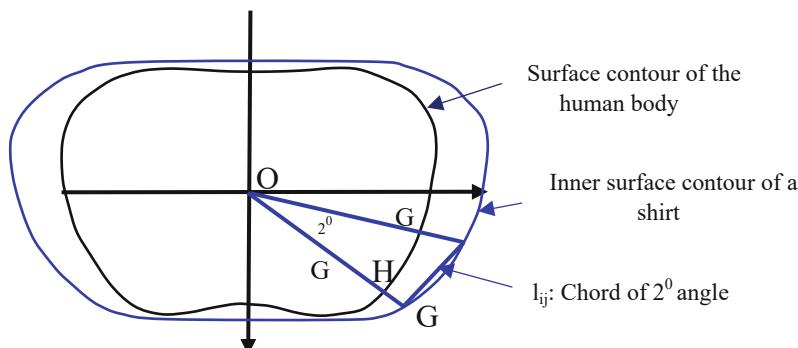
##### **Measure the Distance from the Center of the Cross-section to the Surface of the Human Body**

Cut the human body model and the shirt model with horizontal slices, parallel to each other and 2 mm apart. Divide each slice of the human body model into 180 parts, each with an angle of 2 degrees around the center of the slice. The contour of each slice has 180 points, so each half-slice has 90 points. Number the points from 1 to 90 and the numbers are placed in sequence from the front center then the side and to the back of the human body. From the center of each human body model slice, measure the distance to points on the body surface contour.

##### **Measure the Distance from the Center to the Inner Surface of the Shirt**

Each half cross-section of the shirt has 90 points. Number the points from 1 to 90 same as on the human body. Because when modeling a shirt, the outer surface of the shirt will be modeled, so the thickness of the shirt when modeled is equal to the thickness of the actual fabric and the inner surface of the model is now the inner surface of the shirt. From the center of each shirt model cross-section, measure the distance to points on the shirt inner surface contour.

##### **Determine the Radial Distance Between the Surface of the Human Body and the Inner Surface of the Shirt at each Point of each Model**



**Fig. 1.** Human body surface and shirt inner surface cross-section

Let  $O$  be the center of the cross-section (Fig. 1). From  $O$  draw a line that cuts the body surface contour at  $H$  and the shirt inner surface at  $G$ . Then,

$$HG = OG - OH \quad (1)$$

## 2.5 Use Statistical and Computational Methods to Calculate the Radial Distance from the Surface of the Human Body to the Inner Surface of the Shirt

Since the surface of the human body and the shirt are complex curved surfaces, the space between them also varies at locations. This study used the method of data collection and processing and study the relationship between the data to divide this space into small regions with radial distances tending to be similar on shirts to explore data for later shirt design.

**Region Division.** To zone the radial distance, the most reasonable way is to divide the region by cross-section and zone by point. By representing the distance of the points on each cross-section for 3 shirts, then, visually dividing the points on the cross-section. Points with radial distances between the human body surface and the inner surface of the shirt tend to vary similarly and they are adjacent slices so form groups of slices for the three shirts. Similarly, cross-sections. Combine cross-section groups and point groups to form a region matrix of point and cross-section.

### Applying a Multivariable Linear Regression Model of each Data Region to Build a Model and Calculate the Radial Distance from the Surface of the Human Body to the Inner Surface of the Shirts.

In this study, the input variables are half of the ease bust, cross-section, and point. The output variable is the radial distance from the surface of the human body to the inner surface of the shirts at points. To determine the model type, use the *scatterplot* function in R to represent the correlation between the radial distance from the surface of the human body to the inner surface of the shirts and the positions of those points in each data region. Use the *pairs.panels* function to represent multivariate correlation one at a time between the ease bust, cross-section, point, and radial distance.

To determine the relationship between the ease bust, cross-section, point, and radial distance between human body surface—shirt, a multivariable linear regression model with matrix form was used as follows:

$$D = X\beta + \epsilon \quad (2)$$

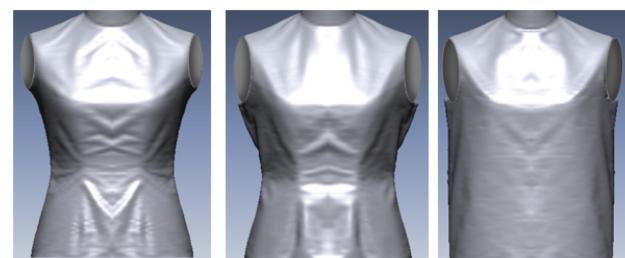
In which, D is the vector including the values of the radial distance of the body surface—shirt. X is the matrix of the ease bust, cross-section, point.  $\beta$  is the vector of regression coefficients,  $\epsilon$  is the vector of residual values.

Use Microsoft Excel software to determine the body-surface radial distances from the selected multivariable linear regression model. R software is used to determine the relationship between the ease bust, cross-section, point, and radial distance between human body surface—shirt to choose the optimal model, through the function *lm*, then use the *summary* command to select the optimal model to view search results, verify results to ensure high reliability and accuracy.

### 3 Results and Discussion

#### 3.1 Result of Redesigning the Shirt Model and Measuring the Radial Distance Between the Surface of the Human Body and the Inner Surface of each Shirt

The study scanned 17 female human bodies, each sample was scanned in 4 cases: wearing 3 shirts and no shirt. The coordinate origin selected in this study is the intersection of the slice passing through the protruding bottom, the center front vertical slice, and the vertical slice passing through two points under the armpit. The complete shirt model (Fig. 2). From the center of each body model cross-section, measure the distance to points on the body surface contour and the distance to points on the shirt's inner contour. Determine the radial distance between the surface of the human body and the inner surface of the shirt at each point of each model according to formula (1).



**Fig. 2.** Shirt models

### 3.2 Result of Radial Distance Division

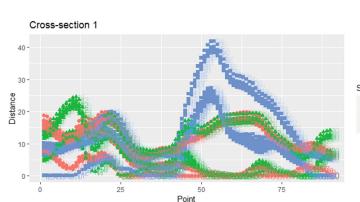
#### Cross-section Grouping Results

To show the change in radial distance values between points on each cross-section and between cross-sections of shirts, number the cross-sections from hem to neck, from 1 to 241. Use the ggplot function in R to represent the distances of points on each cross-section for 3 shirts A1, A2, A3 (Fig. 3).

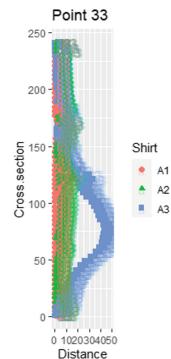
By observation, divide the cross-sections with the radial distance between the surface of the human body and the inner surface of the shirt at each point with a similar tendency to change and the cross-sections adjacent to each other. The results are divided into 16 groups of cross-sections as follows: From cross-sections. 1 to 11; 12 to 23; 24 to 47; 48 to 63; 64 to 87; 88 to 99; 100 to 111; 112 to 123; 124 to 135; 136 to 147; 148 to 163; 164 to 175; 176 to 195; 196 to 215; 216 to 227; 228 to 241.

#### Result of Grouping Points

To show the change in the distance value between the points of the shirts, number 90 points on each half-cross-section from center front to side to center back: 1 to 90. Use the ggplot function in R to represent the distance of each point on the cross-sections of the shirts A1, A2, A3 (Fig. 4).



**Fig. 3.** Distance of points on cross-section 1 of shirts A1, A2, A3



**Fig. 4.** Distance of point 33 on the cross-sections of shirts A1, A2, A3

By observation, divide the points with the radial distance between the surface of the human body and the inner surface of the shirt at each point that tend to vary similar and adjacent to each other. The results are divided into 12 groups of points as follows: From points 1 to 7; 8 to 15; 16 to 19; 20 to 24; 25 to 35; 36 to 39; 40 to 44; 45 to 55; 56 to 62; 63 to 69; 70 to 81; 82 to 90.

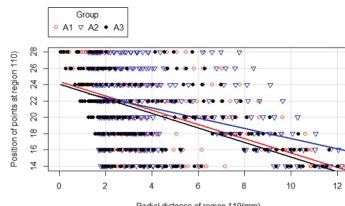
Combining cross-section groups and point groups form a matrix of 192 regions. Points and cross-sections region matrix follow as Table 2.

### 3.3 Model for Calculating the Radial Distance from the Surface of the Human Body to the Inner Surface of the Shirts

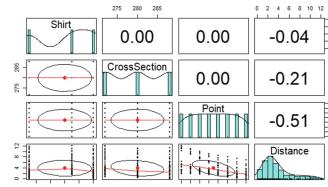
In this study, the value of half-bust ease for A1, A2, A3 is 10, 35, and 50 mm, respectively. Call the input variables for the research as S, C<sub>j</sub>, P<sub>i</sub>. In which, S is the shirt with the ease values of the bust half of 10, 35, 50 mm. C<sub>j</sub> is the cross-section position in millimeters from cross-section 1 to 241 corresponding from hem to neck, cross-section spaced 2 mm apart. P<sub>i</sub> is a point position of order from 1 to 90 lying on the cross-section contour and has a position in degrees, the angle from the center between two adjacent points is 2 degrees, point 1 is in the center front. i is the ordinal number of the point on the cross-section contour, i = [1, 90], i ∈ Z; j is the sequence number of the cross-section, j = [1, 241], j ∈ Z. The output variable D<sub>ij</sub> is the radial distance from the surface of the human body to the inner surface of the shirts at points. To represent the correlation between radial distances of the points and the positions of those points in each data region, use the scatterplot function in R to draw a scatter plot representing this correlation (Fig. 5).

#### Correlation Between Half Bust Ease, Cross-section, Point, Radial Distance Data Pairs

To represent at the same time the pairs correlations between bust ease, cross-section, point, and radial distances, a multivariate correlation plot was used. The *pairs.panels* function in R can respond to this goal (Fig. 6).



**Fig. 5.** Correlation graph between radial distance and position of points at region 110



**Fig. 6.** Multivariate correlation chart of region 110

In this research, the results on the multivariate correlation charts show that: there is no value of the correlation coefficient r greater than or equal to 0,95, meaning that the input variables are independent of each other. Therefore, the above quantities can be used as input variables of the multivariable regression model.

#### Multivariable Linear Regression Model of Each Data Region

To determine the relationship between the bust ease, cross-section, point, and radial distance between human body surface—shirt, a multivariable linear regression model with matrix form was used as Eq. (2). The results of the multivariable linear regression model of each data region are as shown in Table 2.

### 3.4 Determine the Radial Distance from the Surface of the Body to the Inner Surface of Shirts with a Customizable Bust Ease

#### Calculate the Radial Distance from the Multivariable Linear Regression Model of Each Data Region

From the multivariable linear regression model of each data region, as shown in Table 2, it is easy to calculate the radial distance ( $D_{ij}$ ) from the surface of the human body to the inner surface of the shirts with a customized half bust ease for each cross-section As Table 1: Put  $S$ ,  $C_{141}$ , and  $P_i$  into the point group models to calculate the distance  $D_{i,141}$  of the points on cross-section 141.

#### Check the Consistency of the Results of Calculating the Radial Distance $D_{ij}$ According to the Model with the Average Radial Distance $D_{Mean}$ of the Observed Data

Let  $D_{Mean}$  be the average radial distance from the measured samples. To compare the results of calculating the radial distance  $D_{ij}$  and  $D_{Mean}$  of each point, at each region use the t-test between  $D_{ij}$  and  $D_{Mean}$  for that region. The results follow as: In the region 109 to 120, the t-test results for paired variables show that the calculated distance from the  $D_{ij}$  model is 0,16 mm lower than that of  $D_{Mean}$ , the 95% confidence interval is from -0,02 mm to 0,34 mm, with  $p\text{-value} = 0,08 > 0,05$ . Thus, the distance difference calculated from the  $D_{ij}$  model with the mean of the  $D_{Mean}$  sample is not statistically significant, that is, each pair of these distances has no difference. Doing the same for other regions, the results also show that there is no significant difference between the radial distances calculated from the  $D_{ij}$  and the  $D_{Mean}$ .

#### Calculate the Distance from the Center to the Inner Surface of the Custom Shirt

As shown in Fig. 1, set the distance from the cross-section center to the body surface contour is  $OH = H_{ijMean}$  ( $H_{ijMean}$  is the distance from the cross-section center to the mean body surface contour observed). Set the distance from the cross-section center to the inner surface contour of shirt OG =  $G_{ij}$ ; radial distance from the body surface to the inner surface of the shirt HG =  $D_{ij}$ . So,

$$G_{ij} = H_{ijMean} + D_{ij} \quad (3)$$

From formula (3), it is easy to calculate the distance  $G_{ij}$  from the center of the cross-section to the inner surface contour of the shirt for each point on each cross-section.

#### Evaluation of the Model for Each Region

To evaluate the built model, calculate the length of the inner surface contour of the  $L_{i,141}$  shirts from the model of the regions and compare it with the bust circumference and bust circumference ease of the real shirt, the cross-section inner surface of this shirt through over the bust points. From Eq. (3), calculate the distance from the center of the cross-section to the inner surface contour of the shirt for each point on each  $G_{ij}$  cross-section. The length of the inner surface contour of a  $L_{ij}$  shirt is the set chords of 2-degree between 2 points on that cross-section contour, the coefficient of difference between the arc and the chord of the 2-degree angle can be calculated as  $k = 0,000051$ . Let the chords of the 2-degree angle between 2 points on the cross-section contour be  $l_{ij}$ . Then,

$$l_{ij} = \sqrt{G_i^2 + G_{i+1}^2 - 2G_i G_{i+1} \cos \theta} \text{ in which } \theta = 2^0 \quad (4)$$

From (3) and (4), we can calculate chords of 2-degrees on each cross-section of each shirt. The circumference of the half bust of the shirt is

$$L_{i,141} = \sum_1^{90} l_{i,141} \cdot k \quad (5)$$

The ease of half bust circumference of the shirt  $\Delta_{i,141}$  is:

$$\Delta_{i,141} = \frac{1}{2}(L_{i,141} - L_{i,141\text{Mean}})$$

Compare the ease  $\Delta_{i,141}$  at the bust cross-section calculated from the model with the ease  $\Delta$  of the custom shirt and the real shirt. The results are as in Table 1.

Double-pair t-test of half bust ease calculated from model  $\Delta_{i,141}$  with customized shirt ease  $\Delta$ . The results show that the average difference between the customized ease  $\Delta$  and the calculated ease  $\Delta_{i,141}$  is 0,19 mm, 95% confidence interval is from -1,29 mm to 1,67 mm with  $p = 0.7738 > 0.05$ , it is not statistically significant, that is, the difference between  $\Delta$  and  $\Delta_{i,141}$  is trivial. Thus, the ease of circumference bust of the shirt calculated from the model has a value that is not significantly different from the ease of circumference bust of the shirt customized as the design idea. Therefore, calculating the radial distance between the surface of the human body and the inner surface of the shirt is generally feasible within the scope of the research.

**Table 1.** Ease of half bust circumference of shirts

Ease of half bust circumference of the custom shirts $\Delta$ (mm)	Half bust circumference is calculated from model $L_{i,141}$ (mm)	Half bust circumference of average human body $L_{i,141\text{Mean}}$ (mm)	Ease of half bust is calculated from the model $\Delta_{i,141}$ (mm)	Ease of half bust of real shirts (mm)
10	456,2716	442,2551	14,0165	10
15	458,7143	442,2551	16,4592	
20	461,9312	442,2551	19,6761	
25	465,8488	442,2551	23,5937	
30	470,3660	442,2551	28,1109	
35	475,3820	442,2551	33,1269	35
40	480,8120	442,2551	38,5569	
45	486,5890	442,2551	44,3339	
50	492,6604	442,2551	50,4053	50

## Estimating the Importance of Each Input Variable Ease Bust, Cross-section, and Point in the Model

The model created for each region as shown in Table 2 depends on 3 input variables, which are the ease circumference bust of the shirt, the position of the cross-section, and the points on the cross-section contour. The model has shown the impact of all three variables on the variation of radial distance through the coefficient of determination  $R^2$  but has not specified the level of impact of each variable on the radial distance. Calculating the radial distance when designing a 3-dimensional shirt will be much simpler and more efficient when determining the impact of each input variable.

To examine the impact of each of these input variables on the variation in the radial distance, the LMG method can be used to determine the coefficient of determination  $R^2$  of the model for each input variable. Use the *calc.relimp* function and the *boot.relimp* function in R to assign the model's coefficient of determination  $R^2$  for each input variable. The results are as in Table 2. The results show that, in what region of the shirt where the coefficient of determination  $R^2$  of the ease bust S is high, that is, the ease plays an important role to the variation of the radial distance. Therefore, it should influence the change and allocate ease for the shirt when it is necessary to change the width of the shirt. The region from point 0 to 15 has a high  $R^2$  of the point or cross-section, which means that the position of the point or cross-section in this region is the main factor affecting the variation of the radial distance.

**Table 2.** Region matrix of radial distances from the surface of the human body to the inner surface of the shirts and the distribution of the coefficients determined in the model

Region	Cross-section group	Point group	The multivariable linear regression model						
			Model	$R^2$	Distribution of $R^2$			p-value	
					S	C <sub>j</sub>	P <sub>i</sub>		
109	136 to 147	1 to 7	Dij = -11.79 - 0.02S + 0.08Cj - 0.33Pi	0,054	0,0020	0,0075	<b>0,0445</b>	0,01505	
110	136 to 147	8 to 15	Dij = 37,87 - 0,01S - 0,1Cj - 0,33Pi	0,3037	0,0018	0,0454	<b>0,2565</b>	< 2,2e-16	
111	136 to 147	16 to 19	Dij = 43,26 + 0,001S - 0,14Cj - 0,06Pi	0,4675	0,0003	<b>0,4576</b>	0,0096	1,069e-14	
112	136 to 147	20 to 24	Dij = 27,7 + 0,01S - 0,09Cj - 0,02Pi	0,4429	0,0658	<b>0,3727</b>	0,0044	< 2,2e-16	
113	136 to 147	25 to 35	Dij = 9,53 + 0,12S - 0,09Cj + 0,25Pi	0,6103	<b>0,3784</b>	0,0267	0,2052	< 2,2e-16	
114	136 to 147	36 to 39	Dij = 31,43 + 0,25S - 0,2Cj + 0,3Pi	0,7855	<b>0,7214</b>	0,05	0,0144	< 2,2e-16	
115	136 to 147	40 to 44	Dij = -48,7 + 0,42S - 0,08Cj + 0,88Pi	0,914	<b>0,8279</b>	0,0037	0,0824	< 2,2e-16	
116	136 to 147	45 to 55	Dij = 27,99 + 0,59S - 0,07Cj - 0,05Pi	0,7817	<b>0,7798</b>	0,0013	0,0007	< 2,2e-16	
117	136 to 147	56 to 62	Dij = 91,8 + 0,35S - 0,07Cj - 0,59Pi	0,5229	<b>0,4636</b>	0,0022	0,0571	< 2,2e-16	
118	136 to 147	63 to 69	Dij = 41,7 + 0,126S - 0,09Cj - 0,1Pi	0,2132	<b>0,1961</b>	0,012	0,0054	8,295e-10	
119	136 to 147	70 to 81	Dij = 52,525 + 0,14S - 0,08Cj - 0,19Pi	0,6395	<b>0,4974</b>	0,019	0,123	< 2,2e-16	
120	136 to 147	82 to 91	Dij = -17,26 + 0,14S - 0,07Cj + 0,23Pi	0,289	<b>0,2306</b>	0,0074	0,051	< 2,2e-16	

## 4 Conclusions

This method allows a multivariable linear regression model of 192 radial distance regions. From this model, the research has shown a method to calculate the radial distance from the surface of the human body to the inner surface of the shirt with custom width. From there, calculate the ease of circumference bust and give results equivalent to the ease of circumference bust on the real shirt. The research also showed that the variation in the radial distance at the bust side and back is mainly affected by the ease of circumference bust; the center front, this radial distance is mainly influenced by cross-section position and point position.

This method also allows determining the effects of input variables that vary by region, so that when designing a 3-dimensional shirt on a computer, it is possible to adjust these variables to change the width of the shirt quickly, and effectively. Other regions and other factors that can influence the 3-dimensional shirt design will be presented in future research.

As can be seen, this method calculated, modeled, and statistically described the radial distances from the surface of the human body to the inner surface of the shirt for points on the cross-sections. Other studies used the method of calculating the normal distance at the cross-section passing through the anthropometric points. Therefore, the method in this paper has more complete research than previous studies on the space between the surface of the human body and the inner surface of the shirt.

## References

1. Thomassey, S., Bruniaux, P.: A template of ease allowance for garments based on a 3D reverse methodology—Univ.Lille Nord de France, F-59000 Lille, France (2013)
2. Liu, Z., He, Q., Zou, F., Ding, Y., Xu, B.,: Apparel ease distribution analysis using three-dimensional motion capture, *Text. Res. J.* 0(00) 1–13, <https://doi.org/10.1177/0040517519832842> (2019)
3. Xu, J., Zhang, W.: The vacant distance ease relation between body and garment. In: Second International Conference on Information and Computing Science (2009)
4. Xu, J., Zhang, W., Xiao, P.: A study on impact factors of the distance eases between body and garment, 978-1-4244-3291-2/08/\$25.00 ©2008 IEEE (2008)
5. Zhang, F., Little, T.J.: Dynamic ease evaluation for 3D garment design. *J. Fashion Mark. and Mgmt.*, Vol. 22 No. 2, 2018, pp. 209–222 © Emerald Publishing Limited 1361–2026 (2018)
6. Lagè, A., Ancutiené, K., Pukienė, R., Lapkovska, E., Dāboliņa, I.: Comparative study of real and virtual garments appearance and distance ease, ISSN 1392–1320 MATERIALS SCIENCE Vol. 26, No. 2. 2020 (2020)
7. Zhu, Y., Peng, Y., Boodaghian Asl, A.: As, dual adaptive adjustment for customized garment pattern, Hindawi, Scientific Programming, Volume 2019, Article ID 8069373, 12 pages, <https://doi.org/10.1155/2019/8069373> (2019)
8. Lee, Y., Ma, J., Choi, S.: Automatic pose-independent 3D garment fitting, *Comput. Graphics*, **37**(7), November 2013, 911–922 (2013)
9. TCVN 5782:2009: Standard sizing systems for clothes. (2009)
10. Thuy, D.T., Ngoc, N.T.T., Trung, N.C., Le, N.T.: Mannequin modeling from 3D scanning data, intelligent systems and networks, scopus, ISBN 978–981–16–2093–5 ISBN 978–981–16–2094–2 (2021)

11. ISO 8559–1:2017(E): Size designation of clothes—Part 1: Anthropometric definition for body measurement (2017)
12. Dalgaard: Introductory statistics with R, Springer (2004)



# Takagi-Sugeno Fuzzy Control for Stabilizing Nonlinear Inverted Pendulum

Duc-Binh Pham, Duy-Thi Pham, Quy-Thinh Dao, and Van-Anh Nguyen<sup>(✉)</sup>

School of Electrical and Electronic Engineering, Hanoi University of Science  
and Technology, Hanoi, Vietnam

{binh.pd181343, thi.pd181765}@sis.hust.edu.vn  
{thinh.daoquy, anh.nguyenthivan1}@hust.edu.vn

**Abstract.** Fuzzy is an intelligent control technique that is suitable for uncertainty and nonlinear systems. In this work, an advanced fuzzy logic control system is designed for an inverted pendulum, an unstable nonlinear system. First, the dynamic characteristics of the system are expressed through Takagi-Sugeno fuzzy model. Then, a parallel distributed compensation (PDC) controller is developed based on the definition of fuzzy sets. The purpose of this paper is to keep the stability of the pendulum angle. Besides, the linear matrix inequalities (LMI) is used for solving stability problem. Lastly, the efficiency and advantages of the proposed fuzzy controller are verified by simulation results.

**Keywords:** Linear matrix inequalities · Takagi-Sugeno · Inverted pendulum · Parallel distributed compensation

## 1 Introduction

An inverted pendulum has attracted great attention from researchers in the field of control theory and application. The main aim of the control system is to keep the pendulum at the desired vertical position. This aim can be obtained by manipulating the movement of the cart. Many applications work on the principle of inverted pendulum such as modeling and simulating human balance system [1], a self-balancing electrical vehicle named Segway [2].

In the literature, there are many research related to the inverted pendulum system such as Proportional Integral Derivative (PID) controller [3], Linear Quadratic Regulators (LQR) [4], model predictive controller (MPC) [3], Artificial Neural Networks [5], fuzzy logic controller [6–12]. Fuzzy control plays an important part in modern control theory and applications. While conventional

---

D.-B. Pham, D.-T. Pham, Q.-T. Dao and V.-A. Nguyen—Contributed equally to this work.

control methods such as PID and LQR often show their limitation when encountering high complexity and large nonlinear systems, fuzzy logic controllers are superior in solving those problems. A fuzzy logic system uses linguistic variables such as “big”, “medium”, “small” to express a partially “true” and “false”. The membership functions of these linguistic variables have a range from zero to one in which zero and one denote absolute false and true, respectively. Linguistic variables and their membership functions are used for designing rules and determining the controller’s output. These variables and fuzzy output are mainly used in the Mamdani system [13]. Besides, the output of the Takagi-Sugeno fuzzy model can be computed by using the weighted average method. Both Fuzzy Inference Systems above have their advantages and constraints. For example, Takagi-Sugeno FIS is more flexible in the system design than Mamdani FIS.

In this paper, the Lyapunov method is used to ensure the stability of the proposed system and the linear matrix inequalities (LMI) [14] is utilized to find out the state feedback gains. Then the controller can be synthesized by feeding these gains to the parallel distributed compensation (PDC) [15]. A comparison with a conventional PID controller shows that the proposed fuzzy controller can achieve better performance in both transient and steady states.

To the end, the paper is divided into 5 sections. Section 1 shows the brief literature review about control strategies for the proposed system. The inverted pendulum system modeling is given in Sect. 2. The control design for Takagi-Sugeno fuzzy system is illustrated in Sect. 3. Section 4 shows the Matlab-Simulink model and simulation results. Section 5 provides a conclusion of this paper.

## 2 System Modeling

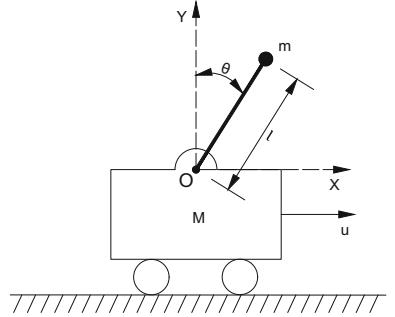
Figure 1 shows a simple inverted pendulum model used in this research. In which  $m$  (kg) and  $M$  (kg) are the mass of a pendulum and a cart, respectively.  $l$  represents the length of a connecting rod (m),  $\theta$  is the rotational angle of the pendulum from the y-axis (rad),  $u$  is the force in the x-axis that impacts the cart,  $g$  is gravitational acceleration vector. Let the coordinates of a pendulum be  $(\bar{x}, \bar{y})$ . It is determined as follows:

$$\begin{cases} \bar{x} = x_M + l \sin \theta \\ \bar{y} = l \cos \theta \end{cases} \quad (1)$$

where  $x_M$  is the travel distance of the cart.

The connecting rod has a very small mass compared with the cart so the system has the following total energy:

$$L = \frac{1}{2}(M+m)\dot{x}_M^2 + \frac{1}{2}ml^2\dot{\theta}^2 + ml\dot{\theta}\dot{x}_M \cos \theta - mg l \cos \theta. \quad (2)$$



**Fig. 1.** Modeling of an inverted pendulum.

The kinematic equation of the inverted pendulum is found through the Euler - Lagrange equation:

$$\frac{d}{dt}\left(\frac{\delta L}{\delta \dot{q}_m}\right) - \frac{\delta L}{\delta q_m} = Q_m. \quad (3)$$

From Eq. (3), the kinematic equation of the system can be written as:

$$\begin{cases} (M+m)\ddot{x}_M + ml\ddot{\theta} \cos\theta - ml\dot{\theta}^2 \sin\theta = u \\ l\ddot{\theta} + \ddot{x}_M \cos\theta - g \sin\theta = 0 \end{cases} \quad (4)$$

From a set of Eqs. (4) we can derive the equation describing the angular dynamics of the pendulum as follows:

$$\ddot{\theta} = \frac{(M+m)g \sin\theta - ml\dot{\theta}^2 \sin\theta \cos\theta - u \cos\theta}{l[m(1 - \cos^2\theta) + M]}. \quad (5)$$

We define the state variables  $\mathbf{x}$  as  $\mathbf{x} = [\theta \ \dot{\theta}]^T$  and the Eq. (5) can be expressed as:

$$\dot{\mathbf{x}} = A\mathbf{x} + B\mathbf{u} \quad (6)$$

where

$$A = \begin{bmatrix} 0 & 1 \\ \frac{(M+m)g \sin\theta}{l[m(1 - \cos^2\theta) + M]} & \frac{-ml\dot{\theta}^2 \sin\theta \cos\theta}{l[m(1 - \cos^2\theta) + M]} \end{bmatrix},$$

$$B = \begin{bmatrix} 0 \\ \frac{-u \cos\theta}{l[m(1 - \cos^2\theta) + M]} \end{bmatrix}.$$

### 3 Designing Fuzzy Controller

#### 3.1 Takagi-Sugeno Fuzzy Model

In order to develop the controller of the system, the inverted pendulum system is expressed in the form of a T-S fuzzy model. The fundamental of this model is to describe the given nonlinear system through several linear system models. To construct fuzzy models, there are two commonly used methods: the first one is to use input and output data to identify and the other method is derivation from the given system equations. The second one is chosen to obtain the inverted pendulum model in which the  $m^{th}$  rules for the continuous fuzzy system (CFS) are:

$$\begin{cases} \dot{\mathbf{x}}(t) = A_m(z)\mathbf{x}(t) + B_m(z)\mathbf{u}(t) \\ \mathbf{y}(t) = C_m\mathbf{x}(t) \end{cases} \quad m = 1, 2, \dots, 2^p \quad (7)$$

where  $z_m(t) (i = 1, 2, \dots, p)$  are premise variables,  $p$  and  $2^p$  are the number of premise variables and model rules, respectively.

Then:

$$\dot{\mathbf{x}}(t) = \sum_{m=1}^{2^p} \omega_m(z(t)) \{A_m(z)\mathbf{x}(t) + B_m(z)\mathbf{u}(t)\} \quad (8)$$

where

$$\omega_k(z) = \prod_{j=1}^p h_j^{k_j}(z_j) \quad (9)$$

and  $\omega_k$  denote the membership functions of corresponding premise variables  $z$ ,  $1 \leq k \leq 2^p$ ,  $k_1, \dots, k_p \in \{0, 1\}$ .

The parallel distributed compensation fuzzy controller is given:

$$\mathbf{u}(t) = -F_m\mathbf{x}(t) \quad m = 1, 2, \dots, 2^p. \quad (10)$$

Equation 10 can be recasted in the general form as follows:

$$\mathbf{u}(t) = - \sum_{m=1}^{2^p} \omega_m(z(t)) F_m \mathbf{x}(t) \quad (11)$$

where the feedback gain for  $m^{th}$  rule is  $F_m$ .

Substituting (11) into (8), the closed-loop equation is obtained:

$$\dot{\mathbf{x}}(t) = \sum_{m=1}^{2^p} \sum_{n=1}^{2^p} \omega_m(z(t)) \omega_n(z(t)) [A_m(z) - B_m(z)F_n] \mathbf{x}(t) \quad (12)$$

The following Theorem derives the conditions to ensure that the system is stable.

**Theorem 1.** *The closed-loop system (12) is asymptotically stable via PDC controller (11) if the following LMI constraints are satisfied with a common positive definite matrix  $X$  and matrices  $G_n$ :*

$$\begin{cases} \varepsilon_{mm} < 0, & \forall m \in \{1, 2, \dots, 2^p\} \\ \varepsilon_{mn} + \varepsilon_{nm} < 0, & \forall m, n \in \{1, 2, \dots, 2^p\}, m < n \end{cases} \quad (13)$$

where  $\varepsilon_{mn} = A_m X - B_m G_n + X A_m^\top - G_n^\top B_m^\top$ . Consequently, the control gains of the PDC controller can be inferred as follows:

$$F_n = G_n X^{-1} \quad (14)$$

Proof:

Consider a candidate of the Lyapunov function with a positive definite matrix  $P$ :

$$V(\mathbf{x}) = \mathbf{x}^\top(t) P \mathbf{x}(t). \quad (15)$$

The derivation of the above function can be inferred as follows:

$$\dot{V}(\mathbf{x}) = \dot{\mathbf{x}}^\top(t) P \mathbf{x}(t) + \mathbf{x}^\top(t) \left( \dot{P} \mathbf{x}(t) + P \dot{\mathbf{x}}(t) \right). \quad (16)$$

From Eq. (16) and (12), we have:

$$\begin{aligned} \dot{V}(\mathbf{x}) &= \sum_{m=1}^{2^p} \sum_{n=1}^{2^p} \omega_m(z(t)) \omega_n(z(t)) \mathbf{x}^\top(t) \left[ (A_m - B_m F_n)^\top P + P(A_m - B_m F_n) \right] \mathbf{x}(t) \\ &= \sum_{m=1}^{2^p} \omega_m^2(z(t)) \mathbf{x}^\top(t) \left[ (A_m - B_m F_m)^\top P + P(A_m - B_m F_m) \right] \mathbf{x}(t) \\ &\quad + 2 \sum_{m=1}^{2^p} \sum_{m < n}^{2^p} \omega_m(z(t)) \omega_n(z(t)) \mathbf{x}^\top(t) \\ &\quad \times \left[ \left( \frac{(A_m - B_m F_n) + (A_n - B_n F_m)}{2} \right)^\top P + P \left( \frac{(A_m - B_m F_n) + (A_n - B_n F_m)}{2} \right) \right] \mathbf{x}(t). \end{aligned} \quad (17)$$

Therefore  $\dot{V}(\mathbf{x}) < 0$ ,  $\mathbf{x} \neq 0$  is equivalent to

$$\begin{cases} A_m^\top P + P A_m - F_m^\top B_m^\top P - P B_m F_m < 0 \\ A_m^\top P + P A_m - F_n^\top B_m^\top P - P B_m F_n + A_n^\top P + P A_n - F_m^\top B_n^\top P - P B_n F_m < 0 \end{cases} \quad (18)$$

Let us take  $X = P^{-1}$ , and multiply  $X$  on both left and right of (18)

$$\begin{cases} X A_m^\top + A_m X - X F_m^\top B_m^\top - B_m F_m X < 0 \\ X A_m^\top + A_m X - X F_n^\top B_m^\top - B_m F_n X + X A_n^\top + A_n X - X F_m^\top B_n^\top - B_n F_m X < 0 \end{cases} \quad (19)$$

For  $X > 0$  and let  $G_n = F_n X$ , we have  $F_n = G_n X^{-1}$ . Substituting into the above inequality shows:

$$\begin{cases} X A_m^\top + A_m X - G_m^\top B_m^\top - B_m G_m < 0 \\ X A_m^\top + A_m X - G_n^\top B_m^\top - B_m G_n + X A_n^\top + A_n X - X F_m^\top B_n^\top - B_n F_m X < 0 \end{cases} \quad (20)$$

with  $\gamma_{mn} = A_m X - B_m G_n + X A_m^\top - G_n^\top B_m^\top$ , we obtain the LMI conditions.

### 3.2 T-S Controller Design

We first set the premise variables as follows:

$$z_1 = \frac{1}{L[m(1 - \cos^2\theta) + M]}, \quad z_2 = \frac{\sin\theta}{\theta}, \quad z_3 = \cos\theta, \quad z_4 = \dot{\theta}\sin\theta. \quad (21)$$

Substituting the premise variables  $z$  into the matrices  $A, B$  above we get:

$$A = \begin{bmatrix} 0 & 1 \\ (M+m)gz_1z_2 - mlz_1z_3z_4 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ -z_1z_3 \end{bmatrix}.$$

Let us define:

$$h_{i0} = \frac{z_{imax} - z_i}{z_{imax} - z_{imin}}, \quad h_{i1} = 1 - h_{i0} \quad (i = 1, 2, 3, 4) \quad (22)$$

Then, the membership functions of the T-S fuzzy model are:

$$\begin{aligned} \omega_1 &= h_{10} * h_{20} * h_{30} * h_{40}, & \omega_9 &= h_{10} * h_{20} * h_{30} * h_{41}, \\ \omega_2 &= h_{11} * h_{20} * h_{30} * h_{40}, & \omega_{10} &= h_{11} * h_{20} * h_{30} * h_{41}, \\ \omega_3 &= h_{10} * h_{21} * h_{30} * h_{40}, & \omega_{11} &= h_{10} * h_{21} * h_{30} * h_{41}, \\ \omega_4 &= h_{11} * h_{21} * h_{30} * h_{40}, & \omega_{12} &= h_{11} * w_{21} * h_{30} * h_{41}, \\ \omega_5 &= h_{10} * h_{20} * h_{31} * h_{40}, & \omega_{13} &= h_{10} * h_{20} * h_{31} * h_{41}, \\ \omega_6 &= h_{11} * h_{20} * h_{31} * h_{40}, & \omega_{14} &= h_{11} * h_{20} * h_{31} * h_{41}, \\ \omega_7 &= h_{10} * h_{21} * h_{31} * h_{40}, & \omega_{15} &= h_{10} * h_{21} * h_{31} * h_{41}, \\ \omega_8 &= h_{11} * h_{21} * h_{31} * h_{40}, & \omega_{16} &= h_{11} * h_{21} * h_{31} * h_{41}. \end{aligned} \quad (23)$$

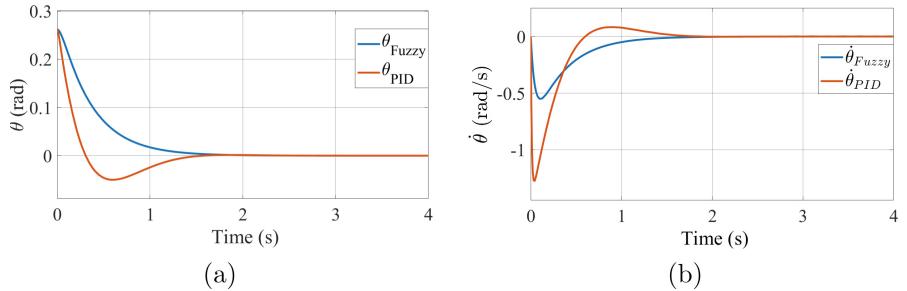
## 4 Result

The parameters of the pendulum are selected as:  $m = 0.2 \text{ kg}$ ,  $g = 9.8 \text{ m/s}^2$ ,  $M = 1 \text{ kg}$ ,  $l = 1 \text{ m}$ ,  $\theta_0 \in [-\pi/2; \pi/2]$ . With the given system parameters,  $z_{imax}$  and  $z_{imin}$  are determined, thereby calculating the values of  $h_{i0}$  and  $h_{i1}$  (22) and the membership functions (23). In addition, the feedback gains  $F_m$  are also acquired through the calculation of the LMI using MATLAB's Robust Control Toolbox. After finding out the  $\omega_m$  and  $F_m$ , the control signal  $\mathbf{u}$  is obtained (11). Below are computed simulation results with some initial angles  $\theta_0$  within the limits given above.

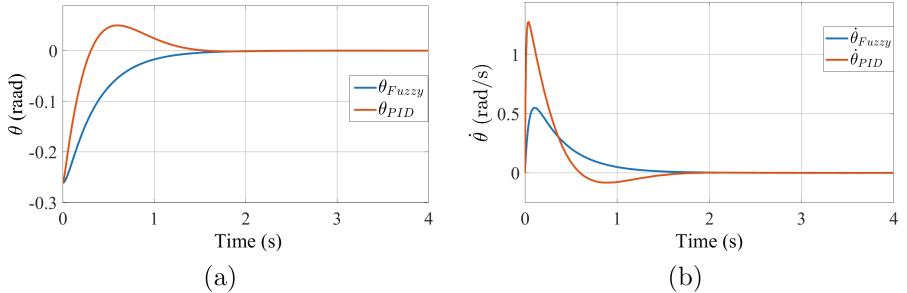
Positive definite matrix  $X = \begin{bmatrix} 8 \times 10^{-14} & -2.4 \times 10^{-13} \\ -2.4 \times 10^{-13} & 7.2 \times 10^{-13} \end{bmatrix}$ .

Feedback gain  $F_m$ :

$$F_1 = [-77.6411 \ -31.5568], \dots, F_{16} = [-77.6411 \ -31.5568].$$

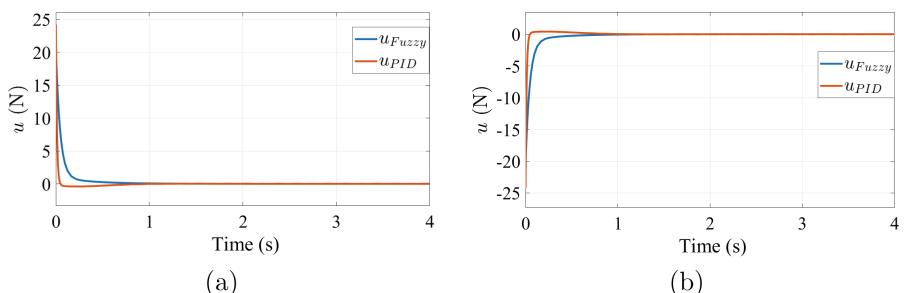


**Fig. 2.** Angle (a) and angular velocities (b) of pendulum with  $\theta_0 = 15^\circ$ .



**Fig. 3.** Angle (a) and angular velocities (b) of pendulum with  $\theta_0 = -15^\circ$ .

In these figures, the red and blue lines are the angle and its velocity in two cases: the first case uses fuzzy controller and the other case uses PID controller. Figures 2, 3 show that the time the pendulum returns to the equilibrium position of both control algorithms takes more than 1.5 s, the steady-state errors are almost zero. However, the fuzzy control algorithm has shown its advantage when it does not have overshoot. The figures about angular velocity show that when controlled with a T-S fuzzy controller, the pendulum oscillates more gently than with a PID controller.



**Fig. 4.** Force  $u$ : (a)  $\theta_0 = 15^\circ$ . (b)  $\theta_0 = -15^\circ$ .

Figure 4 shows that the initial forces of both controllers are at the allowable level and both rapidly approach zero when the pendulum returns to equilibrium. The simulation shows that whether the  $\theta_0$  is negative or positive (the pendulum is to the left or right of the Y-axis), the pendulum also quickly reaches equilibrium.

## 5 Conclusion

In this paper, the problem of maintaining the balance of the cart-mounted inverted pendulum was solved by a Takagi-Sugeno fuzzy controller. The state feedback gains are fed to the parallel distributed compensation to obtain the controller. By using Lyapunov method with linear matrix inequalities, the stability of the closed-loop system has been verified. Through the simulation results, we can see that the proposed fuzzy controller has better performance in both steady states and transient.

**Acknowledgement.** This research is funded by Hanoi University of Science and Technology (HUST) under project number T2021-TT-002.

## References

1. Kot, A., Nawrocka, A.: Modeling of human balance as an inverted pendulum. In: Proceedings of the: 15th International Carpathian Control Conference (ICCC), pp. 254–257 (2014)
2. Draz, M.U., Ali, M.S., Majeed, M., Ejaz, U., Izhar, U.: Segway electric vehicle. In: 2012 International Conference of Robotics and Artificial Intelligence, pp. 34–39 (2012)
3. Bakaráč, P., Klaučo, M., Fikar, M.: Comparison of inverted pendulum stabilization with PID, LQ, and MPC control. In: 2018 Cybernetics and Information (K and I), pp. 1–6 (2018)
4. Lee, H.W.: Performance the balance of circular inverted pendulum by using LQR controlled theory. In: 2017 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW), pp. 415–416 (2017)
5. Korkmaz, D., Bal, C., Gökbüllüt, M.: Modeling of inverted pendulum on a cart by using Artificial Neural Networks. In: 2015 23rd Signal Processing and Communications Applications Conference (SIU), pp. 2642–2645 (2015)
6. Wang, H., Bai, Y.: Application of fuzzy control in the inverted pendulum. In: Proceedings of 2013 2nd International Conference on Measurement, Information and Control, pp. 1354–1357 (2013)
7. Li, H., Liu, Q.: Single inverted pendulum system design based on genetic algorithm and reduced fuzzy control. In: 2010 8th World Congress on Intelligence Control and Automation, pp. 4633–4638 (2010)
8. Wu, J., Jin, Z.: Research on the adaptive fuzzy sliding mode control method of inverted pendulum system. In: Proceedings of the 6th International Forum on Strategic Technology 2011, pp. 1261–1265 (2011)
9. Liu, H., Duan, F., Gao, Y., Yu, H., Xu, J.: Study on fuzzy control of inverted pendulum system in the Simulink environment. In: 2007 International Conference on Mechatronics and Automation, pp. 937–942 (2007)

10. Marzi, H.: Fuzzy control of an inverted pendulum using AC induction motor actuator. In: 2006 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, pp. 109–114 (2006)
11. Yu, L.H., Jian, F.: An inverted pendulum fuzzy controller design and simulation. In: 2014 International Symposium on Computer, Consumer and Control, pp. 557–559 (2014)
12. Nguyen, A.-T., Dequidt, A., Nguyen, V.-A., Vermeiren, L., Dambrine, M.: Fuzzy descriptor tracking control with guaranteed  $L_\infty$  error-bound for robot manipulators. *Trans. Inst. Meas. Control* **43**(6), 1404–1415 (2020)
13. Mamdani, E.H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Man-Mach. Stud.* **7**(1), 1–13 (1975)
14. Yaz, E.E.: Linear matrix inequalities in system and control theory. In: Proceedings of the IEEE, pp. 2473–2474 (1998)
15. Wang, H.O., Tanaka, K., Griffin, M.: Parallel distributed compensation of nonlinear systems by Takagi-Sugeno fuzzy model. In: Proceedings of 1995 IEEE International Conference on Fuzzy System, pp. 531–538 (1995)



# A Comparative Study of Model Selection for High Dimensional Data

Thanh Trinh<sup>1,2(✉)</sup>

<sup>1</sup> Faculty of Computer Science, Phenikaa University, Yen Nghia, Ha Dong,  
Hanoi 12116, Vietnam

[thanh.trinh@phenikaa-uni.edu.vn](mailto:thanh.trinh@phenikaa-uni.edu.vn)

<sup>2</sup> Phenikaa Research and Technology Institute (PRATI), A&A Green Phoenix Group  
JSC, No. 167 Hoang Ngan, Trung Hoa, Cau Giay, Hanoi 11313, Vietnam

**Abstract.** Classification and high dimensional data are the challenging problems in machine learning. In this paper, we carry out an empirical study of model selection for high dimensional data. We select three single classifiers, i.e., decision tree, support vector machine, and naive bayes, and two ensemble methods, i.e., bagging and random forest, to compare the performance of these five methods through real high dimensional data. Six real gene datasets with high correlation are used in this empirical study. Experimental results have illustrated that ensemble methods are better than single classifiers in terms of accuracy metric. In which, decision tree model is the weakest one, and bagging is the best model. The performance of random forest is much more efficient in terms of computing time metrics compared with bagging. From this empirical study, we can conclude that ensemble methods are an effective solution for high dimensional data.

**Keywords:** High dimensional data · Supervised learning · Classification

## 1 Introduction

Classification has been considered one of the main tasks and studied in many works [1]. Many techniques are proposed for classification. Recently, a volume of data is being generated in different fields, such as social networks. Moreover, classifiers need data as detailed as possible, such as gene data or bioinformatic data. These kinds of data are treated as high dimensional data with thousands of features compared with a few dozen of samples, in other words,  $p$  (features)  $\gg n$  (samples). Many machine learning methods are developed and combined in order to cope with the problem of high dimensional data [2,9]. Decision tree model is a famous example of classifiers. Different tree models [13] are developed to classify high dimensional data. Random forest [8,11,24,28] and bagging models [12] are improved in predictions of high dimensional data. In addition, support vector

machine (SVM) [19] models are developed to deal with the data. For example, a combination of bagging and support vector machine is developed by Su et al. [22]. Naive classifier [27] is integrated with Dirichlet to solve the high dimensional imbalanced data. Hence, a common solution of these methods is to select a small group of informative features that are trained to generate better results compared with the whole features.

We present a comparative study of five classifiers through high dimensional data in this paper. Five classifiers are decision tree, SVM, naive bayes, bagging, and random forest. Six real gene datasets are gathered in our experiments. Particularly, features of each gene dataset are highly correlated. We use the metrics of accuracy and computing time to evaluate the performance of five comparative classifiers through the six datasets. The experimental results have shown that ensemble methods perform better than single classifiers. Bagging and random forest take the first place and the second one, respectively. In which bagging gives the best accuracy results; however, random forest is much more efficient in computing time metric.

## 2 Comparative Models

In this section, we present five famous models: decision tree, SVM, naive bayes, bagging and random forest.

### 2.1 Decision Tree

Various algorithms are proposed to construct decision tree models, such as ID3 [18], C5.0 [20], and CART [6]. ID3 and C5.0 use information gain to build trees. CART takes Gini measure to generate trees. However, the basic idea behind the process of these algorithms can be explained as following steps:

1. Choosing the best attribute to divide the remaining cases, and the attribute is selected as a decision node.
2. Repeating Step 1 recursively for each child.
3. The process stops when the stop criteria meet; for example, all the cases have the same target class.

### 2.2 Support Vector Machine

SVM is proposed by [7], and this method is very useful to deal with the classification problem. The main idea of this algorithm is to find the best hyperplane which divides data into its separating classes. Given a data set  $D = \{x_i, y_i\}_1^N$ , SVM method divides D into a set of high dimensional features by finding an optimal hyperplane. A set of support vectors is combined to creating this hyperplane. Then, SVM assigns each vector of features  $x_i$  to a specific class. The process of forming the hyperplane is expressed as follows:

Each support vector can be illustrated as the following equation:

$$w \cdot x + b = 0 \quad (1)$$

where  $w$  is treated as a weight vector,  $b$  is considered a bias. So, the hyperplane is expressed as:

$$f(x) = \text{sign}((w, \phi(x)) + b) \quad (2)$$

Equation 2 is be written as:

$$f(x) = \text{sign}\left(\sum_{i=1}^N a_i y_i k(x_i, x) + b\right) \quad (3)$$

where  $N$  is a number of samples,  $x_i$  is a vector of  $m$  features, and  $y_i$  is a class.  $k(x_i, x)$  is defined as the kernel function, for example, linear and radial basis. SVM can be used and improved with one of these functions.

### 2.3 Naive Bayes

Naive bayes algorithm is a wellknown classification method based on the bayes rules. This method assumes that all features ( $x$ ) have an equal effect on the outcome (given  $y$ ). In other words, the features are conditionally independent of each other. The method can be explained as follows:

Given  $X$  with  $n$  features and  $Y \in \{c_1, c_2..c_C\}$ , bayes rules can be written as:

$$P(X | Y) = \prod_{i=1}^n P(X_i | Y) \quad (4)$$

The probability of  $Y$  that  $Y$  takes the class  $c_j$  is expressed as:

$$P(Y = c_j | X) = \frac{P(Y = c_j) P(X | Y = c_j)}{P(X)} \quad (5)$$

$$P(Y = c_j | X) = \frac{P(Y = c_j) \prod_{i=1}^n P(X_i | Y = c_j)}{P(X)} \quad (6)$$

where  $P(Y = c_j)$  is prior,  $P(X)$  is evidence and  $\prod_{i=1}^n P(X_i | Y = c_j)$  is likelihood. Gaussian, multinomial and bernoulli naive bayes are often used to estimate the likelihood.

Hence, given a vector of feature  $X$ , we can identify the class of  $X$  by the following equation:

$$\hat{y} = \text{argmax}(P(Y = c_j) \prod_{i=1}^n P(X_i | Y = c_j)) \quad (7)$$

## 2.4 Bagging

Bagging (**Bootstrap aggregating**) that is proposed by Breiman [3] is an ensemble method. Bagging is described as the following steps:

1. Bootstrapping: Given a training data set  $D = \{(x_i, y_i)\}_{i=1}^N$  with  $N$  samples where  $x_i$  is a vector of  $m$  features and  $y_i \in Y = \{l_1, l_2, \dots, l_C \mid C \geq 2\}$  is a class label. Bagging creates  $K$  subsets by sampling the dataset  $D$  with replacement.
2. Parallel training: For each bootstrap, build a base classifier, for example, decision tree. All these classifiers are combined into a bagging model.
3. Aggregation: Finally, a new sample comes to bagging model; its class is identified by the majority vote of all classifiers in the bagging model.

## 2.5 Random Forests

Random Forest [5] is a non-parameter and ensemble method. The main idea of this method is to use bagging and to randomly select a subset of the whole features. Similar to Bagging method, the process of random forest model includes the three main steps. However, In Step 2 -Parallel training, Random forest only uses decision tree as base classifier. The tree model is built based on Gini measure. And,  $\sqrt{m}$  features (mtry) are randomly selected to compare Gini of them at each node of the tree.

## 3 Empirical Study

### 3.1 Data

Table 1 describes the six real high dimensional datasets that are used in experiments, and this table indicates the references of these datasets. This kind of data is about gene data. In more detail, features in gene data are often highly correlated to each other. In which the number of features varies from 4869 to 12600. The samples changes from 42 to 203, and the number of classes is from 2 to 9 in these datasets.

**Table 1.** Dataset statistics.

Dataset	Samples	Features	Classes	Reference
Brain	42	5597	5	[17]
Tumors	60	5726	9	[10]
Lung Cancer	203	12600	5	[10]
Prostate	102	6033	2	[21]
DLBCL	77	5469	2	[10]
Breast	95	4869	3	[26]

### 3.2 Experimental Settings

We set up five classifiers to carry out a comparative study about high dimensional data as follows.

- Decision tree (DT): CART method is adopted to construct decision trees in experiments.
- SVM: Radial basis kernel function (RBF) is selected to be used in Eq. 3.
- Naive bayes: Gaussian is chosen to estimate the likelihood in Eq. 6.
- Bagging: Decision tree classifier is used as the base classifier. We generate 1000 bootstraps for each real dataset by sampling with replacement.
- Random forest: We set  $K$  number of bootstraps be 1000. And,  $mtry$  is  $\sqrt{m}$  where  $m$  is the number of features in each dataset.

Each dataset was divided randomly into two parts: 70% for training and 30% for testing. Training part is trained in each comparative model, then testing part is taken to evaluate the performance of five comparative models. This process is run 100 times in order to generate 100 results of accuracy for comparative models.

Five models were implemented in Python and run on a machine (2.9 GHz dual-core CPU and 32 GB memory).

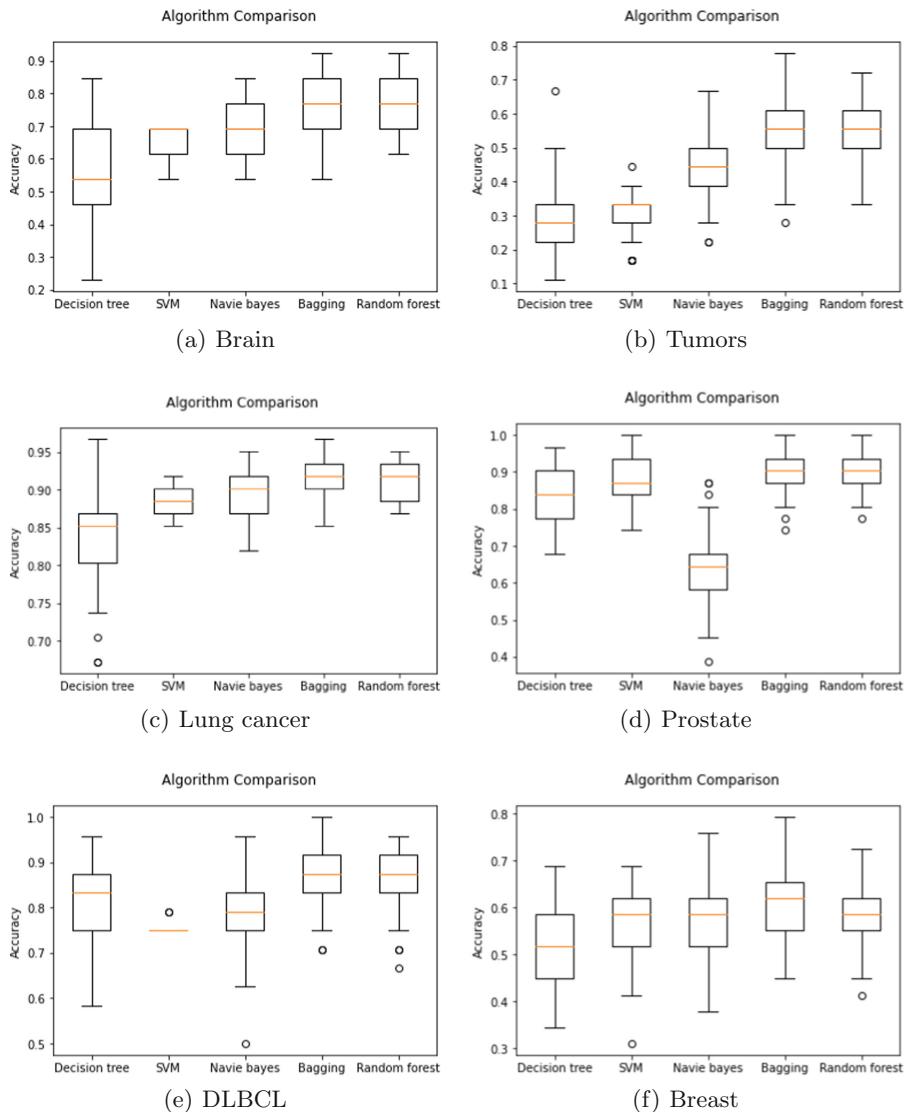
### 3.3 Experimental Results

**Accuracy.** Table 2 and Fig. 1 show the statistical results of five methods through the six datasets. Specifically, Table 2 describes the performance of average accuracy of 100 times for five methods. And, Fig. 1 illustrates the distribution of accuracy of 100 times for each dataset.

We can observe that bagging outperforms the other methods in all datasets, seen in Table 2. And random forest takes the second place. It is clear that single classifiers (decision tree, SVM, and naive bayes) do not perform better than ensemble methods (bagging and random forest). And, decision tree is the weakest model among the five comparative ones.

**Table 2.** Average accuracy of five comparative models.

Dataset	Decision tree	SVM	Naive Bayes	Bagging	Random forest
Brain	0.555	0.660	0.697	0.776	0.766
Tumors	0.283	0.304	0.443	0.554	0.547
Lung Cancer	0.838	0.881	0.896	0.918	0.911
Prostate	0.842	0.877	0.644	0.911	0.907
DLBCL	0.823	0.751	0.794	0.881	0.864
Breast	0.515	0.564	0.565	0.609	0.583



**Fig. 1.** Performance of the five comparative models through six real gene datasets

Figure 1 provides the different statistical metrics, i.e., min, max, median, quantiles, and outliers of the five methods. It is observed that the accuracy results of random forest are more stable than those of bagging. However, SVM yields the most stable results in the five classifiers. In contrast, the results of decision tree are the most unstable in those of the five models, seen in Fig. 1.

In experimental settings, decision tree is set as a base classifier of bagging. A set of weak classifiers forms a more robust model, for example, random forest and

bagging. The results of decision tree, bagging, and random forest are consistent within previous works [3–5]. SVM and Naive Bayes do not perform as well as the two ensemble methods. However, SVM still retains the strength in predictions of two datasets: Lung cancer and Prostate. And, Naive bayes yields good results for Lung cancer dataset.

**Computing Time.** In experiments, we only compared the computing time between ensemble methods, i.e., random forest and bagging. Table 3 demonstrates the differences in computing time between the two models through six datasets. It is clear that the average computing times of random forest in different datasets are much less than those of bagging. Obviously, at each node of decision trees of bagging method, all features need to be scanned; hence, it takes a lot of time to construct trees. However, Random forest only takes  $\sqrt{m}$  (number of features) to compute and choose the best split on each node.

**Table 3.** Average computing time (in seconds) of two comparative models.

Models	Brain	Tumors	Lung Cancer	Prostate	DLBCL	Breast
Random forest	1.01	1.28	3.26	1.28	1.06	1.28
Bagging	12.22	31.27	181.99	25.00	13.98	25.58

## 4 Related Work

Many classifiers are developed and combined with each other to address high dimensional data. Decision tree, SVM, and naive bayes models are famous examples of single classifiers. These models are developed by many works to deal with several kinds of data, such as social data, text data [16, 23, 25], and high dimensional data. Muralidharana et al. [14] presented a comparative study of naive classifier and Bayes net classifier to analyze the data of vibration. In another work [27], the problem of high dimensional imbalanced data was solved by a combination of multinomial naive bayes and Dirichlet priors. Lin et al. [13] combined the supervised hashing technique and decision tree model to deal with the problem of image data. SVM [7] is well-known for classification and regression tasks [19]. Nefedov et al. [15] presented a comparative study of SVM with different kernel functions, i.e., quadratic and linear criteria. Ensemble methods, for example, random forest, bagging, and boosting are developed in many work [8, 11, 22, 24, 28]. Pes [16] presented a study about random forest for high dimensional imbalanced data, in which the authors used gene, text, and image data in their experiments.

## 5 Conclusions

In this study, we have presented the evaluation of model selection for high dimensional data. Five well-known classifiers and six gene datasets are selected in this study. Experimental results have shown that ensemble methods perform better than single classifiers in the six datasets. Moreover, ensemble methods (such as bagging and random forests) outperform a base classifier (decision tree) inside these ensemble methods. However, SVM and Naive Bayes still retain the effectiveness through Prostate and Lung cancer datasets. That means there exists a subset of features that can perform better than the whole features. These things motivate us in future work about feature selection in high dimensional data.

## References

1. Aggarwal, C.C.: Data Mining. Springer International Publishing, Cham (2015). <https://doi.org/10.1007/978-3-319-14142-8>
2. Bellman, R.E.: Adaptive Control Processes. Princeton University Press, Princeton, December 1961
3. Breiman, L.: Bagging predictors. *Mach. Learn.* **24**(2), 123–140 (1996)
4. Breiman, L.: Bias, variance, and arcing classifiers (1996)
5. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001)
6. Breiman, L., Friedman, J. H., Olshen, R., Stone, C.: Classification and regression trees. Pacific Grove (1984)
7. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
8. Do, T.-N., Lenca, P., Lallich, S.: Classifying many-class high-dimensional finger-print datasets using random forest of oblique decision trees. *Vietnam J. Comput. Sci.* **2**(1), 3–12 (2014). <https://doi.org/10.1007/s40595-014-0024-7>
9. Donoho, D.L., et al.: High-dimensional data analysis: the curses and blessings of dimensionality. AMS Math Challenges Lecture, pp. 1–32 (2000)
10. GEMS. GEMS: Gene Expression Model Selector (2005)
11. Kursa, M.B.: Robustness of random forest-based gene selection methods. *BMC Bioinform.* **15**(1), 8 (2014)
12. Lango, M., Stefanowski, J.: The Usefulness of Roughly Balanced Bagging for Complex and High-Dimensional Imbalanced Data, pp. 93–107 (2016)
13. Lin, G., Shen, C., Shi, Q., Van Den Hengel, A., Suter, D.: Fast supervised hashing with decision trees for high-dimensional data. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (ii), pp. 1971–1978 (2014)
14. Muralidharan, V., Sugumaran, V.: A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis. *Appl. Soft Comput. J.* **12**(8), 2023–2029 (2012)
15. Nefedov, A., Ye, J., Kulikowski, C., Muchnik, I., Morgan, K.: Comparative analysis of support vector machines based on linear and quadratic optimization criteria. In: 8th International Conference on Machine Learning and Applications, ICMLA 2009, pp. 288–293 (2009)
16. Pes, B.: Learning from high-dimensional and class-imbalanced datasets using random forests. *Information (Switzerland)* **12**(8), 286 (2021)

17. Pomeroy, S.L., et al.: Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* **415**(6870), 436–442 (2002)
18. Quinlan, J.R.: Induction of decision trees. *Mach. Learn.* **1**(1), 81–106 (1986)
19. Rashid, A.M., Midi, H., Slwabi, W.D., Arasan, J.: An efficient estimation and classification methods for high dimensional data using robust iteratively reweighted SIMPLS algorithm based on nu-support vector regression. *IEEE Access* **9**, 45955–45967 (2021)
20. Salzberg, S.L.: C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers Inc, 1993. *Mach. Learn.* **16**(3), 235–240 (1994)
21. Singh, D., et al.: Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* **1**(2), 203–209 (2016)
22. Sun, Y., Platoš, J.: High-dimensional data classification model based on random projection and Bagging-support vector machine (2021)
23. Trinh, T., Wu, D., Huang, J.Z., Azhar, M.: Activeness and loyalty analysis in event-based social networks. *Entropy* **22**(1), 119 (2020)
24. Trinh, T., Wu, D., Salloum, S., Nguyen, T., Huang, J.Z.: A frequency-based gene selection method with random forests for gene data analysis. In: 2016 IEEE RIVF International Conference on Computing and Communication Technologies: Research, Innovation, and Vision for the Future, RIVF 2016 - Proceedings, pp. 193–198 (2016)
25. Trinh, T., Wu, D., Wang, R., Huang, J.Z.: An effective content-based event recommendation model. *Multimedia Tools and Applications* **80**(11), 16599–16618 (2020). <https://doi.org/10.1007/s11042-020-08884-9>
26. van 't Veer, L.J., et al.: Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**(6871), 530–536 (2002)
27. Wong, T.-T., Tsai, H.-C.: Multinomial naïve Bayesian classifier with generalized Dirichlet priors for high-dimensional imbalanced data. *Knowl.-Based Syst.* **228**, 107288 (2021)
28. Xu, B., Huang, J.Z., Williams, G., Wang, Q., Ye, Y.: Classifying very high-dimensional data with random forests built from small subspaces. *Int. J. Data Warehous. Min.* **8**(2), 44–63 (2012)



# Extract Forest Height from PolInSAR Image Based on Decomposition Scattering Component and Polarization State Optimization Technique

Nghia Pham Minh<sup>1</sup>(✉), Cuong Thieu Huu<sup>2</sup>, Ky Le Van<sup>3</sup>, and Dung Duong Quoc<sup>1</sup>

<sup>1</sup> Le Qui Don Technical University, Hanoi, Vietnam  
nghiapm2018@mta.edu.vn

<sup>2</sup> Telecommunication University, Nha Trang, Vietnam

<sup>3</sup> Factory X56, Technical Department of Vietnam Navi, Hai Phong, Vietnam

**Abstract.** PolInSAR technology is a promising technique for building biomass and forest height maps currently. However, there are some major drawbacks in estimating forest heights using PolInSAR images including the assumption that the interferometry coherence coefficient of the surface scattering component is ideal and that the interaction between surface scattering and volume scattering is ignored. This paper proposes a manner by the combination of decomposition scattering component and optimizing polarimetric state technique to overcome the above limitation. The proposed method, in addition to improving the accuracy of forest parameter estimation, also faithfully reflects the interaction of radar waves in the natural environment. The effectiveness of the suggested method was evaluated with the UAV-SAR data received from the AfisAR project of NASA/JPL001 and the reference LiDAR data. Results indicate that RMSE of suggested method approximate 3.2 m and  $R^2$  is 0.85.

**Keywords:** PolInSAR · UAV-SAR · Complex polarimetric interferometry coherence coefficient · Mean extinction · Ground decorrelation

## 1 Introduction

The inspection, monitoring and protection of forest resources are particularly cared in each country all over the world these days. There are many parameters for the management and monitoring of forest resources such as: canopy structure index (CST), volume scattering index (VSI), biomass index (BMI), attenuation index, etc. radar scattering in the forest environment (FDI), altitude, density, tree species, terrain slope.... In which, the biomass index plays a pivotal role in assessing and monitoring forest resources. Currently, there are many methods applied to determine forest biomass [2] and each method has also yielded certain results. However, determining biomass through forest height is a particularly typical and highly accurate solution. Therefore, forest height can be considered as one of the important parameters for inspecting, monitoring and managing

forest ecosystems. Forest height is not only used for biomass estimation, but also can be used to determine other forest's characteristics such as: age, coverage, timber volume, etc.

The strong development of science and technology has created many modern remote sensing technologies for the task of inspecting, monitoring and protecting forest resources. In which, the PolInSAR technique is one of the prominent and promising applications that are being strongly researched and developed for classifying, identifying and extraction forest parameters [1–4]. This technique has proved outstanding advantages in providing information about the height, structure, density and type of forest [4]. In addition, the PolInSAR system can also be mounted on unmanned aerial vehicles (UAV-SAR) to perform observation missions in a flexible and low-cost way. Up to now, there have been a number of forest height inversion models from PolInSAR images, which have been researched and developed with the aim of improving the accuracy of extracting forest elevation from PolInSAR images.

Particularly, S.R. Cloude and Papathanassious are pioneers in building an inverse algorithm for forest elevation retrieval from Pol-InSAR image [5, 6]. In 2003, the authors mentioned a 3-stage inversion method for extracting forest elevation [6] and this method had been widely used for many years since. In which, the forest height is determined through the method of minimizing the difference between the PolInSAR data and the predicted forest model. Therefore, the accuracy of the algorithm is greatly dependent on the building of the predictive model. To overcome these limitations, Tayebe et al. have proposed a way to optimize the complex polarimetric interferometry coherence coefficient for retrieving forest elevation PollInSAR data (Tayebe manner) [7] to improve the efficiency of the 3-stage manner. Tayebe technique has significantly improved the accuracy of estimated forest elevation compared to the 3-stage manner. However, Tayebe still uses the assumption that the studied forest area is relatively flat ( $\mu(\vec{\omega}) = 0$ ) and that the mean extinction is constant. These assumptions are the cause of large errors for the forest parameters evaluated by this way.

In order to improve the accuracy of retrieving forest elevation a manner is proposed, which is based on the combination of decomposition scattering component and polarization state optimization technique using UAV-SAR images. In the suggested manner, firstly, a polarization state optimization algorithm is used to determine the direct scattering coefficient from the surface to overcome the assumption that this coefficient is ideal. Then, we add some power constraints for determining the optimal pair of polarization vectors ( $\vec{\omega}_{i\_opt}$ ,  $\vec{\omega}_{j\_opt}$ ) combined with the decomposition scattering component technique to improve the precision of estimating the interferometry coherence coefficient of the surface scattering component  $\tilde{\gamma}_{g\_opt}$  as well as the topographic phase. Next, an iterative conditional method is applied to retrieve the forest elevation as well as the mean extinction in the forest environment. The results show that the accuracy of forest elevation is appreciably enhanced by the proposed manner.

## 2 Methodology

### 2.1 Surface Phase Estimation

In PolInSAR systems, the complex interferometry coherence (CIC) coefficient is one of the main parameters that contains a lot of information about the target [6]:

$$\tilde{\gamma}(\vec{\omega}) = \frac{\vec{\omega}_1^H \cdot \Omega_{ms} \cdot \vec{\omega}_2}{\sqrt{(\vec{\omega}_1^H \cdot T_m \cdot \vec{\omega}_1)(\vec{\omega}_2^H \cdot T_s \cdot \vec{\omega}_2)}} = \frac{\vec{\omega}^H \cdot \Omega_{ms} \cdot \vec{\omega}}{\vec{\omega}^H \cdot T_{ms} \cdot \vec{\omega}} \quad (1)$$

where  $\Omega_{ms}$  is the cross-coherence matrix of the two data acquisitions of the PolInSAR system.  $T_m$ ,  $T_s$  are the coherence matrices of the primary and secondary PolSAR system, and  $T_{ms} = (T_m + T_s)/2$ .  $\vec{\omega}_1 = \vec{\omega}_2 = \vec{\omega}$  and  $(\cdot)^H$  are the polarization vector and the complex conjugate transpose operator, respectively.

Corresponding to the two component decomposition technique for PolInSAR data, the pola-coherence matrix and the interferometry coherence matrix can be represented through the linear combination of the scattering and volume scattering matrices are as follows [8]:

$$\begin{aligned} T_{ms} &= f_v T_v + f_g T_g \\ \Omega_{ms} &= \tilde{\gamma}_v T_v + \tilde{\gamma}_g T_g \end{aligned} \quad (2)$$

where  $\tilde{\gamma}_m, f_m \{m = v, g\}$  are the complex interferometry coherence factor and the scattering intensity parameter of the volume and surface scattering components, respectively.  $T_g$ ,  $T_v$  are the polarization coherence matrices for the surface and volume scattering contribution, respectively. The polarization coherence matrix for the volume scattering component is defined as Eq. 3.

$$T_v = \frac{1}{2 + 2e + 2\gamma} \begin{bmatrix} \gamma + 1 + 2\text{Re}(\rho)\sqrt{\gamma} & \gamma - 1 - 2\text{Im}(\rho)\sqrt{\gamma} & 0 \\ \gamma - 1 + 2\text{Im}(\rho)\sqrt{\gamma} & \gamma + 1 - 2\text{Re}(\rho)\sqrt{\gamma} & 0 \\ 0 & 0 & 2e \end{bmatrix} \quad (3)$$

where:

$$\gamma = \frac{\langle |s_{hh}^1 s_{hh}^{2*}| \rangle}{\langle |s_{vv}^1 s_{vv}^2| \rangle}; \quad e = \frac{2\langle |s_{hv}^1 s_{hv}^{2*}| \rangle}{\langle |s_{vv}^1 s_{vv}^{2*}| \rangle}; \quad \rho = \frac{\langle s_{hh}^1 s_{vv}^{2*} \rangle}{\sqrt{\langle |s_{hh}^1 s_{hh}^{2*}| \rangle \langle |s_{vv}^1 s_{vv}^{2*}| \rangle}} \quad (4)$$

where  $s_{pq}^j \{p, q = h, v; j = 1, 2\}$  are the complex scattering coefficients at the primary and secondary PolSAR systems with  $p, q$  polarimetric states, respectively. Based on the Freeman two component decomposition technique, we can completely determine the parameters  $f_v, f_g$  and the matrix  $T_g$  [8]. With a single-baseline PolInSAR system under a reversible scattering condition, the complex integer vector is represented as below [9]:

$$\vec{\omega} = [\cos a \cos a \sin b \cdot e^{j\eta_1} \sin a \sin b \cdot e^{j\eta_2}]^T \quad (5)$$

where,  $a, b, \eta_1$  and  $\eta_2$  are the real coefficients and their values range is determined as in document [9]. By changing the parameters  $a, b, \eta_1$  and  $\eta_2$  in their range and according

to formula (6) we can build a set of  $N$  vectors  $\vec{\omega}$ . In this paper, a complete searching method is proposed to determine an optimal CIC factor that representing the surface scattering component. This means we choose pairs of polarization vectors  $(\vec{\omega}_i, \vec{\omega}_j)$  so that  $\vec{\omega}_i^H T_v \vec{\omega}_j^H = 0$ . Then we find out that the direct scattering component from the canopy has been completely eliminated. By combining formulas (2) and (3) for each pair of polarization vectors  $(\vec{\omega}_i, \vec{\omega}_j)$ , it is possible to calculate the CIC factor for the dispersion component from the surface as Eq. 6.

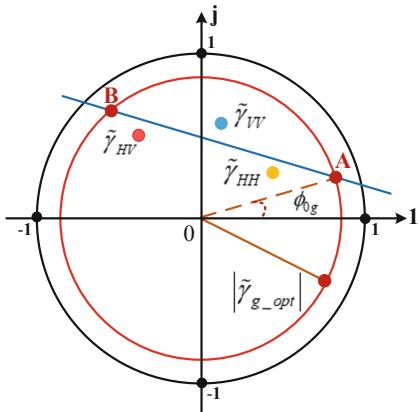
$$\tilde{\gamma}_g = \frac{\vec{\omega}_i^H \Omega_{ms} \vec{\omega}_j + \vec{\omega}_j^H \Omega_{ms} \vec{\omega}_i}{\vec{\omega}_i^H T_g \vec{\omega}_j + \vec{\omega}_j^H T_g \vec{\omega}_i} \quad (6)$$

Suppose that we construct  $L$  pairs of vectors  $(\vec{\omega}_i, \vec{\omega}_j)$  satisfying the condition  $\vec{\omega}_i^H T_v \vec{\omega}_j^H = 0$ , with  $L \leq N/2$ . Then we can determine the optimal pair of polarimetric vectors according to  $(\vec{\omega}_{i\_opt}, \vec{\omega}_{j\_opt})$  based on the following conditions:

$$\|1 - |\tilde{\gamma}_g|\|_{min} \quad (7)$$

Then  $\tilde{\gamma}_{g\_opt}$  will be determined based on the pair  $(\vec{\omega}_{i\_opt}, \vec{\omega}_{j\_opt})$  as follows:

$$\tilde{\gamma}_{g\_opt} = \frac{\vec{\omega}_{i\_opt}^H \Omega_{ms} \vec{\omega}_{j\_opt} + \vec{\omega}_{j\_opt}^H \Omega_{ms} \vec{\omega}_{i\_opt}}{\vec{\omega}_{i\_opt}^H T_g \vec{\omega}_{j\_opt} + \vec{\omega}_{j\_opt}^H T_g \vec{\omega}_{i\_opt}} \quad (8)$$



**Fig. 1.** Schematic representation of the ground phase estimation by suggested method.

Then apply the ATLS method to determine the best fit line [9]. This line will intersect the circle with center O and radius  $|\tilde{\gamma}_{g\_opt}|$  on the complex plane at two points (Fig. 1), and the surface phase is the point at which the distance from the CIC factor of the HV polarization channel is longest. As shown in Fig. 1, the ground phase corresponds to point A.

## 2.2 Optimal Iterative Method for Forest Height Estimation

The previous retrieved forest elevation manners [5–7] usually assumed that there was always some CIC factor that had only the direct scattering component from the canopy. In reality, each polarimetric channel is the mixing of many different scattering components, and this is one of the main causes of errors in extracting forest parameter of manners in the past. In this article, a solution for seeking a CIC factor for the direct scattering component from the canopy is proposed to overcome the limitations of the previous inversion manners. We know that, if the CIC factor for the canopy scattering component is known in advance, it is possible to estimate a CIC factor for any polarization channel as:

$$\tilde{\gamma}_{\text{est}}(\vec{\omega}) = \frac{\tilde{\gamma}_v(h_v, \sigma) + \tilde{\gamma}_{g_{\text{opt}}} \mu(\vec{\omega})}{1 + \mu(\vec{\omega})} = e^{j\phi_0} (\tilde{\gamma}_{0v}(h_v, \sigma) + L(\vec{\omega})(|\tilde{\gamma}_{g_{\text{opt}}}| - |\tilde{\gamma}_{0v}(h_v, \sigma)|)) \quad (9)$$

where:

$$\tilde{\gamma}_{0v}(h_v, \sigma) = \frac{2\sigma}{\cos\theta_{\text{inc}} \left( e^{\frac{2\sigma h_v}{\cos\theta_{\text{inc}} - 1}} \right)} \int_0^{h_v} e^{jk_z z} \cdot e^{\frac{2\sigma z}{\cos\theta_{\text{inc}}}} dz; \quad L(\vec{\omega}) = \frac{\mu(\vec{\omega})}{1 + \mu(\vec{\omega})} \quad (10)$$

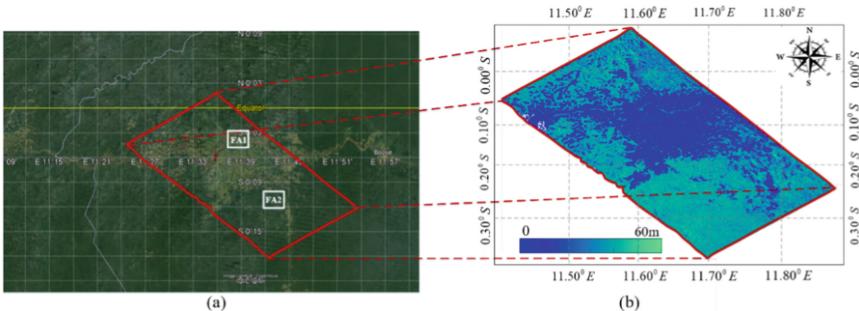
inwhich  $h_v, \sigma$  are the forest elevation and mean extinction factor, respectively. Two parameters  $k_z, \theta_{\text{inc}}$  are the vertical wavenumber and the incident angle. To limit the assumptions of previous forest height inversion manners, a conditional optimal iterative technique is offered for calculating forest height. By letting variables  $h_v \in [0, 2\pi/k_z]$ ,  $\sigma \in [0, 1]$  and  $L(\vec{\omega}) \in [0, 1]$  run in their range and combining formulas (10) with (11) we build a set of M complex polarimetric interferometry coherence coefficient  $\tilde{\gamma}_{\text{est}}$ . In order to excerpt the parameters  $h_v, \sigma, \mu(\vec{\omega})$  simultaneously, we calculate the total distance from the middle of  $\tilde{\gamma}_{\text{est}}$  with the complex interferometry coherence coefficient of polarization channels HH, HV, VV as Eq. 11.

$$D = |\tilde{\gamma}_{hh} - \tilde{\gamma}_{\text{est}}| + |\tilde{\gamma}_{hv} - \tilde{\gamma}_{\text{est}}| + |\tilde{\gamma}_{vv} - \tilde{\gamma}_{\text{est}}| \quad (11)$$

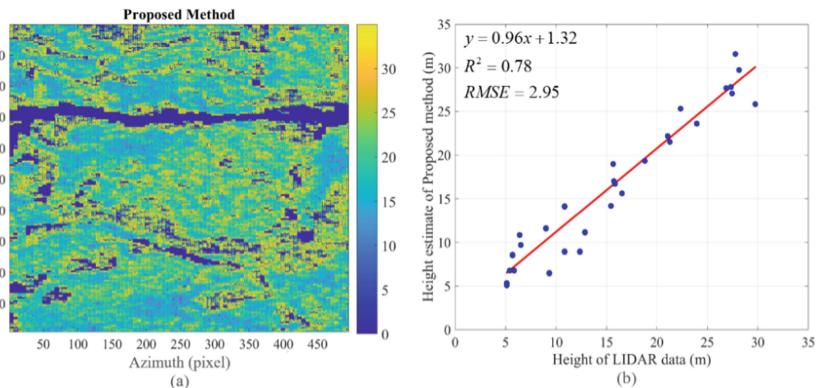
Finally, the optimum  $h_v, \sigma, \mu(\vec{\omega})$  parameters will be extracted when the total distance D is minimized. It can be seen that the finest CIC coefficient is determined based on different conditions to increase the efficiency of forest elevation estimation of the proposed manner. The solution of finding the best volume coherence coefficient of the suggested approach has overcome the shortcomings of the Tayebe one and has also improved the effectiveness of the estimated forest parameters at the same time.

### 3 Experimental Results

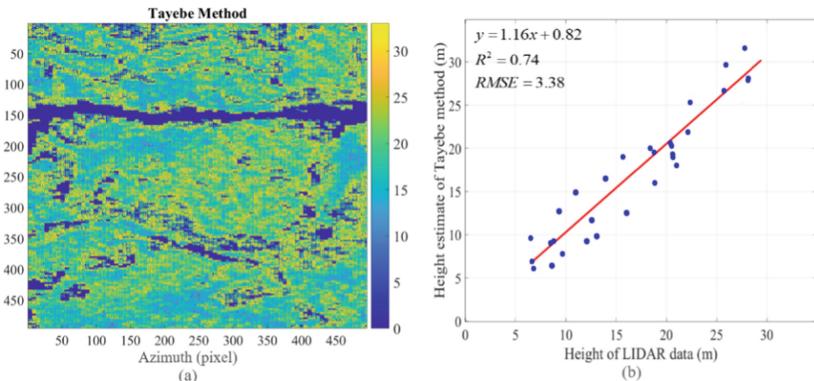
In the section, the proposed manner and Tayebe approach are also applied to the data set collected by NASA/JPL's open-surface synthetic radar on unmanned aerial vehicles (UAV-SAR) of NASA/JPL under the project AfriSAR. The radar operates in the L-band, with angle of incidence is ( $21^\circ$ – $65^\circ$ ) and baseline is 0 m–160 m. Accordingly, this dataset provides PolInSAR images to estimate forest height in Lope National Park in Gabon, USA. Both of them were measured on February 25, 2016 by NASA's partnership with ESA and the Babon Space Agency. LiDAR measurement data will be used as a reference to appraise the performance of the proposed manner. LiDAR data downloaded from the Distributed Active Archive Center for Biogeochemical Dynamics (ORNL DACC) are respectively shown in Fig. 2(b).



**Fig. 2.** The study forest area: (a) optical image from Google Earth, (b) LiDAR image.



**Fig. 3.** Results of proposed manner in forest area 1 (a) forest elevation mapping, (b) scatterplots between LiDAR and estimated forest elevation



**Fig. 4.** Results of Tayebe manner in forest area 1 (a) forest elevation mapping, (b) scatterplots between LiDAR and estimated forest elevation

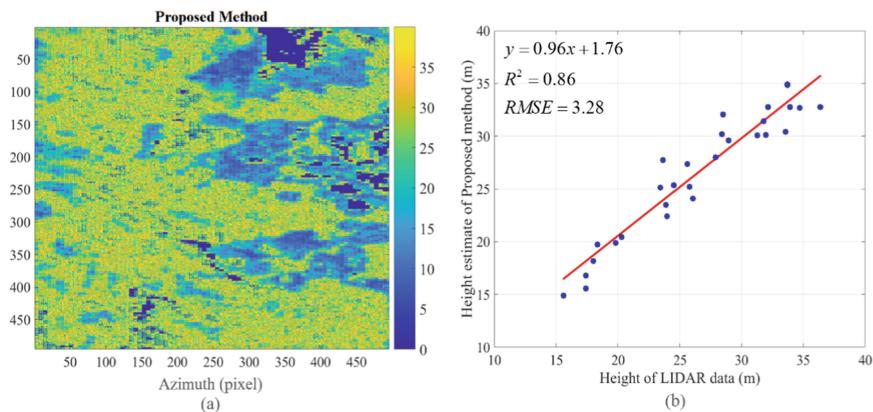
Figures 3(a), 4(a) show the results of forest elevation estimation of the proposed manner and the Tayebe's in the FA1 area. The 2D results show that this is a relatively flat area with rivers and lakes. It also has a sparse tree density with uneven distribution and diverse tree heights, most of the trees are below 20 (m). Both methods give relatively consistent results with the topography and tree height of the actual forest area, however, the height of Tayebe method in FA1 area tends to be lower than the proposed one.

Figure 3(b) presents the comparison of forest elevation estimation results by proposed manner and Tayebe approach (Fig. 4(b)) with forest heights extracted from LiDAR data at FA1 area. In which, the correlation coefficient determined by the new method is 0.78 and the Tayebe's has a lower value of 0.74. In addition, the root mean square error (RMSE) value of the proposed manner is 2.95 (m) and Tayebe's is 3.38 (m). It can be concluded from Figs. 3(b), 4(b) that the forest elevation estimated by the suggested technique has a higher accuracy than Tayebe approach, when compared with LiDAR data.

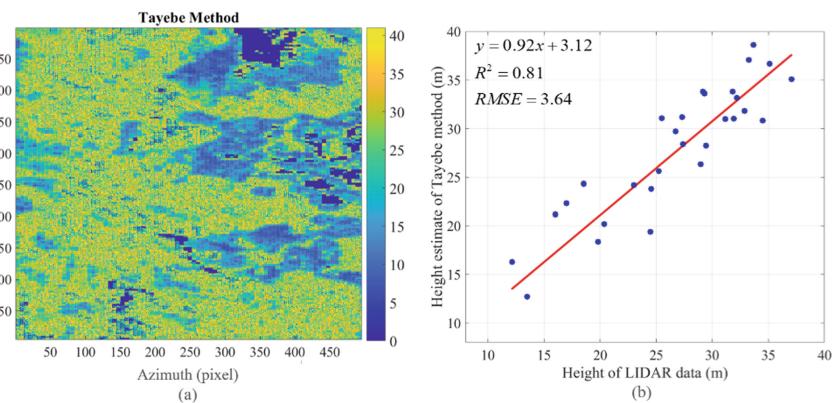
The results of evaluating forest elevation in the FA2 area by the recommended method (Fig. 5(a)) and Tayebe means (Fig. 6(a)) show dense tree density, the average height is above 25 (m) and the terrain has a relatively steep slope. These two figures show that the forest elevation estimated by Tayebe manner tends to be higher than that of the proposal one.

Figures 5(b), 6(b) show the comparison of forest elevation provided by the proposed and Tayebe approaches with LiDAR data in the FA2 area. This result shows that the correlation coefficient  $R^2$  determined by the offered approach is 0.86 and the Tayebe method is 0.81. In addition, the achieved RMSE coefficients of the two manners are 3.28 (m) and 3.64 (m), respectively.

Table 1 shows the forest height and mean extinction concluded by the two methods in the two studied forest areas (FA1, FA2). The average estimated forest elevation in the FA1 area by the proposed manner is 16.1 (m) and by the Tayebe approach is 15.6 (m). This result shows similarities with the results of the two methods in the simulation data and is lower than the elevation extracted from the LiDAR data (17.2 (m)). However, the forest elevation estimated by the proposed manner shows higher accuracy.



**Fig. 5.** Results of proposed manner in forest area 2 (a) forest elevation mapping, (b) scatterplots between LiDAR and estimated forest elevation



**Fig. 6.** Results of Tayebe manner in forest area 2 (a) forest elevation mapping, (b) scatterplots between LiDAR and estimated forest elevation.

**Table 1.** Results of two methods compared to LiDAR data.

Forest area	Forest parameters	LiDAR data	Tayebe manner	Proposed manner
FA1	$\bar{h}_v$ (m)	17.2	15.6	16.1
	$\sigma$ (dB/m)	—	0.2	0.16
FA2	$\bar{h}_v$ (m)	28.9	30.8	27.4
	$\sigma$ (dB/m)	—	0.29	0.24

In the FA2 area, the average forest elevation retrieved by Tayebe approach is 30.8 (m) which is 1.9 (m) higher than the LiDAR data. This represents an error exceeding the threshold, which leads to a significant decrease in the reliability of the Tayebe's. This error occurs because the Tayebe approach assumes that the forest surface is relatively flat when calculating the forest height on terrain with steep slopes. In this area, the forest elevation determined by the proposed manner is 27.4 (m), which is approximately 1.5 (m) lower than the forest height extracted from the LiDAR data. In addition, the mean extinction coefficient assumed by the suggested technique in the two researched forest areas (FA1, FA2) is lower than that of Tayebe means. From Table 1, it is shown that the forest parameters extracted from the proposed manner have higher accuracy and reliability than the Tayebe one.

## 4 Conclusion

This paper has developed a method to increase the accuracy of estimating forest parameters using L-band polarimetric interferometry UAV-SAR data. The proposed approach, in addition to overcomes the limitations of the former, also considerably improves the precision of forest elevation. Through the evaluation with PolInSAR, UAV-SAR and LiDAR data, it is suggested that the new method has achieved high efficiency and fixed the unprecise estimations of the 3-state inversion and the Tayebe manners. However, this new approach is relatively complicated in calculation, because of adding and developing constraint conditions for the optimal solution. This also means that the processing time will be prolonged considerably, while in three-state inversion manner it is much shorter.

## References

1. Tan, N.N., Nghia, P.M., Cuong, T.H., Nhu, L.V.: Volume coherence function optimization method for extracting vegetation and terrain parameters from polarimetric synthetic aperture radar interferometry images. *J. Appl. Remote Sens.* **14**(4), 1–20 (2020)
2. Liao, Z., He, B., Bai, X., Quan, X.: Improving forest height retrieval by reducing the ambiguity of volume only coherence using multibaseline PolInSAR data. *IEEE Trans. Geosci. Remote Sens.* **57**(11), 8853–8865 (2019)
3. Yamada, H., Onoda, H., Yamaguchi, Y.: On scattering model decomposition with PolInSAR data. In: Proceedings of EUSAR2008, Friedrichshafen, Germany, pp. 1–4 (2008)
4. Noelia, R.P., Juan, M.L.S., Ballester-Berman, J.D.: Estimation of RvoG scene parameters by mean of PolInSAR with TanDem-X data: effect of the double-bounce contribution. *IEEE Trans. Geosci. Remote Sens.* **58**(10), 7283–7304 (2020)
5. Treuhhaft, R.N., Madsen, S.N., Moghaddam, M.: Vegetation characteristics and underlying topography from interferometric radar. *Radio Sci.* **31**(6), 1449–1485 (1996)
6. Cloude, S.R., Papathanassiou, K.P.: Three-stage inversion process for polarimetric SAR interferometric. *IEE Proc. Radar Sonar Navig.* **150**(3), 125–134 (2003)
7. Tayebe, M., Maghsoudi, Y., Zanj, M.J.V.: A volume optimization method to improve the three-stage inversion algorithm for forest height estimation using PolInSAR data. *IEEE Geosci. Remote Sens. Lett.* **15**(8), 1214–1218 (2018)
8. Alberto, A.G., Konstantinos, P.P.: Polarimetric change analysis in forest using PolInSAR ground and volume separation techniques. In: 13th European Conference on Synthetic Aperture Radar of EUSAR 2021, pp. 609–614 (2021)
9. Cuong, T.H., Nghia, P.M., Nhu, L.V.: Forest parameters inversion by mean coherence set from single-baseline PolInSAR data. *J. Adv. Space Res.* **68**(7), 2804–2818 (2021)



# Fuzzy Portfolio Selection with Flexible Optimization via Quasiconvex Programming

Tran Thi Thanh Tuoi, Truong Tuan Khang, Nguyen Thi Ngoc Anh,  
and Tran Ngoc Thang<sup>(✉)</sup>

School of Applied Mathematics and Informatics, Hanoi University of Science  
and Technology, 1 Dai Co Viet Street, Hanoi 100000, Vietnam  
[{anh.nguyenthingoc.thang.tranngoc}@hust.edu.vn](mailto:{anh.nguyenthingoc.thang.tranngoc}@hust.edu.vn)

**Abstract.** In this study, we consider a single objective fuzzy portfolio optimization with flexible goal and constraints, in which the Sharpe ratio is chosen as the goal and the portfolio's mean and variance are included in the constraints. Although this problem has much significance in finance, it is difficult to solve because of the nonconvexity of the objective function. Based on fuzzy theory and flexible optimization, the fuzzy portfolio problem is transformed to the crisp form which is proved to be a semistrictly quasiconvex programming problem for any decreasing membership functions. This property of the equivalent problem is the basis to solve the main problem efficiently by available convex programming algorithms. The computational experiments with the SP500 data set is reported to show the performance of the proposed model.

**Keywords:** Fuzzy portfolio selection · Flexible optimization · Soft constraints · Sharpe ratio · Semistrictly quasiconvex programming

## 1 Introduction

In financial technology, portfolio selection is a significant process, specially in business investments. It is the process of assigning a predetermined amount of money to various assets in order to construct a well-balanced portfolio. This process helps determine a mix of securities from a large number of possibilities. Its goal is to help investors get the most out of their money. In recent years, portfolio selection has gotten much attention from many individuals and institutions.

Mean-Variance model or Markowitz model is popular and used widely in modern portfolio theory, where portfolio selection is constructed as a bi-objective optimization problem for minimizing risk of portfolio as well as maximising expected portfolio profit. As usual, this problem is transformed to a single-objective programming problem, such as minimizing risk of portfolio for a chosen level of anticipated return or maximising expected portfolio profit for a chosen

level of risk. Because both the risk and profit objective functions are convex, the transformed problem is a convex programming problem.

In practice, the actual data, however, we can see that a number of variables impact the stock market, resulting in frequent price changes as well as our desire to find the best answer to the difficult situation (see [1]). As a consequence, we can see that tackling the problem of fuzzy portfolio selection is a great idea. Numerous more results that applied fuzzy theory to portfolio problems were presented recently (see [2–4]).

Besides mean and variance, some other measures are also used to evaluate portfolios (see [2, 5, 6]). This research considers a variant Markowitz model by adding an objective function called Sharpe Ratio (SR) (also see [7]). SR is a prominent risk-adjusted return performance metric that indicates the ratio of the expectant profit to the standard deviation of the portfolio. SR has much applied and economic significance, but it has some disadvantages when solving the programming problem with the objective SR because of its non-convexity. Therefore, recently there only some works related to this index, such as Sharpe Ratio - VaR Ratio model with fuzzy coefficients solved using genetic algorithm (see [4]), the “fuzzy Sharpe Ratio” and new risk measure with fuzzy random variables in the fuzzy modeling environment (see [2]). These works used the heuristic approach such as genetic algorithms for solving the programming problem. This approach only helps find the local optimal solutions and cannot guarantee the algorithmic convergence to global optimal solutions. Moreover, for the computational performance aspect, it may have to execute a huge computational load to generate a big enough population to find a good enough approximated solution.

In this research, we show the semistrictly quasiconvex property of Sharpe Ratio function. By utilizing the properties of semistrictly quasiconvex functions, the fuzzy portfolio selection problem using SR is transformed to a semistrictly quasiconvex programming, which can solve globally and efficiently by available convex programming algorithms (also see [8, 9]). The globally convergence of the proposed procedure is guaranteed. Last but not least, we also prove that the equivalent crisp problem is still a semistrictly quasiconvex programming even when the membership functions are arbitrary decreasing functions, while the previous works only consider some specific forms of membership functions (see [10]). This helps one get more flexible, easily and suitable options for fuzzification in the modelling process.

The remainder of the paper is arranged as follows. In Sect. 2, we present the portfolio selection problem which uses the Sharpe Ratio index. Section 3 introduces the fuzzy portfolio selection issue and the solution approach. The findings of the computational experiments are shown in Sect. 4. In the final portion, there are a few conclusions.

## 2 Portfolio Selection Problem

Markowitz's portfolio theory is founded on two key assumptions: 1) investors are wary about risks and want the highest possible expected profit, and 2) investors pick their portfolios based on desirable return and variance of return.

Recall the random vector  $\mathbf{R} = (R_1, R_2, \dots, R_n)^T \in \mathbb{R}^n$  denotes random returns of the  $n$  assets. Suppose that  $p(\mathbf{R})$  is the probability distribution of  $\mathbf{R}$ . Calling mean vector of  $\mathbf{R}$  is  $\mathbf{L} = (L_1, \dots, L_n)^T$  and covariance matrix of  $\mathbf{R}$  is  $\mathbf{Q} = (\sigma_{ij})_{n \times n}$ , where  $\sigma_{jj}^2$  is variance of  $R_j$  and  $\sigma_{ij}^2$  is correlation coefficient between  $R_i$  and  $R_j$ ,  $i, j = 1, \dots, n$ . With  $\mathbf{x} = (x_1, \dots, x_n)^T$  is a portfolio, note that  $\sum_{j=1}^n x_j = 1$  and  $x_j \geq 0$  for all  $j = 1, \dots, n$ . We have the expected return  $\mathcal{E}(\mathbf{x}) = E[\mathbf{R}^T \mathbf{x}] = \sum_{j=1}^n L_j x_j$ , and variance of profit  $\mathcal{V}(\mathbf{x}) = \text{Var}(\mathbf{R}^T \mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_j x_i = \mathbf{x}^T \mathbf{Q} \mathbf{x}$ .

The expected return is used to represent the investment in the future, whilst the variance of return is used to estimate the risk of the investment. As a result, the return is calculated by the projected profit or the average of the profits, while the profit variance throughout the whole portfolio is used to estimated the risk. Investors are best served by two objectives: maximum lucrative and minimum the risk with portfolio value bound constraints.

Set  $M = \{\mathbf{x} \in \mathbb{R}_+^n \mid \sum_{j=1}^n x_j = 1\}$  where  $\mathbb{R}_+^n$  is nonnegative orthant of real  $n$ -dimentional space. Then, the portfolio optimization problem is therefore presented as follows.

$$\begin{aligned} \max \quad & \mathcal{E}(\mathbf{x}) = E[\mathbf{R}^T \mathbf{x}] = \sum_{j=1}^n L_j x_j \\ \min \quad & \mathcal{V}(\mathbf{x}) = \text{Var}(\mathbf{R}^T \mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_j x_i = \mathbf{x}^T \mathbf{Q} \mathbf{x}, \\ \text{s.t.} \quad & \mathbf{x} \in M. \end{aligned} \quad (\text{MV})$$

The Sharpe Ratio function, on the other hand, is the aim of our article - an index that measures a portfolio's risk-adjusted returns with the formulation as

$$\mathcal{S}(\mathbf{x}) = \frac{\mathcal{E}(\mathbf{x}) - r_f}{\sqrt{\mathcal{V}(\mathbf{x})}}$$

where  $r_f$  represents the risk free rate. The higher the SR value, the higher the portfolio return in comparison to the risk the decision-maker is willing to take.

In this research, we propose the single-objective deterministic portfolio optimization model as

$$\begin{aligned} \min \quad & f(\mathbf{x}) = -\mathcal{S}(\mathbf{x}) = -\frac{\mathcal{E}(\mathbf{x}) - r_f}{\sqrt{\mathcal{V}(\mathbf{x})}} \\ \text{s.t.} \quad & \mathcal{E}(\mathbf{x}) = E[\mathbf{R}^T \mathbf{x}] = \sum_{j=1}^n L_j x_j \geq \alpha, \\ & \mathcal{V}(\mathbf{x}) = \text{Var}(\mathbf{R}^T \mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_j x_i = \mathbf{x}^T \mathbf{Q} \mathbf{x} \leq \beta, \\ & \mathbf{x} \in M. \end{aligned} \quad (\text{MVS1})$$

The basic theory of semistrictly quasiconvex programming will be presented later in Sect. 3.3. In this section, there are some propositions related to semistrictly quasiconvexity that can be inferred from the proposed model.

Recall that if  $\varphi_1(\mathbf{x})$  is concave and  $\varphi_2(\mathbf{x})$  is convex defined on  $C \subset \mathbb{R}^n$  satisfied  $\varphi_1(\mathbf{x}) \geq 0, \varphi_2(\mathbf{x}) > 0$ , then  $\phi = \frac{\varphi_1}{\varphi_2}$  is semistrictly quasiconcave function on  $C$  (see [11], Table 5.4, page 165). Therefore, we have the propositions below.

**Proposition 1.** *The Sharpe Ratio function  $\mathcal{S}(\mathbf{x})$  is semistrictly quasiconcave on  $\mathbb{R}^n$ .*

**Proposition 2.** *Problem (MVS1) is a semistrictly quasiconvex programming.*

In actuality, there are some inaccuracies in the parameters, as well as some errors in estimating the expected profit and the risk of the portfolio. If we continue to work on the original objective function, we may not be able to find a satisfactory optimal solution. Furthermore, fuzzicating constraints expands the feasible set, increasing the likelihood of obtaining a satisfied solution with a better optimal value. Therefore, considering the Problem (MVS1) in the fuzzy context is reasonable. In the next section, we will go over some fundamental concepts in fuzzy optimization, the procedure for converting Problem (MVS1) to a fuzzy form, and how to solve the fuzzy problem.

### 3 Fuzzy Portfolio Problem

#### 3.1 Fuzzy Optimization

In this article, we use the fuzzy optimization model with fuzzy components in which the variables and parameters are crisp numbers, which is also called flexible optimization (see [12]). There are two types of fuzzy components. The first component is associated with the objective function ( $\widetilde{\min}, \widetilde{\max}$ ), while the second component is a fuzzy relation ( $\preceq, \simeq, \succeq$ ). Without loss the generality, a fuzzy optimization problem with fuzzy components can be formulated as

$$\begin{aligned} & \widetilde{\min} f(\mathbf{c}, \mathbf{x}) \\ & \text{s.t. } z_i(\mathbf{x}) \preceq 0, \quad i = 1, \dots, k, \\ & \mathbf{x} \in X, \end{aligned} \tag{P_f}$$

where  $X \subset \mathbb{R}^n$  is a set that is not empty. The symbol “ $\widetilde{\min}$ ” and “ $\preceq$ ” stand for a fuzzy version of “minimize” and “ $\leq$ ”, respectively, indicating “the objective function should be reduced as much as feasible” and “the constraints should be possibly well acceptable.”

In the following, we review the fundamentals of fuzzy theory.

**Definition 1. (Fuzzy set ([10])).** Given an universal set  $X$ . To define a fuzzy set  $\tilde{A}$  of  $X$ , a function  $\mu$  is used.

$$\mu_{\tilde{A}} : X \longrightarrow [0, 1]$$

It is called membership function, which sets a real number  $\mu_{\tilde{A}}$  to each  $x \in X$  in order to represent the level of membership of  $x$  in  $\tilde{A}$ .

In general case, a membership function has the form as

$$\mu_i(z_i(x)) = \begin{cases} 0 & \text{if } z_i(x) \geq z_i^0, \\ d_i(x) & \text{if } z_i^0 \geq z_i(x) \geq z_i^1, \\ 1 & \text{if } z_i(x) \leq z_i^1, \end{cases}$$

where  $d_i(x)$  is a monotonic non-increasing function with respect to  $z_i$ ,  $z_i^0$  and  $z_i^1$  represents the value of  $z_i$  such that the grade of  $\mu_i(z_i(x))$  is 0 or 1. Therefore, membership functions are decreasing functions. Membership function for objective function and constraints present the grade that the objective function is minimized and the constraints are satisfied, respectively.

In many other previous studies, researchers propose solving methods for only some particular types of membership functions, but our approach can be used for arbitrary strictly decreasing functions.

### 3.2 Fuzzy Portfolio Problem

In this paper, we adopt vagueness and transform Problem (MVS1) to its fuzzy version as

$$\begin{aligned} \widetilde{\min} f(\mathbf{x}) &= -\mathcal{S}(\mathbf{x}) = -\frac{\mathcal{E}(\mathbf{x}) - r_f}{\sqrt{\mathcal{V}(\mathbf{x})}} \\ \text{s.t. } \mathcal{E}^*(\mathbf{x}) &= -\mathcal{E}(\mathbf{x}) = -\sum_{j=1}^n L_j x_j \preceq -\alpha, \\ \mathcal{V}(\mathbf{x}) &= \mathbf{x}^T \mathbf{Q} \mathbf{x} \preceq \beta, \\ \mathbf{x} &\in M. \end{aligned} \quad (\text{MVS2})$$

By applying arbitrary strictly decreasing membership functions  $\mu_i, i = 0, 1, 2$  (\*) and following the fuzzy decision of Bellman and Zadeh, Problem (MVS2) is transformed to Problem (FP1).

$$\begin{aligned} \max \min \{ &\mu_0(f(\mathbf{x})), \mu_1(\mathcal{E}^*(\mathbf{x})), \mu_2(\mathcal{V}(\mathbf{x})) \} \\ \text{s.t. } \mathbf{x} &\in M. \end{aligned} \quad (\text{FP1})$$

Because  $f(\mathbf{x})$  is not convex, Problem (FP1) is a nonconvex programming problem. However, it belongs to the class of semistrictly quasiconvex programming problems. The reason will be presented in Subsect. 3.3.

### 3.3 Semistrictly Quasiconvex Programming

**Definition 2. (Semistrictly quasiconvex function (see [11])).** Let  $X \subset \mathbb{R}^n$  be a convex set,  $f$  is defined on  $X$ . For all  $\mathbf{x}^1 \in X, \mathbf{x}^2 \in X, 0 < \lambda < 1$ , if

$$f(\mathbf{x}^1) > f(\mathbf{x}^2) \text{ implies that } f(\lambda \mathbf{x}^1 + (1 - \lambda) \mathbf{x}^2) < f(\mathbf{x}^1),$$

then  $f$  is semistrictly quasiconvex on  $X$ .

**Definition 3. (Semistrictly quasiconvex programming problem).** Given a convex set  $X \subset \mathbb{R}^n$ ,  $X \neq \emptyset$ . Then the formulation of a semistrictly quasiconvex programming problem is

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t. } & \mathbf{x} \in X, \end{aligned} \tag{SQP}$$

where  $f(\mathbf{x})$  is semistrictly quasiconvex on  $X$ .

**Proposition 3.** Any local minimum of (SQP) is a global minimum (see [11]). Therefore, (SQP) can be solved by proper convex programming algorithms.

Let's return to Problem (FP1). By Proposition 2,  $f(\mathbf{x})$  is a semistrictly quasiconvex function. In addition,  $\mathcal{E}^*(\mathbf{x})$  is linear and  $\mathcal{V}(\mathbf{x})$  is convex because of the positive symmetric definite matrix  $\mathbf{Q}$ . So we have  $f(\mathbf{x})$ ,  $\mathcal{E}^*(\mathbf{x})$  and  $\mathcal{V}(\mathbf{x})$  are semistrictly quasiconvex functions (\*\*).

Set operator  $\eta_i(y) = 1 - \mu_i(y)$ , then Problem (FP1) is equivalent to

$$\begin{aligned} \min g(\mathbf{x}) = \max \{ & \eta_0(f(\mathbf{x})), \eta_1(\mathcal{E}^*(\mathbf{x})), \eta_2(\mathcal{V}(\mathbf{x})) \} \\ \text{s.t. } & \mathbf{x} \in M. \end{aligned} \tag{FP2}$$

By assumption (\*),  $\eta_i(y)$ ,  $i = 0, \dots, 2$  are strictly increasing functions. Combining with (\*\*), then,  $\eta_0(f(\mathbf{x}))$ ,  $\eta_1(\mathcal{E}^*(\mathbf{x}))$ ,  $\eta_2(\mathcal{V}(\mathbf{x}))$  are semistrictly quasiconvex functions (see [11], Proposition 5.1, page 154). After that, we have Proposition 4, which can be proven by using Definition 2. And finally, all the reasons stated lead to Proposition 5.

**Proposition 4.** The objective function  $g(\mathbf{x})$  of Problem (FP2) is semistrictly quasiconvex on  $\mathbb{R}^n$ .

**Proposition 5.** Problem (FP2) is a semistrictly quasiconvex programming problem and is solved efficiently by using available convex programming algorithms.

The proposed method simplifies the computation process significantly. This approach is different from the other approach of the previous articles which use genetic algorithms.

## 4 Computational Experiment

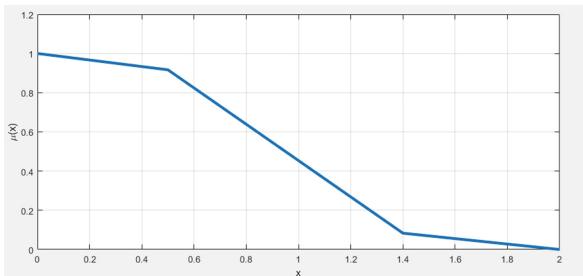
This section uses two examples to describe the performance of our proposal in the application of fuzzy portfolio optimization.

The first example uses stock prices from 1/23/2015 to 6/12/2017 of five stocks with symbols such as ULTA, MLM, NFLX, AMZN and NVDA. The expected profits and covariance matrix of five assets are given in Table 1.

**Table 1.** Data of the Example 1

Stock	$L$	$Q$				
ULTA	0.15672	4.41513	1.12491	2.31042	1.44398	1.39347
MLM	0.15874	1.12491	4.07482	1.96306	1.28708	1.53560
NFLX	0.20462	2.31042	1.96306	9.13912	2.33831	1.98378
AMZN	0.21693	1.44398	1.28708	2.33831	4.43169	1.67068
NVDA	0.34876	1.39347	1.53560	1.98378	1.67068	5.31435

The deterministic problem is (MVS1) model with  $\alpha = 2.5$ ,  $\beta = 0.25$ . Then, the membership functions applied for objective function and constraints are piecewise-linear membership functions which are strictly decreasing functions on the range of  $f(\mathbf{x})$ ,  $\mathcal{E}^*(\mathbf{x})$ ,  $\mathcal{S}(\mathbf{x})$ .

**Fig. 1.** Piecewise-linear membership function

The fuzzy portfolio problem is solved using Matlab's convex programming tool, and the results are presented in Table 2. From Table 2, the optimal value for the fuzzy portfolio problem is 1% greater than the optimal value for the deterministic problem. Flexible optimization that expands the feasible set can look to produce a better objective function value than the initial problem's optimal value.

**Table 2.** Value of (MVS1) and (MVS2) with  $\alpha = 0.25$  and  $\beta = 2.5$ 

Function	$\mathcal{S}(\mathbf{x})$	$\mathcal{E}(\mathbf{x})$	$\mathcal{V}(\mathbf{x})$
Deterministic ( $\bar{\mathbf{x}}$ )	0.14649	0.25363	2.49999
Fuzzy ( $\mathbf{x}^*$ )	0.14795	0.26151	2.62045

The second example uses five stocks ticker symbols such as ALB, AVGO, CHTR, ULTA and MLM. And Table 3 displays the data's expected profits and covariance matrix of stocks cost from 2/2/2015 to 1/25/2017.

**Table 3.** Data of the second example

Stock	$L$	$Q$				
ALB	0.15540	3.72379	1.53106	1.11161	1.05120	1.35014
AVGO	0.15881	1.53106	5.29711	1.08504	1.69505	1.57479
CHTR	0.16339	1.11161	1.08504	4.00362	1.03543	1.20725
ULTA	0.16504	1.05120	1.69505	1.03543	4.39397	1.11849
MLM	0.17814	1.35014	1.574787	1.20725	1.11849	4.06780

To analyze the discovered solutions, we conduct the following: after determining the optimal solution  $\bar{\boldsymbol{x}}$  of the portfolio selection problem, we utilize it to compute the optimal solution  $\boldsymbol{x}^*$  of the fuzzy portfolio problem by applying piecewise-linear membership functions. Then, we calculate the average of Sharpe ratio, mean, and variance values in the next 100 days with each solution  $\bar{\boldsymbol{x}}$  and  $\boldsymbol{x}^*$ . Table 4 displays the results. The SR value of the fuzzy problem is not as good as it is for the deterministic problem, as seen in Table 4. However, When applied to data over the next 100 days, the average of  $S(\boldsymbol{x}^*)$  is 1.12% better than  $S(\bar{\boldsymbol{x}})$ , demonstrating the second meaning of fuzzy.

**Table 4.** Result of the second example

Function	$S(\bar{\boldsymbol{x}})$	$S(\boldsymbol{x}^*)$	$\mathcal{E}(\bar{\boldsymbol{x}})$	$\mathcal{E}(\boldsymbol{x}^*)$	$\mathcal{V}(\bar{\boldsymbol{x}})$	$\mathcal{V}(\boldsymbol{x}^*)$
500 days	0.10576	0.10521	0.16574	0.16426	1.84720	1.82834
Mean of 100 data sets	0.10276	0.10391	0.15802	0.15855	1.75269	1.72776

## 5 Conclusion

In conclusion, this work proposes a portfolio model, in which we optimize the Sharpe ratio and control the value of return and risk, and consider the model in the fuzzy environment. We have stated that the deterministic problem is a (SQP), and proposed a method, which can use just arbitrary strictly decreasing membership functions, to convert it to the flexible version that then was proven to be a (SQP). As the result, the fuzzy portfolio problem can be addressed by just using convex programming methods, which takes less computing effort than genetic algorithms in previous workarounds. Because of its advantages in practical experiments compared to other methods and the appearance of semistrictly quasiconvex functions in many problems in the real world, this approach may be used in a variety of different mathematical models that contain semistrictly quasiconvex functions.

## References

1. Van, N.D., Doanh, N.N., Khanh, N.T., Anh, N.T.N.: Hybrid classifier by integrating sentiment and technical indicator classifiers. *Context-Aware Syst. Appl. Nat. Comput. Commun.* **21**, 25–37 (2018). [https://doi.org/10.1007/978-3-319-77818-1\\_3](https://doi.org/10.1007/978-3-319-77818-1_3)
2. Nguyen, T.T., Gordon-Brown, L., Khosravi, A., Creighton, D., Nahavandi, S.: Fuzzy portfolio allocation models through a new risk measure and fuzzy Sharpe ratio. *IEEE Trans. Fuzzy Syst.* **23**(3), 656–676 (2015). <https://doi.org/10.1109/TFUZZ.2014.2321614>
3. Qin, Z.: Random fuzzy mean-absolute deviation models for portfolio optimization problem with hybrid uncertainty. *Appl. Soft Comput.* **56**, 597–603 (2017). <https://doi.org/10.1016/j.asoc.2016.06.017>
4. Kar, M.B., Kar, S., Guo, S., Li, X., Majumder, S.: A new bi-objective fuzzy portfolio selection model and its solution through evolutionary algorithms. *Soft Comput.* **23**(12), 4367–4381 (2018). <https://doi.org/10.1007/s00500-018-3094-0>
5. Thang, T.N., Vuong, N.D.: Portfolio selection with risk aversion index by optimizing over pareto set. *Intell. Syst. Netw.* **24**3 (2021). <https://doi.org/10.1007/978-981-16-2094-2-28>
6. Huang, X., Jiang, G., Gupta, P., Mehlawat, M.K.: A risk index model for uncertain portfolio selection with background risk. *Comput. Oper. Res.* **132** (2021). <https://doi.org/10.1016/j.cor.2021.105331>
7. Hoang, D.M., Thang, T.N., Tu, N.D., Hoang, N.V.: Stochastic linear programming approach for portfolio optimization problem. In: 2021 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT), pp. 1–4 (2021). <https://doi.org/10.1109/ICMLANT53170.2021.9690552>
8. Thang, T.N., Kumar, S.V., Anh, D.T., Anh, N.T.N., Pham, V.H.: A monotonic optimization approach for solving strictly quasiconvex multiobjective programming problems. *J. Intell. Fuzzy Syst.* **38** (2020). <https://doi.org/10.3233/JIFS-179690>
9. Thang, T.N.: Outcome-based branch and bound algorithm for optimization over the efficient set and its application. In: Dang, Q.A., Nguyen, X.H., Le, H.B., Nguyen, V.H., Bao, V.N.Q. (eds.) Some Current Advanced Researches on Information and Computer Science in Vietnam. AISC, vol. 341, pp. 31–47. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-14633-1\\_3](https://doi.org/10.1007/978-3-319-14633-1_3)
10. Sakawa, M.: Fuzzy Sets and Interactive Multiobjective Optimization, 1st edn. Springer, New York (1993). <https://doi.org/10.1007/978-1-4899-1633-4>
11. Avriel, M., Diewert, W.E., Schaible, S., Zang, I.: Generalized concavity (2010). <https://doi.org/10.1137/1.9780898719437>
12. Lodwick, W.A., Neto, L.L.S.: Flexible optimization. In: Lodwick, W.A., L.L.S.-N. (eds.) Flexible and Generalized Uncertainty Optimization, pp. 141–162. Springer, New York (2021). <https://doi.org/10.1007/978-3-030-61180-4>



# AN-BEATS for Short-Term Electricity Load Forecasting with Adjusted Seasonality Blocks and Optimization of Block Order

Nguyen Anh Tuan<sup>1</sup>, Do Thi Thanh Chau<sup>1</sup>, Le Anh Ngoc<sup>2</sup>,  
and Nguyen Thi Ngoc Anh<sup>1(✉)</sup>

<sup>1</sup> Hanoi University of Science and Technology, Hanoi, Vietnam

[anh.nguyenthongoc@hust.edu.vn](mailto:anh.nguyenthongoc@hust.edu.vn)

<sup>2</sup> Swinburne University of Technology, Hanoi, Vietnam

[ngocla2@fe.edu.vn](mailto:ngocla2@fe.edu.vn)

**Abstract.** For the proper operation of electrical systems, accurate electricity load forecasting is essential. This study focuses on solving the problem that is the optimization of the block order to archive better accuracy of the forecasting model. Furthermore, the seasonality blocks of N-BEATS are adjusted in theory by correctly using the Discrete Fourier Transform. Therefore, AN-BEATS-Adjusted Neural Basic Expand - Analysis Time series model is proposed to forecast short-term power loads based on electricity load history. Experiments show that the proposed model works better than the LSTM model and the order of blocks strongly affects the model's prediction results.

**Keywords:** N-BEATS · Time-series · Forecasting · Electricity load · Seasonal decomposition

## 1 Introduction

Electricity load forecasting at the national level is becoming more crucial in successful power management and is the foundation for corresponding decision making [1]. Deep learning are now chosen by academics for short-term electricity load forecasting due to their various benefits in empirical data [2].

Yaoyao He et al. [3] presented a model to handle an issue of short-term power load probability density forecasting. The kernel-based support vector quantile regression (KSVQR) and Copula theory are used in this model. The short-term electrical load forecasting is solved in the paper [4] by integrating the extreme learning machine (ELM) with a novel switching delayed PSO algorithm. The proposed model outperforms existing cutting-edge ELMs in terms of performance. Song Li et al. present a novel synthetic technique for short-term electrical load forecasting based on wavelet transform, extreme learning machine

(ELM), and partial least squares regression in their publication [5] (PLSR). The wavelet transform was used to decompose the fundamental time series, and ELM was employed for each sub-component resulting from the wavelet decomposition. The findings showed that the suggested approach might improve load forecasting performance. In this paper [6], the authors study the influence of deep-stacked LSTM and Bi-LSTM models on electricity load forecasting. In addition, they tune the hyperparameters of the LSTM model by using the Bayesian optimization method to gain the best performance forecasting. The experiment results show that deep-stacked LSTM layers do not increase prediction accuracy significantly.

In this paper [7], the authors proposed to use a Fourier series form with a detrended linear function what to represent a periodic time series function with a linear trend. The complete Fourier representation is used in the comparative analysis. Diagnostic tests demonstrated that the suggested method outperforms the pure Fourier approach. [8] A novel method forecasts the trend-cycle and seasonal components using fuzzy techniques. The proposed method gets good performance without considering irregular fluctuation for time series with clear trends and seasonal components. [9] Both of the mean price reversion (MRP) and the seasonality component are represented by the Fourier series. In addition, the Fourier series represent the MRP have a stochastic factor. The Kalman Filter is used to estimate the long-term and seasonality components in the representation state space. In another research [10], the authors focus on the ability of Singular Spectral Analysis, Seasonal-Trend decomposition using Loess, and attribute selection pre-processing methodologies combined with neural network models to forecast monthly river streamflows in Turkey's Nallihan stream.

N-BEATS [11] is a deep learning model built only by fully connected layers to solve the univariate time series forecasting problem and performs well with time-series data in many fields of competition. The most important advantage of the model is interpretation. N-BEATS can decompose time series data into three components trend, seasonality, and residual. Another advantage of the model is fast training and inference than another deep learning model. In [1], the authors used the ensemble technique to build N-BEATS models bootstrapped to solve the mid-term electricity load forecasting problem. Compared proposal models with ten other methods (ARIMA, k-NN, MLP,...), the ensemble of N-BEATS has the best performance. In [12], the authors apply a Neural Architecture Search to add Recurrent Neural Network to N-BEATS model called N-BEATS-RNN that help to reduce training time. However, N-BEATS-RNN does not refer to interpretability when compared to N-BEATS. In [13], the authors extend the N-BEATS model in combination with exogenous variables called the N-BEATS<sub>x</sub> model for the multivariate time series forecasting problem.

This paper aims to improve the accuracy of electricity load forecasting by the AN-BEATS model. We made some changes to the orthonormal basis of the Discrete Fourier Transform in the seasonality blocks of the N-BEATS model. In addition, we also considered the influence of block order on the accuracy of electricity load forecasting. Concretely, we create combinations of five blocks that include three trend blocks and two seasonality blocks. Then, for each order obtained, we train the model and evaluate the results of the model error to find the best model.

The left structure of the paper is as follows: Sect. 2 explains the methodology of the N-BEATS model, AN-BEATS's Seasonality blocks, and the Proposed model. The experiments and results have been presented in Sect. 3. The final of this paper is Sect. 4 that provides the conclusion and discussions.

## 2 Methodology

### 2.1 Problem Statement

We consider the problem of univariate time series forecasting in discrete time. Given the observed series history  $[y_1, y_2, \dots, y_T] \in \mathbb{R}^T$ , we need to predict the vector of future values, denoted as  $\hat{\mathbf{y}} = [\hat{y}_{T+1}, \hat{y}_{T+2}, \dots, \hat{y}_{T+N}] \in \mathbb{R}^N$ . The model can be shown as follow:

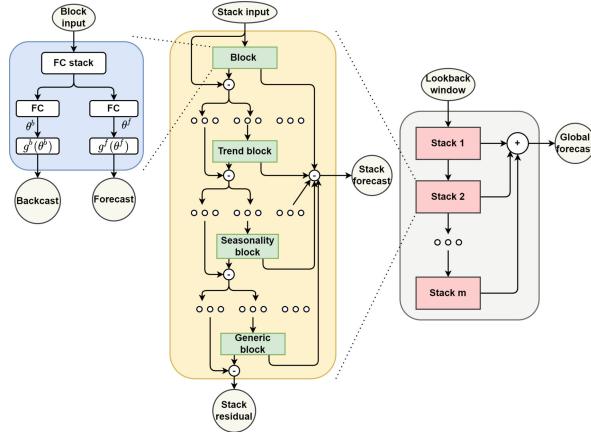
$$\hat{y}_{T+1}, \hat{y}_{T+2}, \dots, \hat{y}_{T+N} = F(y_1, y_2, \dots, y_T) \quad (1)$$

where  $F$  is a model we aim to learn.

### 2.2 N-BEATS

The N-BEATS's structure proposed by the authors in this paper [11]. N-BEATS model consists of stacked blocks which to form stacks. The basic building block includes two-part. The first part is a multi-layer fully connected model with a Relu activation function. Then, we obtain two outputs forward  $\theta^f$  and backward  $\theta^b$ . The second part consists of the backward  $g_l^b$  and the forward  $g_l^f$  basis layers that accept the respective  $\theta_l^f$  and  $\theta_l^b$  expansion coefficient. The idea of double residual stacking is used to arrange the blocks into stacks.

The structure of the model is shown in the following Fig. 1:

**Fig. 1.** N-BEATS architecture

### 2.3 AN-BEATS's Seasonality Blocks

Adjusted Neural Basic Expand - Analysis Time series model called AN-BEATS is proposed by adjusting seasonality blocks in this section. To model seasonality,  $g_{s,l}^b$  and  $g_{s,l}^f$  is periodic functions. Given a length-N forecast horizon, according to the [11], Fourier representation is expressed in the following formula:

$$\hat{\mathbf{y}}_{s,l} = \sum_{i=0}^{\lfloor N/2-1 \rfloor} \theta_{s,l,i}^f \cos(2\pi f_i t) + \theta_{s,l,i+\lfloor N/2 \rfloor}^f \sin(2\pi f_i t) \quad (2)$$

Then, they came up with the following seasonal forecast form:

$$\hat{\mathbf{y}}_{s,l}^{seasonal} = \mathbf{S} \theta_{s,l}^f \quad (3)$$

where  $\theta_{s,l}^f = [\theta_{s,l,0}^f, \theta_{s,l,1}^f, \dots, \theta_{s,l,N-1}^f]^T$  are predicted by a FC network of layer  $l$  of stack  $s$ .  $\mathbf{S} = [1, \cos \frac{2\pi}{N} t, \dots, \cos \lfloor \frac{N}{2} - 1 \rfloor \frac{2\pi}{N} t, \sin \frac{2\pi}{N} t, \dots, \sin \lfloor \frac{N}{2} - 1 \rfloor \frac{2\pi}{N} t]$  is the matrix of sinusoidal waveforms, and  $f_i = \frac{i}{N}$  is the  $i^{th}$  harmonic of the fundamental frequency  $\frac{1}{n}$ .

We can easily see that if  $N$  is an odd number, the Eq. 2 has only  $N - 1$  terms and we are missing  $\cos(\frac{N-1}{2} \cdot \frac{2\pi}{N} t) = \cos(\frac{N-1}{N} \pi t)$  component. In addition, although the Eq. 2 has enough  $N$  terms, we are still missing  $\cos(\frac{N}{2} \cdot \frac{2\pi}{N} t) = \cos \pi t$  component. That leads to the accuracy of the forecasting model can decrease. So, we have made some changes to get the original normal basis of the Discrete Fourier Transform as follows:

If  $N$  is old, we represent  $\hat{y}_{s,l}$  in the following formula:

$$\hat{y}_{s,l} = \theta_{s,l,0} + \sum_{i=1}^{(N-1)/2} \theta_{s,l,i}^f \cos(2\pi f_i t) + \theta_{s,l,i+(N+1)/2}^f \sin(2\pi f_i t) \quad (4)$$

If  $N$  is even, we represent  $\hat{y}_{s,l}$  in the following formula:

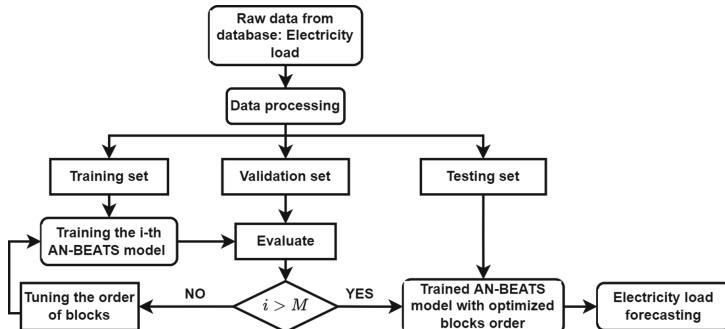
$$\hat{y}_{s,l} = \sum_{i=0}^{N/2} \theta_{s,l,i}^f \cos(2\pi f_i t) + \sum_{i=1}^{N/2-1} \theta_{s,l,i+N/2}^f \sin(2\pi f_i t) \quad (5)$$

Thus, the matrix  $\mathbf{S}$  in the Eq. 3 is also changed.

## 2.4 Proposed Method for Load Forecasting

To solve the electricity load problem with the AN-BEATS model, we have proposed a model as follows (see Fig. 2):

- Step 1: The time-series data is pre-processed and divided into training, validation, and test sets.
  - Step 2: Train the AN-BEATS model on the training set.
  - Step 3: Evaluate the model on the validation set to choose the most optimal parameters set.
- We have all  $M = \frac{(a+b)!}{a!b!}$  combinations of the order of blocks (with  $a, b$  are the number of trend blocks and the number of seasonality blocks, respectively). We train the models with different block orders in turn from  $1^{st}, 2^{nd}, \dots, i^{th}$  combination, until we run out of  $M$  combinations. Thus, after evaluating step 3, we continue to turn the order of blocks and perform step 2 again.
- Step 4: After training  $M$  models, we choose the model with the best performance. The final model with optimized blocks order is used to forecast the testing set.



**Fig. 2.** Proposed model

### 3 Experiments and Results

#### 3.1 Description of Data Set

The data used for the experiment is the Vietnamese National electricity load data set. The time series start from 01/01/2015 to 05/09/2020. It includes 58560 data points (one-hour resolution). Table 1 describes how to divide the data set for training and testing. In that experiment, we predicted the load value for two days ( $N = 48$ ) from the previous ten days ( $T = 240$ ).

**Table 1.** Experiment data set

Data set	From	To	Number of data points
Train	01/01/2015	24/11/2019	52560
Validation	25/11/2019	07/04/2020	2400
Test	08/04/2020	05/09/2020	3600

#### 3.2 Evaluation Criteria

Root mean square error (RMSE) and mean absolute percentage error (MAPE) are used to evaluate the performance of the models. These formulas of criteria are as follows:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|y(t) - \hat{y}(t)|}{y(t)} \quad RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y(t) - \hat{y}(t))^2} \quad (6)$$

#### 3.3 Experimental Scenarios

Three experimental scenarios will be used for comparison:

1. Scenario 1: LSTM model with different hyper-parameters is used for forecasting.
2. Scenario 2: N-BEATS consists of three trend blocks ( $a = 3$ ), two seasonality blocks ( $b = 2$ ), and one generic block with the different trend and seasonality blocks order.
3. Scenario 3: The proposed model is the AN-BEATS model with the number of blocks and changes order similar to N-BEATS in scenario 2 with the new seasonality blocks.

The models in scenarios trained using Adam optimizer, RMSE for the loss function, were trained on 100 epochs and batch size 512. We also applied early stopping to stop training if the validation loss doesn't decrease for 15 consecutive epochs. The best parameters set for each model are selected through RMSE on the validation set with the most optimal value. Finally, the model is evaluated on the test set. We train and inference with NVIDIA GTX 1080Ti 12 GB.

### 3.4 Results

In scenario 1, we fine-tune hyper-parameters: number of layers (Num layers), number of neurons (NN), and learning rate (LR). In scenario 2 and scenario 3, we change the order of trend blocks (T), seasonality blocks (S) and keep the last block as a generic block (G). There are  $\frac{(a+b)!}{a!b!} = \frac{5!}{3!2!} = 10$  models in each scenario 2 and scenario 3. Details of the order of the blocks are depicted in Table 3. In scenarios 2 and 3, we train models with a learning rate is 0.001. For each model in all scenarios, we train and evaluate it 30 times. We then average the results that include the mean and standard deviation of MAPE, RMSE and show them in Tables 2 and 3.

From Tables 2 and 3, we can see that N-BEATS and AN-BEATS model gave superior results when compared with the LSTM model in scenario 1. The LSTM models in scenario 1 also have a longer training time. In addition, the MAPE is above the threshold of 7%, and the RMSE is above the threshold of 6000 (MW). Besides, we see that the order of the blocks also affects the performance of the model N-BEATS and AN-BEATS for the load forecasting problem. With the original order blocks (T, T, T, S, S, G), RMSE and MAPE of the N-BEATS model are 3274 (MW), and 3.79% and RMSE and MAPE of AN-BEATS are 3306 (MW) and 3.71 respectively. Meanwhile, both N-BEATS and AN-BEATS got the best results with the block order (T, S, T, S, T, G). With the block order (T, S, T, S, T, G), the N-BEATS model has RMSE of 3258 and MAPE of 3.63. Besides, RMSE and MAPE of the AN-BEATS model are 3179 and 3.54, respectively. We see that with some orders, the performance of N-BEATS and AN-BEATS decreases. From Table 3, we figure out that for some blocks order, AN-BEATS achieves better results than N-BEATS while for others is worse results. The main reason that interprets AN-BEATS does not improve much compared to N-BEATS is that the load series is pretty clean. Furthermore, the N-BEATS model with the old seasonality blocks can still capture the seasonality of the data even though it is the missing basis in the space like Formula 2.

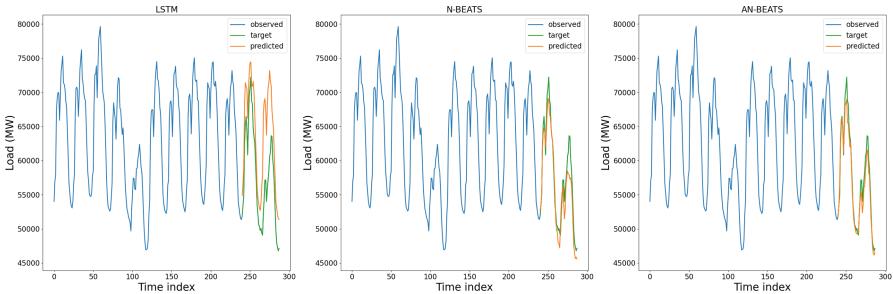
**Table 2.** Results of scenario 1: LSTM with different hyper-parameters

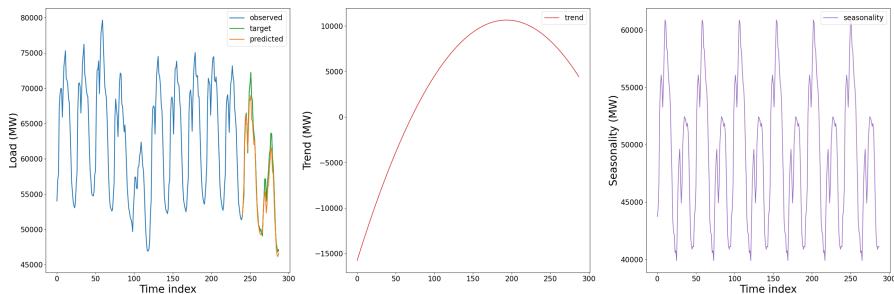
Num layers	NN	LR	MAPE	RMSE (MW)	Training time (s)
1	128	0.001	$7.71 \pm 0.13$	$6474 \pm 76$	2500
1	256	0.001	$7.82 \pm 0.11$	$6490 \pm 57$	2610
1	128	0.002	$7.85 \pm 0.15$	$6531 \pm 80$	2451
1	256	0.002	$7.94 \pm 0.13$	$6598 \pm 110$	2605
2	128	0.001	$7.98 \pm 0.29$	$6605 \pm 197$	2596
2	256	0.001	$8.15 \pm 0.14$	$6722 \pm 42$	2919
<b>2</b>	<b>128</b>	<b>0.002</b>	<b><math>7.62 \pm 0.03</math></b>	<b><math>6409 \pm 120</math></b>	2561
2	256	0.002	$7.69 \pm 0.13$	$6453 \pm 78$	2925

**Table 3.** Results of scenario 2 and scenario 3

Order of blocks						Scenario 2: N-BEATS			Scenario 3: AN-BEATS		
1	2	3	4	5	6	MAPE	RMSE	Training time	MAPE	RMSE	Training time
T	T	T	S	S	G	3.71 ± 0.12	3306 ± 43	1130	3.79 ± 0.01	3274 ± 23	1029
T	T	S	S	T	G	3.78 ± 0.04	3345 ± 21	1085	3.98 ± 0.06	3401 ± 62	1077
T	T	S	T	S	G	3.85 ± 0.02	3354 ± 34	1072	3.91 ± 0.02	3323 ± 40	1068
T	S	S	T	T	G	3.89 ± 0.01	3402 ± 17	1059	4.07 ± 0.04	3434 ± 19	1021
<b>T</b>	<b>S</b>	<b>T</b>	<b>S</b>	<b>T</b>	<b>G</b>	<b>3.63 ± 0.03</b>	<b>3258 ± 30</b>	1038	<b>3.54 ± 0.02</b>	<b>3179 ± 23</b>	1027
T	S	T	T	S	G	3.77 ± 0.06	3343 ± 49	1048	3.66 ± 0.03	3254 ± 10	1009
S	S	T	T	T	G	3.80 ± 0.05	3377 ± 46	1065	3.75 ± 0.03	3277 ± 32	1014
S	T	S	T	T	G	3.71 ± 0.07	3314 ± 43	1071	3.86 ± 0.02	3381 ± 11	1040
S	T	T	S	T	G	3.65 ± 0.05	3272 ± 25	1059	3.60 ± 0.01	3238 ± 18	1028
S	T	T	T	S	G	3.69 ± 0.04	3289 ± 36	1068	3.59 ± 0.03	3268 ± 47	1051

Figure 3 shows an example of hourly load prediction of three models LSTM, N-BEATS, and AN-BEATS (with best parameters from Table 2 and Table 3). The models will observe the load values from 240 h to predict the next 48-h load values. In general, the AN-BEATS model will give the best results in most of the examples in the test set. Figure 4 shows an example of the interpretable AN-BEATS model (with best parameters in Table 3) based on trend and seasonality components. The left chart in the figure shows the original load series and the prediction results of the AN-BEATS model. The center chart shows the trend component, and the right chart shows the seasonality component. The results display polynomial trend components in the data and evident seasonal effects within the days of the load series.

**Fig. 3.** Results of LSTM (left), N-BEATS (center) and AN-BEATS (right)



**Fig. 4.** Example of the interpretable AN-BEATS model

## 4 Conclusion and Discussion

If load forecasting is precise and reliable, power system operations can be safer and more cost-effective. This work introduces AN-BEATS, a new model for short-term load forecasting that focuses on adjusting seasonality blocks and optimization block order. The following are some of the paper's findings:

- Adjusting seasonality blocks to archive AN-BEATS.
- Successfully proposing a new model AN-BEATS with optimization block order.
- Using AN-BEATS for short-term forecasting.
- Applying the proposed approach for forecasting short-term electricity load in the National Load Dispatch Centre of Vietnam Electricity.

In the future, we intend to solve the challenge to optimize the number of blocks in the model, the number of trend blocks, and the order blocks of the AN-BEATS models. Furthermore, because the new AN-BEATS model can only predict with a single time series, expanding AN-BEATS to multiple time series will be investigated.

**Acknowledgments.** This work was supported by the Hanoi University of Science and Technology and the National Load Dispatch Centre of Vietnam Electricity.

## References

1. Oreshkin, B.N., Dudek, G., Pelka, P., Turkina, E.: N-beats neural network for mid-term electricity load forecasting. *Appl. Energy* **293**, 116918 (2021)
2. Serrano-Guerrero, X., Briceño-León, M., Clairand, J.-M., Escrivá-Escrivá, G.: A new interval prediction methodology for short-term electric load forecasting based on pattern recognition. *Appl. Energy* **297**, 117173 (2021)
3. He, Y., Liu, R., Li, H., Wang, S., Lu, X.: Short-term power load probability density forecasting method using kernel-based support vector quantile regression and copula theory. *Appl. Energy* **185**, 254–266 (2017)

4. Zeng, N., Zhang, H., Liu, W., Liang, J., Alsaadi, F.E.: A switching delayed PSO optimized extreme learning machine for short-term load forecasting. *Neurocomputing* **240**, 175–182 (2017)
5. Song, L., Lalit, G., Peng, W.: An ensemble approach for short-term load forecasting by extreme learning machine. *Appl. Energy* **170**, 22–29 (2016)
6. Atef, S., Eltawil, A.B.: Assessment of stacked unidirectional and bidirectional long short-term memory networks for electricity load forecasting. *Electr. Power Syst. Res.* **187**, 106489 (2020)
7. Iwok, I.: Seasonal modelling of Fourier series with linear trend. *Int. J. Stat. Prob.* **5**, 65 (2016)
8. Nguyen, L., Novák, V.: Forecasting seasonal time series based on fuzzy techniques. *Fuzzy Sets Syst.* **361**, 114–129 (2019). Theme: Clustering and Rule-based Models
9. Moreno, M., Novales, A., Platania, F.: Long-term swings and seasonality in energy markets. *Eur. J. Oper. Res.* **279**(3), 1011–1023 (2019)
10. Apaydin, H., Sattari, M.T., Falsafian, K., Prasad, R.: Artificial intelligence modelling integrated with singular spectral analysis and seasonal trend decomposition using loess approaches for streamflow predictions. *J. Hydrol.* **600**, 126506 (2021)
11. Oreshkin, B.N., Carpow, D., Chapados, N., Bengio, Y.: N-beats: neural basis expansion analysis for interpretable time series forecasting. In: International Conference on Learning Representations (2020)
12. Sbrana, A., Rossi, A.L.D., Naldi, M.C.: N-beats-RNN: deep learning for time series forecasting. In: 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA) (2020)
13. Olivares, K.G., Challu, C., Marcjasz, G., Weron, R., Dubrawski, A.: Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with NBEATSx. *Int. J. Forecast.* submitted **Working Paper version available at arXiv:2104.05522** (2021)



# Non-contact Thermometer System for Body Temperature Measurement at Hospitals

Nguyen Phan Kien<sup>1</sup>(✉) , Than Tra My<sup>1</sup>, Tran Anh Vu<sup>1</sup>, and Nguyen Thanh Trung<sup>2</sup>

<sup>1</sup> Hanoi University of Science and Technology, Hanoi, Vietnam

kien.nguyenphan@hust.edu.vn

<sup>2</sup> VietDuc Hospital, Hanoi, Vietnam

**Abstract.** At the end of 2019, a new virus called Covid-19 has spread globally. The pandemic situation is evolving very quickly and complicated, requiring governments to deploy measures to prevent and control the Covid epidemic in a synchronous and widespread way throughout the entire territory. Currently, the most widely used method of mass screening is temperature screening because people infected with Covid will have a higher-than-normal body temperature compared to healthy people. This solution is the fastest and has the lowest cost. However, this is mainly done manually and causes a very high risk of cross-infection if the controller accidentally comes into close contact with F0 and gets infected. To solve this problem, an automatic, non-contact temperature measurement system has been researched and developed. The system will notice users then measure their temperature, warn if the person's temperature exceeds the allowable threshold, from which measures will be taken according to the regulations of the Ministry of Health. After fully researched and designed, the system is demonstrated in many hospitals and events all over Vietnam and proves its ability to ensure continuous operation for a long time, with a local warning function by display, light or buzzer. In the future, we hope that this system will be used in everywhere and become the safest, most common and effective way to prevent the spread of Covid-19.

**Keywords:** COVID—19 · Automatic measurement · Non-contact thermometer · Body temperature

## 1 Introduction

According to the World Health Organization (WHO) statistics, the total cases of COVID-19 all over the world until 5th December 2021 are 265,194,191 cases with 5,254,116 deaths and this statistics in Vietnam are 1,309,092 and 26,260 respectively [1]. Within 2 last quarters of 2021, Vietnam suffers from the biggest wave of COVID-19 due to Delta variant with 39,132 new cases/day at peak [2]. The new variant spreads more than 2 times faster and is much more dangerous in comparison with others previous ones because it can also attack vaccinated people [3]. There are many methods to prevent and control the spread of the virus and the first one that came to mind of all the world leaders is quarantine. However, this attempt must not last long and governments try to bring the

situation back to normal like before the pandemic as soon as possible. Therefore, strict mass screening and early detection are the most crucial parts to keep the situation under control [4].

With the rapid development of science and technology, there are many methods applied to detect suspected infected people such as rapid antigen test or PCR test. However, the most common symptom being used for COVID-19 screening is the raise of body temperature [5] since measuring body temperature is the fastest and most comfortable method, as well as has the most reasonable price. Before entrance of any facility, everyone's body temperature must be measured and only ones who qualified should pass the door. Otherwise, suitable measures will be taken based on the protocol and instructions of the Ministry of Health. However, this method still has a big drawback because it is currently manual therefore it is very dangerous if the thermometer controller comes into close contact with F0 person and get infected. Under these circumstances, we will have a lot of F1 in the process of incubation [6].

As a result, many inventors all over the world have rapidly developed automated temperature measurement systems to minimize cross-contamination in the community. A research [7] proposed a body temperature checking system integrated with automatic door lock web. This tool is designed to use the MLX90614-DCI sensor as a temperature sensor which will be displayed on the LCD if the temperature is normal then the relay will cut off electricity on the electromagnetic door so that the door lock will open and vice versa. However, this design can only be used indoor and there still have difference of 0.3 °C in comparison with conventional infrared thermometers. Research [8] has the same method as [7] but it is also connected to a wireless-control system which allow users to operate it remotely. However, this system will be dependent on internet connection, and it also has low accuracy between 2 sensors. Therefore, [9] developed a non-contact integrated body-ambient temperature sensor with attempt to increase the accuracy of body temperature measurements. The information relative to the ambient temperature is successfully exploited to derive a correction formula for the accurate extraction of body temperature from the measurement provided by the standard infrared sensor. This method achieved an average relative error of 0.47 °C. Since this design requires an algorithm to calibrate the temperature sensor based on ambient temperature, the result may vary from time to time.

Therefore, in this paper, we want to design a system that is totally automatic, do not need operation from human, even from afar and the system must have the accuracy of the best type of standard infrared thermometer without errors.

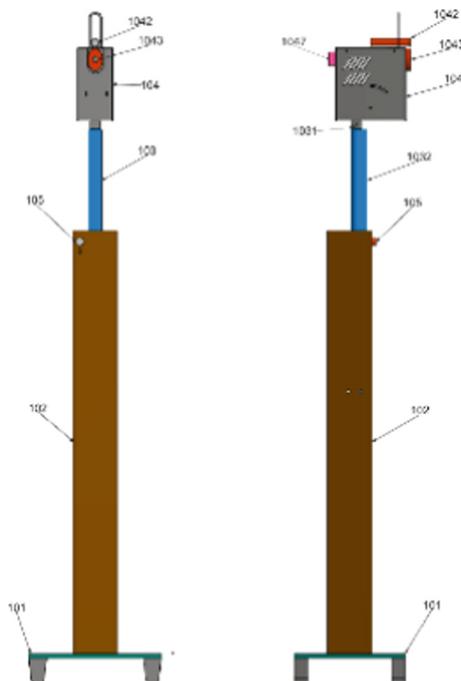
## 2 Materials and Methods

### 2.1 System Overview

As presented, the design idea is a robot capable of automatically pulling the trigger of the forehead infrared thermometer and automatically locate the forehead area to measure the temperature for the human body. Therefore, basically, the device consists of the following parts:

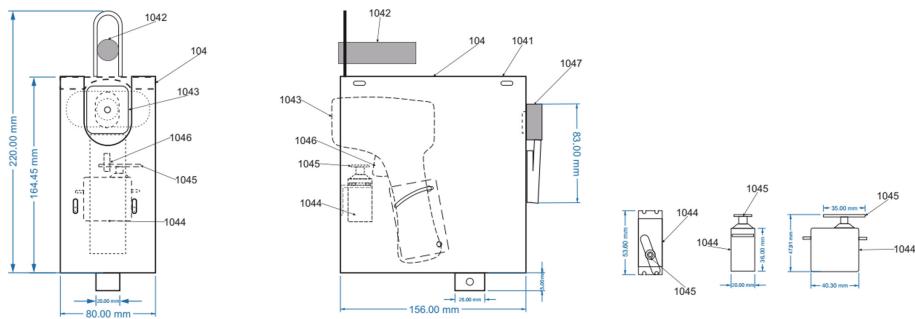
- Mechanical system: Allows lifting to determine the height of the person automatically along with the triggering mechanism system, allowing the system to automatically adjust the heat trigger when the required height position has been reached.
- Electronic control system: Allows the human height sensor to function, receive the signal, locate the right position, and perform the squeeze control
- Display unit of the system: Allows to display the temperature of the infrared thermometer to a larger screen.

## 2.2 Mechanic System

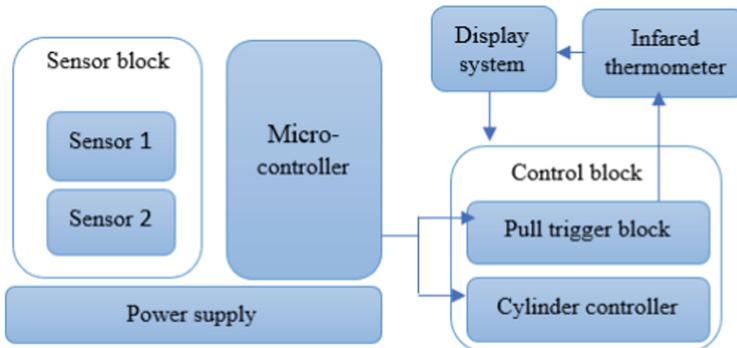


**Fig. 1.** Whole system

Figure 1 shows the whole system. According to [10], the average height of the Vietnamese is 162 cm with male and 151.8 cm with female, compared with highest people in the world with the height around 2 m, we need to choose the electric cylinder with the maximum duration is 50 cm. Figure 1 describes the mechanic structure of the system which includes base part with a height of 145 cm and a maximum travel height of the cylinder – 50 cm to satisfy the requirements. In addition, the inside of the mechanic top part also contains electronic circuits and a infrared thermometer, a camera as well as an automatic triggering system that is described in Fig. 2.



**Fig. 2.** Inside of the top-box



**Fig. 3.** Block diagram of the system

### 2.3 Electronic System

The block diagram of the system is presented in Fig. 3 including: Sensor block: there are 02 main functions that are Human Sensors (Sensor 1): allows to identify people entering the measuring position to perform height control and Height Sensor (Sensor 2): that determines the height of the person. The operating structure control block consists of 02 functions: The electric cylinder control block helps control the electric cylinder going up or down in accordance with the height of the meter and squeezes the heat gauge trigger to pull the trigger to measure the temperature. The display block will show the patient's measured temperature to the computer screen to obtain the results. Finally, the Power Block has the function of powering the circuits, the parts are integrated on the device. The power block can get input from a 220 V grid power source through a 12 V adapter.

#### Sensor Block

The sensor block has the function of identifying the person entering and determining the height of the person entering the measurement. We select a sensor with a transceiver reflector (infrared sensor) E3F-DS30C4 that works as follows:

- For sensor 1, it picks up the reflected signal from the obstruction. Therefore, when someone enters, the sensor will be blocked by people, there is a reflex signal that will recognize whether there is a person or not.
- With the same principle with sensor 1, sensor 2 detects the forehead area. When the sensor has no reflexes, it means that the sensor is above the human head. When the sensor changes state, it means that the sensor has reached the top of the meter's head. If the sensor has a reflex, it means that the sensor is below the height of the person. The up or down direction of the sensor will be able to be selected in accordance with the state of the sensor.

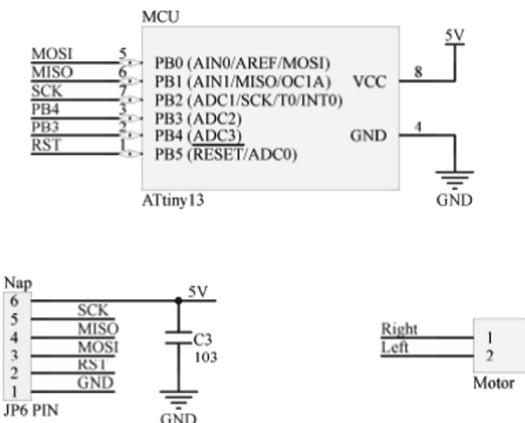
### **Control Block**

*The Electric Cylinder Control Circuit.* Electric cylinders move up and down by reversing current. For the convenience of designing as well as meeting the continuous operating capacity of the machine, the electric cylinder control block uses a 5 Voltage 5-pin relay. The principle of relay is “using a small energy to shut down the larger power source”. If the connection of the electric current is correct, the cylinder goes down. While connected in an opposite way, the electric cylinder runs up. The electric cylinder has a cruise switch built into both ends so that the cylinder will automatically stop at the end of the journey.

*The Trigger-pulling Control Circuit.* To measure the temperature, the system uses an infrared UX-A-01 forehead thermometer that uses high-precision infrared sensors and the latest smart chip. The system can pull the trigger automatically without interfering with the inside of the infrared thermometer, in the subject of a solution that uses a step motor that can control rotation at will based on PWM pulses. The MG966R engine used is an engine with a rotation angle of 0 to 180°. The rotation angle of the motor can be customized at will, so that the gears of the engine will be programmed to hit the trigger automatically.

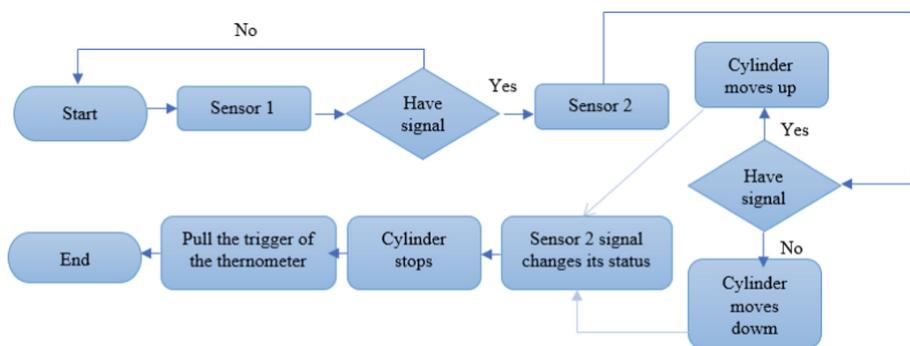
### **Microprocessor**

The AVR Attiny 13 processor used in the circuit is responsible for receiving signal from 02 sensors, exporting control signals to 02 outputs to control the lifting system with electric cylinders and the triggering system. Figure 4 describes the functional diagram of AVR Attiny 13 in which the PB1, PB2 pin of the Attiny 13 processor will be the input signal pin, which is connected to the signal pin (black wire) of sensor 1, sensor 2. PB0, PB3, PB4 pins are output signal and output pins, in which PB0, PB4 pins assume the function of controlling electric cylinders up and down, PB3 legs controlling the engine pulling the trigger of an electronic infrared thermometer.



**Fig. 4.** Microprocessor AVR Attiny 13 functional diagram

The heat measurement algorithm is carried out in the following steps which are described in Fig. 5:



**Fig. 5.** Algorithm flowchart of the system

## Power Source

The device uses an input power source with a voltage of 12 V, which is converted into a 5 V voltage that supplies the components of the electronic circuit. The design uses IC LM7805 which converts 12 V voltage into 5 V voltage.

## Display System

The infrared thermometer will display the user's temperature with an indicator light on its screen. A camera then captures directly from the heat gauge and extracts it from the computer screen so that the medical staff or person responsible for observing the temperature of the person being measured. Because the camera is located very close to the screen of the infrared thermometer, we need to choose the type that can adjust the focus and give the sharpest image possible on the computer screen. Finally, the X5

camera was chosen thanks to many advantages and most importantly, the ability to adjust the focus itself.

### 3 Conclusion and Discussion

The equipment after completion has fully met technically all the functions that are set out in this topic. In addition, the device will be upgraded in the next version with several new functions such as: Extracting the temperature value connected to the electronic medical statement, integrating BMI calculation function, etc. In comparison with other automatic non-contact thermometers sold in the market, the device is outstanding because it is certificated by the Ministry of Health. More than that, other devices already on the market use camera technology that measures infrared or thermal radiation. This will lead to big error (approximately 0.5° error) as well as higher price (10 to 100 times higher than the device that we have been developing).

Up to now, the equipment has been deployed and installed in a few units such as: Central Traditional Medicine Hospital, Tuyen Quang General Hospital, Quoc Oai Hospital, etc. At the same time, contactless heat measuring equipment is also trusted and installed by the Central Office of the Ho Chi Minh Communist Youth Union for the events of the Central Committee such as: Young Talent Congress, Vietnam Young Doctors Congress and most recently, the 90th anniversary of the founding of the Ho Chi Minh Communist Youth Union on March 26. At the event, four devices measured and controlled the temperature for about 1,000 delegates. In fact, the device is stable, time temperature fast, simple use, and received a lot of positive reviews from users.

### References

1. WHO Homepage: <https://covid19.who.int/>. Accessed 07 Dec 2021
2. WHO Vietnam page: <https://covid19.who.int/region/wpro/country/vn/>. Accessed 07 Dec 2021
3. Li, B., Deng, A.: Viral infection and transmission in a large, well-traced outbreak caused by the SARS-CoV-2 Delta variant. (2021)
4. Johanna, N., Citrawijaya, H., Wangge, G.: Mass screening vs lockdown vs combination of both to control COVID-19: A systematic review. J Public Health Res. 9(4), 2020 (2011)
5. Goggins, K.A., Tetzlaff, E.J., Young, W.W., Godwin, A.A.: SARS-CoV-2 (Covid-19) workplace temperature screening: Seasonal concerns for thermal detection in northern regions. Appl. Ergon. Vol. 98 (2021)
6. Lippi, G. et. al.: Is body temperature mass screening a reliable and safe option for preventing COVID-19 spread? Diagnosis (2021)
7. Supriana, A., Prakosa, B.A., Ritzkal, R.: Implementation of body temperature checking system with automatic door lock web and Arduino assistance. Jurnal Mantik (2021)
8. Rahimoon A.A., Abdullah, M.N., Taib, I.: Design of a contactless body temperature measurement system using Arduino. Indonesian J. Elect. Eng. Com. Sci. (2020)
9. Costanzo, S., Flores, A.: A non-contact integrated body-ambient temperature sensors platform to contrast COVID-19. Electronics, 9 (10), 2020
10. Venkataramani, A.S.: The intergenerational transmission of height: Evidence from rural Vietnam. Health Economics (2010)



# Classification of 40 Hand Movements Based on Surface Electromyogram Signal

Thuy Le – Nguyen Thi<sup>1,2</sup>, Thuan Nguyen Phan<sup>1,2(✉)</sup>, and Tuan Van Huynh<sup>1,2(✉)</sup>

<sup>1</sup> Department of Physics and Computer Science, Faculty of Physics and Engineering Physics, University of Science, Ho Chi Minh City, Vietnam

{pnthuan, hvtuan}@hcmus.edu.vn

<sup>2</sup> Vietnam National University, Ho Chi Minh City, Vietnam

**Abstract.** Surface electromyogram (sEMG) signals are bioelectric signals recorded on the skin surface and formed by physiological variations during muscle contraction. They contain valuable information essential in many fields, such as medical research, sports training, and rehabilitation. One of the main ways to develop the prosthesis application supporting post-surgery and accident patients is using these signals as the control command of the system. As a result, the processing and classification of sEMG signals become the central issue of these applications. This study focused on hand and wrist movements recognition using sEMG signals. A publicly available dataset from the Ninapro project was used to train and evaluate the proposed system. We chose database DB7 comprising records from 20 able-bodied and two amputee subjects performing 40 hand gestures. Firstly, we applied the windowing technique to divide the EMG signals into various segments. Then, a set of time features was extracted from these segments. After that, two dimensionality reduction methods were used to minimize the computation cost: Principal Component Analysis and Bhattacharyya Distance. In the end, the features were fed to two simple classification models: an Artificial Neural Network and a Fuzzy Logic. Our two proposed models achieved an overall classification performance of 98.23 ( $\pm 0.02\%$ ) and 89.29 ( $\pm 2.55\%$ ), respectively.

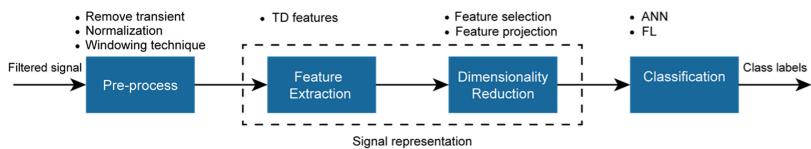
**Keywords:** Electromyogram signals · Ninapro · ANN · Fuzzy logic · Time features · Hand motion classification

## 1 Introduction

The surface electromyogram signals, recorded on the skin's surface, are biomedical signals emanated by muscle fiber in the contraction process. They can be established as an evaluation tool to measure neuromuscular activities. This is fundamental to their use in clinical diagnosis and controlling prosthesis devices for amputees. The condition of the control is that amputees still have the phantom of functions. Therefore, they can generate a repeatable EMG pattern corresponding to each function to “train” the system. Recently, scientists have proposed several methods to design the EMG classification system. Although these systems used various feature sets and classification methods, they had a typical sequence. Before feature extraction, the raw signal was subjected to

pre-processing steps (such as data segmentation, filtering, and rectification). For details, data will be segmented into multiple time windows. Then for each divided segment that has been filtered and rectified, information will be extracted and represented in a feature vector. This part had to be considered carefully to minimize control schemes' error and response time. Generally, there are three main categories of features in the analysis of EMG signals involving time-domain (TD) features, frequency domain (FD) features, and time-frequency (TFD) features [1]. After feature extraction, the next step is dimensionality reduction to reduce the complexity of computing while maintaining the system's accuracy. For the last stage, pattern recognition, researchers have developed different kinds of mathematical models and techniques such as K-nearest neighbor (KNN), Bayesian classifier (BC), Support Vector Machine (SVM), Fuzzy Logic (FL), Linear Discriminant Analysis (LDA), and Artificial Neural Networks (ANN).

This study determines to classify forty movements, including hand, wrist, grasping, and functional hand gestures. Our system diagram is shown in Fig. 1, consisting of four stages: pre-process, feature extraction, dimensionality reduction, and classification. The structure of this paper is as follows: Sect. 2 describes the material and methodology used in the proposed models. Section 3 presented experiments of changing parameters in each stage and comparing two classifier's results. In the end, conclusions were provided in Sect. 4.



**Fig. 1.** The pipeline of the proposed EMG-pattern recognition system.

## 2 Materials and Methods

### 2.1 Database

This paper used the Ninapro DB7 dataset from the Ninapro project, one of the largest publicly available sEMG databases for robotic and prosthetic hands controlled with artificial intelligence [2]. This dataset used 12 electrodes channels (Delsys EMG sensors) to record signals. The sampling frequency of signals was set at 2 kHz. This frequency is ideal for preventing signal loss due to the aliasing effects [3]. To remove the 50 Hz power line interference, the Hampel filter was implemented. For signal acquisition, participants were asked to perform a series of 40 movements which were instructed to them on a laptop screen. The record session of each gesture lasted in 8 s, including 5 s contraction and 3 s relaxing. Participants reproduced six times for each hand movement to account for slight variations within the same motion class. One of the amputee volunteers only performed 38 movements due to a power supply failure.

## 2.2 Data Pre-processing

The EMG signal recorded during the muscle contraction process contains two periods: transient and steady. Although the transient period can improve real-time prosthesis control performance, it has been shown to reduce the classifier's offline accuracy [4]. Hence, the transient needs to be removed enough to converse the transient information and minimize the classification error. This can be achieved by removing the beginning and end of signals in each repetition. Then, we applied the normalization process to the cutting signals. Because one major drawback of any EMG analysis is that the amplitude data are strongly affected by the given acquisition condition (it can change significantly depending on subjects and record setup), we need to overcome this uncertainty by normalizing the signals to a reference value [3]. The normalization stage only affects the amplitude and does not change the shape of the data. In this study, we used the maximum value of each movement as the reference value. Finally, we applied the windowing technique to split signals into many segments.

## 2.3 Feature Extraction

Recent pattern classification algorithms using the deep learning approach can take raw signals as input and automatically learn the features. However, these systems generally require a large amount of training data to converge and consume high memory to process. Therefore, it is challenging to run them in embedded systems which limited hardware. This study applied two simple models (Shallow Neural Network and Fuzzy Logic) and used time-domain features to reduce the computing cost. These features include: mean absolute value, zero crossings, slope sign changes, waveform length, energy, root mean square, variance, and standard deviation [5]. They are extracted from each time window for all channels.

## 2.4 Feature Reducing

As reported by Englehart, based on the objective function, there are two types of dimensionality reduction techniques: feature selection and feature projection [1]. This study employed Bhattacharyya distance (BD) as a feature selection approach and Principal Component Analysis as a feature projection method. The BD method uses the separability index (SI) measurement based on Bhattacharyya distance to choose the best feature subset from the original [6]. On the other hand, the PCA method projects the  $d$  initial feature vector into a new space and only selects  $k$  ( $k < d$ ) vectors with large variance values.

## 2.5 Classification

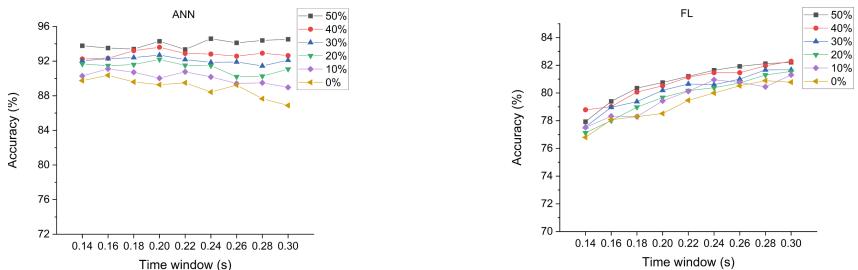
**Artificial Neural Network.** The network proposed in this study was a feedforward network with a shallow architecture: one input layer, one hidden layer, and one output layer. The hidden layer obtains the tansig as the transfer function. The output layer uses the softmax activation function to predict multiple classes. For network training, we implemented the scale conjugate gradient algorithm with the cross-entropy cost

function. Data was used for training, validation, and testing in the proportions of 40%, 20%, and 40%, respectively.

**Fuzzy Logic.** This paper applied a simple fuzzy system with learning capability as a classifier [7]. Firstly, we employed the fuzzy c-means (FCM) clustering to each movement data separately to construct the rule base for recognizing each movement. Then, we used only the winner rule in the inference process for the reasoning method. The classification of a pattern  $x$  is class  $k$  which cluster  $i$  having the largest the membership function value  $m_i(x)$  belongs. After rule base construction, we applied a gradient descent algorithm to optimize the membership function parameters and improve the model accuracy. The details of this model were described in our previous study [5]. We divided data into train and test sets with 60% and 40% ratios, respectively.

### 3 Result and Discussion

#### 3.1 Effect of Window Time and Overlap



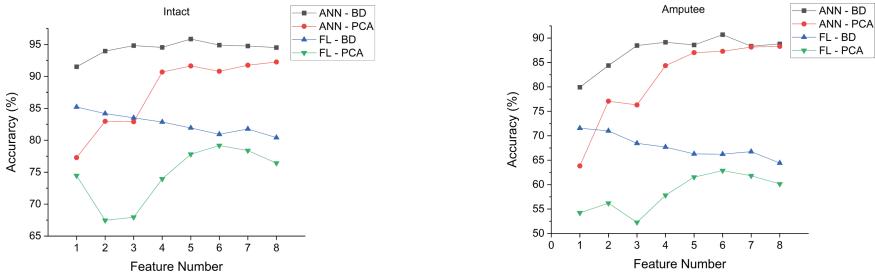
**Fig. 2.** Classification performance with different combinations of the time window and overlap percent for both models ANN and FL.

Large window time and overlap portion might extract more information from signals and increase the performance of models. However, this also increases the system's response time and computation cost. For real-time prosthesis applications, these parameters need to tune effectively to improve the system's recognition rate and response time. In this experiment, the impact of these parameters on the achieved recognition rate was investigated. The window lengths ranged from 0.14 to 0.30 s, and the overlap percentage ranged from 0 to 50%. The overall results from this investigation for all subjects with two proposed models are shown in Fig. 2.

As demonstrated in Fig. 2, choosing a large window size for the ANN model does not show a notable improvement in classification accuracy. On the other hand, the larger the window size is, the higher the performance gain is achieved by the FL model. About the effect of overlap, the results of the two models imply that high overlap would minimize the recognition errors and best performance achieved at 50%. Thus, we selected 50% as overlap percent for both models in the next stage. We set the window size at 0.24 s and 0.3 s for ANN and FL models, respectively.

### 3.2 Effect of Dimensionality Reduction

After the previous experiment, we selected the suitable window size and overlap percentage, which led to the best performance for each model. Next, this experiment investigated how effective feature selection (BD) and feature projection (PCA) were on two classifiers. As mentioned above, we extracted eight time-domain features from each segment; thus, we will reduce the feature set from 8 to 1 and evaluate the recognition rates (see Fig. 3).



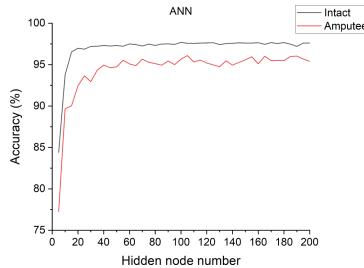
**Fig. 3.** Classification accuracy versus feature number for two dimension reduction approaches (PCA and BD) for both ANN and FL models with intact and amputee groups.

Figure 3 implies a similar trend between the intact and amputee group when applying two feature reducing methods with two proposed models. The best average percentage of correctness was associated with the BD method for both models and all subjects. Based on this result, we suspect that the PCA method might not keep enough information to distinguish the subtle differences between 40 movements; thus, it decreased the classification accuracy. On the other hand, the BD method improved the performance by about 2 and 5% for ANN and FL models. Therefore, we applied this method to continue with the following experiment.

### 3.3 Effect of the Classification Structure

**Artificial Neural Network.** According to the experiment above (see Fig. 3), the overall classification accuracy was 95 and 90% for the intact and amputee groups. This performance was achieved with only one hidden layer. Therefore, increasing the hidden layer is not necessary. Based on this reason, the proposed network model is the Shallow Neural Network instead of a Deep Neural Network (more than one hidden layer), which is more complicated and takes more hardware resources to train. In this experiment, we varied the hidden node number from 5 to 200 and then selected the most appropriate number, leading to high accuracy for each participant.

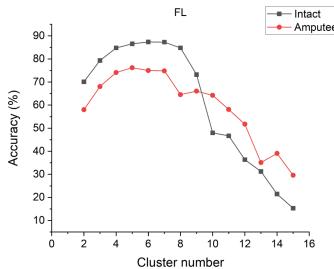
Figure 4 depicts the impact of hidden nodes number on classifiers' recognition rate, averaged across intact and amputee subjects. In the beginning, from 5–20 hidden nodes, the performance tended to rise rapidly by about 12%, and then it remained steady till the end of the axis. The best performance was 97% at 60 hidden nodes for intact subjects. The network requires more hidden nodes for the amputee group to achieve the highest



**Fig. 4.** Experiment result on ANN model using different hidden nodes number.

accuracy (95% with 100 hidden nodes). This result shows that increasing the hidden nodes does not result in a better outcome. However, choosing the number of hidden nodes is a vital stage before training a network model. If the hidden nodes number is too low, the network will have trouble tuning weight parameters, which might result in underfitting. In contrast, if it is too high, the network will require more resources to execute while the performance is not improved significantly.

**Fuzzy Logic.** As mentioned above, we obtained the fuzzy c-mean clustering to construct the rule base. Therefore, the rule number produced was the same as the cluster center number. In this experiment, we varied the cluster centers number for each movement from 2 to 15 and examined the effect of cluster number on the recognition rate of the model across all subjects, see Fig. 5.



**Fig. 5.** Performance of FL model across all subjects on different cluster numbers.

Based on Fig. 5, we suspect that using a large number of cluster centers to build the inference rule base produces redundant rules and lowers classification performance. There is no more information captured from data when increasing the cluster number. In contrast, having too many clusters while input data is limited would lead the fuzzy c-mean algorithm to struggle to find an appropriate cluster center to distinguish many classes (40 movements). In the end, for each participant, we chose the best cluster center number, which balanced between accuracy and training time. We noticed that the suitable cluster number for all subjects, including intact and amputee groups, was 4 to 7.

**Classification Performance Analysis.** We finally chose the best feature and model parameters of two classifiers (ANN and FL) for each subject from the experiments above. In this part, we evaluated the identification capability of the two proposed classifiers based on four factors: accuracy, precision, recall, and prediction time per sample. The prediction time was the summation of feature extraction and the testing time. Table 1 depicts the overall performance of each model. The standard deviation is represented by the numbers preceded by “ $\pm$ ”.

**Table 1.** Overall performance of each classifier in predicting 40 motions across 22 subjects.

	Accuracy (%)	Recall (%)	Precision (%)	Prediction time/window (ms)
ANN	98.23 $\pm$ 0.02	97.59 $\pm$ 1.53	97.56 $\pm$ 1.57	0.71 $\pm$ 0.02
FL	89.29 $\pm$ 2.55	87.01 $\pm$ 1.04	86.56 $\pm$ 5.80	0.87 $\pm$ 0.03

As demonstrated in Table 1, the FL model did not show its advantage over ANN from all perspectives. This can be explained that the learning capability of FL is not as good as ANN and requires a longer time window to extract features. However, its performance is acceptable when compared with the previous studies (see Table 2). Overall, the accuracy for ANN and FL was 98.23( $\pm 0.02$ )% and 89.29( $\pm 2.55$ )%, respectively.

**Table 2.** Comparison of our proposed models and previous studies with the same dataset.

Author	Feature set	Classifier	Accuracy
Agamemnon Krasoulis [2]	Time set	LDA	Intact – 60%, amputee – 42%
Yu Hu [8]	Time set	HMM	79%
Sri-iesaranusorn [9]	Time set	DNN	Intact – 89%, amputee – 88%
This work	Time set	ANN FL	Intact – 97%, amputee - 95% Intact – 90%, amputee – 77%

Furthermore, we compared our proposed system performance with previous studies. To account for a complete and thorough comparison between various classification methods, we selected the studies using the same dataset and the feature’s type with us. The result of this comparison is shown in Table 2. This result indicated that our proposed system was suitable for hand pattern recognition applications.

## 4 Conclusion

This study proposed forty hand movements recognition models based on 12 - sEMG channels. We extracted time features set from signals and then decreased dimension using PCA and BD. In the end, ANN and FL had overall performances of 98.23 ( $\pm 0.02$ )

and 89.29 ( $\pm 2.55\%$ ), respectively. These results suggest the success of implementing a simple neural network and fuzzy system in classifying a large number of movements. Furthermore, because we used simple models (Shallow Neural Network and Fuzzy Logic), the computing cost and processing time were reduced significantly compared with other deep learning algorithms. Therefore, applying these models in embedded systems of prosthesis applications is more suitable and realistic.

**Acknowledgements.** This research is funded by Vietnam National University, Ho Chi Minh City (VNU-HCM) under grant number C2022-18-12.

## References

1. Englehart, K., Hudgins, B., Parker, P.A., Stevenson, M.: Classification of the myoelectric signal using time-frequency based representations. *Med. Eng. Phys.* **21**(6–7), 431–438 (1999)
2. Krasoulis, A., Kyranou, I., Erden, M.S., Nazarpour, K., Vijayakumar, S.: Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements. *J. NeuroEngineering Rehabil.* (2017)
3. Konrad, P.: The ABC of EMG: A practical introduction to kinesiological electromyography (2005)
4. Hargrove, L., Losier, Y., Lock, B., Englehart, K., Hudgins, B.: A real-time pattern recognition based myoelectric control usability study implemented in a virtual environment. In: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Lyon **2007**, 4842–4845 (2007)
5. Thuy Le - Nguyen Thi, Phuc Viet Ho, Tuan Van Huynh: A study of finger movement classification based on 2-sEMG channels. In: 2020 7th NAFOSTED Conference on Information and Computer Science (NICS), 2020
6. Nilsson, N., Ortiz-Catalan, M.: Estimates of classification complexity for Myoelectric pattern recognition. In: 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, pp. 2682–2687, (2016)
7. Chiu, S.L.: Extracting fuzzy rules from data for function approximation and pattern classification. In: Dubois, D., Prade, H., Yager, R.R. (Eds.) *Fuzzy information engineering: A guide tour of applications*, pp.149\_162 (Chapter 9). Wiley, New York (1997)
8. Hu, Y., Chan, Q.: A comprehensive evaluation of hidden Markov Model for hand movement recognition with surface electromyography. In: 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence Shanghai China, pp. 85–91, (2020)
9. Sri-Iesaranusorn et al.: Classification of 41 hand and wrist movements via surface electromyogram using deep neural network. *Front. Bioeng. Biotechnol.* vol. 9, p. 394 (2021)



# An Optimal Path Planning Issue Parameters with Potential Feasible Area Using Evolutions Algorithm

Trong-The Nguyen<sup>1,2(✉)</sup>, Thi-Xuan-Huong Nguyen<sup>3</sup>, Van-Chieu Do<sup>3</sup>,  
and Truong-Giang Ngo<sup>4</sup>

<sup>1</sup> Multimedia Communications Lab, VNU-HCM,  
University of Information Technology, Ho Chi Minh City, Vietnam  
thent@uit.edu.vn

<sup>2</sup> Vietnam National University, Ho Chi Minh City 700000, Vietnam  
<sup>3</sup> Haiphong University of Management and Technology, Haiphong, Vietnam  
{huong\_ntxh, chieudv}@hpu.edu.vn

<sup>4</sup> Faculty of Computer Science and Engineering, Thuyloi University, Hanoi, Vietnam  
giangnt@tlu.edu.vn

**Abstract.** The traditional artificial potential field method's parameter values are usually determined empirically would cause unsMOOTH planned paths, low calculation efficiency, unreachable goal spots. This study suggests optimization parameters in the artificial potential field in finding the shortest route for mobile robots using the differential evolution algorithm (EA). Three parameters of the positive proportional gain coefficient of the gravitational field, the repulsive field gain coefficient, and the influence distance of obstacles to find the shortest path are considered to modify the infeasible ways of search space in optimizing the algorithm in locating the fastest route. A collision-free path model is constructed by using known global environment information for the mobile robot, and the EA algorithm is used to plan the robot's best path. In the experiment section, the obtained results from the suggestion scheme are compared with the other techniques in the literature. The results show that the suggested approach can produce faster convergence and higher satisfaction quality than the other.

**Keywords:** Path planning · Artificial potential field method · Evolution algorithm · Optimization

## 1 Introduction

A mobile robot is a type of robot that autonomously goes from a starting place to a goal point in the presence of obstacles using environmental awareness and behavior planning control [1]. The goal of paths planning is to find a best path from the beginning point to the target location with collision-free path [2, 3]. The following goals of machine path planning [4] have factors: 1) discover a best path for the machine to travel from its current position to the specific location with no clash with obstacles [5]; 2) comply

with the simulated guideline [6]. The road should be as smooth as feasible, and the turning amplitude should be modest, among other things [7]. Scholars have given the global path planning problem of mobile machines a lot of attention, and solutions for many path planning difficulties have been proposed developments. The potential artificial area or feasible field approach assumes that a virtual potential field force exists in the surroundings that influence the robot's movements [4].

Recent studies have found variant repulsive field models, e.g., Dijkstra [8], A\* algorithm [9], genetic calculations algorithm (GA) [10], and gravitational search algorithm (GSA) [11], for the developed path planning techniques. The higher the repulsive force in the potential field, the closer the robot is to the obstacle. A threshold is determined when the distance between the robot and the block exceeds. As a result, the offensive potential field functions are considered models' fitness functions for optimization paths' artificial possible field methods. [12, 13]. The evolution algorithm (EA) [14] is one of the most popular metaheuristic algorithms since that has the advantages such as robustness, simple operation principle, and good optimization performance in the global and parallel search process. Because of the EA algorithm's outstanding performance, the field of evolution becomes a studying hotspot in the optimization domain.

This paper suggests the best dispatching path planning based on combining its solution with the artificial potential field method for solving mobile robots' collision-free shortest path planning problem. A adaptive adjustment parameter is used for the mutation factor of the mutation factor according to the actual optimization process. Simultaneously, the artificial potential field method is used to correct the infeasible solution generated by the differential evolution algorithm's crossover operation, and the corresponding correction strategy is introduced. The obtained outcomes from the experiment suggest that the proposed scheme offers satisfactory results for varied barriers in convergence and solution quality.

## 2 Artificial Potential Field Statement

The path planning purpose is to find a path from the starting and ending positions with the collision-free course of the obstacle environment. A path planning is modeled with the potential field that is closer the moving robot to obstacles would be the repulsive force. The direct distance between the robot and the obstruction is more significant than a specific value that is an offensive force of the barrier to the robot. Therefore, the repulsive potential area as a feasible field function can be calculated as follows.

$$U(q) = \begin{cases} \frac{1}{2} K \left( \frac{1}{\rho(q)} - \frac{1}{\rho_0} \right)^2, & \rho(q) \leq \rho_0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

The field method is a widely used path planning method because its advantages are that the mathematical expressions are concise and clear, the response speed is fast, the real-time performance is good, and the physical meaning is clear. The artificial potential field method also models the repulsive possible field function with  $q$  representing the robot's current position  $t$ .

$$F(q) = \begin{cases} K \left( \frac{1}{\rho(q)} - \frac{1}{\rho_0} \right) \bullet \frac{1}{\rho^2(q)} \bullet \frac{q-q_0}{\rho(q)}, & \rho(q) \leq \rho_0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Analyze the impact of parameter settings in the classic artificial potential field method on the final path planning outcome. The principal parameters of the synthetic potential field method are optimized to design a shorter and smoother path. The traditional possible area of feasible algorithm's path planning is in the position;  $Kr$  is the gain coefficient of the workable area function;  $(q)$  is the distance between the robot and the obstacle, and  $0$  is the most significant distance beyond which the obstruction can impact the robot. The robot's repulsive force in a potential virtual field is calculated with the fundamental premise of the artificial possible field scheme.

### 3 Potential Field Robot Path Planning Using Evolution Algorithm

This section presents the suggested scheme for automobile path planning with the parameter optimization in the potential fields and using differential evolution algorithm. First, we review the differential evolution algorithm (EA) and then describe the optimization parameter for the path planning problem.

#### 3.1 Evolutions Algorithm (EA)

The EA is mainly referred to the several following phases [14].

**The initial population** matrix is generated randomly:  $y_{ij} = (y_{i,1}, y_{i,2}, \dots, y_{i,D})$ ,  $i = (1, 2, \dots, NP)$ , where  $NP$  is the population individual size,  $D$  is the number of dimensions ( $D$  is also the number of path points  $n$ , that is,  $D$  is set to  $n$ ). The robot's initial point and target point are  $(x_0, y_0)$  and  $(x_n + 1, y_n + 1)$  respectively, the initial individual's. The selection method is as follows. Calculate the length of the starting and target points on the  $x$  axis as  $L = x_{n+1} - x_0$ . Compare the size of  $y_0$  and  $y_n + 1$ . The more significant value of the two is  $y_{\max}$ , and the smaller value is  $y_{\min}$ . Then the upper bound of the algorithm population is  $y_{\max} + L/2$ , and the lower bound of the population is  $y_{\min} - L/2$ . Nodes can be randomly selected in turn to form path points by determining  $x = \{x_1, x_2, \dots, x_n\}$ .

**The mutation operation** refers to selecting the different individuals that are formulated as follows.

$$v_{i,g} = y_{r1,g} + F \cdot (y_{r2,g} - y_{r3,g}) \quad (3)$$

where  $g$  is evolutionary algebra,  $r1 \neq r2 \neq r3 \neq i$  and  $i = (1, 2, \dots, NP)$ ,  $F$  is the scaling factor. In the differential evolution algorithm, the function of  $F$  is to scale the differential mutation vector corresponding to each individual in the population to determine the search range of the current individual. At the beginning of the algorithm,  $F$  takes a larger value to keep individuals diverse. In the later stage of the algorithm,  $F$  is close to 0. The value of 5 retains good information, avoids the destruction of the optimal solution, and increases the probability of searching for the optimal global solution.

$$F = F_{\min} + (F_{\max} - F_{\min}) \cdot r \quad (4)$$

**The crossover operation** generates test vectors  $u_{i,j,g}$  by crossing the initial target vectors  $y_{i,j,g}$  and the mutation vectors  $v_{i,j,g}$ , using the binomial crossover method to operate,

and at least one component of the test vector is made up of the mutation vector produced. The specific operation is shown in the following formula.

$$u_{i,j,g} = \begin{cases} v_{i,j,g}, & \text{if } \text{rand}(0, 1) \leq CR \text{ or } j = j_{\text{rand}}, \\ y_{i,j,g}, & \text{otherwise} \end{cases}, \quad (5)$$

where  $j$  is the dimension,  $j = 1, 2, \dots, D$ ;  $CR \in (0, 1)$  is a variable of the crossover, and  $j_{\text{rand}}$  and is an integer generating randomly in  $[1, D]$ .

**The selection operation** is to generate the next-generation population, and the optimal individual is selected according to the fitness value of  $f()$  of the target vector  $y_{i,g}$  and the test vector  $u_{i,g}$ . The specific operation is shown in the following equation.

$$y_{i,g+1} = \begin{cases} u_{i,g}, & f(u_{i,g}) < f(y_{i,g}), \\ y_{i,g}, & \text{else} \end{cases}, \quad (6)$$

where  $f()$  is the fitness function;  $i$  and  $g + 1$  are the target vector of next generation.

### 3.2 Evolution Algorithm Combined Potential Field for Path Planning

The attraction from the robot to the goal point decreases as the distance between them decreases. The robot's potential energy is zero when the distance between it and the goal point is zero, so it reaches the target point. The elastic potential field's properties are comparable to those of the potential virtual area. The gravitational potential field function in the potential gravitational field can be the main steps: the potential energy in the elastic potential field is proportional to the distance, and then the gravitational possible field function in the gravitational field potential field can be the main steps as follows.

Step 1. The obstacles in some cases of the different environment are set initially, e.g., by circles, squares with a radius and lengths; the coordinate origin is set to  $(x_0, y_0), (x_m, y_m)$ ;

Step 2. The initial point coordinates of the robot in the three environments are  $(0, 0)$ , and the target point is  $(10, 0)$ . Parameter settings: For the evolution algorithm, the population size is  $NP$  is set to 100, and the termination condition is IterMax;  $CR$  and  $F$  are set as follows, e.g.,  $CR = 0.9, F_{\min} = 0.5, F_{\max} = 0.9$ ; individual in the population is generated randomly.

Step 3. Calculate the value of the objective function of the individuals in the population to find their optimal individual population.

Step 4. Mutation operation generates mutation vector, and Boundary condition processing is checked by Eq. (3).

Step 5. Cross operation, cross operation on target vector and mutation vector to generate test vector as in Eq. (5).

Step 6. Select in applying Eq. (6), calculate test vector by the objective function; with greedy is the attraction potential and the positive proportional coefficient, and the operation is selected; The distance between the mobile robot and the target point that can get attractiveness.

Step 7. Termination condition: The overall situation is derived from the attractive potential or possible fields and the repulsive potential area fields. Correspondingly, the resultant force Eq. (2), the iterative algebra, increases if the maximum algebra is not reached, go Step 3. Otherwise, the output of the optimal individual and its corresponding objective function value with the global outcome.

## 4 Simulation Experiment and Result Analysis

A simulation experiment of the moving robot path planning travels is carried out to verify the effectiveness of the proposed scheme by setting obstacle environment and setting the coordinates of the starting location and the target points to respectively. The parameters set for the artificial potential area or feasible field method greatly influence the path planned by the artificial potential field method. The gravitational field has a positive proportional gain. Coefficients  $K_r$ ,  $K_a$ , and  $\rho_0$  are the gain coefficient repulsion field, affecting obstacle distance parameters on the planning path. The EA relevant settings are as follows: the initial pop size  $N_p$  is set to 40, the max number of iterations IterMax is 1000, and the crossover probability  $CR$  and the scaling  $F$  factor are set to 0.1 and 0.4, respectively.

**Table 1.** Influence of parameters:  $\rho_0$  intensity;  $K_r$  and  $K_a$  setting for the algorithm performance.

$\rho_0$ intensity			Constants of Kr, and Ka		
Values	Shortest path	Iterations no. for a path	[Kr, Ka]	Shortest path	Iterations no. for a path
1	49.361	41.08	[0.7,0.8]	45.73	52.63
10	48.121	47.03	[0.6,0.9]	45.96	60.23
40	47.792	49.39	[0.6,0.8]	46.29	59.00
80	47.951	52.03	[0.6,0.7]	45.77	62.89
100	46.195	57.16	[0.5,0.9]	47.15	51.68
150	46.686	59.34	[0.5,0.8]	47.73	68.69
200	46.907	60.61	[0.5,0.7]	47.40	51.97

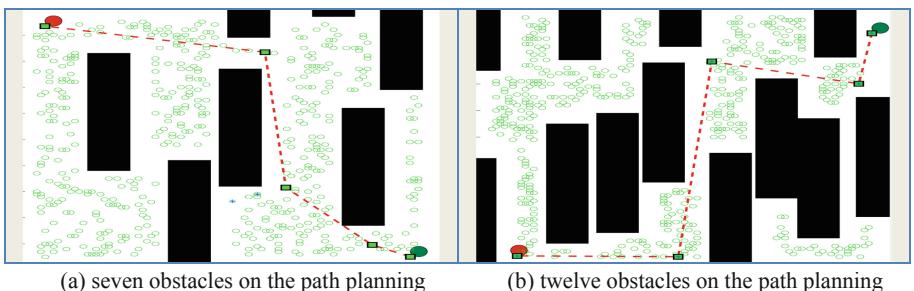
Table 1 shows an example of the influence of parameters, e.g.,  $\rho_0$  intensity;  $K_r$  and  $K_a$  setting for the algorithm performance. The optimized obtained path planning result for the parameters in the artificial potential field method for in the path planned dramatically improves the smoothness of the mobile robot's path. The optimized artificial potential field is compared to the result obtained with the PSO [15], ACO [16], and A\* algorithms [9] is shown in Table 2. A shorter, smoother path is reached, and the scheme of obtained parameter settings is also found out unreachable targets is depicted in Figs. 1a and b.

Table 2 shows the simulation results of the proposed scheme compared with the other algorithms for three scenarios of the setting complex environments in the workspace. From the data in Table 2, it can be shown that the suggested EA's optimal path length

**Table 2.** Simulation results of the proposed scheme are compared with the other algorithms for three scenarios of the setting complex environments in the workspace.

Scenarios	Algorithms	Best	Mean	Std
100 × 100 m complex environment	A*	56.95252	52.96207	1.19232
	PSO	56.00487	53.15206	0.23368
	ACO	57.76607	51.05097	0.14608
	EA	<b>53.02256</b>	<b>50.44013</b>	<b>0.01648</b>
50 × 50 m complex environment	A*	47.30760	44.16887	0.8148
	PSO	46.34112	43.84140	0.30344
	ACO	45.39600	41.06996	0.176
	EA	<b>44.08749</b>	<b>41.95422</b>	<b>0.00784</b>
20 × 20 m complex environment	A*	42.09452	39.20709	0.70064
	PSO	41.70066	39.62108	0.18448
	ACO	42.24799	37.10396	0.17776
	EA	<b>40.72964</b>	<b>36.84022</b>	<b>0.00776</b>

produces superior optimization path planning results than the other approaches. The EA algorithm is more feasible and effective than the original algorithm; however, the algorithm's running time is significantly reduced, the number of inflection points is reduced, and the number of iterations is decreased.



**Fig. 1.** An obtained graphical result of the suggested EA scheme for path planning under artificial potential field with serveral significant obstacles and weights

The robot operating environment space is set to  $M \times M$  m, e.g., 20 × 20 m, 50 × 50 m, and 100 × 100 m, respectively, in applying the EA scheme for optimal path scheduling. Figure 1 shows the obtained graphical result of the proposed EA method in different environment settings, e.g., 3 and 6 obstacles, respectively, for artificial potential field path planning. It can be seen that the obtained results show that parameter optimization can significantly enhance the performance of the potential synthetic field approaching the target.

## 5 Conclusion

This study suggested a modification to the evolution algorithm (EA) for the problem of paths planning optimization using optimized parameters in the potential feasible area in finding the shortest route of moving robots. We combined the traditional artificial potential area of feasible field method with the EA algorithm to optimize the parameters for reducing the path length. The field gain coefficient and the influence distance of obstacles to finding the shortest path are considered by modifying the infeasible search space to optimize the algorithm to locate the fastest route. The simulation results show that parameter optimization can significantly improve the performance of the artificial potential field approaching a target. Compared results show the suggested scheme outperforms the other algorithm, e.g., convergence and solution quality in handling the collision-free path problem for mobile robots.

**Acknowledgments.** This study was partially supported by the VNUHCM-University of Information Technology's Scientific Research Support Fund.

## References

1. Dao, T.K., Pan, J.S., Pan, T.S., Nguyen, T.T.: Optimal path planning for motion robots based on bees pollen optimization algorithm. *J. Inf. Telecommun.* **1**, 1–16 (2017)
2. Dao, T.K., Pan, T.S., Pan, J.S.: A multi-objective optimal mobile robot path planning based on whale optimization algorithm. In: 2016 IEEE 13th International Conference Signal Process. 337–342 (2016)
3. Tsai, P.W., Nguyen, T.T., Dao, T.K.: Robot Path Planning Optimization Based on Multiobjective Grey Wolf Optimizer. In: Pan, J.-S., Lin, J.-W., Wang, C.-H., Jiang, X.H. (eds.) ICGEC 2016. AISC, vol. 536, pp. 166–173. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-48490-7\\_20](https://doi.org/10.1007/978-3-319-48490-7_20)
4. Raja, P., Pugazhenthi, S.: Optimal path planning of mobile robots: A review. *Int. J. Phys. Sci.* **7**, 1314–1320 (2012). <https://doi.org/10.5897/IJPS11.1745>
5. Masehian, E., Sedighizadeh, D.: Classic and heuristic approaches in robot motion planning—A chronological review. *World Acad. Sci. Eng. Technol.* **29**, 101–106 (2007)
6. Tsai, C.F., Dao, T.K., Pan, T.S., Nguyen, T.T., Chang, J.F.: Parallel bat algorithm applied to the economic load dispatch problem. *J. Internet Technol.* **17**, 761–769 (2016)
7. Nguyen, T.T., Pan, J.S., Dao, T.K.: A compact bat algorithm for unequal clustering in wireless sensor networks. *Appl. Sci.* **9**, (2019). <https://doi.org/10.3390/app9101973>
8. Ngo, T.G., Dao, T.K., Thandapani, J., Nguyen, T.T., Pham, D.T., Vu, V.D.: Analysis urban traffic vehicle routing based on Dijkstra algorithm optimization BT—Communication and intelligent systems. Presented at the (2021)
9. Duchoň, F., et al.: Path planning with modified a star algorithm for a mobile robot. *Procedia Eng.* **96**, 59–69 (2014)
10. Castillo, O., Trujillo, L., Melin, P.: Multiple objective genetic algorithms for path-planning optimization in autonomous mobile robots. In: *Soft Computing*. pp. 269–279 (2007)
11. Mirjalili, S.: Knowledge-based systems SCA : A Sine Cosin algorithm for solving optimization problems. **96**, 120–133 (2016). <https://doi.org/10.1016/j.knosys.2015.12.022>
12. Nguyen, T.T., Wang, H.J., Dao, T.K., Pan, J.S., Liu, J.H., Weng, S.W.: An Improved slime mold algorithm and Its application for optimal operation of cascade hydropower stations. *IEEE Access*. **8**, 1 (2020). <https://doi.org/10.1109/ACCESS.2020.3045975>

13. Dao, T., Yu, J., Nguyen, T., Ngo, T.: A hybrid Improved MVO and FNN for identifying collected data failure in cluster heads in WSN. *IEEE Access.* **8**, 124311–124322 (2020)
14. Das, S., Suganthan, P.N.: Differential evolution: a survey of the state-of-the-art. *IEEE Trans. Evol. Comput.* **15**, 4–31 (2011)
15. Masehian, E., Sedighzadeh, D.: A multi-objective PSO-based algorithm for robot path planning. In: 2010 IEEE International Conference on Industrial Technology. pp. 465–470. IEEE (2010)
16. Brand, M., Masuda, M., Wehner, N., Yu, X.H.: Ant colony optimization algorithm for robot path planning. In: 2010 international conference on computer design and applications. pp. V3–436. IEEE (2010)



# Binning on Metagenomic Data for Disease Prediction Using Linear Discriminant Analysis and K-Means

Nhi Yen K. Phan and Hai Thanh Nguyen<sup>(✉)</sup>

College of Information and Communication Technology,  
Can Tho University, Can Tho 900000, Vietnam  
[nthai.cit@ctu.edu.vn](mailto:nthai.cit@ctu.edu.vn)

**Abstract.** Metagenomic data is one of the valuable data resources to predict human disease in personalized medicine. Metagenomic data is very potential and attracted numerous scholars to provide tools and methods to analyze and explore insights in Metagenomics. Binning techniques are promising methods to enhance disease classification on metagenomic data. This study evaluates the integration between Linear Discriminant Analysis and K-Means on preprocessing data before fetching it into prediction models. We perform our experiments on thousands of species abundance metagenomic samples of five diseases have shown that the proposed method can reach 0.913 in accuracy in disease predictions of Liver cirrhosis and obtain promising performance on other four diseases compared to other approaches.

**Keywords:** Linear Discriminant Analysis · Kmeans clustering · Metagenomic · Personalized medicine · Binning · Disease prediction

## 1 Introduction

In recent years, many health services with modern technology have been created because people are paying more and more attention to health care and protection. Personalized medicine (PM) or precision medicine promises to be an effective method of predicting and treating diseases in the future [13]. In contrast to the traditional approach, PM is based on a combination of genetic data, genomics, clinical and environmental factors of each specific patient to provide an effective treatment regimen. Research on metagenomic has been remarkable, most of which have shown that the human gut microbiome provides essential information about their health and diseases. Metagenomic is a non-culturing method of obtaining genetic material based on microbial DNA sequencing techniques. We consider the environment in the human gut to learn more about diseases to make predictions more efficient. However, the proposed diseases are very complex, and the number of observational data samples is limited, so making predictions with diseases with inconsistent results.

Deep learning techniques and artificial intelligence have made great strides in disease diagnosis in the last decade, health care has been demonstrated in many studies [4, 8–10]. Machine learning algorithms are developed to evaluate and recommend effective models for diseases, solving metagenomic-related problems. These issues are covered in [11], for example: Operational Taxonomic Unit clustering [7], mainly prediction genes [14], binning [6], assignment [11] and taxonomic. Our study investigates clustering methods with Linear Discriminant Analysis (LDA) combined with Kmean clustering to improve prediction performance in this study. Ph-CNN [3] introduced by Fioravanti et al. They introduced the OTU hierarchy with the task of metagenomic taxonomy. The authors in [2] propose a binning approach based on high-throughput chromosome conformation capture (Hi-C). The method has high performance and applicability in solving the problem of recovering metagenome-assembled genomes (MAGs). [5] has suggested the GraphBin approach, which is a binning method that uses assembly graphs and applies an algorithm regarding label propagation to enhance the results. GraphBin has good results in identifying mis-binned contigs and binning of contig experimental discarded by existing binning tools. Linear discriminant analysis (LDA) is a data dimensional reduction method for classification problems. The authors' study [12] proposed an LDA-based LEfSe method to support multidimensional class comparison with a specific focus on measurement analysis. K-means is a popular clustering algorithm widely used in research. The authors in [12] performed a comparison of 2 unsupervised binning approaches on the EM algorithm and the K-means algorithm on the metagenomic dataset. Our results revealed that the binning is based on the metagenomic dataset using K-means has better results.

Our contributions include:

- We use a binning method based on Linear discriminant analysis (LDA) and K-means algorithm for data preprocessing and enhancement disease classification performance.
- We run the proposed method experimentally using machine learning models based on metagenomic datasets.
- The proposed binning method has positive results compared to some current approaches.

We briefly describe the six datasets used in Sect. 2 in the remainder of the study. Then, the proposed methodology is presented in Sect. 3, and the experimental results of the proposed methods are illustrated in Sect. 4. Section 5 discusses the results and summarizes the essential points in this study.

## 2 Data Description

In the scope of this study, we conduct the disease classification tasks on six metagenomic datasets related to different specific diseases, such as Liver cirrhosis (CIR), colorectal cancer(COL), Inflammatory Bowel Disease (IBD), Obesity (OBE), Type 2 Diabetes (T2D and WT2). We take more than 1000 species

abundance metagenomic datasets of five diseases, and these samples are for the sum of all cases in this part [6]. The data of the dataset is described in Table 1. The CIR dataset contains 232 samples, including 118 patients and 114 healthy people. The COL dataset owns 48 positive cases, while the IBD dataset was characterized by 465 features of 253 samples, including 164 patients and 89 healthy individuals. The OBE dataset owns 572 features and 344 samples, including 170 patients and 174 healthy. T2D dataset consists of 53 unhealthy individuals and 43 healthy individuals out of 96 samples. WT2 disease dataset with 110 samples and 25 cases positive for the disease.

**Table 1.** Information on six datasets considered in the experiments

	Features	Samples	Patients	Healthy
CIR	542	232	118	114
COL	503	121	48	73
IBD	465	253	164	89
OBE	572	344	170	174
T2D	381	96	53	43
WT2	443	110	25	85

### 3 Binning Approaches

Binning is a data preprocessing method separating continuous real variable intervals into continuous, discrete data intervals. Each data interval is divided into intervals called bins. Types of binning include unsupervised binning and supervised binning. Supervised binning is performing binning on labeled data sets, and this method can achieve higher results than unsupervised binning. However, this method has been understudied due to its complexity.

#### 3.1 EQual Frequency Binning (EQF)

EQual Frequency binning (EQF) [6] is a method of dividing data into n parts (bins), the same number of values is divided approximately to each part. In this case, the break is determined using the training set, so performance in the test phase is poor if the data do not reflect the exact data distribution of the disease under consideration. In addition, the breaks depend entirely on the distribution of the data, so the width of each bin varies considerably.

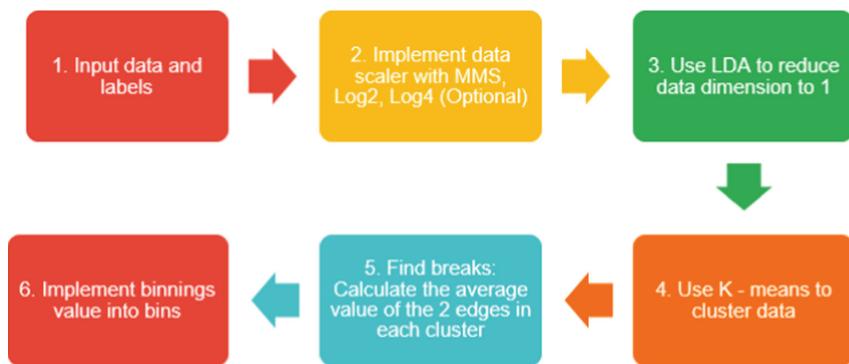
#### 3.2 Binning Based on Linear Discriminant Analysis and Kmean Clustering

Linear Discriminant Analysis (LDA) is considered as a good data dimensionality reduction algorithm. The LDA algorithm reduces the attribute size but retains

the output classes' information. The LDA algorithm searches the decision boundary around each class cluster and projects the data points to new dimensions so that the clusters are as separate as possible and the elements in a cluster are as close to the center as possible. K-means clustering algorithm is a basic algorithm in unsupervised learning. We do not know the label of the data. This algorithm aims to separate the data into  $n$  clusters so that the data in the same cluster have the same properties. The LDA method is implemented by determining the central element of the cluster and the smallest distance between the elements. In this study, we use Euclidean distance as the main metric.

We implement a combination of LDA and K-means algorithms to perform data preprocessing. The steps are as follows:

- **Step 1:** Input is a data set labeled to need binning
- **Step 2:** (Optional) Before binning, the scaler method can be performed with the data. This study implements a scaler with MSS, Log2, Log4.
- **Step 3:** We use the LDA algorithm to reduce the data dimension of step 2. The result obtained is 1-dimensional data because the data set has only two classes (number of dimensions < number of classes)
- **Step 4:** Using K-means algorithm to cluster data, we choose  $k = 10$ . Parameter  $k$  is according to the study [6] to perform the comparison of results
- **Step 5:** In each data cluster, determine the two edges of the cluster as the largest and the smallest value in the cluster. Calculate the mean of the two margins on two adjacent groups.
- **Step 6:** From the breaks found, we take the input data distributed into the corresponding bins.



**Fig. 1.** Steps to perform binning using LDA - Kmean on labeled data.

## 4 Experimental Results

We compare methods of improving the predictive performance of metagenomic datasets. We assess the performances over 10-fold cross-validation with the famous metrics, namely accuracy (ACC) and Area Under the Curve (AUC).

## 4.1 Learning Model Comparison

To assess and compare the efficiency of the proposed methods, we use two popular machine learning models: One-dimensionality convolutional neural network (CNN1D) and MultiLayer Perceptron (MLP) on considered datasets. The models are selected based on the study [6] due to good results and convenience in comparison. In this study, we use the parameter MLP with one hidden layer and 128 neurons like the study in [6]. The CNN1D model in this study includes one-dimensional convolutional layer of 128 kernels followed by a max-pooling of 2 and a fully connected layer.

## 4.2 Evaluation Binning Approaches

Before binning numerical data, we implement scaler method, namely Min-MaxScaler (MMS), Log2, Log4. Table 2 and Table 3 exhibits the performance comparison between our methods and the state-of-the-art including the work of Pasolli et al. (MetAML) [8] and [6]. In the table, None is not executing scaler before binning. The bin number is 10, and this parameter is referenced in the study [1].

**Table 2.** The obtained ACC, AUC and the comparison with the state-of-the-art on 6 datasets. We present the results which are higher than MetAML in **bold** with Multi-layer Perceptron (MLP).

Scaler	Datasets	CIR	COL	IBD	OBE	T2D	WT2
MetAML [8]	<b>val_acc</b>	0.877	0.805	0.809	0.644	0.664	0.703
None	<b>val_acc</b>	<b>0.884</b>	0.761	0.800	<b>0.652</b>	<b>0.660</b>	0.698
MMS	<b>val_acc</b>	<b>0.888</b>	0.744	0.782	<b>0.640</b>	<b>0.677</b>	0.698
Log2	<b>val_acc</b>	<b>0.888</b>	0.753	0.809	0.644	<b>0.668</b>	0.687
Log4	<b>val_acc</b>	<b>0.888</b>	0.753	0.809	0.644	<b>0.668</b>	0.687
MetAML [8]	<b>val_auc</b>	0.945	0.873	0.890	0.655	0.744	0.762
None	<b>val_auc</b>	<b>0.948</b>	0.833	0.847	<b>0.671</b>	<b>0.775</b>	<b>0.786</b>
MMS	<b>val_auc</b>	<b>0.946</b>	0.822	0.796	<b>0.660</b>	<b>0.769</b>	<b>0.789</b>
Log2	<b>val_auc</b>	<b>0.949</b>	0.834	0.844	<b>0.676</b>	<b>0.776</b>	<b>0.778</b>
Log4	<b>val_auc</b>	<b>0.949</b>	0.834	0.844	<b>0.677</b>	<b>0.776</b>	<b>0.786</b>

As observed from the table results Table 2, the ACC values of 3 diseases, CIR, OBE, T2D, when binning with LDA and K-means, gave almost better results than MetAML. CIR disease and T2D disease, when performing binning without scaler and scaler four methods of MMS, Log2, and Log4, both give higher results than MetAML. The ACC value is positive when scaler with MMS, Log2, and Log4. These values are equal and equal to 0.888. For T2D disease, the highest

ACC value was 0.68, and the lowest was 0.666. OBE disease has a higher value without scaler with log2 and log4.

The highest ACC value of this disease is 0.661. For IBD, although there is no higher value than MetAML when performing the proposed binning with Log2 and Log4 callers, the value is equal to MetAML, 0.809. When conducting the binning method with LDA combined with Kmeans, the AUC measure gives better results than the ACC measure. Specifically, the AUC value was higher than MetAML in 4 diseases CIR, OBE, T2D, WT2, and higher in most scaler methods. Considering CIR disease and T2D disease, the disease got the best results when binning with Log2 and Log4. The AUC value of CIR is 0.949 and of WT2 is 0.776, while with the OBE dataset, we performed scaler with log4 with AUC of 0.677 gave the best results. Finally, WT2 disease gave the best results when scaled with MMS with a value of 0.789.

**Table 3.** The obtained ACC and AUC in the comparison with [8] on 6 datasets. The **bold** texts presents the results with Convolutional neural net work (CNN1D). Which are higher than MetAML

Scaler	Datasets	CIR	COL	IBD	OBE	T2D	WT2
MetAML [8]	<b>val_acc</b>	0.877	0.805	0.809	0.644	0.664	0.703
None	<b>val_acc</b>	<b>0.892</b>	0.769	<b>0.827</b>	<b>0.668</b>	<b>0.680</b>	<b>0.707</b>
MMS	<b>val_acc</b>	<b>0.913</b>	0.752	0.800	<b>0.668</b>	<b>0.679</b>	0.687
Log2	<b>val_acc</b>	<b>0.888</b>	0.761	0.800	0.641	<b>0.668</b>	0.686
Log4	<b>val_acc</b>	<b>0.896</b>	0.760	0.809	0.637	<b>0.671</b>	0.686
MetAML [8]	<b>val_auc</b>	0.945	0.873	0.890	0.655	0.744	0.762
None	<b>val_auc</b>	<b>0.950</b>	0.829	0.836	<b>0.700</b>	<b>0.766</b>	<b>0.826</b>
MMS	<b>val_auc</b>	<b>0.953</b>	0.824	0.831	<b>0.693</b>	<b>0.760</b>	<b>0.832</b>
Log2	<b>val_auc</b>	<b>0.949</b>	0.833	0.839	<b>0.699</b>	<b>0.760</b>	<b>0.817</b>
Log4	<b>val_auc</b>	<b>0.951</b>	0.830	0.839	<b>0.694</b>	<b>0.757</b>	<b>0.817</b>

In Table 3, CIR, OBE, T2D, and WT2 diseases with scaler method and similar measure give better results when compared with MetAML method and similar to Table 2. However, the results when running experiments with the CNN model give higher results than the MLP model. When performing non-scaler binning using the proposed method, IBD and WT2 disease give more significant results than MetAML method, and the results are 0.827 0.707, respectively. In both methods (binning with LDA - Kmeans and binning with EQF), we use MultiLayer Perceptron (MLP) model. The bin number is 10 out of 6 metagenomic datasets. The measure used to evaluate is AUC. In general, the proposed method has higher results than EQF in all scaler methods with COL, OBE, T2D disease. The remaining diseases also have better results than EQF from 1 to 2 scaler methods. Considering CIR disease, the LDA - Kmeans method gives better results than the EQF method when scaling with Log2 and Log4, with a

value of 0.949. Similar to COL disease, the AUC value is better when scaler with MMS value of 0.822, Log2 and Log4 with a value of 0.834. IBD disease gives better results than EQF when scaler with Log2, Log4 and has a value of 0.844, greater than EQF 0.068. OBE disease with binning in all methods gave better results than binning with EQF. The best result in this disease is 0.6777 when scaler with Log4. Similarly, the scaler methods all give better scaler results than EQF in T2D disease, which is the highest when the scaler with Log4 has a value of 0.776. In addition, WT2 disease had a better outcome than binning with EQF without scaler and scaler with MMS values of 0.786 and 0.798, respectively.

## 5 Conclusion

This study proposes binning approaches by combining the LDA dimensionality reduction algorithm and the K-means clustering algorithm. With these approaches, we obtain favorable results compared with the frequency-based binning and MetAML. We train the data with various methods on 1D data; when comparing the obtained results, we recognized that performing on the CNN1D model gives higher results. Besides, the results show that the most efficient scaler method is Log4. This study serves as the basis for further studies on combining different approaches to solve the unsupervised binning problem and improve prediction performance. The potential improvements can be archived when applying preprocessing to the proposed algorithm and can obtain positive results when performing in-depth studies and other datasets.

**Acknowledgement.** Phan Kim Yen Nhi was funded by Vingroup JSC and supported by the Master, PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), Institute of Big Data, code VINIF.2021.ThS.21.

## References

1. Auslander, N., Gussow, A.B., Benler, S., Wolf, Y.I., Koonin, E.V.: Seeker: alignment-free identification of bacteriophage genomes by deep learning. *Nucl. Acids Res.* (2020). <https://doi.org/10.1101/2020.04.04.2025783>
2. Du, Y., Sun, F.: Hicbin: binning metagenomic contigs and recovering metagenome-assembled genomes using hi-c contact maps. *bioRxiv* (2021). <https://doi.org/10.1101/2021.03.22.436521>
3. Fioravanti, D., et al.: Phylogenetic convolutional neural networks in metagenomics. *BMC Bioinform.* **19** (2017). <https://doi.org/10.1186/s12859-018-2033-5>
4. LaPierre, N., Ju, C.J.T., Zhou, G., Wang, W.: MetaPheno: a critical evaluation of deep learning and machine learning in metagenome-based disease prediction. *Methods* **166**, 74–82 (2019). <https://doi.org/10.1016/j.ymeth.2019.03.003>
5. Mallawaarachchi, V., Wickramarachchi, A., Lin, Y.: GraphBin: refined binning of metagenomic contigs using assembly graphs. *Bioinformatics* **36**(11), 3307–3313 (03 2020). <https://doi.org/10.1093/bioinformatics/btaa180>
6. Nguyen, T.H., Zucker, J.D.: Enhancing metagenome-based disease prediction by unsupervised binning approaches. In: 2019 11th International Conference on Knowledge and Systems Engineering (KSE), pp. 1–5 (2019). <https://doi.org/10.1109/KSE.2019.8919295>

7. Park, S., Choi, H.s., Lee, B., Chun, J., Won, J.H., Yoon, S.: HC-OTU: a fast and accurate method for clustering operational taxonomic units based on homopolymer compaction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **15**(2), 441–451 (2018). <https://doi.org/10.1109/TCBB.2016.2535326>
8. Pasolli, E., Truong, D.T., Malik, F., Waldron, L., Segata, N.: Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. *PLOS Comput. Biol.* **12**(7), e1004977 (2016). <https://doi.org/10.1371/journal.pcbi.1004977>
9. Reiman, D., Metwally, A.A., Sun, J., Dai, Y.: Popphy-CNN: a phylogenetic tree embedded architecture for convolutional neural networks to predict host phenotype from metagenomic data. *IEEE J. Biomed. Health Inform.* **24**(10), 2993–3001 (2020). <https://doi.org/10.1109/JBHI.2020.2993761>
10. Ren, J., et al.: Identifying viruses from metagenomic data using deep learning. *Quant. Biol.* **8**(1), 64–77 (2020). <https://doi.org/10.1007/s40484-019-0187-4>
11. Sedlar, K., Kupkova, K., Provaznik, I.: Bioinformatics strategies for taxonomy independent binning and visualization of sequences in shotgun metagenomics. *Comput. Struct. Biotechnol. J.* **15**, 48–55 (2017). <https://doi.org/10.1016/j.csbj.2016.11.005>, <https://www.sciencedirect.com/science/article/pii/S2001037016300678>
12. Segata, N., et al.: Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60–R60 (2011)
13. Vicente, A.M., Ballensiefen, W., Jönsson, J.I.: How personalised medicine will transform healthcare by 2030: the ICPerMed vision. *J. Transl. Med.* **18**(1) (2020). <https://doi.org/10.1186/s12967-020-02316-w>
14. Wang, Z., Chen, Y., Li, Y.: A brief review of computational gene prediction methods. *Genom. Proteom. Bioinform.* **2**, 216–221 (2004)



# Letter Recognition in Hand Sign Language with VGG-16

Hai Thanh Nguyen<sup>1(✉)</sup>, Linh Thuy Thi Pham<sup>1,2</sup>, Truc Thuy Nguyen Mai<sup>1,3</sup>,  
Trinh Kieu Ngoc Vo<sup>1</sup>, and Tran Thanh Dien<sup>1</sup>

<sup>1</sup> Can Tho University, Can Tho, Vietnam

{nthai.cit,vnktinh,thanhdien}@ctu.edu.vn

<sup>2</sup> FPT High School-FPT University, Can Tho, Vietnam

<sup>3</sup> Tay Do University, Can Tho, Vietnam

**Abstract.** Instead of spoken language, sign language is the commonly used language by the deaf. To express their thoughts, the deaf people combine handshapes, movements, directions of hands, arms and body, and facial expressions. Currently, there are several methods of hand sign language recognition, such as image recognition or sensor recognition. In this study, we tested the recognition of characters of hand sign language using VGG-16 network architecture to investigate the change in recognition efficiency when adding Dense layers with neural parameters. A test was conducted on a gesture dataset consisting of 26 English letters with 5,391 images, and there are 200 images for each subclass. By investigating the change in recognition efficiency through the change of the VGG-16 model, we have achieved a high recognition rate with an accuracy of 99.902% on the training set and 99.910% on the test set compared to the original model. In addition, the change in the distance when detecting hand gestures also affects the system's accuracy.

**Keywords:** English letters · Gesture recognition · Sign language · VGG-16

## 1 Introduction

Vietnam is one of the Asian countries having a relatively high number of people with disabilities, in which people with hearing disabilities account for nearly 1 million [1]. Deaf people use sign language, specifically hand gesture language with visual hand signals to convey word meanings instead of using the sound of speech which their disabilities limit them. This language is widely used in the deaf community but not in the communication community. This leads to many difficulties for the deaf community to communicate with the people around them.

Sign language is also commonly used in automatic control designs. Some sign language applications include looking up patient records (when doctors are not allowed to touch control devices); controlling robots to assist disabled people; and integration with remote control of television or music player. Sign language

is the language that uses hand gestures instead of voice sounds. People with hearing disabilities create it to help them communicate with one another in the community [2]. This language uses handshapes, movements, directions of hands, arms or body, and facial expressions to express the speakers' thoughts. Wherever deaf communities exist, sign languages have developed as a valuable means of communication, forming the core of their local culture. Although sign language is used primarily by deaf people who are deaf, it is also used by people who can hear, for example, those who cannot speak, those who have difficulties with spoken language due to disabilities, or those who have hard of hearing members in their families. We should not confuse sign language with body language. Currently, there are many methods of image recognition to make it easier for the deaf to communicate. Among those methods, the VGG-16 network [3], a convolutional neural network architecture, is widely applied in deep learning network techniques.

This study focuses on changing the investigation refinement of the VGG-16 network architecture model with epoch and batchsize hyperparameters. We also conduct investigation with different distances between hand and screen recognition in the English sign language dataset, and experimental survey on the entire alphabet.

## 2 Related Work

There are many methods to build a sign language recognition system. According to the review of some recent studies, [4], two main approaches for sign language recognition include image-based and sensor-based sign language recognition. The image-based approach uses image processing (photos or videos) to recognize symbols. The sensor-based approach uses instrumental gloves assembled with sensors to monitor hand joints.

Recognizing sign language through photos (or cameras, videos) has been studied and developed. Applications based on image processing have achieved good results and are overgrowing. In previous studies, high-resolution cameras were used to capture images; however, by the end of 2010, the Kinect device released by Microsoft had changed the method of input data acquisition for the study of sign language recognition. The Kinect device uses 3D webcams, infrared transceivers, and audio receivers. Kinect's software development kit (SDK) can process and allow users to extract data about body movement positions, including two hands, two elbows, head, torso, and two legs, even 3D hand shapes. According to [5], experiments are performed on a dataset of 50 words belonging to 6 different topics, with a total of 2,400 data samples. The results show that the Kinect sensor and recognition support vector machine system achieved an accuracy of 99.46%. The study [6] described a method for American sign language (ASL) alphabet recognition using Convolutional Neural Network (CNN) with the ability to combine multi-methods and inference from depth images captured by Microsoft Kinect. The study's approach was to enhance the original data by generating more views, making training more efficient, and reducing the possibility of overfitting. This approach was evaluated on two public datasets: the ASL

benchmark and the alphanumeric datasets. The results showed the achieved accuracy of 100% and 93% for the half-half test and the leave-one-out test, respectively, on the ASL benchmark dataset, and the recognition accuracy of 100% for both the half-half test and the leave-one-out test on the alphanumeric dataset. The study [7] compared five different CNN models, including CNN, VGG-16 Scratch, VGG-16 Pretrained, InceptionNet, and ResNet50 for hand gestures with a dataset of 87,000 images and 29 layers. The results showed that ResNet50 achieved the best recognition of 99.88%. This paper mainly described the sign language recognition system through images by VGG-16 network architecture. This network architecture consists of 16 layers, including accumulation layers with  $3 \times 3$  filters and stacked up to  $2 \times 2$  aggregate layers. Between these layers, ReLU is applied. Finally, a softmax function is used to generate the probabilities for each layer.

In the study [8], to evaluate the effectiveness of the VGG-11 and VGG-16 models, the authors used two large datasets, including the gestures of the Indian sign language (ISL) and the ASL with a precision of 99.96% for the ISL dataset and 100% for ASL dataset.

A typical study on improving the VGG-16 model can be mentioned in the early diagnosis of hearing loss [9] using a pre-trained deep learning framework by analyzing and evaluating the functional reliability of hearing based on the AEP (auditory evoked potential) response. The proposed method outperformed current studies by improving the classification precision from 57.376% up to 96.879%; it is indicated that the proposed improved VGG-16 architecture can cope significantly with the AEP response. In another field, in the study on plant variety identification, the improved VGG-16 was applied for the task of identifying and classifying 12 peanut varieties [10]. The average precision of the improved VGG-16 model on the peanut shell test set was 96.79%; it was 8.99% higher than VGG-16, and 1.6–12.39% higher than the other classical models such as AlexNet, VGG-16, GoogLeNet, ResNet18, ResNet50, SqueezeNet, DenseNet201, and MobileNetv2.

The mentioned studies show that the improvement of the VGG-16 model helps increase the precision compared to the original models. However, currently few studies on improving VGG-16 have been conducted to increase the accuracy of sign language recognition for hard-of-hearing people. Therefore, identification and improvement of the VGG-16 model in the gesture recognition system are essential.

### 3 Methods

#### 3.1 Datasets

The experimental dataset is based on the English alphabet consisting of 26 letters with a total of 5,391 images, and each letter has more than 200 images collected from [11]. Each image has a resolution of  $64 \times 64$  pixels. The details of the experimental data include image resolution of  $64 \times 64$ , layer number of

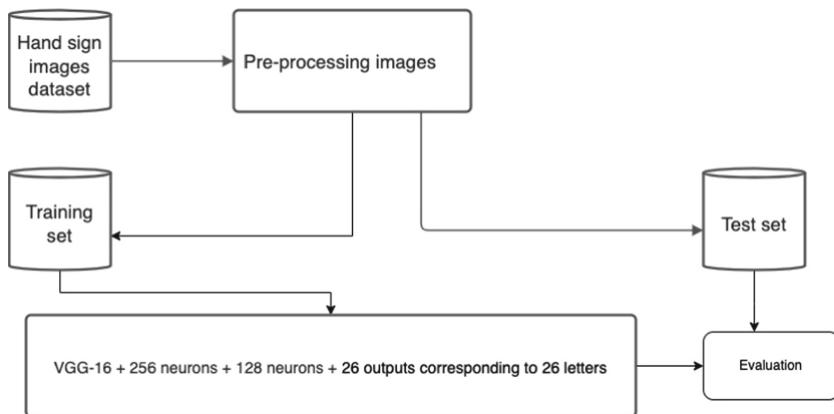
26, and a sample number of 5,391 with more than 200 samples per layer. Some examples of hand sign language recognition data are shown in Fig. 1.



**Fig. 1.** Examples of data samples for hand sign language recognition

### 3.2 Architecture of the Proposed System

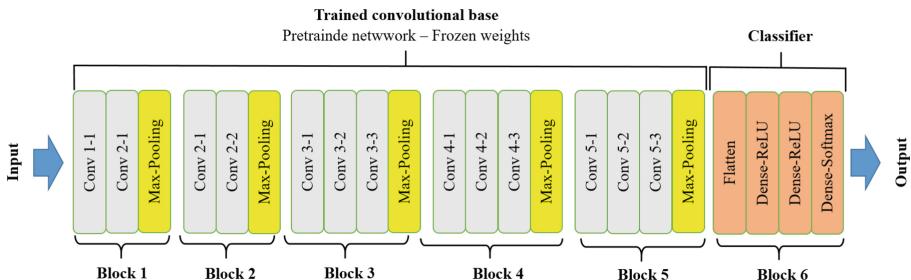
Our overall processing architecture includes preparing the dataset, normalizing the image to  $224 \times 224$  size, and dividing the data into two sets: a training set and a test set with a ratio of 80/20. In addition, the VGG-16 architecture improved with two fully connected layers with 256 and 128 neurons. This architecture is shown in Fig. 2.



**Fig. 2.** The architecture for sign language recognition

In the VGG-16 architecture, as illustrated in Fig. 3, there are 16 layers, including 13 convolutional layers and three fully connected layers. The design consists of 5 blocks. In each block, there are convolutional layers and one max-pooling. In Fig. 3, the input to the architecture is an image with dimensions of  $224 \times 224 \times 3$  and is processed in the following order. In Block 1: firstly, pass the

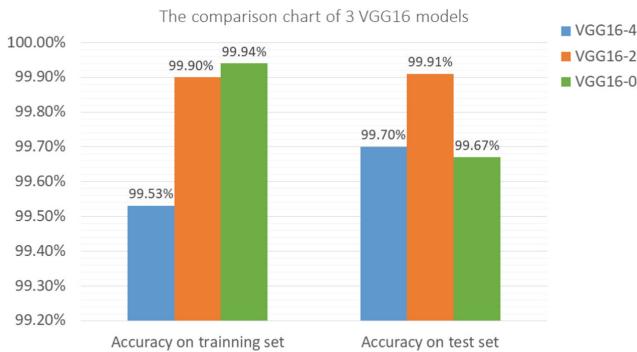
image through 2 convolutional layers having 64 channels with the same  $3 \times 3$  filter and padding; then use max-pooling to reduce the image's remaining input image size to  $112 \times 112 \times 64$ ; reduce the parameters to increase the computational performance. In Block 2: pass the resulting image in Block 1 through 2 convolutional layers having 128 channels with the same  $3 \times 3$  filter and padding; then use max-pooling to reduce the image's remaining input image size  $56 \times 56 \times 128$ ; reduce the parameters. In Block 3: pass the resulting image in Block 2 through 3 convolutional layers having 256 channels with the same  $3 \times 3$  filter and padding; then use max-pooling with sliding window ( $2 \times 2$ ) to reduce the image's remaining input image size to  $28 \times 28 \times 256$ ; reduce the parameters. In Block 4: pass the resulting image in Block 3 through 3 convolutional layers having 512 channels with the same  $3 \times 3$  filter and padding; then use max-pooling with sliding window ( $2 \times 2$ ) to reduce the image's remaining input image size to  $14 \times 14 \times 512$ ; reduce the parameters. In Block 5: pass the resulting image in Block 4 through 3 convolutional layers having 512 channels with the same  $3 \times 3$  filters and padding; then use max-pooling with sliding window ( $2 \times 2$ ) to reduce the image's remaining input image size of the image to  $7 \times 7 \times 512$ . After passing through all 13 convolutional layers, the image will pass through the flatten layer to switch back to the feature vector In Block 6. It will then go through 3 fully connected layers with 256 and 128 neurons. Finally, the image will go through fully connected layers with the softmax result (None, 26).



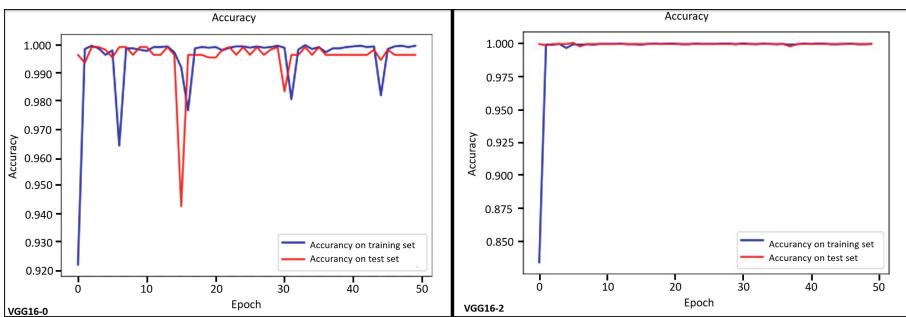
**Fig. 3.** The architecture of the improved VGG-16

## 4 Experiment

We recorded the results of the VGG-16 architecture with an average accuracy of 99.944% for the training set and 99.666% for the validation set. Then, we, in turn, add Fully Connected layers with different neural parameters, including the VGG-16 model added 4 Fully Connected layers with 128 neurons parameters and 1 Dropout layer, and the VGG-16 model added 2 Fully Connected layers with 256 neurons and 128 neurons.



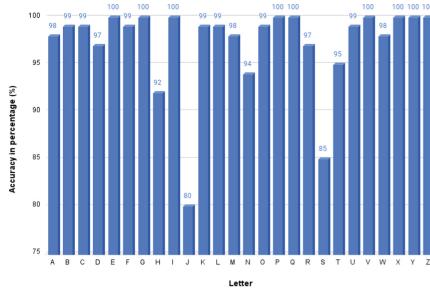
**Fig. 4.** The performance comparison of 3 architectures



**Fig. 5.** An illustration of training phase and evaluation on test set VGG16-0 (the left chart) and the performance of VGG16-2 through epochs (the right one)

We have a graphical comparison, as shown in Fig. 4. The blue curve represents the accuracy of the training set. The red curve represents the accuracy of the test set. Figure 5 (the left chart) shows that for the VGG-16 model, after running 50 epochs, the accuracy on the training set increased from 99.914% to 99.984%. However, it cannot effectively improve the test set. The accuracy on the test set of the model ranges from 99.634% to 99.724%, and the average accuracy is 99.664%. The performance difference between training and test sets is approximately 0.324%; train running time is 32 min 25 s. After that, we improved the VGG-16 model added 4 Fully Connected layers with 128 neurons, the accuracy of the test set of the model ranges from 99.444% to 99.914%, and the average accuracy is 99.704%. The performance difference is approximately 0.234%; train running time is 38 min 24 s, and we need to improve the model one more time. The training process with the second-time improved model shows that after running 50 epochs, both accuracies on the training set and accuracy on test set reach more than 99.900%, final accuracy on test set reaches 99.914% and is stable for many trials, and the loss rate gradually decreases to 0 as illustrated in Fig. 5 (the right chart). Besides, the accuracy rate when training our model increases

to 99.914% after five epochs, and train running time is 30 min 27 s. This shows that our model has a high convergence rate.



**Fig. 6.** Accuracy of letter recognition results through each symbol

The improved VGG-16 model added 2 Fully Connected layers with 256 neurons, and 128 neurons have an average accuracy of 99.911% on the datasets, 0.2441% higher than that of the original VGG-16 model. In addition, we also tested the distances from the hand positions to the recognition areas with the best-chosen model. With 50 different trials from all positions and words, it can be inferred as follows. Hand positions' distances of <80 cm will give accurate results. However, the accuracy will decrease when approaching the threshold of 80 cm. Therefore, hand positions' distances of >80 cm cannot be recognized. Moreover, results are similar between samples due to the influence of hand movements. Alphabetical symbols that are difficult to represent have lower precision. In addition, the results will be affected by the background. If the background has many obstacles in the recognition area, the accuracy will decrease. The accuracy of letter recognition results through each symbol is shown in Fig. 6.

## 5 Conclusions

The study focuses on improving the model extended from the VGG-16 network architecture in the problem of sign language recognition. This is the first step for building a complete system that allows feasible communication between the deaf and normal people through hand gestures. In this paper, we modify the number of neural layers compared to the original VGG-16 architecture and select the best model through the trials of using the VGG-16 architecture added 2 Dense layers with 256 neurons and 128 neurons for a dataset of 5,391 images with 26 sub-classes, train/test ratio is 80/20. Research results show the accuracy of 99.902% for the training set and 99.910% for the test set. In addition, we check the influence of the distance from the recognition area to the hand gesture, showing that the recognition distance affects the accuracy of the model. For example, with a distance of >80 cm, the system will not recognize the gesture. The best distance is ranged from 20 cm to 40 cm. We expect that the results of

this study will contribute to applications using the VGG-16 network model, such as virtual reality interaction systems, support solutions in advertising technology, entertainment, integrative products for better results.

## References

1. Duy, P.A.: Raising the voice of deaf people. DVV International (2017). <https://www.dvv-international.de/en/adult-education-and-development/editions/aed-842017-inclusion-and-diversity/columns/raising-the-voice-of-deaf-people>
2. Sutton-Spence, R., Woll, B.: The Linguistics of British Sign Language: An Introduction. Cambridge University Press, Cambridge (1999)
3. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2015)
4. Nimisha, K., Jacob, A.: A brief review of the recent trends in sign language recognition. In: 2020 International Conference on Communication and Signal Processing (ICCSP). IEEE, July 2020. <https://doi.org/10.1109/iccsp48568.2020.9182351>
5. Khanh, V.H., Khang, P.N.: Sign language recognition using kinect sensor. In: FAIR -Fundamental and Applied IT Research - 2017. Publishing House for Science and Technology, August 2017. <https://doi.org/10.15625/vap.2017.00082>
6. Tao, W., Leu, M.C., Yin, Z.: American sign language alphabet recognition using convolutional neural networks with multiview augmentation and inference fusion. Eng. Appl. Artif. Intell. **76**, 202–213 (2018). <https://doi.org/10.1016/j.engappai.2018.09.006>
7. Bantupalli, K., Xie, Y.: American sign language recognition using deep learning and computer vision. In: 2018 IEEE International Conference on Big Data (Big Data). IEEE, December 2018. <https://doi.org/10.1109/bigdata.2018.8622141>
8. Sharma, S., Singh, S.: Vision-based hand gesture recognition using deep learning for the interpretation of sign language. Expert Syst. Appl. **182**, 115657 (2021). <https://doi.org/10.1016/j.eswa.2021.115657>
9. Islam, M.N., et al.: Diagnosis of hearing deficiency using EEG based AEP signals: CWT and improved-VGG16 pipeline. PeerJ Comput. Sci. **7**, e638 (2021). <https://doi.org/10.7717/peerj-cs.638>
10. Yang, H., Ni, J., Gao, J., Han, Z., Luan, T.: A novel method for peanut variety identification and classification by improved VGG16. Sci. Rep. **11**(1) (2021). <https://doi.org/10.1038/s41598-021-95240-y>
11. AI library - dataset for AI. <https://www.miai.vn/thu-vien-mi-ai/>



# A New Method for IP Camera Based Face Anti-spoofing Systems

Ha Nguyen-Xuan<sup>1</sup> , Dong Hoang-Nhu<sup>2</sup>, Anh Nguyen-Duy<sup>1,2</sup>,  
and Tuan Dang-Minh<sup>2,3</sup>

<sup>1</sup> Hanoi University of Science and Technology, 1 Dai Co Viet, Hai Ba Trung, Hanoi, Vietnam  
[ha.nguyenxuan@hust.edu.vn](mailto:ha.nguyenxuan@hust.edu.vn)

<sup>2</sup> CMC Institute of Science and Technology, 11 Duy Tan, Cau Giay, Hanoi, Vietnam

<sup>3</sup> Posts and Telecommunication Institute of Technology, KM 10 Nguyen Trai, Ha Dong, Hanoi,  
Vietnam

**Abstract.** Detecting face presentation attacks (FPAs) is a critical issue for any face-identification-based application. In this work, a new method for IP-camera-based anti-spoofing systems is proposed. The basis of this method is the detection of objects that are used for face presentation attacks, including printed papers and display devices like smartphones, tablets, and laptops or computer screens. The size and position of the bounding boxes of the detected faces and detected objects are compared through an algorithm to find fake cases. To do this, three deep learning models for object detection were developed, which are based on transfer learning techniques for YOLOv5s6, YOLOv5, and the RetinaFace network. For the face detection task, the RetinaFace with backbone MobileNet0.25 was used. For the computational efficiency evaluation, all developed models and algorithms were deployed to an embedded hardware platform, namely the Jetson Xavier Development Kit. The experiment implementations show that our method outputs excellent results where the YOLOv5s6 has the highest mean average precision mAP50 of 83.91% for the object detection and the processing time is only 25 ms. The detection of FPA has a 99.2% accuracy. The obtained results are very good and can be applied to many practical applications where IP cameras are used.

**Keywords:** Face anti-spoofing · IP-camera · Deep neural network

## 1 Introduction

With the development of deep neural networks, many image processing issues can be efficiently solved, which offers many possibilities for practical applications. Among them, face detection and recognition have received much attention in recent years since they can be applied to verifying services with excellent user experiences. One of the typical applications is the use of face detection and recognition for access control systems. This solution can replace traditional methods like passwords and fingerprints in terms of productivity and convenience. However, one of the biggest problems with systems based on face detection and recognition is face presentation attacks (FPAs) [1].

As reviewed in [1], several types of FPAs can be divided into two catalogs. The first one is the obfuscation attack, and the second one is the impersonation attack. The impersonation attack includes the photo, video replay, and 3D mask attack, whereas the obfuscation attack relates to the extreme makeup, partial occlusion, and plastic surgery. Depending on the types of FPAs, suitable anti-proofing methods are applied, which include both image processing algorithms based on the OpenCV library [2] and deep neural networks [3], and the use of physical devices like infra-image or laser-based sensors.

It is known from the literature that most of the face presentation attack detection methods are developed for generic consumer devices (GCDs) such as smartphones, tablets, or laptops [1]. There are two typical approaches.

In the first approach, the multi-scale model fusion method is considered [3]. The basis of this method is the two-dimensional scaling of the original face frame. The camera frames are processed by a face detection model to obtain the face frame. The edge of the face frame is expanded according to a certain scale to get more information about the edge of the face frame and the background, for example, the edge of smartphones or printed photos. This information will help detect FPAs. The patch from the original frames is passed through the classification model to detect FPAs.

This method can be efficiently used for GCDs like smartphones or tablets, in which the faces of users are very near the front of the device's camera. In the case of IP cameras, they typically have a wide viewing angle and high resolution. Therefore, they are often used for applications where users are far from the camera. This leads to difficulty in determining a suitable scaling factor. If the scaling factor is still small, the edge of the face frame and background cannot be seen. If the scaling factor is too large, other face frames nearby can be considered for the face frame, which leads to the fact that the input image for the classification model contains more than one face frame. Besides, because of the wide viewing angle, when a person moves closer to the camera, the size of the face frame is gradually increased, so fixing the scale factor would not make sense.

In the second approach, only a classification model is used without any scaling factor for the face frame [4–6]. The face's fame is detected via a face detection model. After that, they are aligned and classified to detect FPAs. This approach cannot achieve high precision in the case of IP cameras since the face frame is typically far from the cameras. Furthermore, without the scaling factor, much of the surrounding information is eliminated. Besides, most public datasets are created by the cameras of GCDs, which are not suitable for training models for IP cameras. For this confirmation, we have trained and tested several classification models with our collected dataset. The dataset is collected from public datasets for face anti-spoofing purposes, including CELEB\_A [7], NUAA [8], CASIA-FSD [9], ROSE-YouTu [10], and our self-created dataset. Table 1 shows the results of considering trained models [4–6]. The highest accuracy is only 84.76%, which is still not good enough for practical applications.

It is known that there is a lack of methods that can be efficiently used for anti-proofing systems with IP cameras. Therefore, in this work, we propose a new method to detect typical types of FPAs using IP cameras. Types of FPAs considered in this work include the photo attack and the video replay attack. The 3D mask attack or other unknown attacks are not considered. The overall processing pipeline is proposed. Also,

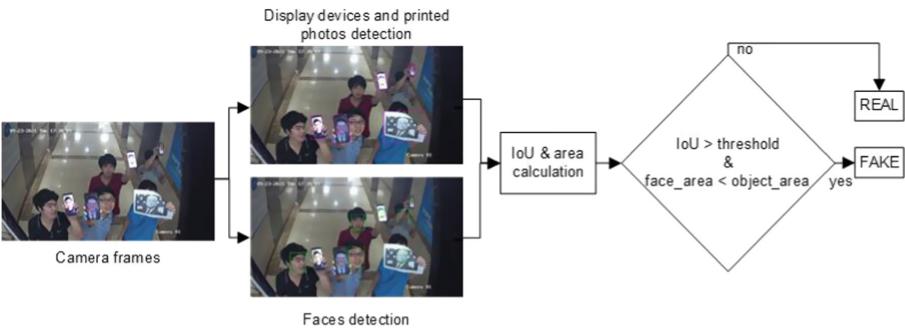
evaluations and deployments of the developed model on the edge computing hardware are presented.

**Table 1.** Accuracy of evaluated models.

Model	Accuracy (%)
SE_ResNet50 + deepix [4]	78.52
SE_ResNet50 + deepix + angular [5]	70.25
SE_ResNet50 + FFT [5]	77.50
ResNet18 + deepix [4]	66.85
SSGD + ResNet18 [6]	84.96

## 2 Method

A typology of face-presentation attacks (PAs) is presented in [1]. In this work, a new method for face presentation attack detection with IP cameras is proposed. Types of considered PAs include printed photo attacks and video replay attacks on mobile phones, tablets, and laptop screen displays. The processing pipeline of our method is shown in Fig. 1.



**Fig. 1.** Overall pipeline for face presentation attack detection.

The basic idea behind this method is the comparison of relative positions of detected faces with those of detected presentation attack devices. First, the Retina face detection model [11] is used to detect all faces appearing in camera frames. Faces in camera frames include both the faces of real people and the faces appearing in presentation attack devices. The Retina model will output the coordinates of the bounding box of detected faces. Similarly, an object detection model was developed for detecting printed papers and displaying devices. Consequently, the bounding box of the detected objects is achieved. In the third step, the bounding boxes of detected faces will be compared

with the bounding boxes of detected objects to find presentation attack situations. If the intersection over union (IoU) of the two bounding boxes is bigger than a threshold, and the bounding box area of the comparing object is bigger than the bounding box area of the corresponding face, the situation is concluded to be a presentation attack.

For the face detection model, a pre-trained model with the Retina network [11] is used without any extension. The object detection model is focused on development. Several models are considered. In the following, details of the dataset and the development of the model are described.

## 2.1 Dataset

A dataset was created for the development of the object detection model. The dataset contains 11683 images of different sizes, which were crawled from the Internet, public datasets, and self-preparation. Details of the created dataset are shown in Table 2. In this work, two types of face presentation attacks, including printed photos and display devices (mobile phone, tablet, laptop, and desktop screen), which often appear in practical applications, are considered. In Fig. 2, the example of these FPAs is illustrated, where Fig. 2a, b, and c present printed papers with a border around them, a very close border, and without a border, respectively. In Fig. 2d and e, the PPA using the mobile phone and the tablet are respectively shown.



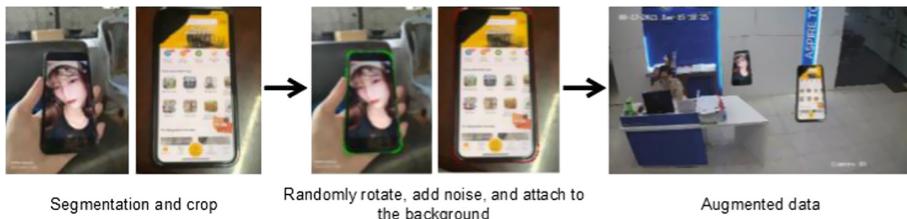
**Fig. 2.** Types of consideration for face presentation attacks: a) printed paper with a border around it; b) printed paper with a very close border; c) printed paper without border; d) face on a smartphone; e) face on a tablet.

To improve the robustness of the dataset, the augmentation for the dataset was implemented, which is illustrated in Fig. 3. First, images of the considered objects are collected. The objects in these images are then segmented and extracted. The extracted objects are randomly rotated. Also, noise is added to the object. After that, the photos of augmented

objects are attached to several backgrounds to generate diverse images for the dataset. All images are resized to dimensions of  $1280 \times 1280$  or  $640 \times 640$ , which is suitable for the input of training networks.

**Table 2.** Statistics of the training dataset.

Dataset	Printed papers	Display devices (tablets, laptops, computer screens, smartphones)
Crawled from the internet	200	1080
Public dataset [7–9]	830	500
Self-preparation and augmentation	3315	5758



**Fig. 3.** Data augmentation.

## 2.2 Object Detection Models

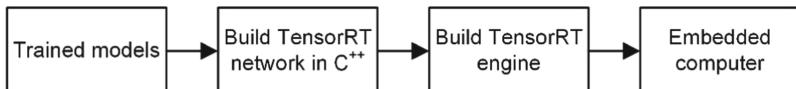
Object detection models aim to detect printed photos and display devices, including mobile phones, tablets, and laptop screens. Models must be developed with high accuracy while using as few computational resources as possible from edge computing devices. Therefore, in this work, three different models were developed and compared. YOLO [12] and the RetinaFace [11] networks were used, with MobileNet0.25 as the backbone. The first model was developed with the YOLOv5s6 backbone [12]. The second model was based on the YOLOv5s [12]. The RetinaFace network [11] is used for the third model. A transfer learning process was applied based on the above-mentioned collected dataset. The finetuning process for the three pre-trained models was implemented. Table 3 shows the parameters for the transfer learning process. The collected dataset is divided into train, test, and validate with a portion number of images of 9735, 1200, and 1948, respectively.

**Table 3.** Transfer learning configurations for training models.

Parameter	YOLOv5s6 [12]	YOLOv5s [12]	RetinaFace [11]
Input image size	$1280 \times 1280$	$640 \times 640$	$640 \times 640$
Batch size	16	96	128
Epochs	100	100	100
Learning rate	0.0032	0.0032	0.001

### 2.3 Models and Algorithm Deployment on Embedded Hardware

It is known that edge computing devices have very limited hardware resources, especially in the case of IP-based artificial intelligence (AI) cameras. Therefore, any deep learning models that are deployed on AI cameras must be computationally efficient. For this reason, in this work, the computational efficiency of the developed models is evaluated. The Jetson Xavier Development Kit [13] (GPU: 512-core Volta GPU with Tensor Cores; CPU: 8-core ARM v8.2 64-bit CPU, 8MB L2 + 4MB L3) was used as an evaluation platform. Figure 4 shows the steps to deploying the models on the hardware. All fine-tuned models on the PyTorch framework are converted to the dictionary format of “.wts”, which contains the weights of the network and the layer information. These parameters are used to build a computational network in the TensorRT format in C++ that can be efficiently run on the Jetson Xavier Kit. The weights are then loaded according to the dictionary onto the TensorRT network to build a TensorRT engine. This engine is run on the hardware for computing tasks. Also, a software environment is installed on the kit with Ubuntu 18.04/cuda10.2/tensorRT7.1/opencv4.5.0.

**Fig. 4.** Deployment pipeline of the proposed algorithms and models on Jetson Xavier [13].

## 3 Results and Discussions

### 3.1 Evaluation of Object Models

The metric Average Precision (AP) with an Intersection Over Union (IoU) threshold of 50 (AP<sub>50</sub>) was used to evaluate the precision of the object detection model. A set of 1200 images, having all classes of considered objects, was used for the evaluation. Results are shown in Table 4. It is seen that the RetinaFace has the lowest precision of all classes of objects, which results in an mAP of only 65.36%. In contrast, the YOLOv5s6 has the highest precision, with an mAP of 83.91%. These results can be explained by looking inside the architecture of the neural network of each model and the input size

of the training images. For the YOLO network, the size of anchor boxes is adaptively calculated based on the size of objects in the whole training dataset, while this one in the RetinaFace is fixed as the network configuration without any further change. If the size of anchor boxes fits well with the size of objects in the image, the precision can be better. In our case, the size of objects in images is subject to much fluctuation. Thus, the adaptively changing size of anchor boxes in the YOLO network could be the reason for the obtained results. Additionally, the input image size of YOLOv5s6 is two times larger than YOLOv5s. Thus, when an image experiences the convolution layers of the network, the YOLOv5s6 can highlight features of an object better, resulting in better precision. For object detection, an mAP of 83.91% is a very impressive result, which can be applied for realistic applications.

**Table 4.** Average precision of the developed models on the testing dataset.

ID	Class	YOLOv5s6	YOLOv5s	RetinaFace
1	Printed papers	87.29	78.49	75.59
2	Tablets and mobile phones	81.65	76.59	57.19
3	Displays on laptop and computer screens	82.78	83.59	63.28
	<b>mAP</b>	<b>83.91</b>	<b>79.56</b>	<b>65.36</b>

### 3.2 Computational Efficiency

The computational efficiency of the developed models and algorithms was evaluated based on the Jetson Xavier Development Kit from NVIDIA [13] as described in Sect. 2.3. The utilization of hardware resources and precision are considered. The results are shown in Table 5. It is obvious that the developed models are very lightweight, as indicated by the computing time and used RAM. The highest processing time for each detection is 25 ms for the case of YOLOv5s6, and the highest allocated RAM is only 1.6 GB for the case of RetinaFace. Since the trained models are converted to a mode of FP16 to reduce the utilization of hardware resources on edge computing devices, the accuracy is slightly reduced by around 0.4%. With these results, YOLOv5s6 can be efficiently deployed on the edge computing hardware of IP-based AI cameras.

**Table 5.** Model deployment results on Jetson Xavier Development Kit [13].

Parameter	YOLOv5s6	YOLOv5s	RetinaFace
Bath size	1	1	1
Mode	FP16	FP16	FP16
Computing time (ms)	25	10	9
RAM usage (GB)	1.3	1.2	1.6
mAP <sub>50</sub> (%)	83.50	79.11	64.95

### 3.3 Practical Application

The developed model and computing algorithms are deployed for our practical applications where IP-cameras, from the HIKVISION model DS-2CD2023G2-IU [14], are used. The system is used to detect spoofing cases where people use printed papers or display devices as sources of FPAs. The accuracy Acc of the system is calculated according to Eq. 1. Where  $T_p$  is the number of true positives (rate of corrected detection);  $T$  is the total number of tests. A test with a  $T$  of 4417 was performed, and the result was excellent, as shown in Table 6. It is completely logical with the results of Table 4 and Table 5. The YOLOv5s6 has the highest accuracy of 99.2%. With this accomplishment and the evaluation of hardware resource use, our system can be used for many practical applications.

$$Acc = \frac{T_p}{T} \quad (1)$$

**Table 6.** Accuracy of the system in practical applications.

	YOLOv5s6	YOLOv5s	RetinaFace
Acc (%)	99.2	98.2	91.9

## 4 Conclusions and Outlook

In this work, a new method for IP-camera-based anti-spoofing systems has been successfully developed. Methods used for FPAs with cameras of GCDs are not suitable and reliable for FPAs with IP cameras. The presented results prove the efficiency of our method in terms of accuracy and computation. Thus, our method offers many practical applications of FPA detection with IP cameras.

In the future, the deployment of our algorithm on IP-based AI cameras is planned. Further optimization for computational efficiency and robustness under lighting conditions will be implemented.

One limitation of our method is that it still cannot detect 3D mask attacks. Thus, methods to overcome this limitation are planned to be studied.

**Acknowledgment.** This work was funded by the CMC Institute of Science and Technology, CMC Corporation, Hanoi, Vietnam.

## References

1. Ming, Z., Visani, M., Luqman, M.M., Burie, J.-C.: A survey on anti-spoofing methods for face recognition with rgb cameras of generic consumer devices. [arXiv:2010.04145v1](https://arxiv.org/abs/2010.04145v1) (2020)
2. OpenCV library homepage. <https://opencv.org/>. Accessed 9 Nov 2021
3. Silent-face-anti-spoofing. <https://github.com/minivision-ai/Silent-Face-Anti-Spoofing>. Accessed 10 July 2021
4. George, A., Marcel, S.: Deep pixel-wise binary supervision for face presentation attack detection. [arXiv:1907.04047v1](https://arxiv.org/abs/1907.04047v1) (2019)
5. Hossain, M.S., Ruptry, L., Roy, K., Hasan, M., Sengupta, S., Mohammed, N.: A DeepPixBis: attentional angular margin for face anti-spoofing. In: Digital Image Computing: Techniques and Applications (2020)
6. Jia, Y., Zhang, J., Shan, S., Chen, X.: Single-side domain generalization for face anti-spoofing. [arXiv:2004.14043v1](https://arxiv.org/abs/2004.14043v1) (2020)
7. Zhang, Y., Yin, Z., Li, Y., Yin, G., Yan, J., Shao, J., Liu, Z.: CelebA-spoof: large-scale face anti-spoofing dataset with rich annotations. [arXiv:2007.12342v3](https://arxiv.org/abs/2007.12342v3) (2020)
8. Tan, X., Li, Y., Liu, J., Jiang, L.: Face liveness detection from a single image with sparse low rank bilinear discriminative model. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6316, pp. 504–517. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-15567-3\\_37](https://doi.org/10.1007/978-3-642-15567-3_37)
9. Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.Z.: A face anti-spoofing database with diverse attacks. In: 5th IAPR international conference on biometrics, pp. 26–31 (2012)
10. Li, H., Li, W., Cao, H., Wang, S., Huang, F., Kot, A.C.: Unsupervised domain adaptation for face anti-spoofing. In: IEEE Transactions on Information Forensics and Security (2018). <https://doi.org/10.1109/TIFS.2021.9667073>
11. Deng, J., Guo, J., Zhou, Y., Yu, J., Kotsia, I., Zafeiriou, S.: RetinaFace: single-stage dense face localisation in the wild. [arXiv:1905.00641v2](https://arxiv.org/abs/1905.00641v2) (2019)
12. YOLOv5. <https://github.com/ultralytics/yolov5>. Accessed 1 Dec 2021
13. Jetson AGX Xavier developer kit homepage. <https://developer.nvidia.com/embedded/jetson-agx-xavier-developer-kit>. Accessed 9 Nov 2021
14. HiK vision home page. <https://www.hikvision.com/en/products/IP-Products/Network-Cameras/Pro-Series-EasyIP-/ds-2cd2023g2-i-u-/>



# Vietnamese Spelling Error Detection and Correction Using BERT and N-gram Language Model

Dong Nguyen Tien<sup>1</sup>, Tuoi Tran Thi Minh<sup>1</sup>, Loi Le Vu<sup>1,3</sup>, and Tuan Dang Minh<sup>1,2(✉)</sup>

<sup>1</sup> CMC Institute of Science Technology, No. 11 Duy Tan Street, Hanoi, Vietnam  
dmtuan1@cmc.com.vn

<sup>2</sup> Posts and Telecommunications Institute of Technology, Km10, Nguyen Trai Street,  
Hanoi, Vietnam  
tuandm@ptit.edu.vn

<sup>3</sup> Hanoi University of Science and Technology, No. 1 Dai Co Viet Street, Hanoi, Vietnam

**Abstract.** Vietnamese spelling error detection and correction is a crucial task in Natural language processing, it plays an important role in many different real-world applications. Although there is a lot of research on it, dealing with diverse types of errors in Vietnamese is still a challenge. In this paper, we propose a model to help detect and correct some specific Vietnamese spelling errors by combining a pre-trained neural network-based Vietnamese language model and N-gram language model. We also provide a clear definition of handleable error types, error generation rules in the training set and evaluate our proposed model on a Vietnamese benchmark dataset at the word level. The experimental results show that our model achieves higher than from 1% to 14% f1-score than other neural network-based pre-trained language models in detection and make comparisons with bi, tri and 4-g language models to choose the best model for correction.

**Keywords:** Vietnamese spell correction · Error detection · Deep learning · Language model

## 1 Introduction

Spelling error detection and correction are an important task and are applied in many systems such as word segmentation [1], named-entity recognition [2], OCR [3]. This is also a problem with great challenges because of the variety of spelling errors and the complexity of the language. To solve the problem, spelling correction systems require a high level of language understanding.

Spelling detection and correction task aims at detecting and converting misspelled words into the correct form in a text of a given language. It plays a part in pre-processing or post-processing of many Natural language processing and Computer Vision. For example, spell correction system can give correction suggestions when there are typographical errors in queries, improve information retrieval results of search engines. It can also be integrated into the post-processing step of the existing Automatic Speech

Recognition (ASR) or Optical Character Recognition (OCR) to increase the recognition quality.

In this paper, we consider open-domain Vietnamese spelling correction. Spelling errors are divided into two main categories: non-word error and real-word error. Non-word error is an error that does not exist in the dictionary, and vice versa, real-word error is an error that exists in the dictionary. Therefore, a non-word error can be easily identified by using a dictionary to check if a word is in the dictionary while identifying a real-word error depends on the context.

There is a lot of research on spelling correction for other languages such as English, Chinese, etc. However, as far as we know, there have not been many studies on spelling correction in Vietnamese. Along with the recent development of language representation models such as BERT to help systems understand languages at a better level, Vietnamese spelling correction is still a problem with great potential in research and development. Therefore, in this work, we introduce a Vietnamese spelling error system comprising two steps: error detection and error correction using pre-trained language model based on BERT and N-gram language model.

Finally, our contributions are:

- A 2-stage Vietnamese spelling error detection and correction method using both Vietnamese neural networks based pre-trained language model and n-gram language models
- Clarify specific types of errors that can occur in practice and define them specifically. Including those kinds of errors that other methods couldn't handle yet
- A public dataset of human-made spelling errors, this dataset is evaluated the proposed method on the various approach

## 2 Related Works

There are a number of approaches proposed for the Vietnamese spelling correction task. Early works followed the steps: error detection, candidate generation and candidate selection. In 2008, PH. Nguyen et al. [4] proposed a two-phase spelling correction method: error identification and correction. In the first phase, the authors used dictionaries such as syllable, proper name, etc. to identify non-word errors. For real-word errors, the authors used a word-level bi-gram combined with a bi-gram POS and some heuristics such as co-occurrence probabilistic and the stopwords list. In the correction phase, the authors generated candidates by employing typical spelling errors and confusion sets specified by Hoang Phe dictionary then ranking candidates using SoundEx, Minimum Edit Distance algorithms and some other measures such as word occurrence frequencies and word lengths.

Nguyen Thi Xuan Huong et al. [5] built an N-gram language model for Vietnamese spell correction. Each syllable is assigned a score calculated based on the frequency of occurrence. Correction candidates of each syllable are generated using edit distance and chosen previous candidates. One syllable is considered as an error if at least one of its candidates has a higher score than its. Another N-gram approach proposed by Quoc-Dung Nguyen et al. [3] was used for Vietnamese OCR post-processing. Correction candidates

are generated and scored based on linguistic features. Those papers show that N-gram is still effective in solving spelling correction task in spite of being an old approach.

H. Nguyen, 2015 [6] built a tool to normalize Vietnamese tweets on Twitter. The spelling correction model consisting of 2 steps: error identification and correction. The authors built a dictionary of Vietnamese words and defined a word as an error if it does not exist in the dictionary. The correct form of each error is considered as the word in dictionary that has the highest similarity score with the error.

Most recently, H Tran et al., 2021 [7] proposed a Hierarchical Transformer model for Vietnamese spelling correction. The model uses two Transformer encoders: Character-level encoder and Word-level encoder to represent words on two levels of words and characters. Then the word representation goes through the detection classifier to determine if a word is misspelled or not and finally a correction classifier to suggest the replacement word for the misspelled word.

### 3 The Proposed Model

#### 3.1 Problem Formulation

Vietnamese spelling error detection and correction (VSEDC) is formulated as follows: Given a sentence of  $n$  words  $X = (x_1, x_2, \dots, x_n)$ , VSEDC aims at transforming  $X$  into a new sentence  $X'$  of  $m$  words  $X' = (x'_1, x'_2, \dots, x'_m)$ , with  $m \geq n$ . Specifically, each misspelled word in the input sentence  $X$  is replaced with one or more correct words in the output sentence  $X'$ .

In this paper, we proposed a model to solve 3 following types of error:

- **MISSING\_SPACE:** As the name suggests, this error happens when 2 consecutive words are stucked together, example: *chính tá* - *chínhtá*. Another example for this error is *chính tá* - *chinhstar*, which happens very regularly when typing. We consider this error type as non-word error since we observed that there is no word of this error type appearing in Vietnamese vocabulary. A word of **MISSING\_SPACE** error will be replaced by at least 2 correct words.
- **EDITS:** This error happens when a unit in a word is altered, insert, delete, replace or permute, where a unit is defined as a vowel (exclude of accents), i.e.: *a, e, i, o, u, y* or a consonant (include 2 or 3-letter consonants), i.e.: *b, ngh*. For example, *súa lõi* - *xíxa lõi*, *chính tá* - *thính tá*. For this type of error, we only consider real-word error since non-word error can be easily addressed by checking the vocabulary.
- **MISSING\_ACCENT:** This error type happens when a vowel misses its accents, e.g.: *súa lõi* - *sua lõi*, *chính tá* - *chinh ta*. **MISSING\_ACCENT** is separated from **EDITS** in error correction phase since it includes cases in which there are more than 1 word being misspelled.

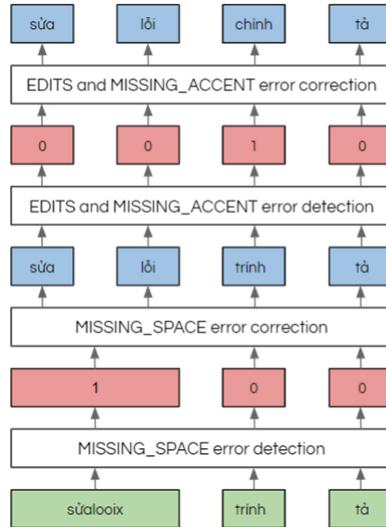
#### 3.2 Data Preprocessing

A characteristic of Vietnamese language is the appearance of accents, e.g.: *á, à, á̂, á̄, á̄̂, á̄̄*, ... These letters are not available on the keyboard. Instead, they are combinations of

several other letters as defined by Telex input method, e.g., *as* - á, *af* - à, *ar* - å. We call such combinations (*as*, *af*, *ar*) telex form and the results displayed by these combinations (á, à, å) accent form. Because of using the Telex input method, there likely exist letters in telex form when people make mistakes in typing. Hence, before detecting and correcting errors, we replace all occurrences of telex form with their corresponding accent form.

### 3.3 Model

We proposed a spell correction model including two phases: error detection and error correction. Figure 1 is the illustration of our proposed model. As shown in Fig. 1.



**Fig. 1.** Overview of Vietnamese spelling error detection and correction (VSEDC)

### 3.4 Error Detection

Input of error detection is a sequence of n words  $X = (x_1, x_2, \dots, x_n)$ , and output is a sequence of n labels  $Y = (y_1, y_2, \dots, y_i, \dots, y_n)$ ,  $y_i \in \{0, 1\}$ , where  $y_i$  is the label of word  $x_i$ . For **MISSING\_SPACE**, we defined a regex pattern to detect **MISSING\_SPACE** errors based on the observation that there are no words in Vietnamese in the format vowel - consonant - vowel. We built a Vietnamese pre-trained language model based on BERT called CMCBert. CMCBert constructs word representations at the subword level using sentencepiece tokenizer. In this work, we fine-tuned CMCBert for binary token classification task on a small amount of data annotated for the purpose of detecting **EDITS** and **MISSING\_ACCENT** errors. Token classifier classifies a subword into one of the two classes: 1 for **EDITS** and **MISSING\_ACCENT** error and 0 otherwise. A word is defined as a error if at least one of its subwords has label 1. Details of our model are shown in Fig. 2.

**BERT** [8] is a bi-directional pre-trained language model that bases on Transformer architecture. Transformer includes two main components, which are encoder and decoder, where each component is a stack of several identical layers. Each layer in encoder contains 2 sub-layers: multi-head self-attention and position-wise fully connected feed-forward. Each layer in decoder also contains these two sub-layers, but has an additional masked multi-head self-attention sub-layer to guarantee that prediction at position  $i$  is not affected by output at positions greater than  $i$ . This is done by removing positions greater than  $i$  from attention mechanism. Multi-head self-attention are defined as follows:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^0 \quad (1)$$

$$\text{where } \text{head}_i = \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right) \quad (2)$$

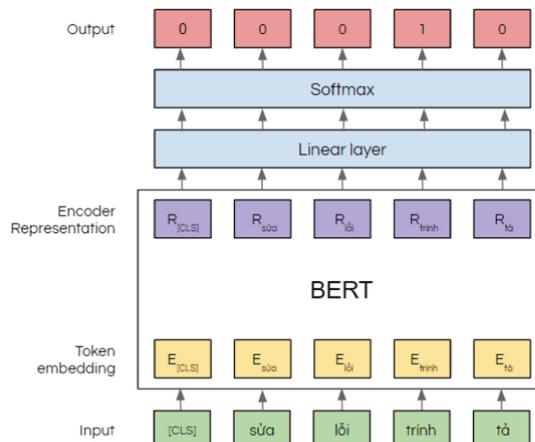
$$\text{FFN}(X) = \max(0, XW_1 + b_1)W_2 + b_2 \quad (3)$$

We denote the sequence of hidden states of the last layer of BERT as  $H = (h_1, h_2, \dots, h_n)$ . Then, the probability of a word  $x_i$  belonging to class  $i, j \in \{0, 1\}$  is defined as:

$$p_{ij} = P(y_i = j|X) = \text{softmax}(Wh_i + b) \quad (4)$$

**Q**, **K**, **V** are matrices representing the inputs of each multi-head self-attention sub-layer.  $W^o, W_i^Q, W_i^K, W_i^V, W_1, b_1, W_2, b_2, W, b$  are parameters and objective function is cross-entropy loss:

$$L = \frac{-1}{N} \sum_{i=0}^N \sum_{j \in \{0, 1\}} y_{ij} \log p_{ij} \quad (5)$$



**Fig. 2.** Error detection model for **EDITS** and **MISSING\_AGENT** errors

### 3.5 Error Correction

Input of error correction is a sequence of n words  $X = (x_1, x_2, \dots, x_n)$  and n corresponding labels  $Y = (y_1, y_2, \dots, y_i, \dots, y_n)$ ,  $y_i \in \{0, 1\}$ . The output is a sequence of m words that are corrected  $X = (x'_1, x'_2, \dots, x'_m)$ ,  $m \geq n$ . **MISSING\_SPACE** errors are corrected before detecting and correcting the other two error types. We built a vocabulary V including telex form of all single-syllable in Vietnamese. This vocabulary is used to check if a word exists, the existence of a word is defined as the appearance of the that word in the vocabulary. Error correction includes the following steps. For each misspelled word, a set of candidates is generated using edit distance and scored by N-gram language model, then top k candidates have highest scores are kept. We set  $k = 1$  for **MISSING\_SPACE** error type and  $k = 10$  for **EDITS** and **MISSING\_ACCENT**.

N-gram language model [9] is a probabilistic language model that assigns probability for a sentence on the basis of Markov assumption. Given a sequence  $P = (w_1, w_2, \dots, w_n)$ , the N-gram language model approximates the probability of the next word in a sequence using conditional probability defined as:

$$P(w_n|w_{1:n-1})P(w_n \vee w_{n-N+1:n-1}) \quad (6)$$

Score of the sequence P is defined as:

$$P(w_{1:n}) = \prod P(w_k|w_{k-N+1:k-1}) \quad (7)$$

We built bi-gram language models ( $N = 2$ ), tri-gram language models ( $N = 3$ ) and 4-g language models ( $N = 4$ ) using KenLM [10] language model library. For each error type, we use different candidate generation methods. For **MISSING\_SPACE** error type, candidates are generated by separating the error word into pairs of subwords at positions j, j + 1, etc. to the end of word, where j is the position of the first vowel in that word. Pairs which contain at least one subword existing in the vocabulary will be kept. This procedure is repeated until all **MISSING\_SPACE** errors are removed. For **EDITS** error type, before candidate generation, we convert words into their telex form. Candidates are generated by inserting, deleting a unit, substituting a unit by another one or permuting positions of 2 consecutive units. Then candidates that exist in the vocabulary will be kept. If there are no valid candidates, we keep the original misspelled word. For **MISSING\_ACCENT** error type, we generate candidates by adding diacritics or tones to the vowels of misspelled words. Like **EDITS**, we only keep candidates that exist in the vocabulary and if there are no valid candidates, we keep the original word. After generating candidates, we convert all candidates to accent form as the results.

## 4 Experiments and Results

### 4.1 Data Preparation

We built a dataset named ARTICLES containing 12,300,000 sentences collected from articles in Health Services, Education, Information Technology, etc. We generated errors in the following ways. The **MISSING\_SPACE** error is generated by removing space

between random adjacent words and converting one or some of them into telex form. For a random word, the EDITS error is caused by insertions, deletions, or substitutions of a single character or permutation of two adjacent characters, the MISSING\_ACCENT error is obtained by removing all accent marks of vowels. The MISSING\_SPACE error type contains 100% non-word error, EDITS and MISSING\_ACCENT error type contains 100% real-word error. The average error rate per sentence is 20%.

This dataset was then divided into three datasets ARTICLES-KENLM, ARTICLES-TRAIN and ARTICLES-TEST. The ARTICLES-TRAIN contains 1,600,000 sentences used to build an Error Detector in the Error Detection step. The ARTICLES-KENLM contains 10,000,000 sentences used to build a language model to rank the candidates in the Error Correction step. We performed evaluations on the ARTICLES-TEST containing 700,000 sentences. The total number of errors in datasets are described in Table 1.

**Table 1.** The total number of errors in the training and test dataset

	Articles-train	Articles-test
MISSING_SPACE	0	0
EDITS	2,365,995	2,365,995
MISSING_ACCENT	3,726,560	3726560

## 4.2 Experiments Settings

In fine-tuning of CMCBert for training Error Detector, we retained the default hyper-parameters and fine-tuned the parameters using Adam optimizer. We used the learning rate scheduler with the warm-up step of **0.1** and the small learning rate of **2e-5**. The model was trained for 300,000 steps (3 epochs). We also built a 2, 3, and 4-g language model using the Kenlm for ranking candidates in the Error Correction step. We utilized recall, precision, and F1-score metrics to evaluate Error Detector performance as in most classification tasks. For Error Corrector, we used two metrics w2t and t2w are defined as follows:

$$w2t = \frac{c_{w2t}}{c_w} \quad t2w = \frac{c_{t2w}}{c_t} \quad (8)$$

$c_{w2t}$  is the number of errors converted into the correct form,  $c_{t2w}$  is the number of non-errors unexpectedly converted into the wrong form,  $c_w$  is the total number of errors, and  $c_t$  is the total number of non-errors.

## 4.3 Results

We performed evaluations on 700,000 sentences in the ARTICLES-TEST dataset. First, we individually evaluate the Error Detection step and Error Correction step, which

means we use ground truth as the input for the Error Correction. The results of the Error Detection step and the Error Correction step are shown in Tables 2 and 3. As shown in Table 2, our model performs slightly better than PhoBert at **MISSING\_ACCENT** error. Additionally, CMCBert does not require word segmentation when pre-processing data, thus significantly reducing inference time when applying in practice. CMCBert and PhoBert greatly outperform the XLM-R for both **EDITS** and **MISSING\_ACCENT** errors. This suggests that monolingual language model is much better than multilingual one in monolingual Spelling Correction task. Table 3 shows that 4-g language model is better than 2 and 3-g on average. Since 4-g model can capture longer distance dependency than the other two. N-grams language model can perform better than N-1 g one but it requires much more data to build and its space complexity grows by an order of magnitude with base N. So, we chose 4-g language model. We report end-to-end results in Table 4. The results indicate that our proposal achieves high accuracy in error correction while minimizing the number of words unexpectedly converted to wrong form. The rate of **EDITS** errors corrected is the lowest because of its diversity, i.e., a word can be converted into many forms of **EDITS** errors.

**Table 2.** Evaluation results of Error Detection. F1-score is reported for our pre-trained language model, XLM-R base and PhoBert base

	CMCBert	PhoBert [11]	XLM-R [12]
MISSING_SPACE	0.95		
EDITS	0.95	0.94	0.81
MISSING_ACCENT	0.97	0.97	0.83

**Table 3.** Evaluation results of Error Correction, ground truth is used as input of Error Correction step, so we ignored the t2w metric

	w2t		
	N = 2	N = 3	N = 4
MISSING_SPACE	0.99	0.99	0.99
EDITS	0.83	0.89	0.86
MISSING_ACCENT	0.78	0.89	0.95

**Table 4.** Evaluation results on combined Error Detection and Error Correction steps

	w2t	t2w
MISSING_SPACE	0.93	0.01
EDITS	0.83	0.00
MISSING_ACCENT	0.90	0.00

## 5 Conclusion

This paper presented Vietnamese spelling error detection and correction using BERT and N-gram language model. The experiment demonstrated that the proposed method takes advantage of both pre-trained and n-gram language models to correct three types of errors: missing space, edits and missing accent. The model was built on 12,300,000 sentences from a large corpus of high-quality resources. In spite of meeting our expectations, the model does not have the ability to solve more diverse error types. For better and more impressive results, in future work, we intend to design a more robust method to solve more Vietnamese spelling error types.

**Acknowledgment.** We gratefully acknowledge the support from the CMC Institute of Science Technology for funding the research project.

## References

1. Nguyen, H.T., Dang, T., Le, C.A., Nguyen, H.T., Dang, T.-T., Le, A.-C.: Adapting Vietnamese word segmentation for microblogs-style data (2014). <https://www.researchgate.net/publication/283502832>
2. Nguyen, V.H., Nguyen, H.T., Snasel, V.: Named entity recognition in Vietnamese tweets. In: Thai, M.T., Nguyen, N.P., Shen, H. (eds.) CSoNet 2015. LNCS, vol. 9197, pp. 205–215. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-21786-4\\_18](https://doi.org/10.1007/978-3-319-21786-4_18)
3. Nguyen, Q.D., Le, D.A., Zelinka, I.: OCR error correction for unconstrained Vietnamese handwritten text. PervasiveHealth Pervasive Comput. Technol. Healthc. 132–138 (2019). <https://doi.org/10.1145/3368926.3369686>
4. Nguyen, P.H., Ngo, T.D., Phan, D.A., Dinh, T.P.T., Huynh, T.Q.: Vietnamese spelling detection and correction: using Bi-gram, minimum edit distance, SoundEx algorithms with some additional heuristics. In: RIVF 2008 - 2008 IEEE International Conference on Research, Innovation and Vision for the Future in Computing and Communication Technologies, pp. 96–102 (2008). <https://doi.org/10.1109/RIVF.2008.4586339>
5. Nguyen, H.T., Dang, T.B., Nguyen, L.M.: Deep learning approach for vietnamese consonant misspell correction. In: Nguyen, L.-M., Phan, X.-H., Hasida, K., Tojo, S. (eds.) PACLING 2019. CCIS, vol. 1215, pp. 497–504. Springer, Singapore (2020). [https://doi.org/10.1007/978-981-15-6168-9\\_40](https://doi.org/10.1007/978-981-15-6168-9_40)
6. Nguyen, V.H., Nguyen, H.T., Snasel, V.: Normalization of Vietnamese tweets on Twitter. In: Abraham, A., Jiang, X.H., Snášel, V., Pan, J.-S. (eds.) Intelligent Data Analysis and Applications. AISC, vol. 370, pp. 179–189. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-21206-7\\_16](https://doi.org/10.1007/978-3-319-21206-7_16)

7. Tran, H., Dinh, C.V., Phan, L., Nguyen, S.T.: Hierarchical transformer encoders for Vietnamese spelling correction, May 2021. <http://arxiv.org/abs/2105.13578>
8. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of Conference, vol. 1, no. M1m, pp. 4171–4186 (2019)
9. Lin, J.: N-Gram Language Models N-Gram Language Models (2009)
10. Heafield, K.: KenLM: faster and smaller language model queries. In: Proceedings of Sixth Workshop on Statistical Machine Translation, pp. 187–197 (2011)
11. Nguyen, D.Q., Nguyen, A.T.: PhoBert: pre-trained language models for Vietnamese, pp. 1037–1042 (2020). <https://doi.org/10.18653/v1/2020.findings-emnlp.92>
12. Ruder, S., Søgaard, A., Vulic, I.: Unsupervised cross-lingual representation learning. In: ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Tutoring Abstracts, pp. 31–38 (2019). <https://doi.org/10.18653/v1/p19-4007>



# Clothing Recognition Using Transfer Learning and Segmentation Techniques

Hai Thanh Nguyen<sup>1</sup>, Thu Anh Thi Nguyen<sup>1</sup>, Khanh Kim Nguyen<sup>1</sup>,  
Chi Le Hoang Tran<sup>2</sup>, Thai Minh Phan Le<sup>1</sup>, Diem Ngoc Thi Pham<sup>1</sup>,  
and Dien Thanh Tran<sup>1</sup>(✉)

<sup>1</sup> College of Information and Communication Technology, Can Tho University,  
Can Tho 900100, Vietnam

{nthai.cit,ptndiem,ttdien}@ctu.edu.vn

<sup>2</sup> Department of Information Technology, FPT Polytechnic,  
Can Tho 900100, Vietnam

**Abstract.** The fashion industry has been having a significant influence on many other fields. In which identification and classification of clothes play a crucial role. It is applied in e-commerce systems to search and recommend suitable outfits for customers, track and capture trending outfits for suitable business solutions, etc. In this study, we present an approach using transfer learning and segmentation techniques to identify types of clothes. We compare and evaluate the classification results of some transfer learning approaches such as InceptionV3, AlexNet, and Transfer Forest. First, we have modified the ImageNet dataset by filtering out mislabeled data. Then we performed data enhancement by changing a few training metrics and increasing the number of training images to 320,000 images, including 45°, 90°, 180° rotated images of the images from the refined dataset. The experimental results have an accuracy of 0.51 with the dataset that we have modified and can reach 0.97 in accuracy in three-class classification tasks. Finally, we have performed segmentation of clothing images with Unet for supporting further analysis on detailed characteristics of clothes.

**Keywords:** Classification of clothes · InceptionV3 · Transfer learning · Fashion

## 1 Introduction

Image classification is no longer a strange problem for many people and is used in many fields, such as identifying objects, animals, people, etc.; clothing identification stands out with many important meanings. Therefore, developers can integrate it into their systems to assist users in finding clothing information through images. At the same time, fashion merchants can also grasp what kind of clothing is of interest and high demand, thereby improving business measures. Besides, it is also applied to suggest clothes for users by personal photos. Thus, they can effortlessly search, choose and buy clothes that are suitable for their fashion style as well as their financial ability. However, this is not an easy task for

many reasons. For example, clothing is too diverse regarding designs and types and input images' quality. Moreover, we need to know how to distinguish the clothing from other objects and identify the color of the clothing in case there are too many different colors in the picture.

The current clothing classification models have not high accuracy when the number of classification classes is too many, and many studies have not mentioned the problem of color classification for clothing. This study classifies the clothing on a multi-layered dataset using the InceptionV3 network. Our contributions include:

- We filter the data by replacing or deleting inappropriate images, shrinking the dataset. Then, we enhanced the image and expanded our dataset by rotating the image in the original dataset.
- We have deployed some transfer learning techniques on clothing classification and compared the results.
- We have also attempted an Unet architecture to perform segmentation tasks on clothes.

## 2 Related Work

There are numerous studies on cloths image recognition have been presented. For example, Lukas Bossard et al. [1] proposed the Random Forest method [2] to classify fifteen types of clothing. Besides, the authors have also introduced a complete process to recognize and classify people's clothes in natural scenes and recognize styles, colors, materials, etc., to get the same clothing description as a human being. Besides, the algorithm also identifies relevant image regions by upper body detectors<sup>1</sup>. The results show that the proposed algorithm gives better accuracy than other methods such as SVM. Another study has been done by Bryan Willimon et al. [3], which offers a multi-class approach to classify clothing items in a laundry. Classification features include color, texture, shape, and edge information from 2D and 3D data from local and global perspectives. Research has introduced a new classification method called L-M-H, specifically L-C-S-H (low-level feature selection and high-level masking), to classify clothes. The multilayer method divides the problem into a high class (H), multiple medium layers (feature (C), selection mask (S)), and a low layer (L). Qing Tian et al. [4] proposed a method to group together categories with high visual and semantic similarity and then perform detection to Improve Apparel Detection with Category Grouping and Multi-grained branches. Test results on DeepFashion2 and OpenImagesV4-Clothing show an overall performance increase of 2.3 mAP for DeepFashion2 and 2.5 mAP for OpenImagesV4-Clothing without the need for additional time cumulative annotation. Another clothing classification study was presented in [5]. The Apparel Classification for Indian Ethnic Clothes is based on Atlas and Generative Indian Fashion datasets, with 106K images and fifteen different cloth categories from various e-commerce websites. The models

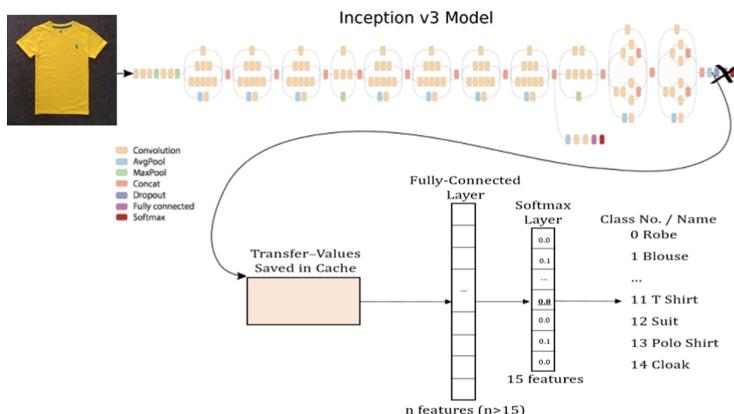
---

<sup>1</sup> [http://calvin-vision.net/bigstuff/calvin\\_upperbody\\_detector/](http://calvin-vision.net/bigstuff/calvin_upperbody_detector/).

used in the research are ResNet-18, ResNet-50, and ResNet-101 networks. The result of ResNet-50 is the highest. Besides, from the classification of clothes, we can also orient fashion and human development as in the study by Wei-Lin Hsiao and Kristen [6] presented. The study used ResNet-18 pre-trained ImageNet on DeepFashion and Mask-RCNN to classify clothes, from which to step by step calculate and predict current and future fashion trends.

### 3 Method

This study uses the InceptionV3 [7] model-based classifier built on Tensorflow<sup>2</sup> [8] as exhibited in Fig. 1. Our classification model is mainly based on the Apparel Classification with Style (ACS) dataset [1], containing 89,484 images collected from ImageNet, divided into fifteen classes (detail in Table 1). For each training category, the image is divided in a ratio of 8:1:1, i.e., 80% for training, 10% for validation, and 10% for testing. Most of the images in the ImageNet set are collected from the Internet, so they are out of sync in size and color. Therefore, the image is refined to a maximum edge length of 320 pixels through the cleaning step, and the histogram of each color channel is normalized [1]. We found that the dataset contains many noisy and misclassified images during the training process, which significantly affects the classification results, so we decided to filter the data again by replacing or deleting inappropriate images, shrinking the dataset to 80,000 images. The dataset after processing is presented in Table 1. We perform data augmentation for the training process to avoid putting too many constraints on the input image when testing the model. We enhanced the image by flipping left, flipping right, cropping, zooming, changing the brightness, and expanding our dataset of 80,000 images to 320,000 images by rotating the image in the original dataset of angles 45°, 90°, 180°. This is a way of improving the



**Fig. 1.** The architecture of the InceptionV3 [8] for clothing classification.

<sup>2</sup> <https://cloud.google.com/tpu/docs/inception-v3-advanced>.

**Table 1.** The dataset description.

Category	The original dataset [1]		Dataset after processing	
	Images	Boxes	Images	Boxes
Long dress	22,372	12,622	12,622	9,368
Coat	18,782	11,338	11,338	10,530
Jacket	17,848	11,719	11,719	9,761
Cloak	15,444	9,371	9,371	8,326
Robe	13,327	7,262	7,262	5,554
Suit	12,971	7,573	7,573	7,966
Undergarment	10,881	6,927	6,927	5,994
Uniform	8,830	4,194	4,194	3,959
Sweater	8,393	6,515	6,515	6,378
Short dress	7,547	5,360	5,360	5,328
Shirt	3,140	1,784	1,784	1,784
T-shirt	2,339	1,784	1,784	1,775
Blouses	1,344	1,121	1,121	1,093
Vest	1,261	938	938	1,207
Polo shirt	1,239	976	976	977
Total	<b>145,718</b>	<b>89,484</b>	<b>89,484</b>	<b>80,000</b>

dataset, which is the most straightforward and most helpful. However, because the number of training images increased by four times compared to the original, the problem of training time may be one of our obstacles; so in this study, we only perform the training 5,000 steps for a set of 80,000 images and 320,000 images when adjusting the model's indexes, for example flip from left to right, cropping, brightness and scaling randomly with 10% for data augmentation.

## 4 Experimental Results

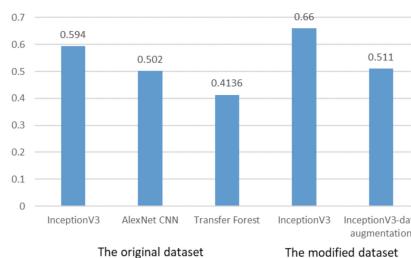
We evaluate and compare the performance of some famous convolutional neural networks such as InceptionV3 [7], AlexNet[9] on the classification tasks with three classes and 15 classes of clothing. We have deployed InceptionV3<sup>3</sup> to train a model for classifying three classes of clothing (sweater, jacket, shirt); each layer consists of 1,000 images extracted from the ACS [1] original dataset, the accuracy achieved when using InceptionV3 architecture [8] in Tensorflow reaches 97.30% with 1,000 steps (Table 2). To identify more clothing types instead of only three types, we use the entire ACS dataset with fifteen classes; the result after training with 5,000 steps is 59.40%, better than AlexNet CNN (50.20%) in [10], Transfer Forest (41.36%) in [1] as shown in Table 2. After reviewing the images in the ACS

<sup>3</sup> <https://github.com/chhari/Classification-of-Clothes>.

**Table 2.** The detailed classification results.

Model	The original dataset			The modified dataset			
	InceptionV3	AlexNet CNN	Transfer Forest	InceptionV3		InceptionV3 with data augmentation	
Number of categories	3	15	15	15	15	15	15
Number of steps	1,000	5,000	5,000	5,000	5,000	10,000	100,000
ACC	0.9700	0.5940	0.5020	0.4136	0.6600	0.6700	0.6940
							0.5110

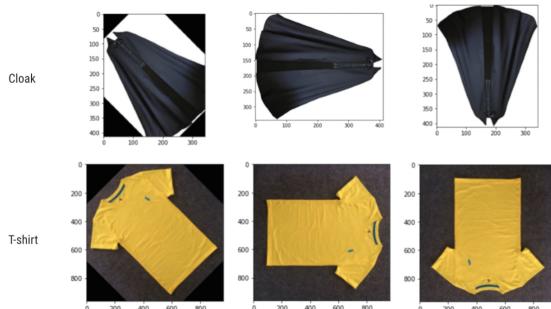
dataset, we noticed that too many images were mislabeled. We believe that this may be the reason for the low accuracy of the classification model. Therefore, we decided to re-filter the dataset by re-labeling the mislabeled images so that the model can better classify. After training the dataset that filtered out the wrong images, the accuracy is 66% with 5,000 steps (Table 2). The classification results increase significantly when training with the filtered dataset, but some images show that the accuracy has increased compared to the previous training but still low. Besides, there are still some misclassification results because this clothing has many similarities in terms of design. Besides, the small details to classify different clothing items have the same color. After increasing the number of steps to train the model to 10,000 and 100,000 steps, the accuracy is 67% and 69.40%, respectively. When the outfit in the photo is placed in the correct position, most of the classification results are correct. To test the classifier ability of the model on images out of the dataset, we perform rotation before classifying it (Fig. 3), the results have shown in Table 3 (Solution 1). There are a lot of false classification results for 45°, 90°, 180° rotated images using the original training model; we also tried to use 10,000 steps and 100,000 steps running model to try to classify the rotated images. However, the classification result does not change but only increases the classification percentage. We found that many factors affect the classification of an image, such as flipping left, flipping right, scaling, cropping, brightness. Moreover, these factors are fully adjustable right on our model and do not increase the dataset. The results for 5,000 steps training with metrics adjusted (Table 1) are shown in Table 3 (Solution 2 row). The classification results are slightly improved compared to the unadjusted indexes, but

**Fig. 2.** Comparison chart of classification results (detail in Table 2).

**Table 3.** Compare results with different solutions.

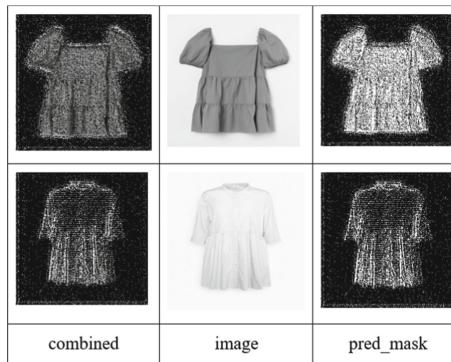
	Category	Short dress			Cloak			T-shirt		
		Rotate	45°	90°	180°	45°	90°	180°	45°	90°
Solution 1	Result	F	F	T	T	T	F	T	F	F
	ACC	0.2663	0.3597	0.4788	0.3103	0.1999	0.1965	0.6316	0.3366	0.3905
Solution 2	Result	F	F	T	T	T	F	T	T	F
	ACC	0.7137	0.3494	0.5433	0.3206	0.2018	0.1639	0.7055	0.4423	0.3434
Solution 3	Result	F	T	T	T	T	T	T	T	T
	ACC	0.3820	0.5161	0.6503	0.2615	0.2304	0.1822	0.6837	0.5331	0.4404

the classification percentage is still relatively low. For the model to learn images when rotating the angles as tested, we cannot capture or collect them from the Internet because it would be very time-consuming to find, filter, and label those images. Therefore, we have three rotation angles, each angle can increase the dataset by 80,000 images to have a dataset of 320,000 images, including the original image, including 45°, 90°, 180° rotated images. The accuracy is 51.10% (detail in Table 2) of 5,000 steps training on this dataset plus adjusting the metrics in Table 1 and Table 3 (Solution 3 row) show the example classify results on images out of the dataset with some improved results (Fig. 2).

**Fig. 3.** The illustration of images for clothing classification.

Besides classifying clothes, color recognition also plays a significant role in increasing accuracy in finding clothes [11]. To identify the color of the outfit, it is necessary first to identify the area of the garment in the image. The goal is to train a model that can divide a classified image into regions belonging to the clothes and regions other than the clothes. From there, remove the unnecessary parts and keep only the clothes. Lung segmentation model<sup>4</sup>, a model built on U-Net [12] to remove the unrelated parts of the lung. The model based on a dataset consisting of 1000 CT images associated with 1000 labels of lung contour images was used to train on a dataset of 100 short-dressed images with epoch = 10 and

<sup>4</sup> <https://github.com/limingwu8/Lung-Segmentation>.



**Fig. 4.** The segmentation results on clothes.

an image size of 128 px. The model, after training, has removed some parts of the clothes instead of just removing the outer parts of the clothes, we, in turn, increase the parameters such as the number of trained images, image size, batch size, the number of epoch, the learning rate, and perform training with each case, but the results received do not change much. After reviewing the results to see why the model was not performing well, we found that each image in the dataset was assigned a mask, the model predicted a pred\_mask; the clothes image is stripped of the outline (combined) is the original image plus pred\_mask. This study reused the available mask set to label the short dress set without changing it, which may affect the results of the trainable model. To test our hypothesis, we replaced all the images in the mask set of the Lung segmentation model with basic short dress images taken from the fashion set MNIST<sup>5</sup>, with the size of each image being 28 × 28 pixels. Results of training the model with 1000 short images as in Fig. 4, each image size is 256 px, epochs = 100, a learning rate of  $10^{-8}$  and a batch size of 8. The results are pretty good when testing on images of simple clothes, without people or background, but give the opposite results with images with many objects. We also tried to increase the training parameters, but the obtained results led us to believe that the Lung segmentation model is not suitable for the problem of determining the clothes region in the image.

## 5 Conclusion

There have been many studies on identification and classification in many fields; clothing is one of the fields that has been receiving attention and has numerous practical applications. We have deployed some transfer learning, including InceptionV3, which achieved an accuracy of 66% with 5,000 steps, 67% with 10,000 steps, and 69% with 100,000 steps on the processed ACS dataset. We also compared the results of the solutions to increase the classification accuracy to find

<sup>5</sup> <https://machinelearningmastery.com/how-to-develop-a-cnn-from-scratch-for-fashion-mnist-clothing-classification/>.

the most optimal solution. We performed data augmentation by changing a few training metrics and increasing the number of training images to 320,000 images, including 45°, 90°, 180° rotated images of the images from the refined dataset. The segmentation results with Unet have revealed that this method works well on medical images and other fields such as fashion and is expected to support further analysis with more details on each type of clothes. Further work can study to analyze and recognize characteristics of clothes, style, material, brand, etc., including color, texture, fashion styles, etc.

## References

1. Bossard, L., Dantone, M., Leistner, C., Wengert, C., Quack, T., Van Gool, L.: Apparel classification with style. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012. LNCS, vol. 7727, pp. 321–335. Springer, Heidelberg (2013). [https://doi.org/10.1007/978-3-642-37447-0\\_25](https://doi.org/10.1007/978-3-642-37447-0_25)
2. Breiman, L.: Mach. Learn. **45**(1), 5–32 (2001). <https://doi.org/10.1023/a:1010933404324>
3. Willimon, B., Walker, I., Birchfield, S.: Classification of clothing using midlevel layers. ISRN Robot. 1–17 (2013). <https://doi.org/10.5402/2013/630579>
4. Tian, Q., Chanda, S., Kumar, K.C.A., Gray, D.: Improving apparel detection with category grouping and multi-grained branches (2021)
5. Rajput, P.S., Aneja, S.: IndoFashion: apparel classification for Indian ethnic clothes. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, June 2021. <https://doi.org/10.1109/cvprw53098.2021.00440>
6. Hsiao, W.L., Grauman, K.: From culture to clothing: discovering the world events behind a century of fashion images (2021)
7. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818–2826 (2016)
8. Inception v3 deep convolutional architecture for classifying acute myeloid/lymphoblastic leukemia. <https://www.intel.com/content/www/us/en/developer/articles/technical/inception-v3-deep-convolutional-architecture-for-classifying-acute-myeloidlymphoblastic.html>
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Commun. ACM **60**(6), 84–90 (2017). <https://doi.org/10.1145/3065386>
10. Lao, B., Jagadeesh, K.A.: Convolutional neural networks for fashion classification and object detection (2015)
11. Rame, A., Douillard, A., Ollion, C.: Core: color regression for multiple colors fashion garments (2020)
12. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)



# An Evaluation of the Multi-probe Locality Sensitive Hashing for Large-Scale Face Feature Matching

Ha Nguyen-Xuan<sup>1</sup>(✉) , Dong Hoang-Nhu<sup>2</sup>, and Tuan Dang-Minh<sup>2,3</sup>

<sup>1</sup> Hanoi University of Science and Technology, 1 Dai Co Viet, Hai Ba Trung, Hanoi, Vietnam  
ha.nguyễnxuan@hust.edu.vn

<sup>2</sup> CMC Institute of Science and Technology, 11 Duy Tan, Cau Giay, Hanoi, Vietnam

<sup>3</sup> Posts and Telecommunication Institute of Technology, KM 10 Nguyen Trai, Ha Dong, Hanoi, Vietnam

**Abstract.** Large-scale face feature matching is a critical issue for many practical applications where the computing process requires a burden of hardware resources and computational time. In this work, the so-called multi-probe locality sensitive hashing (LSH) is exploited to solve the large-scale face feature matching problem. The standard procedure for hashing the dataset into smaller buckets and a matching algorithm with some improvements were implemented. Experimental implementations on a dataset of 12,720,006 images, corresponding to 617,970 person identifications, were carried out. The results show that extended LSH outperforms excellently in terms of accuracy and computational time. In comparison to linear matching, the computing time is reduced from 12446 ms to 267 ms for data type FP32 bits, and from 5,794 ms to 192 ms for data type INT8, while the accuracy is reduced very slightly. The results show the high potential use of the algorithm for practical applications.

**Keywords:** Data matching · Locality sensitive hashing · Face recognition

## 1 Introduction

With the development of image processing techniques and deep neural networks, there have been many image-based applications released. Among them, face detection and recognition have a great advancement in the accuracy that can be used for many practical applications [1]. One of the key issues with face recognition systems is the processing of large-scale databases. Besides ensuring accuracy, fast-processing time is required. For large-scale datasets, the matching time of a query face feature vector with the feature vectors in the database often requires much of the hardware resources and computing time.

For face recognition, each person has an ID, which is accompanied by several face images. Each face image is extracted into a face feature vector  $\mathbf{x}_i$  with a dimension of 512 and a type of float 32 bit (FP32) or integer 8 bit (INT8) via a face inference model. The face feature database contains a set of  $N$  vectors  $S_X \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ . When a face

needs to be detected, its image is also extracted into a feature vector  $\mathbf{q}$ . This face feature vector is then matched to each vector of  $S_N$  using cosine similarity according to Eq. (1). Two vectors are similar when their cosine distance  $d$  is bigger than a similarity threshold  $\tau$ . The best-matched vector  $\mathbf{x}_m$  is the one that has the maximum cosine similarity value. From  $\mathbf{x}_m$ , the person's ID is discovered. This is referred to as "linear matching".

$$d(\mathbf{q}, \mathbf{x}_i) = \frac{\mathbf{q} \cdot \mathbf{x}_i}{\|\mathbf{q}\| \|\mathbf{x}_i\|} \quad (i = 1 \div N) \quad (1)$$

If the dataset  $S_X$  consists of several or even a hundred million vectors, the linear matching requires a burden of computational resources and time. This makes the system cost-ineffective and cannot be used for real-time applications. Thus, fast matching algorithms for large-scale datasets are required.

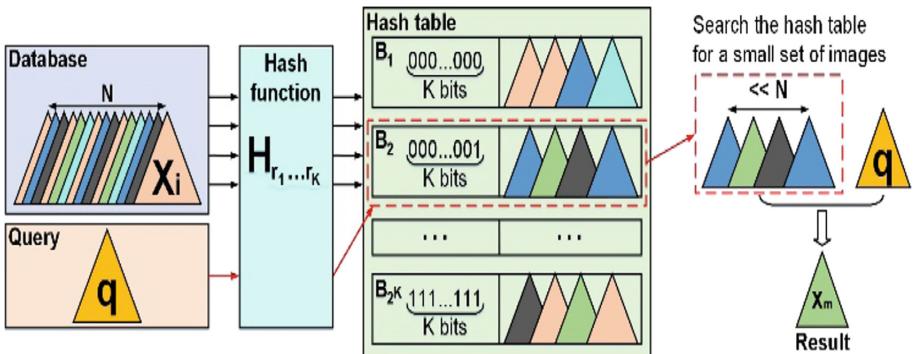
There have been many studies on the topic of high-dimensional large-scale data matching that used both hardware and software approaches [2–12]:

- For the hardware approach [2–5], the computation of the linear matching is accelerated via hardware-based parallel computing. The FPGA (Field Programmable Gate Array) is often used for computational acceleration [2–5]. However, the FPGA is only efficient when the data fits on the chips. Because of the limited hardware resources of edge computing devices, the data must be shuffled on or off chips, which is a bottleneck because of costly merging operations or data transfer time [3]. To solve this issue, a hybrid method that hashes the data into smaller sets and performs the acceleration of the linear matching on FPGA has been introduced [5]. However, for very large-scale datasets with limited hardware resources, software approaches are needed.
- For the software approach, the so-called Locality-Sensitive Hashing (LSH) is often used [6–12]. LSH is an algorithm that allows us to solve the nearest neighbor search problem in high-dimensional, large-scale datasets. This approach has proved to have many advantages in terms of computational efficiency. Recently, the library FALCONN, practical deployment of the LSH, has been introduced [7]. The library provides very efficient and well-tested implementations of LSH-based datasets. A thorough evaluation of the LSH for the similarity matching of large-scale datasets like web-scale query logs has been introduced [8]. However, the use of the LSH for large-scale face feature matching has not been evaluated yet.

In this work, the evaluation of the LSH on large-scale face feature datasets is performed. Based on the work of [8], we proposed an evaluation pipeline that allows us to evaluate the computational efficiency and accuracy. Several improvements to reduce the computational time and increase the accuracy were introduced. Experimental results are analyzed and discussed. Possible applications of this work's findings, as well as future studies, are also discussed.

## 2 Method

In this work, for the large-scale face feature matching, the so-called multi-probe Locality Sensitive Hashing (LSH), which was introduced in [1], is used. The baseline for the matching procedure is shown in Fig. 1. Based on this method, a hash function  $H_{r_1 \dots r_K}$ , which has  $K$  base vectors, is used to calculate the hash value of the vector  $x_i$  ( $i = 1 \div N$ ) of the database  $S_N$ . The hash value of each vector is a conversion of the  $K$  bit binary string to an integer number. The vectors that have an equal hash value are clustered into buckets. Consequently, the dataset  $S_X$  is clustered into  $2^K$  buckets of a hash table. Each bucket in the hash table has a hash value  $(h_1, h_2, \dots, h_{2^K})$ . When a query  $q$  is required, the hash function is used to calculate the hash value  $h_q$ . Vectors in the bucket with the hash value equal to  $h_q$  are extracted to a smaller dataset  $S_q$  as a result. The feature vector  $q$  is then linearly matched with the vectors of  $S_q$  using the cosine similarity to find the best-matched vector  $x_m$ .

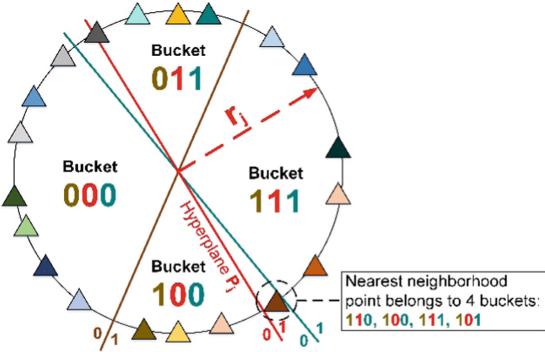


**Fig. 1.** Overall processing baseline of the LSH algorithm [8]

### 2.1 Hashing the Database

Each hash function consists of  $K$  base vec  $r_j$  ( $j = 1 \div K$ ) with a dimension of 512 that are randomly generated according to the Gaussian probabilistic distribution with a mean value of zero and a standard deviation equal to 1. Each base vector represents for a hyperplane  $P_j$ , which is used to hash the database into two binary regions, 0 or 1, as illustrated in Fig. 2. The hash value of each feature vector in the database  $S_N$  is calculated from the base vectors of the hash function according to Eq. (2). Each hash value of the feature vectors is a binary string with a length of  $K$  ( $b_0 b_1 \dots b_K$ ) that is then converted into an integer number in a range of  $[0, 2^{K-1}]$ . The feature vectors, which have a similar hash value, are then clustered into buckets ( $B_1, B_2, \dots, B_{2^K}$ ).

$$b_j = \begin{cases} 1 & \text{if } d(r_j, x_i) \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{with } (j = 1 \div K; i = 1 \div N) \quad (2)$$



**Fig. 2.** Illustration of hash functions

## 2.2 Hashing the Query Vector

In principle, the hashing function is also used to hash the query vector  $\mathbf{q}$  to obtain the hash value  $h_q$ . However, in fact, many data points are very near the hyperplanes. Thus, to make sure not to miss these points, the bits that correspond to the relating hyperplanes must be flipped. As illustrated in Fig. 2, there are three hyperplanes presented by three base vectors. The triangle points represent data points in the database. There appears the so-called nearest neighborhood point, which is very near two hyperplanes (the red and cobalt green). Therefore, when calculating the hash value of this point, the last two bits of the binary string must be flipped. As a result, this data point can belong to four buckets having the hash values 110, 100, 111, and 101.

The word “very near”, mentioned above, is quite quantitative. Therefore, we must define a criterion to determine when the bits of the binary string of the hash value must be flipped. In this work, the so-called “Distance Flip B”, which was introduced in [8], is used. The calculation of the bits for the hash value is given in Eq. (3). For each hyperplane, presented by the base vector  $\mathbf{r}_j$ , the cosine distance between the query vector and the base vector is computed and compared with a threshold  $s$ . If the condition  $-s < d(\mathbf{r}_j, \mathbf{q}) < s$  is satisfied, the query vector is very near the hyperplanes (local sensitive). Thus, the bit of the binary string containing the hash value must be flipped. Otherwise, the bit has only one value: 0 or 1.

$$b_j(\mathbf{q}) = \begin{cases} 0 & \text{if } d(\mathbf{r}_j, \mathbf{q}) \geq s \\ 0 & \text{if } d(\mathbf{r}_j, \mathbf{q}) \leq -s \\ 1 \text{ or } 0 & \text{if } -s < d(\mathbf{r}_j, \mathbf{q}) < s \end{cases} \quad \text{with } (j = 1 \div K) \quad (3)$$

From Eq. (3), we calculate  $m$  as the number of bits that need to be flipped. Then, we can simultaneously flip 1 bit, 2 bits, ...,  $m$  bits. This will result in the number of the hash value  $N_{h_q}$  of the query  $\mathbf{q}$  calculated by Eq. (4).

$$N_{h_q} = 1 + C_K^1 + C_K^2 + \cdots + C_K^m. \quad (4)$$

In fact, if the number of bits that need to be flipped is too large, the hashing process will be ineffective since the number of the feature vectors in  $S_q$  will only be slightly decreased in comparison to the database  $S_X$ , while the accuracy will not be improved.

### 2.3 Linear Matching

To make sure the hashing process does not miss the searching point, we need to either increase the number of base vectors (hyperplanes) or repeat hashing the database and the query vector with additional hash functions. In this work, the hash function is randomly generated  $L$  times. Consequently, the hashing of the query vector by these hashing functions will cluster the database into a dataset  $S_q^L$  which is the summation of  $S_q$  of each hash function. If the number of hash functions is increased, the accuracy can be improved. However, the dataset  $S_q^L$  will be enlarged, which leads to an increase in the computing time. To reduce the computing time when performing the linear matching on  $S_q^L$ , in this work we proposed several optimizations as follows:

First, it is observed that when hashing the database with several hash functions, there are feature vectors in the dataset  $S_q^L$  that appear either once or several times. The higher the appearance frequency of the feature vectors in  $S_q^L$ , the higher the probability of similarity of that vector with the query vector. Therefore, we introduced a parameter  $p$  that presents the times that a feature vector appears in the hashed dataset. This parameter is used to eliminate feature vectors with appearance times that are less than  $p$ . This elimination helps reduce the dataset  $S_q^L$  and thus the computing time for the linear matching. The larger the  $p$  is, the smaller the dataset  $S_q^L$  is. However, we need to choose this parameter empirically.

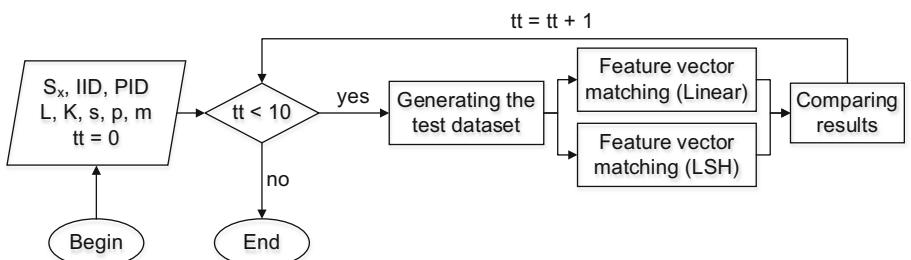
Second, the linear matching using cosine similarity according to Eq. (1) requires the computation of the cosine scalar between the feature vector and the query vector and their modulus. We propose using the library Eigen in C++ [13], which helps reduce the computing time. Furthermore, all feature vectors and the query vector are normalized. Therefore, their modulus is equal to one, and we do not need to compute it.

Third, instead of saving and sorting the database on the hard disk drive using the database management system (PostgreSQL DB), we can load and sort the database on RAM (Random Access Memory). This will make accessing the data quicker.

## 3 Experimental Implementation and Results

### 3.1 Experimental Implementation

Based on the method introduced in Sect. 2, experimental implementations were performed. The overall processing pipeline is shown in Fig. 3. Details of the experimental process are described as follows:



**Fig. 3.** The overall pipeline of the processing of the face feature matching test.

## Loading the Database

The database for the face recognition problem contains records that have several fields of information. For each person to be identified, we need several face images of them. These images are assigned an Image ID (IID) and are linked to a Person ID (PID). Face feature vectors are extracted from these the images via a face recognition model using backbone Resnet 100. The hashed face feature vectors and their corresponding hash values are added to the database. The feature vectors are sorted via their hash value. Vectors that have an equal hash value are clustered into a bucket with an integer index in a range of  $[0, 2^K - 1]$ . Thus, when searching for a bucket with a given hash value, we only need to use the index. This helps avoid performing the “for loop”, which reduces the computing time. In addition, the information from the base vector is also added, which is used to hash the query vector. The sorted database is loaded into RAM.

## Making the Test Dataset

A set of  $N_t$  feature vectors, in which each vector of  $N_t$  belongs to a PID, are randomly extracted from the database to generate the test dataset. After that, the cosine similarity between each extracted feature vector and the remaining feature vectors of the whole database is calculated. The condition for choosing the extracted vector is that the vector pair with the highest similarity must belong to a PID and the similarity value must be greater than a threshold  $d$  equal to 0.55, which is often used in practical applications. This action makes sure that the feature vector of a PID cannot be mistaken for another PID. Finally, we delete the set  $N_t$  out in the database.

## Matching of Feature Vectors

After extracting the test dataset, the purely linear matching and the LSH-based matching are performed and compared. For both methods, each vector of the test dataset  $N_t$  is linearly matched to the database to find the highest-similarity vector and thus the PID of the test vector. In the case of matching using LSH, we need only match the test vectors with the buckets that have an equal hash value to those test vectors. The parameters for this procedure are given in Table 1. It is noticed that these parameters are empirically chosen.

**Table 1.** Parameter configuration for LSH-based matching.

Parameters	Value
Number of hash tables	$L = 100$
Number of base vectors	$K = 20$
Distance threshold	$s = 0.1$
Cosine similarity threshold	$d = 0.55$
Number of repetitions	$p = 2$
Number of flipped bits	$m = 1, 2 \text{ or } 3$

## Result Comparison

Let the results of the linear matching as standard. We compare the result obtained by the LSH-based method with those of this standard. The matching results include the best-similarity image and the found PID. If a matched vector by the LSH method has the result as the standard, the LSH-based matching for this vector is correct. Otherwise, it is wrong. The accuracy of the LSH method is calculated via the ratio of the number of correct matchings to the total number of tests  $N_t$ .

## 3.2 Results

We experimented with a dataset collected from MegaFace [14], which contains a total of 12,720,006 images corresponding to 617,970 PID. A test dataset of 10,000 images of 10,000 PID was randomly extracted and, consequently, the accuracy evaluation was performed. To make sure of the reliability of the experiment, the extraction and evaluation of the test data are repeated ten times. As a result, the accuracy is the average of these repetitions. Furthermore, the data type of the feature vectors is set to either FP32 or INT8. The data type will influence the accuracy of the matching, as well as the required hardware resources, which will also be evaluated. All the experiments were implemented on a workstation computer that has 64 GB of RAM and a CPU, the Intel Core i7-9700K@3.60 GHz. The following aspects will be considered:

- The accuracy,
- The calculation time, and
- The usage of hardware resources.

The first experimental results are shown in Table 2, where the feature vectors have a data type of FP32, and the library Eigen for the calculation of the cosine similarity is not exploited. The matching results of the linear method and the LSH method are presented and compared. First, for the accuracy of the IID and PID, it is obvious that when the number of flipped bits  $m$  is increased, the accuracy of the LSH method is gradually increased. When  $m$  is set to 3, the accuracy is 99.99% for IID and 100% for PID. Thus, it can be concluded that the LSH method can fully ensure accuracy. Experiments for bigger values of  $m$  are not necessary because the accuracy is saturated, requiring more computational effort. Second, it is noticed that the computing time of the LSH method is extremely reduced. While the linear method requires 12,446 ms, the LSH method requires only 267 ms for the case of flipping three bits. This result confirms the efficiency of the LSH method with a large-scale dataset, which allows us to deploy real-time applications. Third, for the usage of hardware resources, the linear method requires less RAM than the LSH method (33 GB vs. 40 GB). This is quite logical because the LSH method needs more RAM for loading the base vectors, hashing tables, and hashing values.

**Table 2.** Experimental results with data type FP32 without the Eigen library.

Matching method	No. of flipped bits	RAM (GB)	Times (ms)	Acc. of matched IID (%)	Acc. of matched PID (%)
Linear		33	<b>12,446</b>	100	100
LSH	1	40	18	82.57	94.38
	2	40	33	98.62	99.86
	3	40	<b>267</b>	<b>99.99</b>	<b>100</b>

Experiments using the computing acceleration library Eigen for calculating the cosine similarity were also performed. The results are shown in Table 3. It is seen that the accuracy and the usage of RAM are quite comparable to those of Table 2. However, the computing time is strongly reduced from 12,446 ms to 1,702 ms for the linear method, and from 267 ms to 125 ms for the LSH method. These results confirm the advancement of the library Eigen and thus its use in practice.

**Table 3.** Experimental results with data type FP32 with the Eigen library for computing acceleration.

Matching method	No. of flipped bits (bits)	RAM (GB)	Times (ms)	Acc. of matched IID (%)	Acc. of matched PID (%)
Linear		33	<b>1,702</b>	100	100
LSH	1	40	16	82.64	94.46
	2	40	26	98.66	99.87
	3	40	<b>125</b>	<b>99.99</b>	<b>100</b>

For very large-scale databases, the use of the data type FP32 will require a lot of RAM. Thus, we can use the type of INT8 for saving RAM. However, the accuracy can be affected. In addition, the acceleration library Eigen does not support the data type INT8. Thus, the computing time can be increased. Table 4 shows experimental results with data type INT8 without using the Eigen library. It is seen that the use of RAM is reduced to 50% in comparison to the case with FP32. The accuracy is slightly lower than that in Table 2. When flipping 3 bits, the accuracy reaches 99.93% for IID and 100% for PID. Since the processing for the INT8 data type is faster than FP32, the computing time in Table 4. is less than that in Table 3 (12,446 ms vs. 5,794 ms, and 267 ms vs. 192 ms), although both experiments do not use the Eigen library. It is easy to see that for the INT8, we need to flip 3 bits. The PID can get 100% accuracy while the computing time is only 192 ms. This result should be used for practical applications.

**Table 4.** Experimental results with data type INT8 without the Eigen library

Search	No. of flipped bits (bits)	RAM (GB)	Times (ms)	Acc. of matched IID (%)	Acc. of matched PID (%)
Linear		14	<b>5,794</b>	100	100
LSH	1	21	17	81.40	94.10
	2	21	29	98.39	99.88
	3	21	<b>192</b>	<b>99.93</b>	<b>100</b>

Experiments on the larger scale of the dataset were also performed. A dataset of 21,732,116 images, accompanied by 1,054,603 PID from [14], was also evaluated. The result is shown in Table 5. The result is very promising, as the accuracy reaches 99.92% for IDD and 100% for PID, while the computing time is only 283 ms. Certainly, the use of RAM has increased to 32 GB.

**Table 5.** Experimental result with data type INT8 without Eigen library (database: 21,732,116 images; 1,054,603 PID)

Search	No. of flipped bits (bits)	RAM (GB)	Times (ms)	Acc. of matched IID (%)	Acc. of matched PID (%)
Linear		24	9681	100	100
LSH	1	32	21	81.65	94.23
	2	32	38	98.40	99.88
	3	32	<b>283</b>	<b>99.92</b>	<b>100</b>

## 4 Conclusions and Outlook

In this work, the evaluation of the LSH method for large-scale face feature matching has been successfully implemented. It is efficient to introduce several improvements to speed up the computation. For large-scale datasets, the use of the linear matching method is not suitable since the computing time is too long, which means it cannot be used in real-time applications. The LSH-based method has shown its efficiency for the large-scale dataset in terms of accuracy and especially computing time. In practice, the data type of INT8 should be used since it helps reduce the use of RAM while still ensuring the matching accuracy and computing time even in the case of not using the Eigen library.

In future work, evaluations for larger-scale datasets will be implemented. Also, the influence of other parameters, including the number of the hash function and the number of the base vector, should be thoroughly investigated.

**Acknowledgment.** This work was funded by the CMC Institute of Science and Technology, CMC Corporation, Hanoi, Vietnam.

## References

1. Wang, M., Deng, W.: Deep face recognition: a survey. [arXiv:1804.06655](https://arxiv.org/abs/1804.06655) (2018)
2. Melikoglu, O., Ergin, O., Salami, B., Pavon, J., Unsal, O., Cristal, A.: A novel FPGA-based high throughput accelerator for binary search trees. [arXiv:1912.01556](https://arxiv.org/abs/1912.01556) (2019)
3. Chen, H., Madaminov, Sergey., Ferdman, M., Milder, P.: FPGA-accelerated samplesort for large data sets. In: 2020 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (FPGA 2020), pp. 222–232. Association for Computing Machinery, USA (2020)
4. Cosine Similarity Using Xilinx Alveo. <https://developer.xilinx.com/en/articles/cosine-similarity-using-xilinx-alveo.html>. Accessed 20 Nov 2020
5. Dong, H.-N., Ha, N.-X., Tuan, D.-M.: A new approach for large-scale face-feature matching based on LSH and FPGA for edge processing. In: Tran, D.-T., Jeon, G., Nguyen, T.D.L., Lu, J., Xuan, T.-D. (eds.) ICISN 2021. LNNS, vol. 243, pp. 337–344. Springer, Singapore (2021). [https://doi.org/10.1007/978-981-16-2094-2\\_42](https://doi.org/10.1007/978-981-16-2094-2_42)
6. Locality Sensitive Hashing. <https://www.mit.edu/~andoni/LSH/>. Accessed 20 Nov 2020
7. FALCONN Github Repository. <https://github.com/FALCONN-LIB/FALCONN>. Accessed 10 Oct 2021
8. Cormode, G., Dasgupta, A., Goyal, A., Lee, C.H.: An evaluation of multi-probe locality sensitive hashing for computing similarities over web-scale query logs. PLoS ONE **13**(1) (2018). <https://doi.org/10.1371/journal.pone.0191175>
9. Lv, Q., Josephson, W., Wang, Z., Charikar, M., Li, K.: Multi-probe LSH: efficient indexing for high-dimensional similarity search. In: Event 33rd International Conference on Very Large Data Bases, VLDB 2007, Vienna, Austria, pp. 950–961. Association for Computing Machinery, Inc., Austria (2007)
10. Loi, T.L., Heo, J., Lee, J., Yoon, S.: VLSH: Voronoi-based locality sensitive hashing. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, Tokyo (2013). <https://doi.org/10.1109/IROS.2013.6697130>
11. Andoni, A., Indyk, P., Laarhoven, T., Schmidt, L.: Practical and optimal LSH for angular distance. [arXiv:1509.02897](https://arxiv.org/abs/1509.02897) (2015)
12. Moran, S., Lavrenko, V., Osborne, M.: Variable: bit quantisation for LSH. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, vol. 2 Short Papers, Sofia, Bulgaria, pp. 753–758 (2013)
13. Eigen Homepage. [https://eigen.tuxfamily.org/index.php?title=Main\\_Page](https://eigen.tuxfamily.org/index.php?title=Main_Page). Accessed 18 Aug 2021
14. MegaFace Homepage. <http://megaface.cs.washington.edu/>. Accessed 18 Aug 2021



# Determination of Fruit Freshness Using Near-Infrared Spectroscopy and Machine Learning Techniques

Duy Khanh Ninh<sup>1</sup>(✉), Kha Duy Phan<sup>1</sup>, Chi Khanh Ninh<sup>2</sup>, and Nhan Le Thanh<sup>3</sup>

<sup>1</sup> The University of Danang—University of Science and Technology, Danang, Vietnam  
nkduy@uts.udn.vn

<sup>2</sup> The University of Danang—Vietnam-Korea University of Information and Communication Technology, Danang, Vietnam  
nkchi@vku.udn.vn

<sup>3</sup> The University of Danang—Danang International Institute of Technology, Danang, Vietnam  
nhan.le-thanh@univ-cotedazur.fr

**Abstract.** This paper addresses the problem of fruit freshness categorization in the context of fruit quality assessment during short storage periods. As it is hard to handle by using computer vision technology, we propose a novel method by using absorbance near-infrared spectroscopy combined with machine learning (ML) techniques. We collected several samples of five popular fruits with various properties and classified them into three degrees of freshness based on the storage duration. We then examined multiple combinations of feature extraction and machine learning techniques. Experimental results show that the proposed Convolutional Neural Network (CNN) architecture were superior to other traditional ML models regardless of the selected feature vector. In particular, the proposed CNN when trained on the concatenated first and second derivatives of the pre-processed absorbance spectrum achieved the highest accuracy of 80.0%. The obtained classification performance was evaluated on a variety of fruits, which shows the potential of our approach.

**Keywords:** Near-infrared spectroscopy · Absorbance spectrum · Feature extraction · Machine learning · Fruit freshness classification

## 1 Introduction

Accurate recognition of fruit freshness is highly important because it affects consumer health as well as economic activities. It is estimated that nearly one-third of fruit costs are spent on fruit spoilage [1]. In addition, spoiled fruit will also damage consumers' health due to the decreased concentration of amino acids, vitamins, sugar/glucose and some nutrients. Fruit quality is often manually examined based on fruit appearance observation or internal destructive analysis methods. However these methods are subjective, laborious, inconsistent, time-consuming and costly.

Currently, fruit freshness grading via computer vision technology is the most common method for solving the automatic freshness determination problem. Machine learning (ML) and deep learning (DL) models trained on large datasets of digital images of a variety of fruits with various textures, colors and shapes can obtain classification results with the accuracy ranging from 70% to 100% [1, 5, 6, 8]. In spite of the high accuracy, this approach does not analyze internal characteristics affecting the freshness of the fruit such as dry matter, acidity, heavy metals, dissolved solids, etc. Rather, it fully depends on the appearance of fruits. Consequently, the performance of computer vision based freshness classification systems is highly dependent on how considerably the color, texture and shape of fruits varies during their growing or spoilage process, making this technology suitable for particular kinds of fruits (e.g., papaya [8], gooseberry [5], palm oil fresh fruit bunch [6]) or for long time spans of fruit maturation/ deterioration process (e.g., in weeks).

In this study, we considered the problem of fruit freshness categorization in the context of fruit quality assessment under usual storage conditions (i.e., without any climate-controlled system) for consumption at homes or for sale at street markets during short time periods. For this aim, we investigated five popular fruits with various appearances and flavors (including western apple, avocado, dragon fruit, guava, and mango) during a short interval of observation of six days. In particular, we classified collected fruit samples into three degrees of freshness based on the storage duration from the time of buying at the supermarket in fresh states to the time of six days later after being kept in room conditions in completely unrefresh states. For fruits such as western apple and dragon fruit, their appearances exhibit little change during this period, making computer vision based methods hard to predict their freshness. We thus leveraged near-infrared (NIR) spectroscopy technology to solve the fruit freshness determination problem.

Recently, several studies have employed NIR spectroscopy techniques for evaluating chemical and micro-biological attributes of food products due to its simplicity, speediness and non-destructive nature [3]. NIR spectra may contain lots of information reflecting internal properties of fruits. Researchers have exploited NIR spectroscopy in combination with ML techniques to classify fruits not only into organic and inorganic categories but also into different geographical origins [4]. Besides, the combined DL models and NIR spectra have shown the effectiveness for the prediction of dry matter in mango fruit [2] or for the recognition of different fruits [7]. However, none of previous studies using the NIR approach focused on determining the freshness of fruits.

This paper presents the first attempt in detecting fruit freshness using absorbance NIR spectroscopy combined with ML techniques. We examined multiple combinations of feature extraction techniques and ML/DL models. The paper has following contents. Previous works on NIR spectroscopy and ML techniques in fruit examination are reviewed in Sect. 2. The collection of NIR data is detailed in Sect. 3. Section 4 presents our methods and Sect. 5 describes experimental results. The last section gives the conclusion and future work.

## 2 Related Works

NIR radiation consists of wavelengths ranging from 780 to 2500 nm in the electromagnetic spectrum. When the radiation hits a fruit sample, it may be reflected, transmitted

or absorbed, resulting correspondingly a NIR spectrum in the reflectance, transmittance or absorbance mode. Each of these spectra can reflect some physical attributes and chemical composition of the sample. Consequently, many studies have leveraged NIR spectroscopy for fruit quality evaluation over the last decades [3].

Gupta et al. [4] presented an approach that uses visible and NIR spectroscopy to identify various properties of products such as variety, farmer and organic category. Combining spectral data of over 75,000 fruit and vegetable samples with support vector machine (SVM) technique gave a high accuracy of 90–98% and 98–99% for organic/inorganic classification and farmer recognition, respectively. Besides, NIR spectroscopy is also combined with DL to predict dry matter in mango [2]. This study gives the root mean square error of prediction (RMSEP) with the lowest is 0.79% and the highest is 0.84%. In addition, Ninh et al. [7] proposed to combine NIR spectra and deep neural networks to recognize several types of fruit including apple, avocado, dragon fruit, guava and mango. By using the derivative features of NIR spectra, these DL based models achieved the recognition accuracy as high as 99.0%.

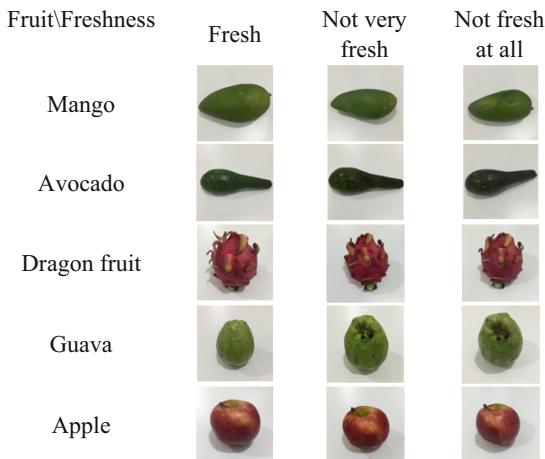
There are many other studies in literature that leveraged NIR technology to analyze and classify food and agricultural products. However, none of the previous researches has employed NIR spectra for classifying the freshness of fruits. In the next sections of the paper, we will present our novel approach in details.

### 3 Data Collection

In this study, we performed the classification of the freshness of five fruits including western apple, avocado, dragon fruit, guava, and mango. For each fruit, we bought eight samples from the same supermarket to ensure they have the same origin as this factor may have some affect on spectral data. However these samples were purchased at different times (at least ten days apart from each purchase) for the diversity of sample data. We created three different levels of freshness: (i) “Fresh” for samples on the day of purchase in their fresh condition; (ii) “Not very fresh” for samples stored in room conditions from two to three days after the day of purchase; (iii) “Not fresh at all” for samples stored in room conditions from four to six days after the purchasing day. Figure 1 illustrates five types of fruit with three freshness labels for several samples. It could be observed that some fruits, particularly western apple and dragon fruit, exhibited almost no change in their appearances during this short storage period.

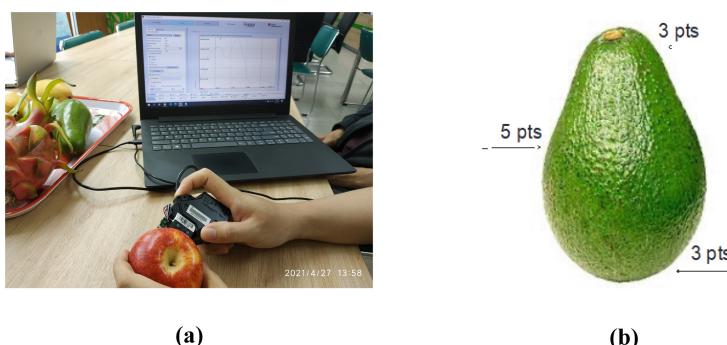
We used a portable NIR spectrometer named DLP NIRscan Nano EVM from Texas Instrument to collect the NIR spectrum of fruits. The fruit sample to be measured is placed close to the sapphire glass of the instrument. During each measurement, the sample absorbs a certain amount of NIR light emitted by the scanner, while the unabsorbed light is reflected into the scanner’s lens and converted by the optical processing module into a spectrum containing 228 wavelengths with a range of 900–1700 nm. The spectral data is transmitted from the measurement instrument to a computer using a USB cable as.csv files containing three types of spectra: intensity, reflectance, and absorbance. Among the three spectra, the absorbance spectrum gained highest performance in fruit freshness classification in our initial experiments. Therefore, we only employed absorbance spectra in this study. Physically, the absorbance spectrum indicates how well the molecules in

fruits absorb electromagnetic waves (i.e., NIR light), thus it can reflect the freshness of fruits to a certain degree. The collection of the NIR spectrum of a fruit sample is shown in Fig. 2a.



**Fig. 1.** Example samples of five fruits with three freshness labels.

For each fruit sample, we collected its absorbance spectrum in three freshness states. Moreover, we measured the spectrum at different positions on the fruit to take into account the measurement position. In particular, the measurements were carried out at 11 different positions on the sample, including five points around the body and six points around the two ends (Fig. 2b). As a result, we have collected 1320 NIR absorbance spectra from forty fruit samples with three freshness states. After removing those suffered from data errors (i.e., containing negative values), we obtained a total of 1252 valid spectra measurements for data modeling as listed in Table 1.



**Fig. 2.** Setup of the NIR spectrum collection (a) and 11 measurement positions (b).

**Table 1.** Number of valid NIR absorbance spectra per fruit and per freshness label.

Fruit\Freshness level	Fresh	Not very fresh	Not fresh at all	Total
Apple	88	88	88	264
Avocado	88	88	86	262
Dragon fruit	88	88	88	264
Guava	88	66	66	220
Mango	88	88	66	242
<b>Total</b>	<b>440</b>	<b>418</b>	<b>394</b>	<b>1252</b>

## 4 Proposed Methods

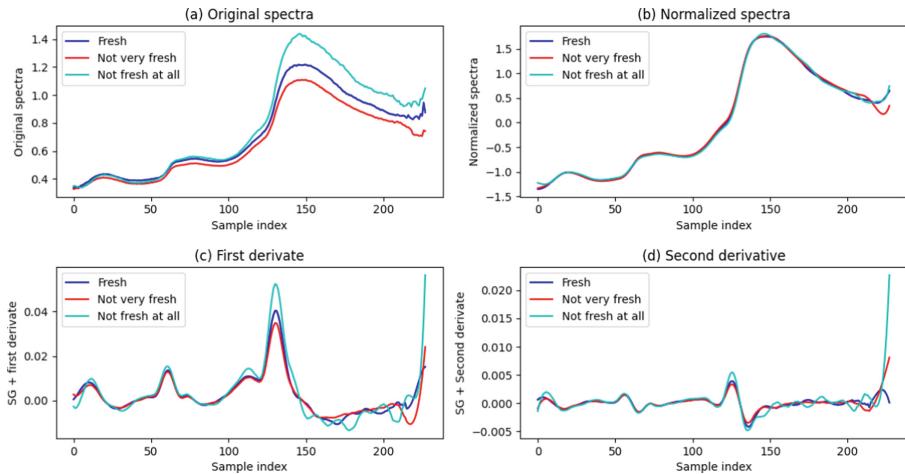
Our fruit freshness classification system was carried out by using supervised ML framework. We examined multiple combinations of feature extraction schemes and ML models to seek for the optimal one. We trained a common ML model for different fruits, thus the spectral data of different fruits were pooled and used for modeling and validation. The system includes three stages: pre-processing, feature extraction, and model building.

### 4.1 Pre-processing

In the pre-processing stage, all of the collected NIR absorbance spectra were smoothed by a Savitzky-Golay (SG) filter (with the window length of 25 and the polynomial order of 5) to remove spectral noise. The smoothed spectra then were calibrated by the standard normal variate correction (i.e., z-score normalization) to eliminate the deviations caused by particle size and scattering. These are two techniques widely used for pre-processing NIR spectra [3].

### 4.2 Feature Extraction

Relevant features need to be extracted for building classification models. For a fruit sample, its pre-processed NIR spectrum is a certain choice as the feature vector for freshness identification. We further examined the derivatives of the pre-processed spectrum to see if they can help to differentiate levels of freshness. Figure 3 illustrates the originally collected spectra, spectra pre-processed by SG-smoothing and normalizing, first and second derivatives of pre-processed spectra of a fruit sample in three freshness states. We investigated four types of feature vector based on the concatenation of the pre-processed spectrum and/or its derivatives as described in Table 2.



**Fig. 3.** Originally collected absorbance spectra (a), spectra pre-processed by SG-smoothing and normalizing (b), first (c) and second (d) derivatives of pre-processed spectra of a fruit sample in three freshness states.

**Table 2.** Four types of feature vector with their vector sizes and descriptions.

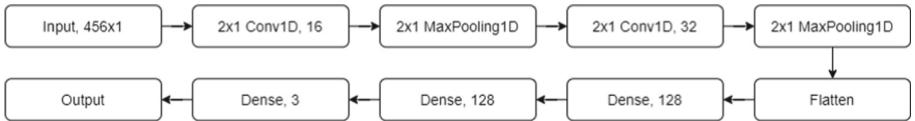
Feature type	Vector size	Description
x	228	Pre-processed spectrum
d1	228	1st derivative of pre-processed spectrum
d1_d2	456	1st and 2nd derivatives of pre-processed spectrum
x_d1_d2	684	Pre-processed spectrum + its 1st and 2nd derivatives

### 4.3 Machine Learning Models

In the training stage, the extracted spectral features of training fruit samples were used as the training data, and the three levels of freshness were used as the labels. We used both the traditional ML and modern DL approaches to build classification models and compared their performances. For the traditional ML approach, five algorithms were experimented including decision tree (DT), random forest (RF), gradient boosted trees (GBT), k-nearest neighbors (KNN), and support vector machine (SVM).

For the DL approach, we proposed a Convolutional Neural Network (CNN) whose architecture is illustrated in Fig. 4. Note that this CNN architecture was utilized regardless of the type of the feature vector. The model includes one input layer contains  $N$  neurons as input data, representing the feature vector of size  $N \times 1$  ( $N$  can be set to 228, 456, or 684 depending on the chosen feature). The CNN includes two convolutional layers with 16 and 32 filters, consecutively. They used kernels of size  $2 \times 1$  and ReLU (Rectified Linear Unit) activation functions. In the following max pooling layers, a pool size of  $2 \times 1$  along with a stride length of 2 was employed. The conversion from a 2D filter matrix into an 1D feature vector with a size of  $N \times 8$  was carried out by a flatten layer.

The 1D feature vector was then fed into two fully connected (FC) layers, both having 128 neurons and a ReLU activation function. Lastly, the third FC layer equipped with 3 neurons and a softmax activation function performs the classification to output the freshness label.

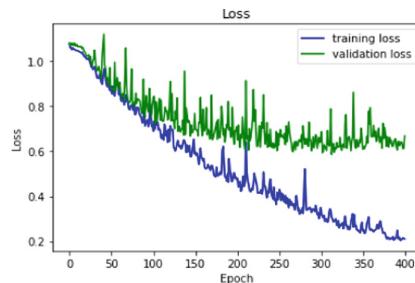


**Fig. 4.** Proposed CNN architecture (the size of input feature vector N is 456 in this case).

## 5 Experimental Results

### 5.1 Results of Proposed CNN Model

We set up experiments with the proposed CNN model by using functions of the deep learning library Keras. The collected spectral dataset were randomly divided into three sets for training, validating, and testing CNN models according to the ratio 3:1:1, correspondingly. In each set, the quantity of samples among fruit types as well as among freshness levels were kept balanced. A loss function based on cross-entropy and an Adam optimizer were utilized for training CNNs. The learning rate was initially set to 0.001 and to be reduced by a factor of 0.8 when the training result was not progressing. Moreover, the validation set was used for stopping the training process. Figure 5 shows the variation of the loss function on the training and validation sets over training epochs when trying with the feature vector d1\_d2. In this case, the training process was stopped after 400 epochs since the loss function on the validation set converged at this point. We applied the same procedure for training CNN models regardless of the feature types to prevent model over-fitting.



**Fig. 5.** Variations of loss function on training and validation sets over epochs of training CNNs.

Table 3 shows the full performance of CNN models. We can find that the model achieved the best performance when the feature vector d1\_d2 was used (the accuracy is 80% on the test set with F1-score of 0.801).

**Table 3.** Performance of CNN models.

Feature vector	Accuracy (%)			Other metrics on test set		
	Training set	Validation set	Test set	Precision	Recall	F1-score
x	70.1	52.3	60.7	0.612	0.623	0.614
d1	93.0	73.0	78.5	0.805	0.792	0.798
d1_d2	91.0	80.1	<b>80.0</b>	0.801	0.800	<b>0.801</b>
x_d1_d2	85.0	61.0	70.0	0.710	0.721	0.710

## 5.2 Results of Traditional ML Models

The proposed CNN models were compared with other five traditional classifiers mentioned in Sect. 4.2. The ML training experiments were conducted with the scikit-learn toolkit. As model parameters can have significance influence on the performance of these ML algorithms, grid search procedures using four fold cross validation on the training set were carried out to produce the optimal models. Then the accuracies as seen on the test set were reported. Here, we combined the training and validation sets for constructing the CNN models into the training set for building the traditional ML models while keeping the test set identical for both cases. Table 4 lists set of parameters used in the grid searching for the optimal ML models.

**Table 4.** Set of parameters used in the grid searching for the optimal traditional ML models.

Model	Set of parameters (values in parentheses are those of the optimal models)
DT	Maximum depth of the tree (9), minimum number of samples at a leaf node (4)
RF	Maximum depth of a tree (5), minimum number of samples at a leaf node (9), number of trees in the forest (29)
GBT	Step size shrinkage (0.1), maximum depth of a tree (3), minimum sum of instance weight needed in a child (10), number of boosting rounds (57)
KNN	Number of neighbors (9)
SVM	Regularization parameter (100), kernel type (polynomial)

Table 5 summarizes the classification accuracy on the test set of the optimal model for each traditional ML algorithm regarding the feature types in comparison with that of the proposed CNN. It can be observed that the performance of all traditional ML models were inferior to that of the CNN model whatever the feature type was used. Among the traditional ML models, the GBT trained on d1 feature achieved the highest accuracy of 64.9%, much lower than the best performance of the proposed CNN when trained on d1\_d2 feature with the accuracy of 80.0%. All of the models meet the real-time classification with the recognition time shorter than 0.4 s when running on Google’s Colab platform.

**Table 5.** Classification accuracies (%) of ML/DL models on the test set.

Feature vector	Model					
	DT	RF	GBT	KNN	SVM	CNN
x	50.9	59.6	56.6	52.6	56.6	60.7
d1	55.7	61.0	<b>64.9</b>	52.2	60.5	78.5
d1_d2	50.0	60.5	62.3	53.1	60.5	<b>80.0</b>
x_d1_d2	53.1	60.5	61.4	50.9	55.7	70.0

## 6 Conclusion

We have presented the first attempt in fruit freshness categorization based on absorbance NIR spectra of fruits. For a short period of six-days storage during consumption time, many fruits exhibit little or no change in their appearances, making computer vision techniques hard to deal with the freshness prediction problem. Consequently, NIR spectroscopy technology shows its advantage in this context thanks to its capability in characterizing internal properties of fruits. We have examined multiple combinations of feature extraction schemes and ML models to seek for the optimal one. Experimental results show that the proposed CNN architecture were superior to other traditional ML models regardless of the feature vector type. In particular, the proposed CNN when trained on the concatenated first and second derivatives of the pre-processed absorbance spectrum achieved the highest accuracy of 80.0%. The obtained classification performance was tested on a variety of fruits, which shows the potential of our approach. The additional use of chemometrics methods (such as those suggested in [2]) would be helpful to improve the capability of ML models in grading the freshness of fruits based on NIR spectra.

**Acknowledgments.** This work was supported by The University of Danang, University of Science and Technology, code number of Project: T2020-02-32.

## References

1. Fu, Y.: Fruit freshness grading using deep learning. Master of computer and information sciences thesis, Auckland University of Technology (2020)
2. Mishra, P., Passos, D.: A synergistic use of chemometrics and deep learning improved the predictive performance of near-infrared spectroscopy models for dry matter prediction in mango fruit. Chemometr Intell. Lab. Syst. **212**, 104287 (2021)
3. Wang, H., Peng, J., Xie, C., Bao, Y., He, Y.: Fruit quality evaluation using spectroscopy technology: a review. Sensors **15**(5), 11889–11927 (2015)
4. Gupta, O., Das, A.J., Hellerstein, J., Raskar, R.: Machine learning approaches for large scale classification of produce. Scientific Reports 8, Article number: 5226 (2018)
5. Castro, W., Oblitas, J., De-La-Torre, M., Cotrina, C., Bazán, K., Avila-George, H.: Classification of cape gooseberry fruit according to its level of ripeness using machine learning techniques and different color spaces. IEEE Access **7**, 27389–27400 (2019)

6. Ibrahim, Z., Sabri, N., Isa, D.: Palm oil fresh fruit bunch ripeness grading recognition using convolutional neural network. *J. Telecommun. Electron. Comput. Eng.* **10**(3–2), 109–113 (2018)
7. Ninh, D.K., Doan, T.N.C., Ninh, C.K., Nguyen-Thi, T.X., Le-Thanh, N.: Fruit recognition based on near-infrared spectroscopy using deep neural networks. In: 5th International Conference on Machine Learning and Soft Computing, pp. 90–95. Association for Computing Machinery, New York (2021)
8. Behera, S.K., Rath, A.K., Sethy, P.K.: Maturity status classification of papaya fruits based on machine learning and transfer learning approach. *Inf. Process. Agric.* **8**(2), 244–250 (2021)



# Vietnamese Named Entity Recognition on Medical Topic

Loi Le Vu<sup>1,3</sup>, Duong Pham Van<sup>1,3</sup>, Dong Nguyen Tien<sup>1,3(✉)</sup>, Tuoi Tran Thi Minh<sup>1</sup>,  
and Tuan Dang Minh<sup>1,2(✉)</sup>

<sup>1</sup> CMC Institute of Science Technology, No. 11 Duy Tan Street, Hanoi, Vietnam  
dongnguyentien1996@gmail.com, dmtuan1@cmc.com.vn

<sup>2</sup> Posts and Telecommunications Institute of Technology, Km10, Nguyen Trai Street,  
Hanoi, Vietnam  
tuandm@ptit.edu.vn

<sup>3</sup> Hanoi University of Science and Technology, No1. Dai Co Viet Street, Hanoi, Vietnam

**Abstract.** Recent advances in language representation using neural networks combined with transfer learning have brought great success in natural language processing tasks, such as named entity recognition (NER), question answering, sentiment analysis, etc. It has been shown that leveraging of pre-trained language models improves overall performance on many tasks and is very beneficial when little data is tagged. For the task of recognizing named entities in the medical field, most of available datasets are English-specific. Therefore, it is of interest to create datasets related to health topics for languages other than English. In this paper, we present the first manually annotated dataset for Vietnamese for named entity recognition (NER) task with newly defined entity types that can be used in future outbreaks. Our dataset also contains the largest number of entities compared to the existing Vietnamese NER dataset. We experimentally conducted our dataset tests on language models for Vietnamese and make comparisons including PhoBERT and our own language model CMCBERT.

**Keywords:** Vietnamese natural language processing · Pretrained language model · Named entity recognition

## 1 Introduction

With the rise of disease issues, the need to learn about articles on health topics is increasing. Today, with the increasing amount of literature on health topics, the exploitation, use and discovery of knowledge at large scale is still limited. As an important step in natural language processing, entity recognition is a well-researched topic that aims at extracting all kinds of meaningful information in unstructured text. Named entity recognition (NER) is the task of identifying text spans that mention named entities (NEs) and classifying them into predefined categories, such as person, organization, location, or any other classes of interest. Despite being conceptually simple, NER is not an easy task. The category of a named entity is highly dependent on textual semantics and its surrounding

context. Furthermore, there are many definitions of named entity and evaluation criteria, introducing evaluation complications.

Current state-of-the-art NER systems employ neural architectures that have been pre-trained on language modeling tasks. Examples of such models are ELMo [1], OpenAI GPT [2], BERT [3], XLNet [4], RoBERTa [5], Albert [6] and T5 [7]. It has been shown that language modeling pretraining has been shown significantly processing the performance of many natural language tasks and reduces the amount of labeled data needed for supervised learning [8].

Applying these recent techniques to the Vietnamese language can be highly valuable, given that annotated resources are scarce, but unlabeled text data is abundant. In this paper, we experimented two Vietnamese pretrained language models, which are PhoBERT [9] and our language model CMCBERT. Since PhoBERT [9] requires an additional word segmentation step compared to regular language models, which takes much more time on large dataset, using CMCBERT is a better alternative.

Compared with other languages, the data resources for the NER task for Vietnamese are extremely limited, including only two public datasets from the VLSP 2016 and 2018 NER shared tasks [10]. Here, the VLSP-2018 NER dataset is an extension of the VLSP-2016 NER dataset with more data. These two datasets only focus on identifying generic entities of person names, organizations, and locations in online news articles. Recently, a dataset of NER Covid-19 [11] has been published, but the label has not really covered all topics in health and medicine.

Finally, our contributions are:

- (i) Develop NER task in the medical domain that potentially impacts research and downstream application.
- (ii) Provide the research community with a new dataset for identify medical related named entities in Vietnamese.
- (iii) Conduct numerous experiments to show efficiency of using pretrained language model in NER task and compare the results from different language models.

## 2 Data Preparation

### 2.1 Collect Data

We collect data from Vietnamese medical websites, including vimec, life health and several other websites. The main body of each article is extracted from the original html content, performing sentence splitting to obtain the dataset. The data is divided into two sets, the training set and the test set. The training set contains a total of 6788 sentences while the test set contains 755 sentences.

### 2.2 Entity Definition

We define 9 different types of entities, includes: (1) ANATOMICAL\_PART; (2) SYMPTOM; (3) DRUG; (4) TREATMENT; (5) DIAGNOSTIC; (6) DISEASE; (7) MEDICINE; (8) SUBJECT; (9) DISEASE\_CAUSE. These entity types are guaranteed

to cover most of the entity types that appear in the healthcare domain, which is synonymous with that any entity will belong to one of these 9 defined entity types. In addition, entities are meant with the goal of avoiding duplication between two different entity types. This goal is set to avoid the situation where an entity can be of many different types of entities, making it impossible for the model to identify the correct entity type and the annotator will not be able to correctly identify it.

### 2.3 Labeling Rule

Table 1 lists the rules to be applied in the annotation process. Different from the problem of entity recognition in the form of proper names (person name, organization name, place name, etc.) which has a clear distinction between types of entities, medical entities can be defined in many different ways, for example in the sentence “Nhiều phương pháp phẫu thuật dây chằng chéo trước được mô tả, nhưng không có phương pháp nào là tốt nhất” (Many techniques of anterior cruciate ligament surgery have been described, but none is absolutely dominant), we can assign “phẫu thuật dây chằng chéo trước” (anterior cruciate ligament surgery) as TREATMENT or “dây chằng chéo trước” (anterior cruciate ligament) as ANATOMICAL\_PART. However, we only choose one of the two ways, specifically in this case the first way (assigning “phẫu thuật dây chằng chéo trước” as TREATMENT) (Table 2).

**Table 1.** Labeling rule

Entity type	Labeling rule
ANATOMICAL PART	Only when standing alone, not under any other label, e.g. “ung thư phổi” (lung cancer) would be labeled as DIASEASE, not “phổi” (lung) as ANATOMICAL_PART Do not attach adjectives such as: “mịn” (smooth), “trắng” (bright), “đẹp” (beautiful), etc.
SYMPTOM	Attach body part in case body part is present, e.g.: “tê nhức chân” (numbness leg pain), “mỏi gối” (knee fatigue), etc. Distinguishing from DISEASE: some diseases are named after symptoms such as “đau dạ dày” (stomach pain), “tê bì tay chân” (numbness of hands and feet), etc. Depending on the case, the same phrase can be SYMPTOM in one context but DISEASE in another
DRUG	If there is a body part or disease name, it must be attached, e.g.: “thuốc trị đau họng” (sore throat medicine), etc. If there is the word “thuốc” (drug), it must be attached, e.g.: “thuốc kháng sinh” (antibiotic drug) instead of “kháng sinh” (antibiotic)
TREATMENT	If have organ or disease name, it must be attached, e.g.: “phẫu thuật khói u đại trực tràng” (colorectal tumor surgery) Assign only specialized method, not phrases such as “quấn khăn quanh chỗ chảy máu” (wrap a towel around the bleeding site)

(continued)

**Table 1.** (*continued*)

Entity type	Labeling rule
DIAGNOSTIC	If there are body parts or disease names, they must be attached, e.g.: “nội soi dạ dày” (gastroscopy)
DISEASE	If there are body parts or name comorbidities, they must be attached, e.g.: “ung thư dạ dày” (stomach cancer) If there is a word “bệnh” (disease), it must be attached, e.g.: “bệnh viêm gan B” (hepatitis B), etc.
MEDICINE	If there is a part of the plant, it must be attached, e.g.: “hột chuối” (banana seed), etc.
SUBJECT	Subject must be accompanied with a disease name, an age, or associated with a specific characteristic, do not assign generic terms such as “bệnh nhân”, “người bệnh” (patient)
DISEASE_CAUSE	If the DISEASE_CAUSE coincides with the DISEASE or SYMPTOM, priority is given to DISEASE or SYMPTOM. For example, in the sentence “ho liên tục gây đau họng hạt” (constant cough cause sore throat county), “ho” (cough) must be labeled as SYMPTOM instead of DISEASE_CAUSE

## 2.4 Statistics Entities

**Table 2.** Quantity of each type of entity in the training set, test set as well as total count are given

Entity type	Train set	Test set	Total
ANATOMICAL_PART	2317	264	2581
SYMPTOM	1914	196	2110
DRUG	1267	131	1398
TREATMENT	1002	143	1145
DIAGNOSTIC	760	83	843
DISEASE	3271	346	3617
MEDICINE	874	108	982
SUBJECT	749	84	833
DISEASE_CAUSE	603	55	658
<b>Total</b>	<b>12757</b>	<b>1410</b>	<b>14167</b>

### 3 Problem Statement

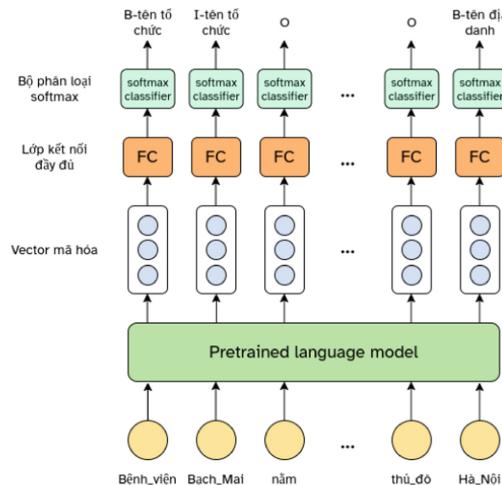
The problem of named entity recognition for the medical domain is stated in the form of a sequence labeling problem with BIO labels. Here, the sequence labeling problem is stated as follows.

**Input:** (1) The set of entity types. For example, set of 3 entity types {PERSON, LOCATION, ORGANIZATION}. (2) A sentence that has been broken down into words. For example: [“Bach”, “Mai”, “hospital”, “locates”, “in”, “Ha”, “Noi”, “city”].

**Output:** The BIO label of each word in the input sentence. For example: [(“Bach”, B-ORG), (“Mai”, I-ORG), (“hospital”, I-ORG), (“locates”, O), (“in”, O), (“Ha”, B-LOC), (“Noi”, I-LOC), (“city”, I-LOC)], where ORG stands for organization and LOC stands for location.

Our approach is to use language model to encode words in the input sentence into embedding vectors. The embedding vector of each word has the feature of containing the contextual information of the whole sentence, thereby giving appropriate labels in each specific context. To be able to label each word in the input sentence, the embedding vector of each word is pushed through a FC layer before being fed to the softmax classifier, as shown in Fig. 1.

We use cross entropy loss and the AdamWeightDecay [12] optimizer.



**Fig. 1.** Neural network architecture for the medical NER problem

## 4 Experiments and Results

### 4.1 Settings

We conducted various experiments to analyze effects of different parameters to the model performance. In subsequent sections, if not explicitly mentioned, the default values of parameters are used, as shown in Table 3.

**Table 3.** Experimental results hyper parameter

Parameter	Value	Interpretation
pretrained_model	PhoBERT	Name of language model used
learning_rate	5e-5	
max_grad_norm	1.0	Maximum value of gradient
batch_size	16	
pool_type	Concat	How to combine embedding vectors from transformers model (each layer of transformers output an embedding vector for each word). <i>concat</i> means concatenating vectors together into one longer vector. Another value is <i>average</i> , which means vectors are averaged to get a single one
ignore_index	0	Label that is ignored when calculating the loss function. Here, default value 0 means no loss function for special characters [CLS], [SEP]
add_special_tokens	True	Whether to add special characters [CLS] and [SEP]
num_hidden_layer	1	The number of hidden layers used to compute the final embedding vector of a word when using the language model

Each experiment was run for 10 epochs. We observed that training model longer tends to be overfitting. For comparison, we use the micro F1-score measure.

### 4.2 Main Results

#### 1. Evaluating the effect of *num\_hidden\_layer* and *pool\_type*

We consider two ways of combining the encoding vectors of a word, corresponding to the values *concat* and *average* of the parameter *pool\_type*, where *concat* is the combination of vectors by concatenation, into a longer vector, and *average* is the average of the vectors. Along with that, the parameter *num\_hidden\_layer* was tested with values of 1, 2, 4, 6, 12 respectively. The evaluation results are shown in Table 4.

As shown in Table 4, *concat* gives better results than *average* in most cases. When *pool\_type* is *concat*, using *num\_hidden\_layer* greater than 1 results in better accuracy. However, the accuracy tends to decrease when the *num\_hidden\_layer* is too large, i.e. 6, 12. In the opposite, when *pool\_type* is *average*, using more hidden layers gives worse

**Table 4.** Micro F1-score with other values difference between *num\_hidden\_layer* and *pool\_type*

num_hidden_layer	pool_type	
	Concat	Average
1	0.6885	
2	0.6934	<b>0.6944</b>
4	<b>0.6937</b>	0.6826
6	<b>0.6908</b>	0.6834
12	<b>0.6897</b>	0.6556

results, except for  $num\_hidden\_layer = 2$ . We can see that combining vectors by concatenating has no loss of information while averaging does. Thus, concatenating usually gives better results than averaging.

## 2. Evaluating the effect of *add\_special\_tokens* and *ignore\_index*.

From Tables 5 and 6, not adding special characters [CLS] and [SEP] decreases the model performance significantly. This behavior occurs because we are using a pre-trained language model. Since most existing pre-trained language models were trained by prepending [CLS] token and appending [SEP] token to each input sentence, we need to do the same when fine-tuning for the model to work successfully.

In Table 6 ( $num\_hidden\_layer = 1$ ), we see that calculating the loss function on the characters [CLS] and [SEP] ( $ignore\_index = -100$ ) brings better results. However, the opposite is observed in Table 7. In both cases, the difference is not remarkable. This implies that computing loss function for special tokens does not affect too much to the model. We must add [CLS] and [SEP] tokens to the input sentence for the model to work correctly, but can optionally computing loss function for these tokens (Table 8).

**Table 5.** Micro F1-score comparison with different values of *add\_special\_tokens* and *ignore\_index* (with  $num\_hidden\_layer = 1$ )

add_special_tokens	ignore_index	Micro F1-score
True	0	0.6865
True	-100	<b>0.6929</b>
False		0.6536

## 3. Evaluating the influence of *learning\_rate* and *max\_grad\_norm*

In this experiment, we adjusted learning speed of the model by changing values of *learning\_rate* and *max\_grad\_norm*.

**Table 6.** Micro F1-score comparison with different values of *add\_special\_tokens* and *ignore\_index* (with *num\_hidden\_layer* = 2)

<i>add_special_tokens</i>	<i>ignore_index</i>	Micro F1-score
True	0	<b>0.6934</b>
True	-100	0.6926
False		0.6657

**Table 7.** Micro F1-score comparison with different values of *learning\_rate* and *max\_grad\_norm* (with *num\_hidden\_layer* = 1)

<i>learning_rate</i>	<i>max_grad_norm</i>	Micro F1-score
5e-5	1.0	<b>0.6865</b>
3e-5	2.0	0.6814
1e-5	3.0	0.6199

**Table 8.** Comparison micro F1-score with different values of *learning\_rate* and *max\_grad\_norm* (with *num\_hidden\_layer* = 2)

<i>learning_rate</i>	<i>max_grad_norm</i>	Micro F1-score
5e-5	1.0	<b>0.6934</b>
3e-5	2.0	0.6747
1e-5	3.0	0.6230

Overall, the decrease *learning\_rate* and increase *max\_grad\_norm* greatly reduces model accuracy. Reducing the *learning\_rate* will certainly reduce the learning rate of the model but increasing the *max\_grad\_norm* does not necessarily increase the learning speed because the derivative pruning is only performed if the derivative value exceeds *max\_grad\_norm*. Thus, *learning\_rate* is equal to 5e-5 and *max\_grad\_norm* is equal to 1.0 which are the optimal values for training the model.

#### 4. Compare different language models.

In this section, we consider two language models for Vietnamese, vinai/phobert-base (PhoBERT [9]) and our CMCBERT, respectively. These language models use different tokenizers for processing input sentences. Vocabulary of vinai/phobert-base contains multi-syllable word (e.g. thành\_phố (city), nhà\_hàng (restaurant), etc.) while vocabulary of CMCBERT does not. Therefore, for the vinai/phobert-base model, we need to take one more step when processing input sentence, which is word segmentation. We

followed guide in official GitHub repository of PhoBERT [9] to do this processing step. Specifically, RDRSegmenter [13] was used to process multi-syllable words.

Results when comparing these two language models are given in Table 9.

**Table 9.** Micro F1-score comparison between the two models *vinai/phobert-base* and *CMC Bert*

num_hidden_layer	Vinai/phobert-base	CMCBERT
1	0.6885	0.6841
2	0.6934	0.6774

*vinai/phobert-base* and *CMCBERT* give almost the same result when using *num\_hidden\_layer = 1*. Whereas *vinai/phobert-base* is 1.6% higher in micro F1-score than *CMCBERT*. Through numerous experiments (not reported here due to length limit), we observed that *vinai/phobert-base* usually outperforms *CMCBERT* by 1–1.5% with respect to micro F1-score. In application, this is not a significant improvement and has the trade-off of additional processing step which is word segmentation. Hence, using our *CMCBERT* is a good alternative in many cases.

## 5. Detailed results analysis on each entity type.

In this section, we perform analysis of recall, precision and F1-score measures on each entity type using the trained model with default parameters. Table 10 shows the evaluation results.

**Table 10.** Experimental results

Entity type	Precision	Recall	F1-score
DIAGNOSTIC	0.6617	0.6286	0.6447
TREATMENT	0.6011	0.6011	0.6011
ATOMICAL_PART	0.6311	0.6457	0.6383
SUBJECT	0.5825	0.6186	0.6000
DISEASE_CAUSE	0.5333	0.5263	0.5298
DISEASE	0.7544	0.7715	0.7629
MEDICINE	0.7314	0.8477	0.7853
DRUG	0.8070	0.8519	0.8288
SYMPTOM	0.5692	0.6136	0.5905

In general, most entity types give higher scores to precision than recall. High recall means that the model properly recognizes entities in the test set. Meanwhile, low precision indicates that among the entities recognized by the model, many entities are

misidentified. The fact that recall is higher than precision suggests that there might be a number of entities in the test set which are not annotated. Therefore, when the model recognizes a phrase as an entity, it is possible that this is the right entity to look for but was not properly labeled in the test set, thus increasing the number of false positives. This greatly affects precision while not affecting the recall at all. In the train dataset, if there are entities that are not annotated, the model still can learn to identify similar entities later if the fraction of unannotated entities is small. Thus, recall is not affected heavily if the number of annotated entities completely dominates the number of unannotated ones.

The micro average of F1-score for all entity types is approximately 70%. Although this number seems not impressive, we saw that our model worked well. The reason is that metrics like recall, precision, F1-score require that prediction results match exactly the annotated data. This strict requirement is not always necessary, especially when the entity to be predicted is very long. When the entities are simple, i.e. entities that have short length and are constructed from frequently appeared words, our model correctly identifies them in almost every case.

## 5 Conclusion

In this article, we have defined entities in the medical topic, analyzed and proposed an appropriate labeling method. From there, evaluate the results of the named entity recognition problem and compare it on two pretrained language models for different Vietnamese and cmcbert and phobert. In the future, we will refine more data, build a pretrained language model specifically for medical topics to further improve the results.

**Acknowledgment.** We gratefully acknowledge the support from the CMC Institute of Science Technology for funding the research project.

## References

- Peters, M.E., et al.: Deep contextualized word representations. In: NAACL HLT 2018 - 2018 Conference on North American Chapter of the Association for Computational Linguistics: Human Language Technologies – Proceedings of Conference, vol. 1, pp. 2227–2237 (2018). <https://doi.org/10.18653/v1/n18-1202>
- Brown, T.B., et al.: Language models are few-shot learners. Adv. Neural Inf. Process. Syst. **2020-December** (2020)
- Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: NAACL HLT 2019 - 2019 Conference on North American Chapter of the Association for Computational Linguistics: Human Language Technologies – Proceedings of Conference, vol. 1, no. Mlm, pp. 4171–4186 (2019)
- Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., Le, Q.V.: XLNet: generalized autoregressive pretraining for language understanding. Adv. Neural Inf. Process. Syst. **32**(NeurIPS), 1–18 (2019)
- Liu, Y., et al.: RoBERTa: A Robustly Optimized BERT Pretraining Approach, no. 1 (2019)
- Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., Soricut, R.: Albert: a lite BERT for self-supervised learning of language representations. arXiv, pp. 1–17 (2019)

7. Raffel, C., et al.: Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* **21**, 1–67 (2020)
8. Howard, J., Ruder, S.: Universal language model fine-tuning for text classification. In: ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of Conference (Long Paper, vol. 1, pp. 328–339 (2018). <https://doi.org/10.18653/v1/p18-1031>
9. Nguyen, D.Q., Nguyen, A.T.: PhoBERT: pre-trained language models for Vietnamese, pp. 1037–1042 (2020), <https://doi.org/10.18653/v1/2020.findings-emnlp.92>
10. Nguyen, H.T.M., Ngo, Q.T., Vu, L.X., Tran, V.M., Nguyen, H.T.T.: VLSP shared task: named entity recognition. *J. Comput. Sci. Cybern.* **34**(4), 283–294 (2019). <https://doi.org/10.15625/1813-9663/34/4/13161>
11. COVID-19 named entity recognition for Vietnamese, pp. 1–6
12. Loshchilov, I., Hutter, F.: F w d r a (2017)
13. Nguyen, D.Q., Nguyen, D.Q., Vu, T., Dras, M., Johnson, M.: A fast and accurate Vietnamese word segmenter. In: LR 2018 - 11th International Conference on Language Resource and Evaluation, pp. 2582–2587 (2019)



# Cooperative Communication in NSA and SA 5G Networks

Sinh Cong Lam<sup>1</sup> and Duc-Tan Tran<sup>2(✉)</sup>

<sup>1</sup> Faculty of Electronics and Telecommunications,  
VNU University of Engineering and Technology, Hanoi, Vietnam  
[conglv@vnu.edu.vn](mailto:conglv@vnu.edu.vn)

<sup>2</sup> Faculty of Electrical and Electronic Engineering,  
Phenikaa University, Hanoi, Vietnam  
[tan.tranduc@phenikaa-uni.edu.vn](mailto:tan.tranduc@phenikaa-uni.edu.vn)

**Abstract.** The 5G cellular system is a promising new radio network to provide an ultra-high user performance and experience by enhancing various advanced techniques from previous network generations, particularly 4G. Cooperative communication is one of the most powerful techniques, that allows each User Equipment (UE) instantaneously to associate with more than one Base Station (BS). In Stand Alone (SA) network architecture where 4G and 5G networks operate independently, cooperative communication technique is reformed as Coordinated Multi-Point Processing (CoMP) in which the UE can have connections to either 4G or 5G BSs. In Non-Stand Alone architecture where the 4G and 5G networks coexist and co-operate together, the cooperative communication technique is officially called EN-DC (Evolved-Universal Terrestrial Radio Access-New Radio) and allows the UE to access both 4G and 5G technologies at the same time. In this paper, the performance of UE in terms of coverage probability is analyzed in an urban area with different scenarios such as CoMP between 4G BSs, EN-DC, and CoMP between 5G BSs.

**Keywords:** Dual connectivity · Coordinated Multi-point Processing · 5G · Poisson Point Process · Coverage probability

## 1 Introduction

The 5G cellular systems are being deployed by most telecommunication companies to provide a wonderful Quality of Services (QoS) such as ultra-reliable low latency data transmission [6, 7]. According to 3GPP documents, the deployment of 5G networks can be separated into two phases [3]. During the first phase, the 5G and its predecessor coexist and co-operate together to allow the UE to perform dual data transmission with both 4G BS (eNodeB) and 5G BS (gNodeB). The network structure is officially called Non-Stand Alone (NSA) while the dual connectivity is called EN-DC (Evolved-Universal Terrestrial Radio Access-New Radio) [4]. From the UE point of view, the EN-DC has a similar function as the

Coordinated Multi-Point Processing (CoMP) that was introduced for 4G networks [11] and is being developed to adapt with the 5G systems [4]. The main difference between EN-DC and CoMP refers to the BSs that participate in the transmission process. In the EN-DC scheme, the BSs are eNodeB and gNodeB. Meanwhile, CoMP schemes require the BSs to utilize the same technologies. For example, the gNodeB only can cooperate with other gNodeB to perform CoMP functions. In general, both EN-DC and CoMP can be called Cooperative Communication Scheme (CCS).

The performance of both EN-DC and CoMP scheme have been studied in the literature [1, 5, 8]. The coverage probability of user in Poisson Point Process (PPP) network utilizing CoMP was derived. References [1] evaluated the EN-DC scheme and compared to hard handover. The analytical results illustrated that both CoMP and EN-DC can provide better network performance. In [9], a new EN-DC scheme was proposed to enhance the network throughput while maintaining the 5G QoS. In [10], the data rate and latency of SA and NSA 5G networks was conducted through preliminary field trials. However, the comparison between these CCS have not been derived. Thus, this paper utilizes stochastic geometry model to analyze and compare the performance of these CCS schemes in term of coverage probability.

This paper utilizes both mathematical model and simulation tool to analyze the UE's coverage probability in the cellular network utilizing CCS. The analytical results illustrate that the EN-DC outperforms both 4G and 5G CoMP due to this scheme can provide coverage probability in both indoor and outdoor environments.

## 2 System Model

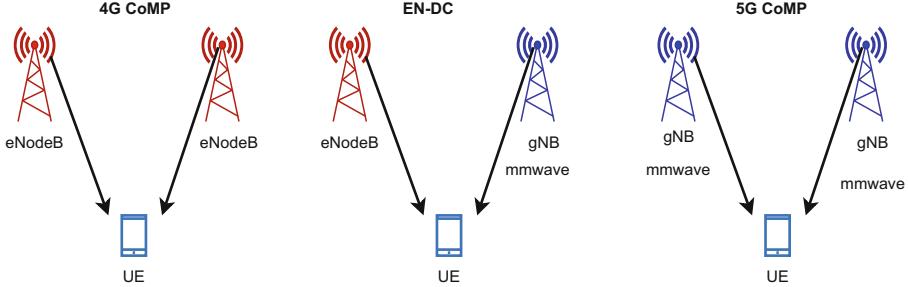
The paper considers CCS in RAN of cellular network schemes, particularly 4G, 5G NSA and 5G SA. It is assumed that all BSs share the whole allocated bandwidth and the network area unit is served by  $\lambda$  BSs, e.g.  $\lambda \text{ BS}/\text{km}^2$ . These BSs are spatially located according to Poisson Point Process. In practical networks, the BSs can work independently in terms of resource allocation. Thus, each UE may experience intercell interference from all adjacent base stations.

Since the number of BSs  $\gg 1$ , it is supposed that the density of interfering BSs is approximated to  $\lambda$ . Denote  $\lambda^{e\text{NodeB}}$ ,  $\lambda^{g\text{NodeB}}$  are the density of eNodeB and gNodeB stations in the networks. Thus,  $\lambda = \lambda^{e\text{NodeB}} + \lambda^{g\text{NodeB}}$ .

Figure 1 illustrates potential cooperative communication schemes

- In 4G networks, RAN comprises of eNodeBd,  $\lambda^{e\text{NodeB}} = \lambda$  and  $\lambda^{g\text{NodeB}} = 0$ . Thus, the CCS refers to the cooperative operation of eNodeBs. The serving BSs are two eNodeB while the density of interfering eNodeBs is  $\lambda$ .
- In 5G NSA network where the eNodeB and gNodeB stations are installed and works together. Each UE can be served by both eNodeBs and gNodeBs. Thus, CCS is called EN-DC. The interfering BSs includes eNodeBs and gNodeBs with density of  $\lambda^{e\text{NodeB}}$  and  $\lambda^{g\text{NodeB}}$ , respectively.

- In 5G SA networks where there are only gNodes in the system, the CCS is taken place between gNodeBs. The serving and interfering stations are two gNodeBs where the density of interfering gNodeBs is  $\lambda$ .



**Fig. 1.** Cooperative communication model in cellular networks

As recommendation of 3GPP, the path loss can be characterized as a dual slope model in which the UE experience either Light of Sight (LoS) or Non LoS [2]. Conventionally, the dual slope model can be characterized by the indicator functions LoS  $p^{LoS}$  and NLoS  $p^{NLoS}$  where  $p^{NLoS} = 1 - p^{LoS}$  and

$$\begin{cases} p^{LoS} = 1 & \text{if the wireless link is LoS} \\ p^{LoS} = 0 & \text{if the wireless link is NLoS} \end{cases} \quad (1)$$

This paper examines an urban network area with a high density of obstacles. The eNodeB works as a picocell with a radius of up 200 m while gNodeB is served as a femtocell with a coverage range of 10 m. In other words, the eNodeB can cover outdoor areas such as street corner, a park, a stadium; the gNodeB provides connections to indoor UEs within a room, small houses or shops. Thus, the path loss of the signal generated from eNodeB and gNodeB are follows UMi-Street Canyon and Indoor-mixed office [2].

## 2.1 Downlink SINR

In dual connectivity network model, the UE can associate with either eNodeB or gNodeB. The signal of the UE from eNodeB and gNodeB can be presented by the following equation

$$S^{(z)} = P^{(z)} \sum_l p^{(l)} g^{(z,l)} L^{(z,l)} \quad (2)$$

where  $z = \{e, g\}$  and  $l = \{LoS, NLoS\}$  correspond to eNodeB and gNodeB, LoS and NLoS, respectively;  $P^{(z)}$  is the transmission power of BS  $z$ ;  $G^{(z)}$  and  $L^{(z)}$  are channel power gain and path loss from the UE to BS type  $z$ .

The UE experiences interference from all adjacent BSs include eNodeB and gNodeB, the total interference of the user is determined by

$$I = \sum_z P^{(z)} \sum_{m \in \theta_e} \sum_l p^{(z)} g_m^{(z,l)} L_m^{(z,l)} \quad (3)$$

where  $G_m^{(z)}$  and  $L_m^{(z)}$  are channel power gain and path loss from the UE to  $m^{th}$  of BS type  $z$ .

Since the UE performs data exchange BS providing the best link, the UE periodically measures signal from the cooperative BSs to find out the best one. In this paper, it is also assumed that the cooperative BSs transmit on the same sub-carrier to serve the UE. Thus, the UE experiences the same intercell interference whatever it connects to gNodeB or eNodeB. Hence, the desired received signal power of the UE is

$$S = \max(S^{(e)}, S^{(g)}) \quad (4)$$

The received SINR of the UE with Gaussian power  $\sigma^2$  is defined as

$$SINR = \frac{S}{I + \sigma^2} \quad (5)$$

### 3 Coverage Probability

#### 3.1 Definition

In this paper, the possibility of a user is under the network service area, i.e. coverage probability, is the primary metrics that will be evaluated to compare the performance of CCS in these networks. Mathematically, the coverage probability is defined as

$$\mathcal{P}(T) = \mathbf{P}(SINR > T) \quad (6)$$

in which  $T$  is the minimum required SINR for the UE to perform successfully data exchange. It is noted that SINR is a function of various random variables such as the distances, channel power gain. Thus, symbol  $\mathbf{P}$  is used in Eq. 6 instead of  $P$  to recall that the probability of  $SINR > T$  is the conditional probability and the coverage probability is obtained by taking the expectation of this conditional probability.

#### 3.2 Stochastic Geometry Approach

The following equations determine the coverage probability in the case of EN-DC. For 4G CoMP and 5G CoMP, the similar approaches can be employed. Substituting SINR definition in Eq. 5 into coverage definition in Eq. 6, we obtain

$$\begin{aligned}\mathcal{P}(T) &= \mathbf{P} \left( \frac{\max(S^{(e)}, S^{(g)})}{I + \sigma^2} > T \right) \\ &= 1 - \mathbf{P} \left( \min(S^{(e)}, S^{(g)}) < T(I + \sigma^2) \right)\end{aligned}\quad (7a)$$

$$= 1 - \mathbf{P} \left( S^{(e)} < T(I + \sigma^2) \right) \mathbf{P} \left( S^{(g)} < T(I + \sigma^2) \right) \quad (8a)$$

$$= 1 - \prod_z \mathbf{E} \left[ \mathbf{P} \left( P^{(z)} \sum_l p^{(l)} g^{(z,l)} L^{(z,l)} < T(I + \sigma^2) \right) \right] \quad (8b)$$

$$= 1 - \prod_z \mathbf{E} \left[ \sum_l \bar{p}^{(l)} \mathbf{P} \left( P^{(z)} g^{(z,l)} L^{(z,l)} < T(I + \sigma^2) \right) \right] \quad (8c)$$

where Eq. 8b is obtained by using Eq. 2; Eq. 8c due to the mutual exclusion of the events  $l = \text{LoS}$  and  $l = \text{NLoS}$ . Thus,  $\mathcal{P}(T)$  is given by

$$\mathcal{P}(T) = 1 - \prod_z \mathbf{E} \left[ \sum_l \bar{p}^{(l)} F_G \left( \frac{T(I + \sigma^2)}{P^{(z)} L^{(z,l)}} \right) \right] \quad (9)$$

where  $F_G(\cdot)$  is the Cumulative Density Function of channel power gain  $g$ .

If  $g$  has an exponential distribution such as the Probability Density of  $g$  is  $f(g) = \exp(-g)$ , the coverage probability  $\mathcal{P}(T)$  equals to

$$\mathcal{P}(T) = 1 - \prod_z \mathbf{E} \left[ \sum_l \bar{p}^{(l)} \left[ 1 - \exp \left( \frac{T(I + \sigma^2)}{P^{(z)} L^{(z,l)}} \right) \right] \right] \quad (10)$$

In a cellular system with a high density of BSs, the BS's transmission power  $P^{(z)} \gg \sigma^2$ . Thus, it is possible to ignore the effects of  $\sigma$  on network performance. Hence,  $\mathcal{P}(T)$  can be approximated by

$$\mathcal{P}(T) = 1 - \prod_z \mathbf{E} \left[ \sum_l \bar{p}^{(l)} \left[ 1 - \exp \left( \frac{TI}{P^{(z)} L^{(z,l)}} \right) \right] \right] \quad (11)$$

Utilizing the definition of  $I$  in Eq. 3, we obtain

$$\mathcal{P}(T) = 1 - \prod_z \mathbf{E} \left[ \sum_l \bar{p}^{(l)} \left[ 1 - \prod_z \prod_{m \in \theta_e} \prod_l \exp \left( \frac{TP^{(z)} p^{(z)} g_m^{(z,l)} L_m^{(z,l)}}{P^{(z)} L^{(z,l)}} \right) \right] \right] \quad (12)$$

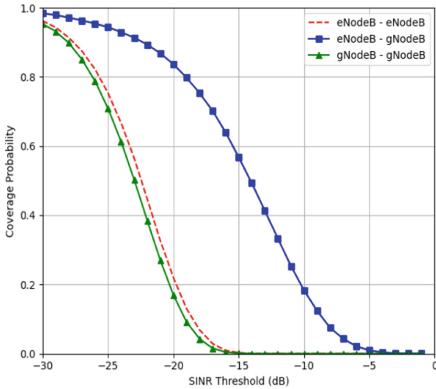
### 3.3 Performance Simulation and Comparison

In this section, the coverage probability of a UE in three cooperative communication scenarios are simulated and compared together.

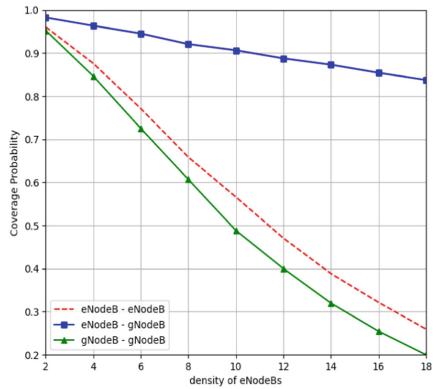
Figures 2 compares the coverage probability of the user with density of eNodeB  $\lambda^{eNodeB} = 2 \text{ BS/km}^2$  and  $\lambda^{gNodeB} = 20 \text{ BS/km}^2$ . The eNodeB and gNodeB are assumed to work on a carrier frequency of 5 GHz and 30 GHz, respectively. Thus, the signals of eNodeB will coverage a large area such as a crowded street while the signals of gNodeB only can cover a small area such as Indoor environment. Hence, it is assumed that the signals of eNodeB and gNodeB suffer UMi-Street Canyon and Indoor-mixed office path loss models [2].

It is seen from the figure that 4G CoMP and 5G CoMP provide a similar performance to UE but significantly lower than EN-DC. For example, when the SINR threshold is  $-25 \text{ dB}$ , the coverage probability of the UE in EN-DC is about 0.95 which is 1.36 times others in 4G CoMP and 5G CoMP. This phenomenon can be explained as follows:

- In 4G CoMP scheme, the eNodeB can cover a large area. There will be a huge amount of interference if a lot of eNodeBs are deployed in this networks. Thus, the UE performance in this network architecture is limited.
- In 5G CoMP scheme, since the gNodeBs in this simulation only cover the indoor UEs. Thus, the outdoor UEs will be out of coverage area. Furthermore, the indoor UEs are also affected by inter-gNodeB interference if more than one gNodeBs are deployed in the same room.
- In EN-DC scheme with deployment of both eNodeB and gNodeB, the outside UEs can be served by eNodeB while inside ones are under responsibility of gNodeBs. Thus, it is reasonable that all the EN-DC can achieve the best performance.



**Fig. 2.** Coverage probability vs SINR threshold with  $\lambda^{eNodeB} = 0.2$  and  $\lambda^{gNodeB} = 2$



**Fig. 3.** Coverage probability vs  $\lambda^{eNodeB}$

Figure 3 plots the coverage probability with different values of density of eNodebs  $\lambda^{eNodeB}$ . As seen from this figure that higher values of density of

eNodeBs and gNodeBs results in lower coverage probability. For example, when  $\lambda^{eNodeB} = 10$  and  $\lambda^{gNodeB} = 100$  ( $BS/km^2$ ), the coverage probability is about 0.5 which is 44% less than when  $\lambda^{eNodeB} = 6$  and  $\lambda^{gNodeB} = 60$  ( $BS/km^2$ ). This is due an increase of intercell interference as density of eNodeBs and gNodeBs increases. Consequently, it can be said that reducing  $\lambda^{eNodeB}$  and  $\lambda^{gNodeB}$  is an possible solution to improve UE coverage probability. However, the selection of  $\lambda^{eNodeB}$  and  $\lambda^{gNodeB}$  depend on various parameters such as number of required connections, bandwidth and so on.

## 4 Conclusion

In this paper, the Cooperative Communication Schemes (CCS), particularly 4G CoMP, 5G CoMP and EN-DC were studied. The UE coverage probability of these schemes were analyzed with different values of SINR thresholds and density of BSs. The simulation in a dense urban area illustrates that the EN-DC scheme that utilizes both eNodeB and gNodeB can provide a significant coverage probability than others. This is due to eNodeB is good at providing outdoor coverage while gNodeB is suitable for serving indoor UEs. It is also noticed that higher density of eNodeB and gNodeB may results in an increase in inter-cell interference and negative impacts on UE performances.

## References

1. Polese, M.: Performance comparison of dual connectivity and hard handover for LTE-5G tight integration in mmWave cellular networks. Master's thesis, University of Padova (2016)
2. 3GPP TR 38.901 version 15.0.0 Release 15: 5G; Study on channel model for frequencies from 0.5 to 100 GHz (2018)
3. 3GPP TS 37.340 version 15.3.0 Release 15: Universal Mobile Telecommunications System (UMTS); LTE; 5G; NR; Multi-connectivity; Overall description (2018)
4. Agiwal, M., Kwon, H., Park, S., Jin, H.: A survey on 4G–5G dual connectivity: road to 5G implementation. IEEE Access **9**, 16193–16210 (2021). <https://doi.org/10.1109/ACCESS.2021.3052462>
5. Bassoy, S., Farooq, H., Imran, M.A., Imran, A.: Coordinated multi-point clustering schemes: a survey. IEEE Commun. Surv. Tutor. **19**(2), 743–764 (2017). <https://doi.org/10.1109/COMST.2017.2662212>
6. Larsson, C.: Fundamentals of 5G Mobile Networks. Academic Press, Cambridge (2018)
7. Erik Dahlman, S.P., Skold, J.: 5G NR, The Next Generation Wireless Access Technology. Wiley, Hoboken (2020)
8. Liao, W.S., Kibria, M.G., Villardi, G.P., Zhao, O., Ishizu, K., Kojima, F.: Coordinated multi-point downlink transmission for dense small cell networks. IEEE Trans. Veh. Technol. **68**(1), 431–441 (2019). <https://doi.org/10.1109/TVT.2018.2880788>
9. Mishra, P., Kar, S., BollaPragada, V., Wang, K.C.: Multi-connectivity using NR-DC for high throughput and ultra-reliable low latency communication in 5G networks. In: 2021 IEEE 4th 5G World Forum (5GWF), pp. 36–40 (2021). <https://doi.org/10.1109/5GWF52925.2021.00014>

10. Mohamed, R., Zemouri, S., Verikoukis, C.: Performance evaluation and comparison between SA and NSA 5G networks in indoor environment. In: 2021 IEEE International Mediterranean Conference on Communications and Networking (MeditCom), pp. 112–116 (2021). <https://doi.org/10.1109/MeditCom49071.2021.9647621>
11. Sesia, S., Toufik, I., Baker, M.: LTE, The UMTS Long Term Evolution: From Theory to Practice. Wiley, Hoboken (2009)



# Cross-Layer Design for Wireless Sensor Networks Using Cooja Tool in Contiki Operating System

Nguyen Canh Minh<sup>1</sup>, Duc-Tan Tran<sup>2</sup>, Quang-Trung Hoang<sup>2(✉)</sup>, Trung Ninh Bui<sup>3</sup>, Gian Quoc Anh<sup>4</sup>, Kolla Bhanu Prakash<sup>5</sup>, and Sayed El-Rabaie<sup>6</sup>

<sup>1</sup> University of Transport and Communications, Hanoi, Vietnam  
ncminh@utc.edu.vn

<sup>2</sup> Faculty of Electrical and Electronic Engineering, Phenikaa University, Hanoi 12116, Vietnam  
{tan.tranduc,trung.hoangquang}@phenikaa-uni.edu.vn

<sup>3</sup> VNU University of Engineering and Technology, Hanoi, Vietnam  
ninhbt@vnu.edu.vn

<sup>4</sup> Nam Dinh University of Technology Education, Nam Dinh, Vietnam

<sup>5</sup> Koneru Lakshmaiah Education Foundation: Vijayawada, Andhra Pradesh, India

<sup>6</sup> Menofia University, Menouf, Egypt

**Abstract.** RPL is a routing protocol for low power and lossy networks (LLNs) or wireless sensor networks (WSNs). In recent years, this protocol is considered more and more by research communities in the world because of its potential benefits on the internet of things (IoT) applications. However, this protocol still has some drawbacks related to balancing packet traffic and saving consumed energy of sensor nodes which needs to be improved. In this paper, we propose an enhanced routing protocol based on ContikiMAC and RPL protocols in the Contiki operating system approached on the cross-layer design, named CL-RPL. By analyzing and evaluating the affection of the proposed protocol using the Cooja tool in the Contiki operating system, we estimated the consumed battery energy of sensor nodes in WSNs and proved that our CL-RPL protocol can support the system performances better than the previous original RPL protocol in terms of network lifetime and packet delivery ratio.

**Keywords:** ContikiMAC · RPL · Cross-layer design · WSN · LLN · IoT

## 1 Introduction

Internet of things (IoT) plays a very important role in the industrial revolution 4.0, and wireless sensor networks are a fundamental component to form real-world IoT applications. One of the issues that has attracted much attention by the research community in recent times is how to make WSNs work effectively when the network nodes often work on built-in battery power. The approach for popularization studies is to focus on the network protocol stack. The main challenges for researches related to solving problems of saving consuming energy, storage capacity, and processing power, and thus LLNs are

introduced. Nodes in this network work with low power consumption, low transmission rate, and complexity. To improve the performance of LLNs, there was some previous research related to the medium access control (MAC) and the routing protocol.

Considering the MAC sub-layer, in the reference [1], the author Adam Dunkels presented the ContikiMAC protocol for LLNs. This protocol allows saving the energy of network nodes based on the radio duty cycling (RDC) mechanism. In addition, this protocol also was known by other authors as in references [2–4].

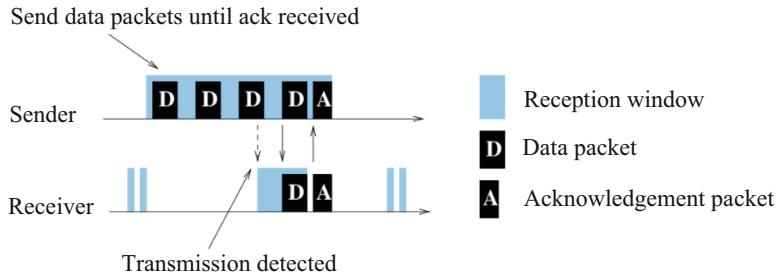
According to another approach, many authors focus on the network layer to improve the performance of LLNs [5]. In where, the IPv6 Routing Protocol for LLNs, so-called the RPL protocol in [6] was known such as the flexible potential routing protocol [7, 8]. There were many research results based on the RPL protocol directed to the objective function (OF) for improving the performance of networks as in references [9–11]. However, most of these previous works mainly consider in distinguishing protocol layers such as the network layer of the MAC layer in the network protocol stack. There were a few authors who exploited the wireless network protocol operation which allows sub-layers can exchange information with each other leading to being able to enhance the network performance [12–15]. Authors in [15] considered a cross-layer objective function, but new information fields must be added in control messages. Depending on the actual network architecture, the cross-layer design should be applied properly so that the network nodes can easily operate and archive the best network performance. In this paper, we propose a cross-layer design between ContikiMAC and RPL for LLNs, so-called the CL-RPL protocol. Therefore, online topology status from the network layer (such as the number of any parent node's child nodes) is informed the RDC sub-layer for adjusting the channel check rate (CCR), and vice versa (for two-way interaction), the network layer can receive related information from the RDC sub-layer to calculate the path metric for routing to the sink node. The result is that the congestion problem or the payload balancing problem can be solved. In addition, the objective function of the proposed protocol is also designed to be aware of the consumed battery energy and the forwarding packet traffic for improving the network lifetime. To implement simulation, we use the Cooja tool in the Contiki operating system. Results showed that our CL-RPL protocol outperforms the original RPL protocol.

The remainder of this paper is organized as follows: Sect. 2 presents a brief on the ContikiMAC and RPL protocol, Sect. 3 details a description of the cross-layer design, the system setup and the simulation results are shown in Sect. 4, and the final section is the conclusion of the paper.

## 2 ContikiMAC and RPL Protocol

### 2.1 ContikiMAC

In the Contiki's network stack for IEEE 802.15.4 [16], RDC is part of the MAC sub-layer which can be one in many standard MAC protocols. The Contiki-MAC solves some disadvantages of previous protocols such as B-MAC, X-MAC, etc. [2]. The mechanism of this protocol is shown as in Fig. 1 [1].



**Fig. 1.** The mechanism of the ContikiMAC.

Further, the ContikiMAC uses a Clear Channel Assessment (CCA) mechanism for detecting the radio activity and Channel Check Ratio (CCR) for waking network nodes up for checking this radio activity. The CCR parameter is also known as the RDC frequency. Depending on this frequency, the consumed energy of the network node can be high or low. For the ContikiMAC protocol, the fixed RDC frequency is configured. This can cause energy waste in low packet traffic situations where the RDC frequency is large, or it can also decrease the ability of packet reception if the packet traffic is large and the RDC frequency is small because the network congestion occurs. Therefore, a flexible mechanism is necessary so that the RDC frequency can be adjusted flexibly to adapt to different conditions or traffic states of the network. This problem can be solved through cross-layer operation between the network layer and the MAC layer as well as lower layers, as the design is given in this paper.

## 2.2 The RPL Protocol

In the architecture of the Contiki operating system, the RPL is built to work on IPv6. It is a distance-vector routing protocol that operates on top of the ContikiMAC. The RPL protocol is driven by two objective functions, namely, the objective function zero (OFO) and the minimum rank with hysteresis objective function (MRHOF). In which, the latter one allows the protocol to select the most reliable path, except that it doesn't have a load balancing mechanism [11].

MRHOF uses the ETX metric which is the number of transmissions a sensor network node expects to make to its destination node to successfully deliver a data packet. The ETX value is formulated by  $ETX = 1/(D_f * D_r)$ . Where  $D_f$  is the measured probability that a data packet is received by the neighbor and  $D_r$  is the measured probability that the acknowledgment packet is successfully received.

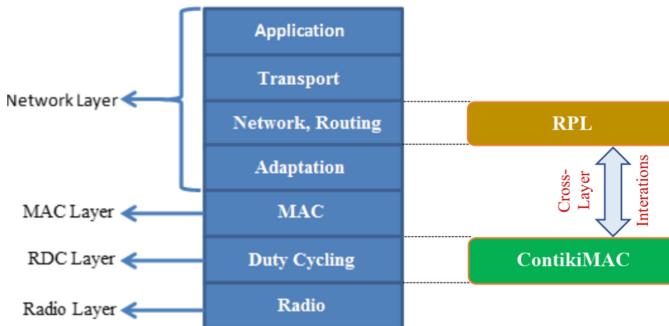
To solve the packet traffic balancing problem mentioned above, network protocol layers need to interact or exchange information with each other to adjust system parameters flexibly.

### 3 Proposal ContikiMAC-RPL Cross-Layer Design

#### 3.1 The Protocol Architecture

Figure 2 shows the enhanced protocol stack for cross-layer design in which the mechanism of the cross-layer interaction is integrated with the Contiki network stack [18]. This design makes it possible to exchange information between the RDC layer and the network layer via ContikiMAC and RPL protocols respectively.

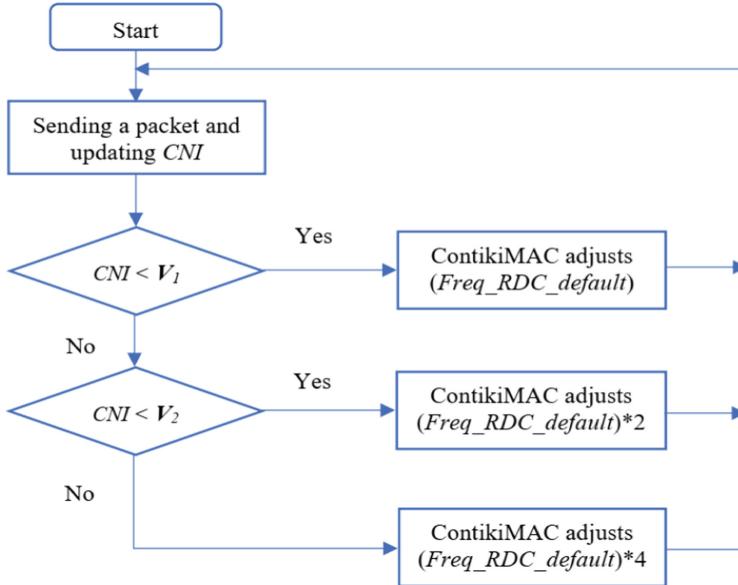
Contiki's network protocol stack is organized by the network modules corresponding to a complete protocol stack covering all traditional OSI layers. There are four layers in the Contiki's network stack model: Radio layer, RDC (Radio Duty Cycling) layer, MAC (Medium Access Control) layer, and Network layer. Where, the RDC plays a role as a part of MAC and the ContikiMAC protocol is used together with the Carrier Sense Multiple Access (CSMA) mechanism. In the Contiki operation system, the RDC layer is modeled by the module `contikimac.c`, and the network and routing layer is modeled by the module `rpl.c`. Therefore, we can design processing algorithms flexibly for enhancing the performance of the network protocol.



**Fig. 2.** Cross-layer interaction enabling network architecture for LLNs.

Interactive operations for interchanging information between the RPL protocol and the ContikiMAC protocol in the protocol layers are implemented according to the algorithm flowchart shown in Fig. 3.

In the Fig. 3, the child node index (CNI) parameter presents the number of any medium or parent node's child nodes. The larger this value is, the easier it is for bottlenecks to occur at intermediate network nodes. Therefore, the ContikiMAC and RPL protocols are designed to be able to exchange information about this value.  $V_1$ ,  $V_2$  respectively are the threshold values of CNI for the ContikiMAC protocol to base on to adjust the RDC frequency to increase or decrease.  $Freq\_RDC\_default$  is the frequency of the radio duty cycle (RDC) or channel check rate (CCR). In the Contiki operation system, this value is set by 8 Hz. To estimate the CNI value, we design to the network layer updates the information of the number of child nodes using the CL-RPL protocol, and this information is transferred to the RDC layer by the ContikiMAC. The intermediate node based on the packet IP address information of the nodes it received a packet



**Fig. 3.** The flowchart of cross-layer operation in the proposed protocol design.

in a certain interval to count and update the number of its child nodes (or CNI index). If the CNI is less than the threshold value  $V_1$ , the ContikiMAC protocol will set the RDC frequency to the default value (Freq\_RDC\_default), otherwise, if the CNI is less than  $V_2$ , the RDC frequency will be set to Freq\_RDC\_default\*2, otherwise again, the RDC frequency is set by Freq\_RDC\_default\*4.

### 3.2 Enhanced Combining Objective Function (ECOF)

In the architecture of the Contiki operating system, Contiki-RPL is an IPv6-based routing protocol for LLNs. It is a distance-vector routing protocol that operates on top of the Contiki-MAC. The Contiki-RPL protocol is driven by two objective functions, namely, the objective function zero (OF0) and the minimum rank with hysteresis objective function (MRHOF). In which, the latter one allows the protocol to select the most reliable path, except that it doesn't have a load balancing mechanism. In the research, we enhanced the MRHOF of the RPL protocol, and it names e-MRHOF.

The e-MRHOF uses a new path metric including three parameters related to the link quality (the ETX parameter as in the original RPL protocol [6]), the child node index (CNI) and the consumed energy index (EI) of the sensor node. Once a node (N) chooses its preferred parent (P), the node computes its rank from the preferred parent's rank as follows

$$Rank(N) = Rank(P) + Rank_{increase} \quad (1)$$

with

$$Rank_{increase} = \alpha * ETX + \beta * CNI + \gamma * EI \quad (2)$$

where,  $\alpha$ ,  $\beta$ , and  $\gamma$  are weighted coefficients. By choosing various values for these coefficients, we can obtain different results for the performance of the network protocol. To get the best objective function, we do a simulation with various coefficients of  $\alpha$ ,  $\beta$ , and  $\gamma$  and choose accordant values, that is  $\alpha = 1$ ,  $\beta = \frac{1}{2}$ , and  $\gamma = \frac{1}{2}$ . Further, the  $EI$  parameter is assigned equal to values 1, 2, 3, 4, and 5 corresponding to the consumed energy percentage of the sensor node is less than or equal to 20%, larger than 20% and less than or equal to 40%, larger than 40% and less than or equal to 60%, larger than 60% and less than or equal to 80%, larger than 80% and less than or equal to 100%.

### 3.3 The Network Lifetime

The lifetime of the sensor node is calculated based on the remaining energy of each node. So, first, we present the expression to calculate the node's energy consumption as follows [17]

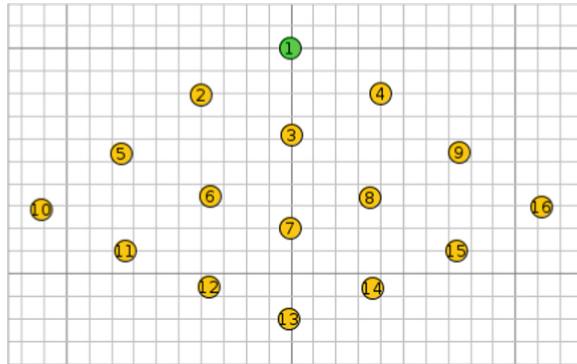
$$\frac{E}{V} = I_{cpu}t_{cpu} + I_{lpm}t_{lpm} + I_{rx}t_{rx} + I_{tx}t_{tx} + \sum_i I_{C_i}t_{C_i} \quad (3)$$

where  $V$  is the supply voltage,  $I_{cpu}$  and  $t_{cpu}$  are the current drawn from the microprocessor when running and the time taken by it in running mode (the CPU is active without using the radio transceiver),  $I_{lpm}$  and  $t_{lpm}$  are the current drawn and the time of the microprocessor in low power mode,  $I_{tx}$  and  $t_{tx}$  is the current drawn and the time of the radio module in transmit mode,  $I_{rx}$  and  $t_{rx}$  are the current drawn and the time of the radio module in receive mode respectively;  $I_{C_i}$  and  $t_{C_i}$  are respectively the current drawn and the time of other components such as sensor boards, LEDs, external flash access, etc. Therefore, the lifetime of any sensor node depends on its own consumed energy, and we can refer that the network lifetime is the duration for the network operating until there is any sensor node in WSN being stopped working due to its battery capacity.

## 4 System Setup and Simulation Results

### 4.1 Simulation Setup

To evaluate the proposed cross-layer design of the CL-RPL protocol, in this paper, we use the Cooja tool in the Contiki operating system. The network simulation topology is shown in Fig. 4 which consists of a sink node and 15 sender nodes.

**Fig. 4.** Network topology for simulation.

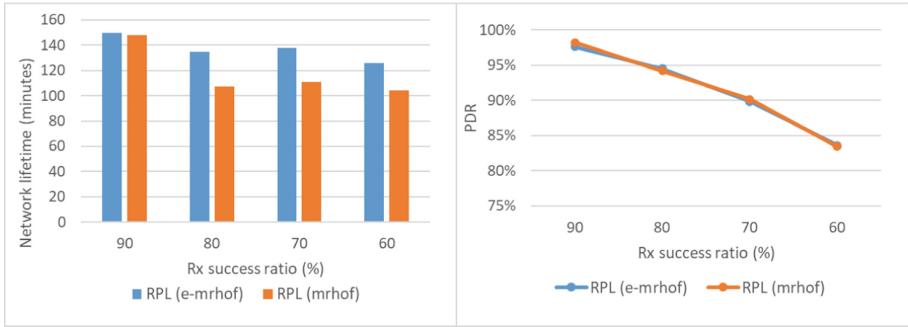
In addition, Tmote Sky platform and CC2420 radio modules are chosen for building the network model. Table shows general simulation configurations (Table 1).

**Table 1.** General simulation configurations

Parameter	Value	Parameter	Value
Operation system	Contiki-2.7	Rx succ. ratio	60%, 70%, 80%, 90%
Simulator	Cooja	Routing	RPL, CL-RPL
Radio medium	UDGM	MAC/RDC Driver	CSMA/ContikiMAC
Mote type	Tmote-Sky	RDC frequency	8, 16, 32 (Hz)
Tx/Int range	50 m/ 100 m	Packet Duration	30 s
Tx succ. ratio	100%	$V_1/V_2$	4/10

## 4.2 Simulation Results of the First Scenario

Firstly, we evaluate the effect of the e-MRHOF. Figure 5 shows a comparison between the RPL protocol based on e-MRHOF and the original RPL protocol [6] based on MRHOF using together with the ContikiMAC protocol [1]. The simulation result shows the network lifetime and PDR respectively overall values of receiving success ratio of packets (Rx success ratio). we implemented simulation 10 times to get the average value, each simulation was run is 90 min. In this case, the RDC frequency is set to 8 Hz.



**Fig. 5.** Comparison of average energy consumption at the PDR.

Observing Fig. 5, we clearly see that the network lifetime of the RPL protocol using e-MRHOF is longer than the RPL protocol using MRHOF (the left window) while the PDR of both two protocols is approximately equal (the right window). Especially at Rx success ratio corresponding to 80, 70, and 60%, the time difference is more evident. The reason is that the e-MRHOF objective function is designed to be aware of the parameters related to the remaining battery power of each sensor node and the number of its children at instantaneous times. When a parent node has a deep battery reduction or/and its number of children increases rapidly, its path metric value will be increased, leading to its child node also having an increased path metric, and this child node tends to move to another parent node. Then, the forwarding packet traffic by intermediate or parent nodes can be decreased, and the possibility of network congestion is reduced as well as the number of retransmissions is also decreased. Therefore, the lifetime of intermediate or parent nodes lasts or the network lifetime is longer.

#### 4.3 Simulation Results of the Second Scenario

To compare the CL-RPL protocol and the RPL protocol [6] using the ContikiMAC [1], in this scenario, we set up values  $V_1 = 4$  and  $V_2 = 10$ , the default RDC frequency is set to 8 Hz. The simulation result is shown in Fig. 6. We can see clearly that the PDR of both two protocols is decreased overall as the link quality, specifically, the Rx success ratio decreases from 90% to 60%. However, the CL-RPL protocol obtains a higher PDR than the original RPL protocol. Thanks to the cross-layer design, the ContikiMAC driving the RDC layer receives the status information of CNI for adjusting the RDC frequency higher when any medium node has more children or more forwarding packet traffic as shown in the flowchart of Fig. 3. This results in increasing the number of received packets or the PDR is increased. At the Rx success ratio of 90%, the difference is insignificant while At lower values of the Rx success ratio, the difference of the PDR is clear.

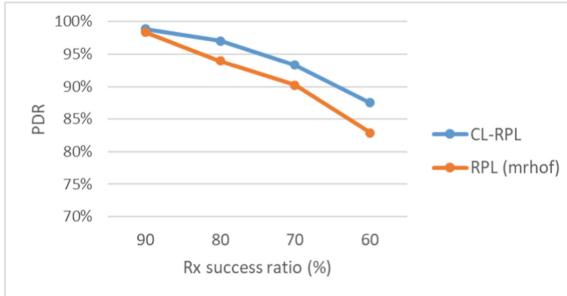


Fig. 6. Comparison of PDR between the CL-RPL and the RPL protocol.

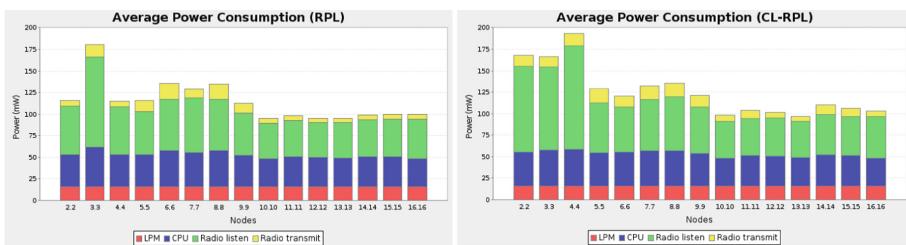


Fig. 7. Average power consumption at Rx success ratio 60%.

Observing Fig. 7, we see that the average power consumption of leaf nodes is lower than medium nodes while medium nodes consume more battery capacity. For both protocols, sensor nodes labeled from 5 to 16 have the approximately same power consumption. At average, sensor nodes being near to the sink (node 2, 3, 4), they consume more battery energy due to forwarding more packet from their child nodes. However, the CL-RPL protocol archives the energy balance among these nodes compared with the original RPL protocol. In addition, for the CL-RPL protocol, these nodes being near to the sink node consume slightly more power than the RPL protocol because the higher RDC frequency is used for these nodes to send and receive packets successfully and obtain higher PDR. Further, by implementing simulation using Cooja tool, we can estimate the energy consumption of each node in WSNs for sensor nodes in actual networks [18–20].

## 5 Conclusion

In this paper, we have demonstrated the effectiveness of a protocol designed with a cross-layer approach to exchange information dynamically. The results of analysis and computer simulation using the Cooja tool in the Contiki operating system have shown that our proposed CL-RPL protocol achieves better performance than the original RPL protocol in terms of the network lifetime and PDR. Especially, it is possible to flexibly adjust the parameters of the protocol to achieve system performances as well as to determine the network configuration before implementing it in the fact. In future research, we will consider cooperation among sensor nodes for improving transmission reliability and interaction at the radio layer.

## References

1. Dunkels, A.: The ContikiMAC Radio Duty Cycling Protocol, SICS Technical Report (2011)
2. Homayouni, S., Javidan, R.: ERA-ContikiMAC: an adaptive radio duty cycling layer in Internet of Things. In: 2018 9th International Symposium on Telecommunications (IST) (2018)
3. Youssef, M.F., Elsayed, K.M.F., Zahran, A.H.: Contiki-AMAC—The enhanced adaptive radio duty cycling protocol: Proposal and analysis. In: 2016 International Conference on Selected Topics in Mobile Wireless Networking (MoWNeT) (2016)
4. Wymore, M.L., Qiao, D.: RIVER-MAC: A receiver-initiated asynchronously duty-cycled MAC protocol for the Internet of Things. In: 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC) (2019)
5. Dinesh, P.S., Palanisamy, C.: A survey on design and implementation of routing protocol for low power and lossy networks. In: 2019 International Conference on Advances in Computing and Communication Engineering (ICACCE), (2019)
6. Thubert, P., et al.: Winter, RPL: IPv6 routing protocol for low-power and lossy networks., RFC 6550, IETF, March 2012
7. Xie, H., Zhang, G., Su, D., Wang, P., Zeng, F.: Performance evaluation of RPL routing protocol in 6lowpan. In: 2014 IEEE 5th International Conference on Software Engineering and Service Science (2014)
8. Mohamed, B., Mohamed, F.: Experimental evaluation of RPL protocol. In: 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST) (2016)
9. Sousa, N., Sobral, J.V.V., Rodrigues, J.J.P.C., Rabélo, R.A.L., Solic, P.: ERAOF: a new RPL protocol objective function for Internet of Things applications. In: 2017 2nd International Multidisciplinary Conference on Computer and Energy Science (SpliTech) (2017)
10. Janani, P., Diniesh, V.C., Joseph Auxilius Jude, M.: Impact of path metrics on RPL's performance in low power and lossy networks. In: 2018 International Conference on Communication and Signal Processing (ICCPSP) (2018)
11. Ghaleb, B., et al.: A survey of limitations and enhancements of the IPv6 routing protocol for low-power and lossy networks: a focus on core operations. IEEE Commun. Surv. Tutorials **21**(2), 1607–1635 (2019)
12. Hoang, Q.T., Tran, X.N.: Improved cross-layer cooperative MAC protocol for wireless ad hoc networks. In: Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific (2014)
13. Gonizzi, P., Medagliani, P., Ferrari, G., Leguay, J.: RAWMAC: a routing aware wave-based MAC protocol for WSNs. In: 2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) (2014)
14. Jalal, D., Jeridi, M.H., Ezzedine, T.: Low power and lossy networks LLNs: cross-layer optimization. In: 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM) (2019)
15. Safaei, B., Monazzah, A.M.H., Ejlali, A.: ELITE: an elaborated cross-layer RPL objective function to achieve energy efficiency in Internet-of-Things devices. IEEE Internet Things J. **8**(2), 1169–1182 (2021)
16. Samarth Pai, M.: 9 September 2016. [https://anrg.usc.edu/contiki/index.php/Network\\_Stack](https://anrg.usc.edu/contiki/index.php/Network_Stack). Accessed Oct 2021
17. Kumar, J.M., Venkateswaran, R.P., Kini, G.N.: Software-based energy estimation for localization algorithms in wireless sensor networks using Contiki-Cooja. Int. J. Pure Appl. Math. **119**(12), 12655–12663 (2018)
18. Nguyen, D.-C., Tran D.-T., Tran, D.-C.: Application of compressed sensing in effective power consumption of WSN for landslide scenario. In: Asia Pacific Conference on Multimedia and Broadcasting, pp. 1–5. IEEE (2015)

19. Gian Quoc, A., Nguyen Dinh, C., Tran Duc, N., Tran Duc, T., Kumbesan, S.: Wireless technology for monitoring site-specific landslide in Vietnam. *Int. J. Electr. Comput. Eng.* **8**(6), 4448–4455 (2018)
20. Chinh, N.D., Tan, T.D.: Development and implementation of a wireless sensor system for landslide monitoring application in Vietnam. *Int. J. Inf. Commun. Technol.* **13**(2), 227–242 (2018)



# Mobile Robot Motion Control Using a Combination of Fuzzy Logic Method and Kinematic Model

Anh-Tu Nguyen<sup>1</sup>(✉) and Cong-Thanh Vu<sup>2</sup>

<sup>1</sup> Faculty of Mechanical Engineering, Hanoi University of Industry, Hanoi 159999, Vietnam  
tuna@haui.edu.vn

<sup>2</sup> Faculty on Mechanical Engineering, University of Economics - Technology for Industries, Hanoi 159999, Vietnam  
vcthanh@uneti.edu.vn

**Abstract.** Mobile robots have been widely used in various aspects of human life. When a robot moves between different positions in the working area to perform the task, controlling motion to follow a pre-defined path is the primary task of a mobile robot. Furthermore, the robot must remain at its desired speed to cooperate with other agents. This paper presents a development of a motion controller, in which the fuzzy logic method is combined with a kinematic model of a differential drive robot. The simulation results are compared well with experimental results indicate that the method is effective and applicable for actual mobile robots.

**Keywords:** Fuzzy logic control · Kinematic model · Trajectory tracking

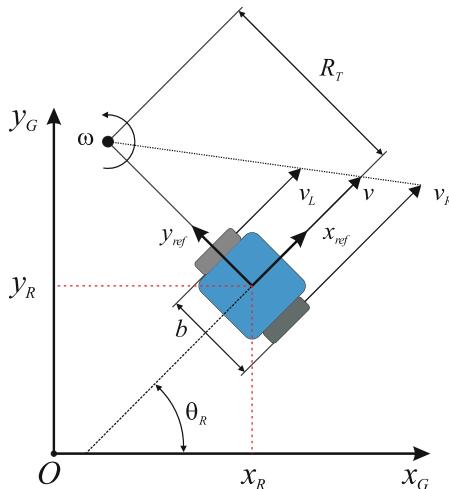
## 1 Introduction

Motion control for the mobile robot has been an interesting topic and received the attention of numerous researchers. The objective of robot motion control is to follow pre-defined trajectories represented by its position, velocity, or acceleration profile as a function of time. The trajectory is normally divided into basic segments such as straight lines, circle pieces, or curves. The control problem is thus to precompute a smooth trajectory-based online and circle segments which drive the robot from the start point to the goal point [1]. Conventional control methods are mainly developed using the kinematic and dynamics governing equation system. The kinematic control system is represented by kinematic equations and the control inputs are the pure velocity [2, 3]. This method has shown several disadvantages because the robot will not automatically adapt to the change of the environment or the unsMOOTH transition between trajectory segments. This results in a discontinuous acceleration of the motion. To solve the limitations regarding the kinematic control method, the dynamics control has been developed, in which both torque and force are used as the input parameters and the dynamics model is necessary for designing the controller [4–6]. However, it is almost impossible to build a perfect dynamics model of a robot in real applications. As such, fuzzy logic control

methods were introduced to cope with parameter uncertainties, therefore, improve the motion control quality [7–10]. The fuzzy logic controller is normally applied in combination with the PID controller. The fuzzy logic controller detects the PID parameter to adapt to the working environment. This method help robots tract the pre-defined path better [8]. The simulation results showed a better control quality compared with the traditional control method [9]. Another application of fuzzy logic control that should be cited here is the method based on Z-number for tracking trajectory. This approach helps coding works become easier and deal with missing observations [10]. Recently, with the development of technologies in hardware such as vision sensors and artificial intelligence. Robots become smarter and can autonomously perform complex tasks, cooperate with other agents with high precision, high speed, and stable safety [11].

This paper presents an approach for controlling mobile robot motion, in which the fuzzy method is incorporated with the kinematic model to manage the linear velocity while following the trajectory. The proposed method then is applied in a differential drive mobile robot for verification. The good comparison between simulation and experiment proves that the robot can follow the pre-defined path and manage desired linear velocity while remains an acceptable angular velocity.

## 2 Kinematic Model



**Fig. 1.** The demonstration of the robot concerning reference and global coordinate system.

This section addresses the relationship between different geometrical factors and kinematic parameters of a robot model. Because the reference coordinate system ( $Ox_{ref}$ ,  $y_{ref}$ ) is fixed at the center of the robot, therefore the position ( $x_R$ ,  $y_R$ ) and orientation ( $\theta_R$ ) are the relative motion between this coordinate system and the global coordinate system ( $Ox_G$ ,  $y_G$ ) as illustrated in Fig. 1.

The tangential velocity of two wheels ( $v_L$ ,  $v_R$ ) are in the relation with the angular velocity ( $\omega_L$ ,  $\omega_R$ ) as following:

$$v_R = \omega_R R \quad (1)$$

$$v_L = \omega_L R \quad (2)$$

Therefore, the robot linear velocity can be determined as:

$$v = \frac{v_R + v_L}{2} \quad (3)$$

$$v = \frac{(\omega_R + \omega_L)R}{2} \quad (4)$$

Supposing the distance between two wheels is  $b$ , the angular velocity of the robot is:

$$\omega = \frac{v_R - v_L}{b} \quad (5)$$

$$\omega = \frac{(\omega_R - \omega_L)R}{b} \quad (6)$$

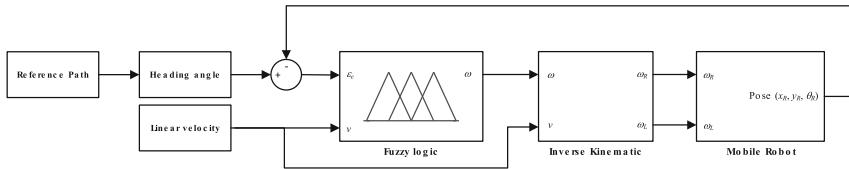
From the Eqs. (4) and (6), the angular velocities of each wheel in the relation with  $v$  and  $\omega$  can be expressed:

$$\omega_R = \frac{v + b\omega}{R} \quad (7)$$

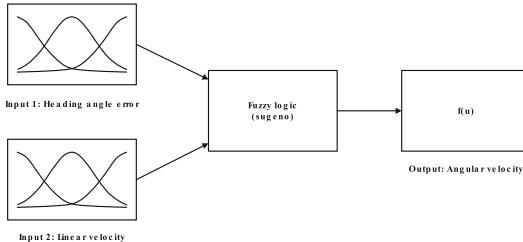
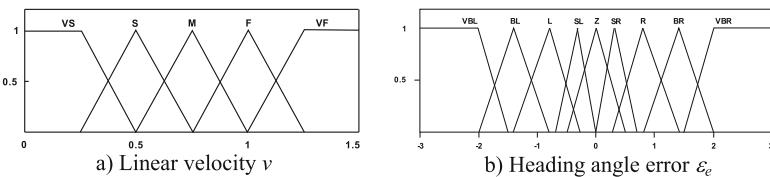
$$\omega_L = \frac{v - b\omega}{R} \quad (8)$$

### 3 Fuzzy Control Method

To enhance the robot operation, this section presents an approach in which the controller is designed by using the kinematic model and fuzzy logic method. The control system is shown in Fig. 2, in which inputs of the fuzzy block involve the heading angle error ( $\varepsilon_e$ ) and linear velocity ( $v$ ) of the robot. After defuzzification, the output of the fuzzy controller, the angular velocity ( $\omega$ ), is used along with the linear velocity for calculating the angular velocity of each robot's wheel via inverse kinematic equations. This helps detect appropriate angular velocity to remain the linear velocity stable while still following the path.

**Fig. 2.** The block diagram of the control system.

The fuzzy controller involves three main units as in Fig. 3: fuzzification, fuzzy inference, and defuzzification. Figure 4 and Table 1 show the input membership functions and the fuzzy control outputs. If the heading angle error is small, the output angular velocity is low. Otherwise, it is high. This output angular velocity will also depend on the setup linear velocity. Table 2 describes the definition of the fuzzy rules.

**Fig. 3.** Configuration of fuzzy control.**Fig. 4.** Membership function plotting of  $v$  and  $\varepsilon_e$ .**Table 1.** Membership function plots of output variable angular velocity.

Information	Name	Value
Very big left deviation	VBL	- 5
Big left deviation	BL	- 3
Left deviation	L	- 1.75
Small left deviation	SL	- 0.5
Zero deviation	Z	0
Small right	SR	0.5
Right deviation	R	1.75

(continued)

**Table 1.** (continued)

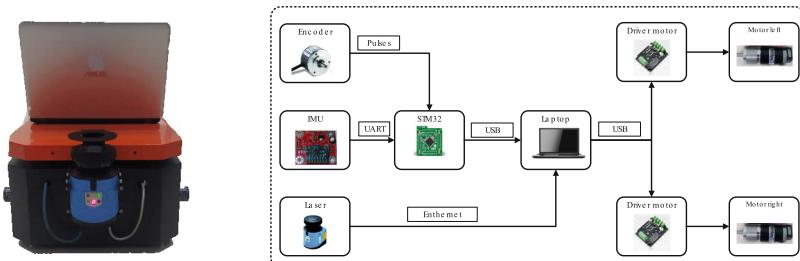
Information	Name	Value
Big right deviation	BR	3
Very big right deviation	VBR	5

**Table 2.** Fuzzy rules for motion control.

Linear Velocity		VS	S	M	F	VF
Heading angle error	VBL	BR	VBR	VBR	VBR	VBL
	BL	BR	BR	VBR	VBR	VBR
	L	R	R	BR	BR	BR
	SL	SR	SR	R	R	R
	Z	Z	Z	Z	Z	Z
	SR	SL	SL	L	L	L
	R	L	L	BL	BL	BL
	BR	BL	BL	VBL	VBL	VBL
	VBR	BL	VBL	VBL	VBL	VBL

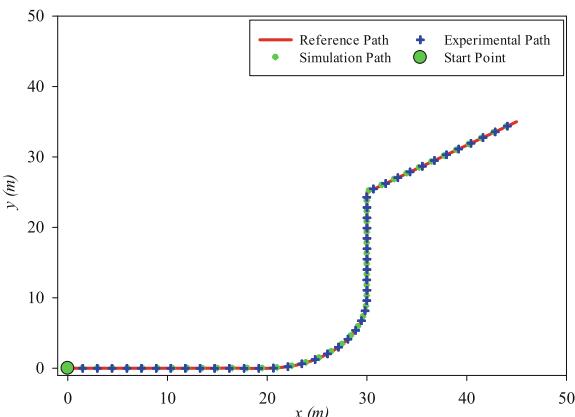
## 4 Results and Discussion

To examine the effectiveness and feasibility of the combination of kinematic model and fuzzy logic control for trajectory tracking of the differential wheeled mobile robot. The control system is demonstrated in Fig. 5, including a laptop that takes the role of center processing, a microcontroller, STM32F407VET6, receives and transfers signals among the laptop, drivers, and sensor system. The active wheels were driven using two DC motors. The sensor system is designed using both relative and absolute measurement methods, encoders, a digital compass IMU, and a laser scanner NAV245.

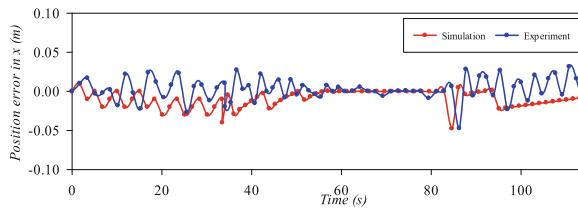
**Fig. 5.** The hardware system of the experimental robot

The trajectory is designed to include distinct profiles: a straight line along an axis, a circle, and a bias line with a total distance of 60 m. The simulation results are validated by experiments. As mentioned above, this method control robot to follow the path while remains the desired speed. Therefore, the velocity is varied to estimate the repose of the fuzzy controller.

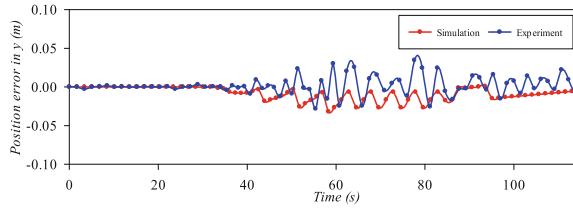
The robot's simulation and experimental performance of trajectory following are shown in Fig. 6 in the case of  $v = 0.6$  m/s. The results show that the robot can track the reference path over a long distance with different profiles. Figure 7 and Fig. 8 illustrate the specific position and orientation variation among the consideration case studies. When the robot moves in the x-direction, the error in the y-axis is pretty small while the error in the x-axis fluctuates stronger. The inversion in the error occurs when the robot runs along the y-direction. In the bias path, the position derivation in both two-axis is similar. Because the present method prioritizes to control robot's speed, therefore the simulation result of linear velocity is closed the reference value. However, there is a small oscillation in the experimental results (Fig. 9). This is understandable because the experimental performance is affected by different factors. The specific angular velocity of the robot is demonstrated in Fig. 10. The simulation and experimental results are compared well in almost the motion. Nevertheless, when the robot runs in the transition path, the head angle fluctuates stronger than the linear velocity. The estimation of the standard error is evaluated by Root Mean Square Error (RMSE) method in position, the linear and angular velocities after 30 trials are addressed in Table 3. In all consideration case studies, the error increases slightly in all parameters, however, the average error of the robot's linear velocity is less than 9 mm while the other parameters remain under the acceptable values for the robot activities.



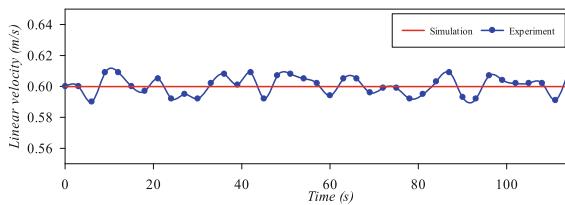
**Fig. 6.** Trajectory tracking at  $v = 0.6$  m/s.



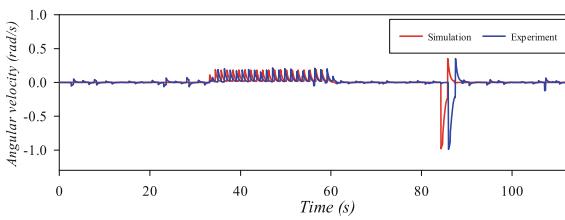
**Fig. 7.** The specific position derivation in x direction at  $v = 0.6$  m/s.



**Fig. 8.** The specific position derivation in y direction at  $v = 0.6$  m/s.



**Fig. 9.** The simulation results of linear and angular velocities at  $v = 0.6$  m/s.



**Fig. 10.** The experimental results of linear and angular velocities at  $v = 0.6$  m/s.

**Table 3.** The error estimation at different velocity.

No	Parameter	$v = 0.3 \text{ m/s}$		$v = 0.6 \text{ m/s}$		$v = 1.2 \text{ m/s}$	
		Simulation	Experiment	Simulation	Experiment	Simulation	Experiment
1	Position error in x-axis (m)	0.006	0.009	0.014	0.015	0.019	0.025
2	Position error in y-axis (m)	0.006	0.008	0.012	0.013	0.013	0.017
3	Linear velocity error (m/s)	0	0.006	0	0.007	0	0.009
3	Angular velocity error (rad/s)	0.05	0.064	0.082	0.085	0.11	0.124

## 5 Conclusion

This paper presents an approach of fuzzy logic along with the kinematic model to develop a controller for a wheeled mobile robot. Since the robot's pose plays an important role in doing tasks and is determined by angular and linear velocities, therefore, the angular error and the reference velocity are both used as inputs of fuzzy logic controller. The experiments indicates that the robot can successfully follow the trajectories and operate in desired linear velocity while remains acceptable angular velocity.

## References

1. Birol, K.: Motion control of wheeled mobile robots. *Interdiscip. Descr. Complex Syst.* **13**(1), 41–47 (2015)
2. Sandeep, K.M., Jharna, M.: Kinematics, localization, and control of differential drive mobile robot. *Glob. J. Res. Eng. Robot. Nano-Tech* **14**(1), 2–7 (2014)
3. Salem, F.A.: dynamic and kinematic models and control for differential drive mobile robots. *Int. J. Curr. Eng. Tech.* **3**(2), 253–263 (2013)
4. Wang, T.Y., Tsai, D.A. Wang, D.A.: Dynamic control of an omnidirectional mobile platform. *J. Nan Kai* **7**, (2010)
5. Ollin, P.M., Luis, A.M.M., Joaquin A., Miguel, G.V.C., Ramón, G.H.: Motion Control Design for an Omnidirectional Mobile Robot Subject to Velocity Constraints. *Mathematical Problems in Engineering* (2015)
6. Boru, D.H., Wang, Z.: Design and control for differential drive mobile robot. *Int. J. Eng. Res. Tech.* **6**(10), 327–334 (2017)
7. Tian, Q., Cheng, L., Wang, K., Liu, Y.: Research on motion control of mobile robot with fuzzy PID Arithmetic. In: *The Ninth International Conference on Electronic Measurement & Instruments*, pp. 3–363 (2009)

8. Abatari, H.T., Tafti, A.D.: Using a fuzzy PID Controller for the path following of a car-like mobile robot. In: International Conference on Robotics and Mechatronics, pp. 189–193 (2013)
9. Wang, T.Y., Chang, C.D.: Hybrid fuzzy PID controller design for a mobile robot. In: Proceedings of IEEE International Conference on Applied System Innovation 2018 IEEE ICASI 2018- Meen, Prior & Lam (Eds) (2018)
10. Abdelwahab, M., Parque, V., Ahmed, M.R.F.E., Abouelsoud, A.A., Sugano, S.: Trajectory tracking of wheeled mobile robots using Z-number based fuzzy logic. *IEEE Access* **8**, 2169–3536 (2020)
11. Rihem, F., Khaled, A.J., Saad, A., Mohammad, T.Q., Moufida, K.: A deep learning approach for the mobile-robot motion control system. *Intell. Autom. Soft Comput.* **29**(2), 423–435 (2021)



# Examining Topology and Dynamics-Related Modular Robustness in a Large-Scale Signaling Network

Cong-Doan Truong<sup>(✉)</sup> and Duc-Quynh Tran

International School, Vietnam National University, Hanoi, Vietnam  
`{doantc, quynhtd}@isvnu.vn`

**Abstract.** The discovery of the role of structural properties and in-/out-module robustness for predicting important modules of biological networks is a challenge. Many existing studies have been proposed to detect modules in biological network significantly. However, the topology and robustness of modules have not been well discovered yet. Therefore, there is a pressing need to conduct that. In particular, we detected all modules by using an existing algorithm in a large-scale signaling network. Next, the in-/out-module robustness computation was implemented using the OpenCL library for parallel computation to reduce computation time. Then, for the purpose of evaluating the position of modules in the network, we use five central metrics. Moreover, all modules were divided into two groups (Group 1, Group 2) based on a mean value of module size in which Group 1 contained larger modules and Group 2 consisted of smaller modules. We compared gene ontology (GO) analysis of Group 1 with Group 2. We also examined the proportion of crucial genes such as disease, drug-target, and essential genes in the former and that in the latter. Consequently, module sizes were positively correlated with centrality values. It meant that the group of larger modules was located in center of network. Regarding to in-/out-module robustness, the average value of in-module robustness of Group 1 was slightly higher than that of Group 2. On the other hand, the average value of out-module of Group 1 was smaller than that of Group 2. Additionally, the GO of Group 1 was enriched more significant than that of Group 2. Finally, the number of important genes in group 2 was less than in group 1. Taken together, the group of larger modules is more important than that of smaller modules in large-scale signaling network.

**Keywords:** Boolean dynamics · Module detection · In/out module robustness · Network structure · Essential genes · Gene ontology · Biological signal network

## 1 Introduction

In fact, the sustainability of the biosignaling network and the structure of the biosignaling network have attracted the attention of many studies in recent times. The structure of the biological network can be found in the following studies. In study [1], for example, they showed that biological networks have a hierarchical or scale-free structure. Regarding the sustainability of a biological network, how can it measure the sustainability and

stability of the network is one of the challenges. In previous studies [2, 3], it has been shown that the robustness of the network can resist mutations that occur. Besides, there are also many studies showing the relationship between the strength of the network and the biological network structure [4, 5]. However, all these studies have the common feature of focusing only on the network level, not considering the module level of the network.

Recently, there have been many studies on network module finding algorithms [6, 7]. These studies have a common feature of focusing on improving the module search accuracy of the network instead of studying the structure of the modules. In addition, there are some studies on the internal and external durability of the module when mutations occur [2, 3]. The common feature of these studies is that they focus on internal and external sustainability depending on the module's internal and external mutations. In summary, the previous studies have many limitations, have not focused on research to find important modules in the biological network or have not studied the structural relationship of these modules with the sustainability of the modules. Thus, researching on the network structure of modules, such as determining the position of modules in the biological network. What are the characteristics of the important genes in modules? These are research questions that need to be addressed. Therefore, in this paper, we focus on studying module structure characteristics and the durability and biological function of modules in bio-signal networks.

This scientific document is divided as follows. Section 2 will mention the methods applied in this paper such as module strength measurement, center measurement. Next section is the results part. Finally, conclusion section will be mentioned.

## 2 Preparations and Methodology

### 2.1 Datasets

In this study, we used human signal network as data [8]. This biological network has been commonly used in many previous studies [2, 3]. The structure of network is as follows, and is composed of 5,443 genes and 24,469 interactions.

### 2.2 Module Detection

In this article we use a well-known algorithm to find all the modules in the bio-signal network [6]. The bio-signal network is seen as a directed graph with  $V$  considered the vertex set and  $A$  being the set of oriented edges. Algorithm divides graph into completely separate modules. To increase the precision of modularity, we performed the algorithm 30 times and took the average of those 30 times for the modularity value.

### 2.3 Definition of Network Robustness and Internal and External Durability of the Module

In this research paper, the robustness of the biological signal network has been evaluated using model. This model is called as the network which is Boolean. This Boolean model

is represented as a network which is directed. The network has a direction that is  $G(V, A)$ . The group of nodes is stored in  $V$ . The group of edges is called  $A$ . Each node has two states, 0 or 1. Each edge of the graph has two types: activating and inhibiting from  $v_i$  to  $v_j$ . In this model, the state of the model at time  $t$  changes to state  $t + 1$  through a pre-existing set of rules. There are two types of convergence states that are convergent to a point. The second is the converging state to form a closed loop. The state  $s(t)$  will converge to one of these two states. The network is considered stable when the mutation occurs at the node of  $v_i$  if the states before and after the mutation are equal. More generally, the strength of the network is calculated using the following formula:

$$\gamma(G) = \frac{1}{N \cdot |S|} \sum_{s \in S} \sum_{i=1}^N I(\langle s \rangle = \langle s_{\bar{v}_i} \rangle), \quad (1)$$

In case of calculating the stability of in-/out-module of module  $V_i$  in graph  $G$ . Graph  $G$  is divided into discrete modules  $V_1, V_2, V_3, \dots, V_M$  as follows:

$$\gamma_{in}(V_i) = \frac{1}{|\tilde{S}|} \sum_{s \in \tilde{S}} \sum_{v \in V_i} \frac{H(\prod_{V_i} \langle s \rangle, \prod_{V_i} \langle s_{\bar{v}} \rangle)}{|V_i|} \quad (2)$$

and

$$\gamma_{out}(V_i) = \frac{1}{|\tilde{S}|} \sum_{s \in \tilde{S}} \sum_{v \in V_i} \frac{H(\prod_{V \setminus V_i} \langle s \rangle, \prod_{V \setminus V_i} \langle s_{\bar{v}} \rangle)}{|V_i|}, \quad (3)$$

In general, the durability of in-module and out-module can be calculated as follows:

$$\gamma_{in}(G) = \frac{1}{M} \sum_{i=1}^M \gamma_{in}(V_i) \quad (4)$$

and

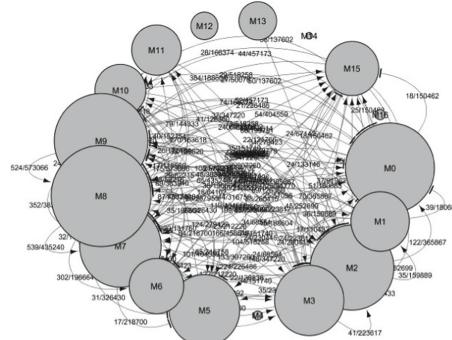
$$\gamma_{out}(G) = \frac{1}{M} \sum_{i=1}^M \gamma_{out}(V_i) \quad (5)$$

## 2.4 GO Analysis and Definition of How to Evaluate the Center of the Module

Several studies have shown that the location of genes/protein is related to their function. The genes or protein, if at the center, are often more important than the gens or proterins in the outer position. To test the position of modules in the network, we apply the measures degree [9], closeness [10], betweenness [11], stress [12] and eigenvector [10] of modules. Besides, we also apply MORO [2] to analyze the structural and functional characteristics of the two groups of modules.

## 3 Results

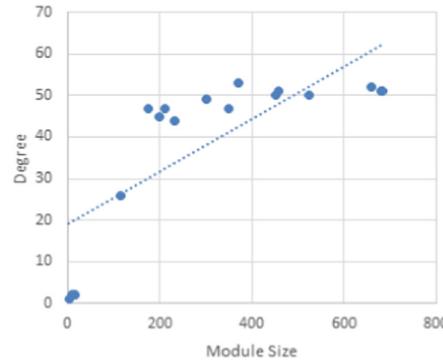
We apply the algorithm to find all the modules of the large signal network. The result is 16 discrete modules.



**Fig. 1.** This is the simulation result using the MORO tool. The results show the simulation of each relationship between the modules. Each node in the figure is considered a module. The number of genes determines the size of that respective node. Between two nodes there is a connecting edge [2].

### 3.1 The Relationship of Module Size and Module Centrality

There are five measures of centrality of gene in biological networks. We use five measures which are stress, closeness, degree, betweenness, and eigenvector. These five metrics are used to determine the position of modules in the bio signaling network. In Fig. 1, the results show as follows for the degree method, the correlation coefficient is 0.7677 and the P-value is 0.00032 using t-test (Fig. 2).



**Fig. 2.** The correlation between module size and degree centrality

Similarly, for the betweenness measure, the correlation coefficient value is 0.4716, but the P-value is not less than 0.01. The correlation coefficients of closeness and stress were 0.4696 and 0.6586, respectively (P-value < 0.05). Finally, eigenvector has a correlation coefficient which is 0.7047 and a P-value is 0.001. This indicates that modules with larger sizes are likely to be more centrally located in biological signal networks.

Previous studies have shown that important genes often have large degrees in biological networks. Therefore, the results in this paper are completely consistent with the previous results.

### 3.2 Comparison of In-Module Robustness Between Two Different Groups of Modules

In the second experiment, there are two groups in network. The total number of modules of the two groups is 16. Divide the group based on the threshold value. This threshold value is the average of the module size. In the first group, the module sizes are larger than the module average. In contrast, in the second group, the size of the modules is less than the mean value of the module. Next, we analyze the in-module and out-module durability value for all modules in these two groups. In-module robustness value is calculated by formula 4 and out-module robustness value is calculated by formula 5, resistant to initial states mutation. The results showed that the average value of the in-module robustness of the first group was greater than the average value of the in-module robustness of the second group (Group 1 was 0.9077 and group 2 was 0.9067). This suggests that larger groups of modules are generally more resilient to the mutations that occur within the biological signaling network. In contrast, the out-module robustness of the first group was less than the out-module robustness value of the second group (group 1 was 0.9026 and group 2 was 0.9027). This shows that the group of modules is larger than the group of modules with smaller size when mutations occur outside the module of the bio-signal network. This result is quite reasonable when compared with the results of previous studies on in/out-module robustness [2].

### 3.3 Gene Ontology Analysis

In Table 1, we conducted gene ontology analysis for two groups of modules, group 1 and group 2, using MORO software on Cytoscape [13] platform. The purpose of this analysis is to evaluate the locational and functional characteristics of two groups of genes belonging to two different module groups. The results in Table 1 show several GO terms such as Immune system process, Negative regulation of cell proliferation, Translational initiation, Early endosome membrane, DNA binding, Serine-type endopeptidase activity, and GTPase activator activity of group 1 and group 2. The percentage of GO in group 1 is higher than the percentage in group 2. In conclusion, the function between the two groups of modules is completely different.

**Table 1.** Below is the analysis result of HSN signalling network. This result gives GO analysis of two groups using P-value and t-test.

Types of GO analysis to analyze	Term of GO	Term Annotation % per Group 1	Term Annotation % per Group 2	P-value
Biological process	Immune system process	28	2.776	0.00017
	Negative regulation of cell proliferation	37	2.895	0.00119
	Translational initiation	33	0.79	0.00153
Component	Early endosome membrane	22	0.526	0.00974
Function	DNA binding	55	5.169	0.00213
	Serine-type endopeptidase activity	44	0.981	0.00967
	GTPase activator activity	33	0.766	0.00586

### 3.4 Essential Genes Analysis

We use probabilistic analysis to analyze the number of important genes. These important genes are located in two groups of modules. The purpose of this experiment is whether large modules often contain many important genes or not? We investigated the following three genes: the essential gene, the drug-target gene, and the disease gene. (Table 2)

**Table 2.** This table shows the percentage of essential genes of module group 1 and module group 2

Type of gene	Essential	Drug-target	Disease
Average of group 1	26.55%	18.50%	26.21%
Average of group 2	17.40%	13.52%	25.24%

The results showed that the percentages of the essential gene, drug-target gene, and disease gene of the module group 1 had a higher percentage of the number of genes than the module group 2. The results suggest that larger groups of modules are often important modules in the biological signaling network because important genes are more likely to appear than smaller modules.

## 4 Conclusion

There have been many studies regarding the identification of important modules in biological networks. However, there has been very little research related to the structural characteristics, dynamics, and statistics of important genes in the modules of the signaling network. In this article, we have performed analysis to find 16 modules of large network of human biological signals. The interesting result is that larger networks are often located at the center of the biological network. Another interesting result is that the robustness value of the in-module relative to a larger module group is larger than a smaller one. However, the out-module robustness of the larger module group is less than the out-module robustness of the smaller module group. We then obtained another result related to Gene Ontology. In the end, important genes such as essential, drug-target, and disease genes often appear in the large module group. In short, larger modules are often more important than smaller modules in the bio signal network.

## References

- Wuchty, S., Ravasz, E., Barabási, A.L.: The architecture of biological networks. In: Deisboeck, T.S., Kresh, J.Y. (eds) Complex Systems Science in Biomedicine. Topics in Biomedical Engineering International Book Series. Springer, Boston, MA, pp. 165–181 (2006). [https://doi.org/10.1007/978-0-387-33532-2\\_5](https://doi.org/10.1007/978-0-387-33532-2_5)
- Truong, C.-D., et al.: MORO: a Cytoscape app for relationship analysis between modularity and robustness in large-scale biological networks. *BMC Syst. Biol.* **10**, 122 (2016). <https://doi.org/10.1186/s12918-016-0363-3>
- Truong, C.-D., et al.: Investigation on changes of modularity and robustness by edge-removal mutations in signaling networks. *BMC Syst. Biol.* **11**, 125 (2017). <https://doi.org/10.1186/s12918-017-0505-2>
- Trinh, H.-C., et al.: PANET: a GPU-based tool for fast parallel analysis of robustness dynamics and feed-forward/feedback loop structures in large-scale bio-logical networks. *PLoS ONE* **9**, e103010 (2014). <https://doi.org/10.1371/journal.pone.0103010>
- Trinh, H.-C., et al.: R package for a Boolean sensitivity analysis against various types of mutations. *PLoS ONE* **14**, e0213736 (2019). <https://doi.org/10.1371/journal.pone.0213736>
- Leicht, E.A., Newman, M.E.J.: Community structure in directed networks. *Phys. Rev. Lett.* **100**, 118703 (2008). <https://doi.org/10.1103/PhysRevLett.100.118703>
- Chen, W., et al.: Multitask module identification for biological networks. *IEEE Trans. Evol. Comput.* **24**, 765–776 (2020). <https://doi.org/10.1109/TEVC.2019.2952220>
- Kim, J.-R., Heslop-Harrison, P., Cho, K.-H.: Reduction of complex signaling networks to a representative kernel. *Sci. Signal.* **4**, ra35 (2011). <https://doi.org/10.1126/scisignal.2001390>
- Jeong, H., et al.: Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001). <https://doi.org/10.1038/35075138>
- Wuchty, S., et al.: Centers of complex networks. *J. Theor. Biol.* **223**, 45–53 (2003). [https://doi.org/10.1016/s0022-5193\(03\)00071-7](https://doi.org/10.1016/s0022-5193(03)00071-7)
- Freeman, M.E.J., et al.: A set of measures of centrality based on betweenness. *Sociometry* **40**, 35–41 (1977). <https://doi.org/10.2307/3033543>
- Shimbel, A., et al.: Structural parameters of communication networks. *Bull. Math. Biophys.* **15**, 501–507 (1953). <https://doi.org/10.1007/BF02476438>
- Shannon, C., et al.: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003). <https://doi.org/10.1101/gr.123930>



# Implementation of SMOTE and VGG-16 for COVID-19 Radiography

Xuan Tho Dang<sup>1</sup>(✉) and Nam Anh Dao<sup>2</sup>

<sup>1</sup> Faculty of Information Technology, Hanoi National University of Education,  
Hanoi, Vietnam  
[thodx@hnue.edu.vn](mailto:thodx@hnue.edu.vn)

<sup>2</sup> Faculty of Information Technology, Electric Power University, Hanoi, Vietnam  
[anhdn@epu.edu.vn](mailto:anhdn@epu.edu.vn)

**Abstract.** This article presents a method for the analysis of chest X-ray images to automatically predict COVID-19. It relies on neural networks of deep convolutions for constructing features and analysis of class distribution for balancing the distribution. The use of VGG-16 - a pre-trained model allows us to achieve accurate features. The assumption of having the same distribution for all classes in dataset may be not suitable for our case study when the number of chest images which are affected by other diseases is larger than the number for COVID-19. We show the problem of data imbalance in order to improve the efficiency of predicting the minority class of COVID-19 patients. Thus, two different classification ways were proposed to carried out to check the effect of imbalanced distribution of data samples in each class and to demonstrate the advantages of the join classification solution.

**Keywords:** COVID-19 radiography · VGG-16 · SMOTE · CNN · SVM

## 1 Introduction

According to WHO, the corona virus belongs to a large family, from the common cold to dangerous diseases, and can infect both humans and animals. In recent studies, many unusual features on chest X-ray images are signs of COVID-19 patients. Specifically, Shi et al. [1] marked that many of the COVID-19 patients in their experiment had abnormal radiographic features; Bacellar et al. [2] also identified many anomalies such as background blur in X-ray images of COVID-19 patients. In the problem of medical image prediction and classification, to extract and classify image features, deep learning method is used very effectively by many scientists. Therefore, the application of artificial intelligence to exploit, analyze and identify patterns from medical data is becoming more and more a trend, and the experimental results are also very positive.

---

N. A. Dao—These authors contributed equally to this work.

In this paper, the proposed method uses deep learning network on X-ray images to automatically predict COVID-19. For this purpose, the features of the X-ray images were constructed using VGG16 trained models for highly accurate classification results.

## 2 Related Work

Currently, to diagnose COVID-19 patients, many countries apply the Polymerase Chain Reaction (PCR) method. However, the disadvantage of this method is that the high level of false negatives leads to difficulty in identifying COVID-19 positive patients [3]. Therefore, the method of diagnosis of COVID-19 derived from Computed Tomography (CT) and X-ray images is suggested by many current studies to improve accuracy in prediction [4]. Many studies have recently developed a machine learning system to help doctors classify patients with suspected COVID-19 infection on X-ray images [5, 6]. Apostolopoulos et al. [7] classifies X-ray images of COVID-19 into three categories: COVID-19, bacterial pneumonia, and healthy individuals. The dataset included a total of 1427 X-ray images.

In this study, we propose a method that combines deep learning and data rebalancing to promote the strengths of the two methods to achieve better prediction results. The imbalance problem is directly associated with the distribution of data into layers, can be solved in two main approaches, (1) by balancing the data then applying the traditional classifier; or (2) modify the classification algorithm to find the exact classification resulting from unbalanced data. In this article we focus on the first approach, which can be called the data level approach, where only the data is changed and no change is made in the classifier. Some methods that stand out in this approach, such as the synthetic minority over-sampling technique (SMOTE) [8], are notable methods in the area of over-sampling used in many applications. In addition, the methods SPY [9] and Mask-S [10] indicate that the majority class elements near the boundary are more misclassified than others and call them spy elements. From there, this method balances the data of the class labels by adjusting the class labels of some majority class elements to the minority class, thereby increasing the classification efficiency.

## 3 The Method

In the context of chest image analysis, the classification of an image requires a set of features to express the observed chest region and to detect patterns from the features. With a chest image dataset our learning method includes feature extraction (A), imbalance checking (B) and the classification (C).

**A. Feature Extraction.** The task starts with convolution operation conducted for creating basic layers in the Convolutional Neural Network (CNN): the operation is capable to capture the high-level properties such as textures and edges through sum of the product of a kernel  $h$  and image region:

$$s(x) = \sum_{i=-d}^d h(i) * I(x+i) \quad (1)$$

where  $I$  is the mark for the observed image,  $x$  is pixel location and  $d$  is the size of the filler  $h$ . To assess different levels of abstraction from image pixel, convolutional layers which are created by particular kernels are added to the CNNs. The adding order is specified in accordance to layers' ability of abstraction. The final layer summarized the condensed properties allowing to represent initial image by its high-level features suitable for further analysis. VGG-16, the strong CNN of the K. Simonyan and A. Zisserman [11] is implemented in our model for extracting the features for the chest images.

**B. Imbalance Checking.** This classification of a class  $c$  to an image sample  $s$  with features extracted from previous task (A) is expressed by using the Bayes's rule [12] which shows inference with conditional dependence of the class and the sample:

$$p(c | s) = \frac{p(s | c)p(c)}{p(s)} \quad (2)$$

After evaluating the posterior  $p(c | s)$  for labeled samples  $s$  and the evidence  $p(s)$  of the samples which are available in a training data set, we can proceed to estimate the prior  $p(c)$  for (2):

$$p(c) = \sum_{s \in S_{train}} p(c | s)p(s) \quad (3)$$

Each class such as class  $c_1$  is studied as in (3) and the prior that is corresponding to them is specified in a detailed formula:

$$p(c = c_1) = \sum_{s \in S_{train}} p(c = c_1 | s)p(s) \quad (4)$$

Assume that the posterior in our case study is binary with two classes  $c_1, c_2$ . To search a relevant class for a test sample  $s$ , the posterior  $p(c | s)$  for each class can be evaluated and compared, showing which posterior is the largest:

$$\frac{p(c = c_1 | s)}{p(c = c_2 | s)} = \frac{p(s | c = c_1)}{p(s | c = c_2)} * \frac{p(c = c_1)}{p(c = c_2)} \quad (5)$$

Now consider the second element of product in (5). It is defined by the ratio of the priors for the two classes:

$$r = \frac{p(c = c_1)}{p(c = c_2)} \quad (6)$$

As prior of each class was addressed by (4) we can estimate the ratio  $r$  by expanding (6) to remove the dependency of the evidence  $p(s)$ .

$$r = \frac{\sum_{s \in S_{train}} p(c = c_1 | s)p(s)}{\sum_{s \in S_{train}} p(c = c_2 | s)p(s)} = \frac{\sum_{s \in S_{train}} p(c = c_1 | s)}{\sum_{s \in S_{train}} p(c = c_2 | s)} \quad (7)$$

Note that, the lower number of samples labeled by class  $c_1$  in the training set, which is shown by the sum in the top of (7), in comparison with the number by

class  $c_2$ , the ratio  $r$  is estimated smaller as result of the imbalance. In this case, the possibility of recognizing a sample to be in class  $c_1$  turns to very low and the accuracy of the classification can be affected. It is convenient for denoting  $S_{c_1}$  for the set of samples in training with assigned label of  $c_1$ :

$$S_{c_1} = \{s, c(s) = c_1\} \quad (8)$$

In order to narrow down the difference of class distribution in the training dataset a small set of samples can be generated for minority class  $c_1$  by implementation of SMOTE proposed by Chawla N.V. et al. [8]:

$$S_{c_1} = S_{c_1} \cup \{s^{gen}, c(s^{gen}) = c_1\} \quad (9)$$

**C. Classification.** Once the tasks above have been conducted, data with both features and classes were checked and ready for learning. The way in which border line between samples of different classes in the feature space to be drawn is suggested by Support-vector machine (SVMs) [13] where the width  $\delta$  of the border is maximized.

$$c(s) = c_1 \text{ if } \left( \frac{p(c = c_1 | s)}{p(c = c_2 | s)} > 1 + \delta \right) \text{ else } c_2 \quad (10)$$

The leaning method of K-nearest neighbors (KNN) and random forest (RF) are also implemented in our model. The performance of classification can be measured with assistance of *Accuracy*, *Recall*, *Specificity* and *Gmean* [14].

## 4 Experimental Results

This experiment is with radiography images for evaluation the described learning method for dataset with unbalanced class distribution. A full set of radiography images is provided by COVID-19 Radiography Dataset [15]. As with the chest X-ray images, there are four classes covering Lung Opacity, COVID-19, Normal, and Viral Pneumonia.

Our first experiment is designed for binary classification that includes samples of COVID-19 class and group other classes with class distribution ratio of 1 : 9. The second experiment covers all four classes with the ratio of 1 : 3 : 3 : 3, see Table 1. Following cross validation, each test has 5 different data splits where a split has total 3136 images selected for training and 1346 images for test. The number of images for each class in a split is set in accordance with the mentioned above ratios.

**Table 1.** Data structure

Experiment	Class	Versus other classes	Ratio
Two classes	COVID-19	Group (Normal, Lung opacity, and Viral Pneumonia)	1:(3:3:3)
Four classes	COVID-19	Normal, Lung opacity, and Viral Pneumonia	1:3:3:3

**Table 2.** Performance report by VGG-16 features with two classes

Without SMOTE	Accuracy	Gmean	Recall	Specificity	Precision
KNN	93.39	61.12	37.61	99.55	90.34
RF	92.57	54.83	30.30	99.45	86.17
SVM	93.77	63.52	40.60	99.65	92.91
With SMOTE	Accuracy	Gmean	Recall	Specificity	Precision
KNN	81.35	83.97	87.46	80.68	33.65
RF	92.67	66.95	45.82	97.85	70.32
SVM	94.96	82.70	70.00	97.72	77.39

**Table 3.** Performance report by VGG-16 features with four classes

Without SMOTE	Accuracy	Gmean	Recall	Specificity	Precision
KNN	87.81	83.35	75.62	91.87	75.62
RF	91.40	88.34	82.79	94.27	82.79
SVM	92.94	90.46	85.88	95.29	85.88
With SMOTE	Accuracy	Gmean	Recall	Specificity	Precision
KNN	88.14	83.81	76.27	92.09	76.27
RF	91.54	88.55	83.09	94.36	83.09
SVM	93.03	90.59	86.06	95.35	86.06

By measuring performance with metrics for all splits, the averaged results were derived and Table 2 shows the performance of the first experiment with two classes. The *Accuracy* of KNN is 93.39 and this is down to 81.35 with SMOTE. RF and SVM show improvement of the *Accuracy* when use SMOTE. Note that the scores of *Gmean* are increased for all KNN, RF and SVM by SMOTE. This is explained by sensitivity of the metric to groups of samples of the same class.

Consider the performance of the second experiment in Table 3 by *Accuracy* scores and *Gmean* rates. This is clear that both *Accuracy* and *Gmean* have higher scores for SMOTE implementation. Using specific preparation and ResNet, chest images were studied for discovering COVID-19 in [16] and [18] while DenseNet was addressed in [20] with different datasets. Table 4 shows the results of these algorithms by *Accuracy* metric that is commonly used for performance. Our results by SVM, SMOTE and VGG-16 are list in the end of the table.

**Table 4.** Comparison with related methods

Method	Data	Accuracy
Combination of ResNet50 and MobileNet [16]	Dataset of COVID-19 chest X-ray, and Kaggle dataset of chest X-ray images	84.35
Five convolutional layers CNN [17]	2 databases	90.64
ResNet and Location attention [18]	219 COVID-19, 224 Viral pneumonia, 175 Healthy	86.70
Dual sampling attention network ResNet34 [19]	4982 images	87.90
Class balancing by SPY [6]	244 images of COVID-19 image data collection	89.88
DenseNet-169 [20]	COVID-CT with masks	89.00
Influence of batch size and image size [21]	COVID-19 Radiography database	87.00
VGG-16, SMOTE, SVM, 2 classes (our)	COVID-19 Radiography database	94.96
VGG-16, SMOTE, SVM, 4 classes (our)	COVID-19 Radiography database	90.03

## 5 Conclusions

In this paper, a deep learning-based method was described for detecting patients infected with COVID-19 by X-ray imaging incorporated with data rebalancing to improve performance. It can be first concluded that features extracted with pre-trained VGG-16 can cover high abstraction patterns and the use of VGG-16 is suitable for chest X-ray images analysis. By implementing convolution operators on layers, the features are condensed and appropriate for further analysis.

Note that, the imbalance of class distribution can lead to miss-classification for identification of COVID-19 from chest X-ray images. The clarification of the imbalance problem assisted for improvement of prediction performance. The association between feature engineering by VGG-16 and class balancing by SMOTE is used in our learning model. Tests were carried out by KNN, RF and SVM for two classes and four classes. The overall *Accuracy* and *Gmean* are promising on a benchmark real medical image database. The future work can cover study of other imbalance method for improvement of performance.

**Acknowledgment.** This research project was assisted by the Vietnam Ministry of Education and Training, with project code B2022 - SPH - 04.

## References

1. Shi, H., Han, X., et al.: (2020) Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect Dis.* **20**(4), 425–434 (2020)
2. Bacellar, G., Chandrappa, M., et al.: COVID-19 Chest X-Ray Image Classification Using Deep Learning, *medRxiv* 2021.07.15.21260605 (2021)
3. Alizad-Rahvar, A.R., Vafadar, S., et al.: False negative mitigation in group testing for COVID-19 screening. *Front. Med.* **8**, 579 (2021)
4. Ortiz, A., Trivedi, A., et al.: Effective deep learning approaches for predicting COVID-19 outcomes from chest computed tomography volumes. *Sci. Rep.* **12**, 1716 (2022)
5. Tho, D.X., Anh, D.N.: Deep learning-based imbalanced data classification for chest X-Ray image analysis. In: The International Conference on Intelligent Systems & Networks, pp. 109–115 (2021)
6. Tho, D.X., Anh, D.N.: Imbalance in learning chest X-ray images for COVID-19 detection. In: Phuong, N.H., Kreinovich, V. (eds.) *Soft Computing: Biomedical and Related Applications*. SCI, vol. 981, pp. 107–119. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-76620-7\\_9](https://doi.org/10.1007/978-3-030-76620-7_9)
7. Apostolopoulos, I.D., Mpesiana, T.A.: Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Phys. Eng. Sci. Med.* **43**(2), 635–640 (2020). <https://doi.org/10.1007/s13246-020-00865-4>
8. Chawla, N.V., Bowyer, K.W., et al.: SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002)
9. Dang, X.T., Tran, D., et al.: SPY: a novel resampling method for improving classification performance in imbalanced data. In: 2015 Seventh International Conference on Knowledge and Systems Engineering (KSE), pp. 280–285 (2015)
10. Hung, B.D., Anh, D.N., Tho, D.X.: Relabeling with mask-S for imbalanced class distribution. In: Satapathy, S.C., Bhateja, V., Nguyen, B.L., Nguyen, N.G., Le, D.-N. (eds.) *Frontiers in Intelligent Computing: Theory and Applications*. AISC, vol. 1013, pp. 31–41. Springer, Singapore (2020). [https://doi.org/10.1007/978-981-32-9186-7\\_4](https://doi.org/10.1007/978-981-32-9186-7_4)
11. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556* (2015)
12. Barber, D.: Bayesian Reasoning and Machine Learning. Cambridge University Press (2012)
13. Cortes, C., Vapnik, V.N.: Support-vector networks (PDF). *Mach. Learn.* **20**(3), 273–297 (1995). CiteSeerX 10.1.1.15.9362
14. Akosa, J.S.: Predictive accuracy : a misleading performance measure for highly imbalanced data. In: Proceedings of the SAS Global Forum (2017)
15. COVID-19 Radiography Database. <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>
16. Nandi, R., Mulimani, M.: Detection of COVID-19 from X-rays using hybrid deep learning models. *Res. Biomed. Eng.* **37**(4), 687–695 (2021). <https://doi.org/10.1007/s42600-021-00181-0>
17. Ahmed, F., Bukhari, S.A.C., et al.: A deep learning approach for COVID-19 & viral pneumonia screening with X-ray images. *Digit. Gov. Res. Pract.* **2**(2), 18 (2020)
18. Ozturk, T., Talo, M., et al.: Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput. Biol. Med.* **121**, 103792 (2020)

19. Xi, O., Jiayu, H., et al.: Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia applications. *Stud. Comput. Intell.* **981** (2020)
20. Xi, Y., He, X., et al.: COVID-CT-Dataset: A CT Scan Dataset about COVID-19 (2020)
21. Hasmadi, A.B.: Interactive automation Of COVID-19 classification through X-Ray images using machine learning. *J. Independent Stud. Res. Comput.* **18**(2) (2020)



# Research Complex Fuzzy Inference System in Early Warning Credit Risk at Commercial Banks in Viet Nam

Hong Thi Chu Hai<sup>(✉)</sup> and Trieu Thu Huong

Faculty of Management Information Systems, Banking Academy, Hanoi, Vietnam  
[{Haict, huongtrieu}@hvnh.edu.vn](mailto:{Haict, huongtrieu}@hvnh.edu.vn)

**Abstract.** Credit risk can adversely affect all areas of socio-economic life, especially in Finance and Banking. The risk can cause unexpected large losses to the profit as well as the safety of a bank. Therefore, prediction is a solution to hedging risks and limiting losses in which banks are particularly interested. Today, the Mamdani Complex Fuzzy Inference System (M-CFIS) has been known as an outstanding decision support tool by evaluating the amplitude and phase values of an event leading to timely and accurate decision making. The experiments show that this method is objective and effective. As a result, it can be regarded as a theoretical, scientific, and dependable basis for commercial banks' early detection and warning of credit issues.

**Keywords:** Complex fuzzy · Mamdani (M-CFIS) · Credit risk

## 1 Introduction

As the capital demand of the economy increases, the bank's risks increase, putting great pressure on credit risk management activities. After the financial crisis of 2007–2008, credit risk reduction has arisen as a hot topic, prompting financial institutions to collect several of the decision-making data. As a consequence, risk analysts are faced with the challenging task of precisely estimate of a client's creditworthiness.

In Vietnam, the report of commercial banks in the first 6 months of 2021 has reflected a significant increase in bad debt ratio; a number of large banks sharply increased debt in group 5 (debts that could lose capital) such as VietinBank, Vietcombank, MB bank. In addition, SHB, HDBank, ABBank, PGBank have a smaller bad debt ratio [13]. These figures are the clearest sign for the ineffectiveness of early warning of credit risks. One of the core reasons is that early warning models are incomplete with insufficient information and inaccuracy, leading to many limitations in practical application in Vietnam. Early warning of credit risks is an effective method to help commercial banks identify and minimize the percentage of customers with bad debts and overdue debts by regularly evaluating investment portfolios, helping to maintain credit quality. In 2016, VietinBank successfully researched and built a credit risk early warning system based on stored information about customers in the bank [12]. Up to now, this system has improved

and expanded manually collected data sources from other relevant departments in the bank [10]. However, VietTinBank is still among the top of banks with high bad debt; the system lacks information about customers, so warning is still difficult. In 2021, author Dao Thanh Binh and his colleagues have proposed three models of early warning of credit risk, focusing on proposing policies in credit lending [11]. Author Nguyen Van Huan et al., also gave research results on this topic with the use of the Z-SCORE model. The Z-score model only shows its superiority when the database in the financial statements of Vietnamese enterprises must be clear and transparent [14]. It can be seen that the above research results have helped organizations receive warnings if the organization has explicit data and has a time period, the accuracy depends on the amount of explicit data collected from customers. In reality, collecting accurate and explicit client information is difficult and time-consuming, resulting in delayed and inefficient decision-making. In 2019, the author Selvachandran and colleagues researched and published: M-CFIS is more effective when solving problems involving time series data. CFIS applies complex numbers throughout the calculation, giving much more flexibility in performing unexpected, non-linear oscillations, which is new to CFIS [3].

At the same time, there was also a publication on the research results of M-CFIS for multi-attribute decision problems. There are a lot of researches on Mamdani CFIS for decision support, and they all show that it works. However, the research topic on Mamdani CFIS in early warning of credit risks in banks has not been interested by the authors and the research results have been published. It is clear that researches on early warning of credit risks have attracted the interest of many researchers, and many research findings have been published, all of which have contributed to banks' risk ratio is limited. But the needs and goals in banks are increasing day by day, and the competition is getting fiercer, so it is also necessary to have new methods to support the implementation of the organization's goals in a more accurate and timely manner. That is why the research team uses complex fuzzy theory, specifically the Mamdani Deductive Model, to analyze and empirically determine the level of financial risk faced by customers in order to provide early warning of a customer's bad debt condition. In this article, the research team presents the Mamdani application method (M-FIS) on the bank's customer data set to identify the customer's credit level and give suggestions warnings to credit institutions based on customer information.

## 2 Background

### 2.1 Fuzzy Set

L.A. Zadeh has represented fuzzy, uncertain information in nature such as: young, fast, tall-low, beautiful [9],..., through a mathematical concept, called fuzzy set. The fuzzy set is defined as follows:

Given the universe set  $U$ . The set  $P$  is determined by the formula:  $P = \{\mu_p(x) | x : x \in U, \mu_p(x) \in [0, 1]\}$  is called the fuzzy set on  $U$ .

The variable  $x$  that takes a value in  $U$  is called the base variable and therefore the set  $U$  is also called the reference set or the base domain. Function  $\mu_p : U \rightarrow [0, 1]$  is called membership function and value  $\mu_p(x)$  at  $x$  is called the member value of element  $x$  in fuzzy set  $P$ .

## 2.2 Complex Fuzzy Set [3]

Then, complex fuzzy logic (CFL) [1, 2] and complex fuzzy set (CFS) developed by Ramot and et al in 2002, with the addition of phase factor to determine the degree of dependence of element  $x$  on fuzzy set  $A$  for time and period data.

*Definition 1:* A fuzzy set  $S$  over the universe  $U$  is characterized by  $\mu_S(x)$ , attach any  $x \in U$ , with complex value membership degree in  $S$  lies in the unit circle on the complex plane:

$$\mu_S(x) = r_S(x)e^{j\omega_S(x)} \quad (1)$$

where  $j = \sqrt{-1}$ ,  $r_S(x)$  and  $\omega_S(x)$  are real values with  $r_S(x) \in [0, 1]$ ,  $\omega_S(x) \in [0, 2\pi]$ .  $r_S(x)$  is the amplitude and  $\omega_S(x)$  is phase. The amplitude is defined by the phase factor adds some extended information related to the space-time period in the fuzzy set.

*Definition 2:* Let  $S$  and  $T$  are two complex fuzzy sets

i. The complement set  $C$  of  $S$  is:

$$C = \{(x, \mu_c(x))|x \in U\} = \left\{ \left( x, r_c(x)e^{j\omega_c(x)} \right) |x \in U \right\} \quad (2)$$

where  $r_c(x) = 1 - r_S(x)$  và  $\omega_c(x) = 2\pi - \omega_S(x)$

(ii) The union of two sets  $S$  and  $T$  ( symbol  $S \cup T$ )

$$S \cup T = \{(x, \mu_{S \cup T}(x))|x \in U\} = \left\{ \left( x, r_{S \cup T}(x)e^{j\omega_{S \cup T}(x)} \right) |x \in U \right\} \quad (3)$$

The intersection of two sets  $S$  and  $T$  ( symbol  $S \cap T$ )

$$S \cap T = \{(x, \mu_{S \cap T}(x))|x \in U\} = \left\{ \left( x, r_{S \cap T}(x)e^{j\omega_{S \cap T}(x)} \right) |x \in U \right\} \quad (4)$$

*Definition 3.* The implication rule for complex fuzzy set:

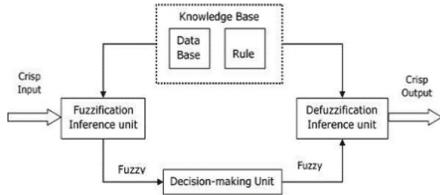
$$\mu_{S \rightarrow T}(x, y) = r_{S \rightarrow T}(x, y)e^{j\omega_{S \rightarrow T}(x, y)} \quad (5)$$

where  $r_{S \rightarrow T}(x, y)$ ,  $\omega_{S \rightarrow T}(x, y)$  are the amplitude and phase components of complex fuzzy implication, respectively.

## 2.3 Mamdani Complex Fuzzy Inference System (M-CFIS)

Fuzzy inference system (FIS) is a system that takes clear input data, clear output have real values, the processing inside the inference system is performed based on fuzzy rules and fuzzy numbers.

In the FIS, the Knowledge Base (Rule and Database) is an important component of the inference system (Fig. 1).

**Fig. 1.** Fuzzy inference system [15]

- Rule- Contains fuzzy IF-THEN rules. Rule bases can be built from experts or exploit rules derived from other systems.
- Database - Define membership functions of fuzzy sets.

M-CFIS proposed by Selvachandran et al. [3] is based on the Mamdani fuzzy inference system and Ramot's complex fuzzy theory. It towards the problem of classification and predictive output prediction. The steps in the Mamdani Complex Fuzzy Inference System

Let  $x_1, x_2, \dots, x_m \in \mathbb{C}$  be the inputs of this model

Stage 1: Establish a set of complex fuzzy rule

Stage 2: Fuzzification of the inputs

Using Gaussian fuzzy function

$$\begin{aligned} Re(cGaussian(x, m, \sigma)) &= \exp\left[-0.5\left(\frac{x-m}{\sigma}\right)^2\right] \\ Im(cGaussian(x, m, \sigma)) &= -\exp\left[-0.5\left(\frac{x-m}{\sigma}\right)^2\right] \times \left(\frac{x-m}{\sigma^2}\right) \end{aligned} \quad (7)$$

Stage 3: Establish the firing strength of a rule

Stage 4: Calculate the consequence of the complex fuzzy rules

In Mamdani CFIS, the value of the consequence of the complex fuzzy rules is obtained by using the Mamdani implication rule.

$$\mu_{A \rightarrow B}(x, y) = (r_A(x).r_B(y)).e^{j2\pi\left(\frac{\omega_A(x)}{2\pi} \cdot \frac{\omega_B(y)}{2\pi}\right)} \quad (8)$$

Stage 5: Aggregation

In this stage, the output distribution is calculated as follows:

$$D(y) = \Gamma_1(y) + \Gamma_2(y) + \dots + \Gamma_k(y)$$

Stage 6: Defuzzification

Choose a function  $\Phi : \mathcal{F}(\mathbb{C}, \mathbb{C}) \rightarrow \mathbb{C}$ . Determine the value of the output  $y_{op} = \Phi(D)$ . For example, we can choose the trapezoidal approximation such as

$$\Phi(D) = \frac{\int_{-\infty}^{\infty} y|D(y)| dy}{\int_{-\infty}^{\infty} |D(y)| dy}$$

### 3 Using M-Cfis to Determine Client's Financial Risk

#### 3.1 Problem

As previously stated, credit risk in banks refers to customers' failure to pay their loans to the bank when they are due. So to warn of financial risks early, banks need to identify financial risks for each customer when a customer makes a loan. Thanks to the results from M-Cfis, credit officers also have more information to advise customers on a loan that is suitable for their existing financial situation. The bank also reduces risk as a result of this. Because the imaginary part is observed, the approach of using Complex Fuzzy Inference System produces better results than standard methods.

With that goal, the research team conducted an experiment on a credit risk data set on Kaggle consisting of 4378 records and 14 attributes. The dataset is divided into two sets:

- + Training set: 3502 records, accounting for 80%
- + Test set: 876 records, accounting for 20%

**Table 1.** Attributes and description 's credit risk data set

	Attributes	Description		Attributes	Description
1	Seniority	0–48	8	Expenses	35–173
2	Home	1–6	9	Income	0–959
3	Time	6–72	10	Assets	0–250000
4	Age	18–68	11	Debt	0–30000
5	Marital	0–5	12	Price	105–11140
6	Records	1–2	13	Status	0- Not risk; 1-Risk
7	Job	0–4			

#### 3.2 Implementation Model

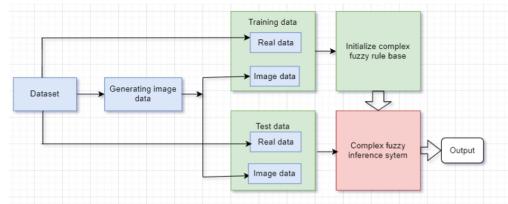
The complex fuzzy inference system we make follows the M-CFIS model

Where:

- The process of generating phase data is done by the method of collecting expert opinions.
- The FCM (Fuzzy C-means) method is used to generate complex fuzzy rules from the training data set (Fig. 2)

With problem "**Determining financial risks of customers**" the mathematical model of the problem is:  $y_n = f(x_1, x_2, \dots, x_m)$ , where:

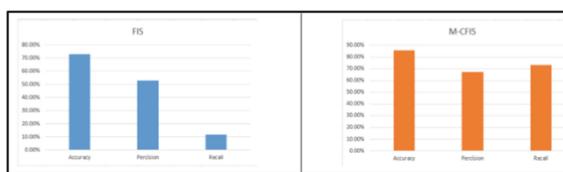
- $x_1, x_2, \dots, x_m$ : là các thuộc tính của đối tượng cần xếp loại
- $y_i$ : the client's financial risk



**Fig. 2.** M-Cfis empirical model for customer risk determination problem

### 3.3 Experimental

The experimental process on the data set obtained the following results (Fig. 3).



**Fig. 3.** Experimental results obtained with FIS and M-CFIS

Model accuracy increased significantly compared to using FIS and especially Precision and Recall values also increased. With this result, we find that using more phase factor in fuzzy model, specifically Mamdani complex fuzzy inference system, produces better inference results for data with time and period factors. or data with an attribute binding element.

## 4 Conclusion and Future Research Scope

The application of fuzzy logic in fields such as remote control, engineering, and environmental science has been a huge success in general and complex fuzzy logic in particular. According to the published scientific literature, fuzzy logic has been particularly effective when dealing with problems involving uncertainty and ambiguity, which are two of the most common characteristics associated with analysis. Business analysis in general and the problem of identifying target customers in particular or determining customer financial risks. experimental results of early warning of credit risk for banks, applying Mandani complex fuzzy system on the bank's customer data set, the results have an accuracy of up to 89,5621% in 666 In this paper, in addition to publishing the experimental results of early warning of credit risk for banks, applying Mandani complex fuzzy system on the bank's customer data set, the results have an accuracy of up to 89,5621%. There are 666 cases that have been tested. If the case is tested, the research team also creates a complex fuzzy theory statement that can be applied to a group of problems in the field of business analysis, finance and banking, and so on, using a series of requirements generated based on data mining, the data is divided into real and imaginary parts to improve the accuracy of the inferential model.

## References

1. Ramot, D., Milo, R., Friedman, M., Kandel, A.: Complex fuzzy sets. *IEEE Trans. Fuzzy Syst.* **17**, 171–186 (2002). [https://doi.org/10.1007/978-3-476-05004-5\\_1](https://doi.org/10.1007/978-3-476-05004-5_1)
2. Ramot, D., Friedman, M., Langholz, G., Kandel, A.: Complex fuzzy logic. *IEEE Trans. Fuzzy Syst.* **11**, 450–461 (2003). [https://doi.org/10.1007/978-1-349-16956-6\\_1](https://doi.org/10.1007/978-1-349-16956-6_1)
3. Selvachandran, G., et al.: A New design of mamdani complex fuzzy inference system for multi - attribute decision making problems. *IEEE Trans. Fuzzy Syst.* **29**(4), 716–730 (2019)
4. Tuan, T.M., et al.: M-CFIS-R: Mamdani complex fuzzy inference system with rule reduction using complex fuzzy measures in granular computing. *Mathematics* **8**(5), 707 (2020)
5. Hien, P., Phung, T.K.: Fuzzy logic and applied problems in finance. *Manag. Inform. J.* **27**–44 (2014)
6. Duan, J.C., Shrestha, K.: Statistical credit rating methods. *Glob. Credit Rev.* **1**, 43–64 (2011)
7. Alizadeh Zoeram, A., Karimi Mazidi, A.R.: A new approach for customer clustering by integrating the LRFM model and fuzzy inference system. *Iranian J. Manag. Stud.* **11**(2), 351–378 (2018)
8. Tram, N.T.M.: Application of fuzzy logic techniques in corporate credit rating; Financial Magazine (2020 )
9. Zadeh, L.A.: Fuzzy sets. *Inf. Control* **8**, 338–353 (1965)
10. Do Thi, T.H.: Early warning of risks to the granting of credit to customers and related persons of Vietnamese commercial banks. *J. Bank. Sci. Train.* **208** (2019)
11. Dao, T.B.: Credit risk early warning model for commercial banks. *J. Finance* (2021)
12. Cao, T.T.: 2016. Loss reduction thanks to early warning of credit risks (2021). <https://vietnamnet.vn/vn/Kinh-doanh/tai-chinh/giam-ton-that-nho-can-bao-som-rui-ro-tin-dung-310376.html>
13. Thuy, L.: Handling bank bad debts: It's time to need a higher legal corridor (2021). <https://www.vietnamplus.vn/xu-ly-no-xau-ngan-hang-da-den-luc-can-hanh-lang-phap-ly-cao-hon/733436.vnp>
14. Nguyen, V.H., Nguyen, Q.T.: Studying Z-Score model in early warnings of credit risk at vietnam commercial banks. *J. Trade Sci. Code 154.1FiBa.11* (2021)
15. [https://www.tutorialspoint.com/fuzzy\\_logic/fuzzy\\_logic\\_inference\\_system.htm](https://www.tutorialspoint.com/fuzzy_logic/fuzzy_logic_inference_system.htm)



# A Proposed Architecture of Intelligent Operations Center for Smart Cities Development

Minh Tan Ngo<sup>1</sup>, Ngoc Hai Nguyen<sup>2</sup>, Cong Duan Truong<sup>3</sup>, Duy Hung Tran<sup>4</sup>,  
and Cong- Doan Truong<sup>5(✉)</sup>

<sup>1</sup> Swinburne Vietnam, FPT University, Hanoi, Vietnam

tanmseh191048@fpt.edu.vn

<sup>2</sup> FPT University, Hanoi, Vietnam

hai19mse13029@fsb.edu.vn

<sup>3</sup> Academic Department, Swinburne Vietnam, FPT University, Hanoi, Vietnam  
duantc@fe.edu.vn

<sup>4</sup> Hanoi Open University, Hanoi, Vietnam

hungtd@hou.edu.vn

<sup>5</sup> International School, Vietnam National University, Hanoi, Vietnam  
doantc@isvnu.vn

**Abstract.** Smart cities with the construction of Intelligent Operations Center (IOC) have been the key development of countries around the world, including Vietnam. Since cities are facing several problems, such as aging infrastructures and disparate data storage systems, the city authorities are unable to efficiently offer services sustainably, protect citizens or promote future economic growth without an integrated platform. In this paper, we introduce an Intelligent Operations Center architecture which is developed on smart application systems with the focus on critical sectors such as city infrastructure, urban transit, the environment, economics and public opinion to address these problems in Vietnam. The IOC integrates monitoring, regulating, and other systems into a unified platform, allowing more efficient information delivery, better operational management, communication, cooperation, and decision-making. Being considered as the brain and central system of a smart city, we develop the IOC architecture with five main layers: (1) data integration layer, (2) storage layer, (3) data analytics layer, (4) data visualization layer and (5) application layer. Having been implemented in more than 30 cities in Vietnam, the results highlight the efficiency and effectiveness of the IOC system not only in the decision-making process and city development planning but also in the economic development and increasing living standards of citizens. Further research into the IOC system should focus on optimizing the system process and integrating Artificial Intelligence models in the application layer.

**Keywords:** Smart city · Intelligent Operations Center · Intelligent systems · Big Data · Command centers · Decision-making process

## 1 Introduction

A smart city incorporates information and communication technologies into the planning, design, construction and operation of a city's infrastructure to address urban issues

and builds a connected, technologically integrated and long-lasting infrastructure [1]. There are six important components of a smart city [2]: Smart Government includes solutions for efficiency and optimization of administrative units; Smart Economy emphasizes investment solutions—efficient production and flexible labor market; Smart Environment describes solutions for clean energy, low-energy consumption and smart building; Smart Living focuses on improving the living standard for residents in terms of health-care, consumption, housing conditions and safety; Smart Mobility aims to improve the efficiency and quality of transportation, including cost-saving mobility solutions and emission reduction; Smart People sets an open and inclusive society to increase a city's or community's economy and innovation. Information regarding buildings, residents, gadgets and assets is analyzed for this purpose to effectively regulate urban flows via real-time reactions.

To enhance city operations, smart city development needs a powerful “brain” to simplify data silos, execute information exchange, convergence and integrate the physical and digital worlds. The development of an Intelligent Operations Center (IOC), to solve the problem of urbanization management is one of the most optimal solutions. IOC, which is constructed on smart application systems, focuses on critical sectors such as city infrastructure, urban transit, the environment, public security, economics, and public opinion. As the smart brain and central nervous system of a city, IOC provides an executive board to support municipal officials in gaining insight into all parts of the city [3]. IOC helps city departments exploit information and data systems to change the ways of operating, managing and promoting creativity and instrumentation to meet the needs of city residents, improve the efficiency of coordinating cross-agency resources and mitigate the effect of service and operations disruptions.

IOCs have been implemented in different regions and countries around the world; however, they are developed to solve city problems themselves. Rio de Janeiro, for example, had developed the Rio Operations Center to enable geologists, field operations and security to collaborate to improve emergency responsiveness, risk prevention and management [4]. Copenhagen in Denmark focuses on developing a laboratory for applications in the city with the collaboration of the residents, decision-makers and businesses in Copenhagen Solutions Lab [5]. Seoul Transport Operation & Information Service (TOPIS) operates and regulates the city's overall traffic with the purpose of minimizing severe traffic problems [6]. Such approaches will make these IOCs support effectively in a certain aspect, yet show some limitations, such as shared data has not been considered thoroughly, deployment components are stand-alone and single solution-oriented, with no emphasis on building the platform. Consequently, these applications are difficult for mutual exchange and cannot inherit existing systems.

In recent years, IOC has been deployed in provinces and cities in Vietnam. Depending on each city's characteristics, these IOCs have different functions and tasks, although they followed the same architectural foundation. The Intelligent Environment Operations Center in Hue, for example, was constructed to solve environmental and infrastructure problems [7]. Da Lat was one of the first cities to develop IOC in Vietnam that helped the city's authorities better administer and regulate social activities [8]. Each government agency focuses on their specialized data, such as the Ministry of Public Security developed an IOC for Residential Database, the Ministry of Finance built an IOC for the

Insurance sector. The government also developed an IOC with the goal of connecting all these above systems in a single platform.

Cities in Vietnam are confronted with aging infrastructures, limited resources, shifting populations and growing dangers. Unshared data information limits the ability to coordinate and cooperate between sectors, as well as between the government and society, creating a cumbersome and slow system that is difficult to expand. In most cities in Vietnam, critical information is frequently stored in separate systems across multiple, separated divisions, making situation awareness difficult and coordinating agency activities challenging. A city may not offer services sustainably, protect residents or drive future economic growth without an integrated picture of events, incidents, or potential crises. Addressing these problems with purely traditional approaches is no longer effective.

In this paper, we propose an IOC architecture that solves the mentioned limitations with six main principles: Putting citizens and businesses as a core; increasing the data effectiveness; prioritizing deployment of platform solutions; ensuring flexible scalability; providing high integration ability; assuring safety and security. This architecture has been deployed to the government and 35 provinces and cities in Vietnam, which initially brought positive results such as promoting the growth of the ICT Index, PAPI, more information is shared to citizens, data is more connected between departments. The IOC helps city authorities better monitor and manage city services by providing comprehensive information about city daily operations through centralized data management, especially during the COVID-19 period.

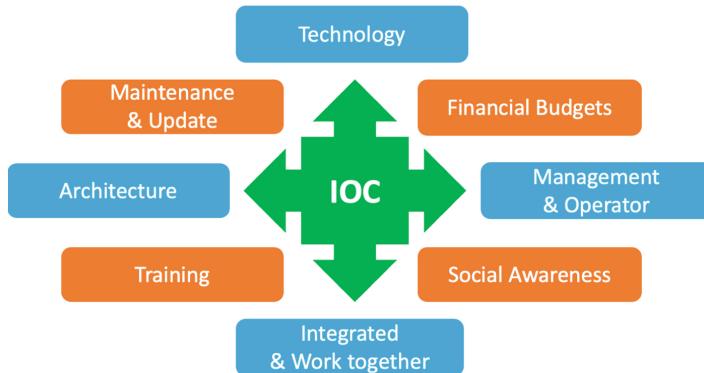
## 2 Methodology

### 2.1 Background

The Intelligent Operations Center incorporates measuring, controlling and other systems into a single platform and more efficiently provides information, improving operational management and contributing to collaboration, cooperation and decision-making [9]. The IOC helps to share information across departments and agencies, allowing them to cooperate, predict and respond to problems such as weather, natural disasters and crises.

There are some key components in the construction of an IOC, which interplay and are considered as standard in implementing technologies to develop the IOC as shown in Fig. 1:

- Technology: The implementation depends on the city's existing architecture, accompanied by data traffic, servers, IDCs, operating procedures and human resources, which directly affect the reliability of the system.
- Financial budgets: They are one of the most concerned categories. The system will bring advantages to the city yet be considered from a cost perspective.
- Management & Operator: Close coordination among people, processes, policies and technology should be considered. Transforming gradually from a vertical (silo-type) model to a centralized one.
- Social Awareness: Training needs to be conducted for those involved in the system including authorities, business and citizens.



**Fig. 1.** Key components in developing an IOC system

- Integrated & Work together: Integrating different applications and technologies requires complex systems and highly qualified human resources. The state agencies, who are mainly responsible for management tasks, are not compatible with the capabilities required for human resources. Hence, these systems will be technically supported by providers.
- Training: The training process is carried out in transfer method (on the job training), so that city agencies can operate the system independently, then proceed to be in charge and connect future projects themselves.
- Maintenance & Update: The process of maintaining and upgrading the system should be invested in, to ensure the system works efficiently and keeps up to date with the latest technology.
- Architecture: The system is built based on the existing architecture and is suitable for oriented development. A completely new system can affect the integration and transformation of an ongoing one. As a result, it should be inherited from an existing architecture that easily develops and scales in the future.

## 2.2 IOC Architecture

We propose an IOC architecture solution with the components and functions described in Fig. 2.

The IOC architecture includes five layers:

- Data Integration layer: This layer gets the input data to the system and can forward notifications, alerts, KPI measurements to other systems. An Enterprise Service Bus (ESB) receives data from sources by different methods, it is also an infrastructure for consolidating services in the application environment [10]. Data integration lies in shared multi-dimensional integration through ESB where data are linked via protocol conversion.

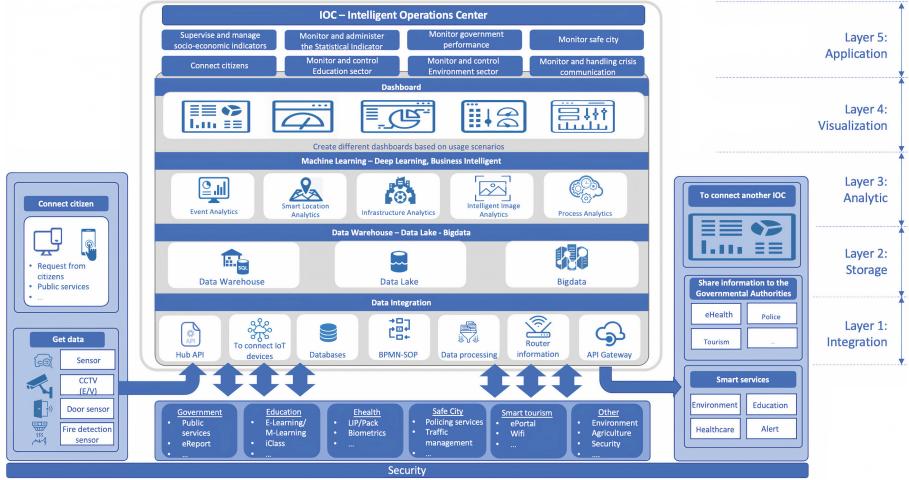


Fig. 2. Intelligent Operations Center Architecture

- **Storage layer:** The data system is stored in this layer, including Data Warehouse—Data Lake—Big Data. These storages will be deployed based on stage and customer demands. The storage layer contains three main components: Specialized Database, Open Data and Data Warehouse.
- **Specialized Database:** The database is built by units within a city themselves and can be transformed into a shared database and open database
- **Open Data:** The shared database, such as Geographic Information System (GIS) database [11], is used for search and display information purposes. City citizens can freely get access to and share the data but must ensure the origin of such data is preserved. Open Data is particularly important for smart cities as citizens can view the operational status and access information easily.
- **Data Warehouse:** Different types of data are gathered and collected from city departments via data replication and synchronization to serve the needs of data sharing between units. Such component helps to connect and to build integrated services for citizens and businesses but requires strong authentication processes as sensitive data can be stored in the shared database.
- **Data Analytics layer:** This layer connects to the storage layer, which is used to analyze the centralized data in the system, creating new values for the end-users.
- **Data Visualization layer:** Analyzed data will be presented and visualized by dashboards and charts which show the overall status of a city. The dashboard helps city executives to obtain crucial insight into current challenges as it captures data from different sources, such as traffic reports, environment conditions and citizen satisfaction indices, and turns it into usable data. The visualization can be concretized by mobile applications, websites or physical display systems at operation centers. It also includes Standard Operating Procedures (SOPs) [12], situation reports and activity data that city authorities can evaluate properly and assess the situation for further improvements.

- Application layer: The arrogated, analyzed data from underneath layers are monitored and displayed in this layer. The city authorities need to prepare for both foreseen and unexpected situations, such as public health problems and traffic challenges. This will help city authorities in the decision-making process by providing ad-hoc reporting patterns and insights.
- Security layer: The security layer runs throughout the system. It covers other components in the system to ensure data security and information safety from the physical layers to the application layers, standards, policies and procedures. Information security systems operate on strategic policies, frameworks and implementation principles that are compliant with ISO 27001 [13].

## 2.3 IOC Architecture Characteristics

Three main characteristics of this IOC architecture are connection, sharing and data analysis that are expressed in the following advantages:

- The infrastructure platform utilizes Cloud Computing technology with flexible scalability that can be shared across departments to optimize investment and operating costs.
- Data is centralized from the IoT database systems to form an open Data Warehouse. A centralized information-sharing platform optimizes costs for storage systems, serves the needs of sharing and analyzes data which facilitate the development of integrated and intelligent services.
- This architecture can help city agencies and departments integrate information from different systems to develop an interactive, connected platform that encourages collaboration, improves efficiency and promotes better decision-making. The city authorities can be kept informed, address public concerns, plan for events and respond to unanticipated emergencies.

## 2.4 Technology Used

The IOC system is a combination of diverse and complex technologies, provided by technology corporations after analysis, comparison and testing. The IOC architecture is developed on shared computing infrastructure; a multi-cloud model is used accordingly to maximize the system security. The architecture is developed following Software-as-a-Service (SaaS) model on the Cloud Foundry platform [14] combined with the Zenko open-source code [15] (Table 1).

**Table 1.** Major components in IOC architecture

No	Layer	Technology
1	Data integration	Based on the open-source Apache Nifi and Apache Sqoop [16] with Oracle Data integration, we build the data integration at different levels for types of users
2	Data Warehouse, Data Lake, Big Data	Customized system to meet the requirements of the city by collaborating with Cloudera [17]
3	Data analytics	A set of tools were created to analyze data by combining the systems on the Cloudera and Apache Lucene platform
4	Data visualization	Power BI was used [18] for data visualization

### 3 Results

With the integration and visual presentation of information and data from different fields, IOC has provided the city authorities and the government with a comprehensive overview of all socio-economic activities. From that, they can make timely and accurate decisions, saving a lot of time and effort in the process of monitoring and operating the work. Through the IOC system, the interaction between citizens and the government is also improved, contributing to improving the level of satisfaction of citizens and businesses with the operation of the government.

The IOC enhances the citizens' experience by making infrastructure navigation simpler. In the cities and provinces that have deployed the IOC, the quality-of-life indicators have significantly improved. As shown in Fig. 3, Binh Phuoc operated a comprehensive IOC system with 10 services in 2020. After more than 2 months of trial operation, the authorities received over 150 complaints, petitions, calls and messages daily from citizens in the security, fire prevention, traffic accidents, public administration sectors. The rate of receiving and handling level 3 and level 4 online public services has been increased from 9% at the beginning of the year to over 90% in September.

Officially operated in December 2019, Da Lat IOC showed significant results as shown in Fig. 4. The citizens' feedback to the city agencies were processed slowly from a few days to a few months; it takes about 2 to 24 h from receiving complaints to complete processing in some cases. After operating for 6 months, the Da Lat authorities received an average of 30 to 40 complaints every day in different sectors [19]. In the tourism field, the Da Lat IOC has continuously updated data from 1,176 hotels and accommodation places, 772 restaurants, 103 tourist attractions, 75 shopping stores and 489 entertainment spots. The Dalat.vn tourism portal and smart travel application have connected with the IOC to update and track data from the system. Visitors can search and update information about the city, book tickets, hotel rooms online or look up cultural events, food, traffic and weather easily.



**Fig. 3.** VNPT's technical team support the operation of Binh Phuoc IOC



**Fig. 4.** Da Lat Intelligent Operations Center

The IOC also provides the city agencies and the government with a comprehensive status of the city activities, which increases the effectiveness of public services and city safety. A positive result of the Binh Phuoc IOC is through the traffic safety monitoring and operating system integrated with traffic cameras: about 15,000 traffic violations have been recorded with 13 billion VND fine. The camera system has helped detect many cases of security and traffic violations, importantly contributing to ensure public security in the province. The government and city agencies can update the development in the socio-economic situation and other problems in various fields, which helps them to forecast, assess the situation then accurately handle the problems.

Economic developments have been demonstrated in the implementation of the IOC system. The National Document Communication Axis has connected with 100% of ministries, city agencies and organizations with over 2.2 million electronic documents sent and received that saves the national budget more than 1,200 billion VND per year [20]. The National Public Service Portal opened in December 2019; after 8 months of operation, it has connected with 18 ministries, agencies, 63/63 cities and provinces, with nearly 56.4 million accesses, more than 220,000 registered accounts, over 14 million synchronized records. The total costs saved when operating public services are estimated at more than 13,000 billion VND per year, of which, the National Public Service Portal contributes over 6,700 billion VND per year [20].

## 4 Conclusion

Smart city is the development objective of many countries in the world including Vietnam, with the construction of IOC as a smart brain that collects and operates the cities' activities by collecting and standardizing data. This paper proposes an IOC architecture for smart cities in Vietnam with five main layers which includes data integration, storage, data analytics, data visualization and application layer. Acting as the smart brain of a smart city, the IOC helps local organizations and authorities in integrating data from many systems to create an interactive, connected platform that provides a comprehensive overview of a city' activities. The results show that this architecture increases the efficiency of the decision-making process by reducing the time and enhances the citizens' experience. Future work is planned to optimize the system process and implement Artificial Intelligence models in the application layer.

**Acknowledgement.** This research is funded by International School, Vietnam National University, Hanoi (VNUIS) under project number CS.2021-02.

## References

1. TWI Global: What is a smart city? - Definition and examples. TWI Ltd, Granta Park, Great Abington, Cambridge, UK. <https://www.twi-global.com/technical-knowledge/faqs/what-is-a-smart-city>
2. Giffinger, R., Haindlmaier, G., Kramar, H.: The role of rankings in growing city competition. *Urban Res. Pract.* **3**(3), 299–312 (2010). <https://doi.org/10.1080/17535069.2010.524420>
3. Gontar, B., Gontar, Z., Pamula, A.: Deployment of smart city concept in Poland : selected aspects. *Manag. Organ. System. Res.* **67**(67), 39–51 (2013). <https://doi.org/10.7220/MOSR.1392.1142.2013.67.3>
4. Rio Operations Center: integrating data and monitoring utilities in a truly intelligent system. Guangzhou International Award for Urban Innovation, 31 Dec 2014
5. Copenhagen Lab: Street lab | Copenhagen solutions lab, 20 Dec 2017. <https://cphsolutionslab.dk/projekter/labs/street-lab>
6. Seoul transport operation & information service. [https://topis.seoul.go.kr/eng/page/about\\_1.jsp](https://topis.seoul.go.kr/eng/page/about_1.jsp)
7. Intelligent environment operations center opened. <https://thuathienhue.gov.vn/en-us> (2019). Accessed 19 Aug 2019
8. Smart city monitoring and operating center in Da Lat. <https://en.qdnd.vn>
9. Zheng, W., chen, H., Li, C., Yu, N., Chen, J.: Intelligent operation center for Hengqin new area smart grid. *J. Int. Counc. Electr. Eng.* **4**(3), 216–219 (2014). <https://doi.org/10.5370/JICEE.2014.4.3.216>
10. Enterprise service bus. <https://www.oracle.com>
11. Ali, E.: Geographic information system (GIS): definition, development, applications & components, pp. 1–12. Department of Geography, Ananda Chandra College, Jalpaiguri (2020)
12. Akyar, I.: Standard operating procedures (what are they good for?). Latest research into quality control, IntechOpen (2012) <https://www.intechopen.com/chapters/37593>
13. Beckers, K., Faßbender, S., Heisel, M., Schmidt, H.: Using security requirements engineering approaches to support ISO 27001 information security management systems development and documentation. In: 2012 Seventh International Conference on Availability, Reliability and Security, pp. 242–248 (2012). <https://doi.org/10.1109/ARES.2012.35>
14. Cloud foundry—Open source cloud native application delivery. <https://www.cloudfoundry.org/>
15. Zenko - Multi-cloud data controller - manage your data without cloud lock-in. <https://www.zenko.io/>
16. Welcome to the apache software foundation! <https://www.apache.org/>
17. Cloudera | The hybrid data cloud company. <https://www.cloudera.com/>
18. Data visualization | Microsoft power BI. <https://powerbi.microsoft.com/en-us/>
19. Xã Hội Thông Tin, XHTT. <http://xahoithongtin.vnmedia.vn/tin-noi-bat/202007/trung-tam-dieu-hanh-thong-minh-da-lat-hieu-qua-sau-06-thang-hoat-dong-67e03bd/>
20. <http://vpcp.chinhphu.vn/Home/He-thong-Thong-tin-bao-cao-quoc-gia-Diem-nhan-trong-phat-trien-Chinh-phu-dien-tu/20208/28421.vgp>. Accessed 6 Sept. 2021



# Greedy Detection for ReIQ System

Trung Tan Nguyen<sup>(✉)</sup>

Le Quy Don Technical University, Ha Noi, Vietnam

trungtan68@gmail.com

**Abstract.** This study introduces an enhanced index modulation scheme called repeated inphase (I) and quadrature (Q) index modulation-OFDM (ReIQ). This system uses activated subcarriers in each I and Q component to transmits the same data symbol. By doing so, ReIQ can attain improvement in both spectral efficiency and transmission reliability. The advanced detector, called as greedy detection (GD) for ReIQ is proposed to provide optimal reliability of the maximum likelihood (ML) detector and considerably decrease the computational complexity. Simulation results are introduced to demonstrate the superiority of ReIQ as well as low-complexity detector compared to benchmark systems even when the estimated channel errors appear at receivers.

**Keywords:** Index Modulation (IM) · Transmit diversity · OFDM-IMIQ · ML detection · Greedy detection

## 1 Introduction

In recent years, Index Modulation (IM) is a novel solution that attracted much attention of researchers. IM with OFDM is known as the effective IM model in the frequency domain. The IMO-FDM system utilizes only a number of subcarriers to transmit data via complex symbols and the active subcarrier index [1].

To improve the reliability of IMO-FDM, many studies have been introduced. By introducing the joint I/Q IM [2], each OFDM frame can convey more bits to obtain improved spectral efficiency. Recently, the dual-mode OFDM was introduced where inactive subcarriers are utilized as a supplement to convey more data bits. Besides, the multi-mode OFDM was introduced in [3].

Considering the diversity issue, I and Q parts of complex data symbols [4] were transferred over repeated distinguished sub-carriers to attain the diversity gain. IM-OFDM with simultaneous diversity reception and transmit to improve diversity gain and decrease the complexity can be found in [5]. Besides, the IMO-FDM principle was put into practice Multi-Input Multi-Output (MIMO) system [6]. Moreover, low complexity detectors for IMO-FDM to reduce the complexity and still obtaining the optimum reliability of the ML detection was also introduced. Motivating the idea of IMO-FDM [2], the repeated I/Q-IMO-FDM scheme referred to as ReIQ is proposed, where the repeated transmission was put into all active subcarriers in I/Q parts to achieve improved reliability and data rate over the conventional IMO-FDM. Further- more, greedy detection with

small complexity is introduced in this work to allow the ReIQ system to be much more simply. The influence of imperfect channel estimation process on the error performance of system is also investigated.

The paper is divided into five sections. Section 2 is the proposed scheme. Section 3 proposes the GD detector. The performances of ReIQ system with proposed detector and imperfect CSI condition is evaluated in Sect. 4. Conclusions are presented in Sect. 5.

## 2 Proposed Scheme

Assuming that the OFDM system has  $N_c$  subcarriers to transmit  $m_t$  bits, which are subdivided into  $B$  subblocks with  $m_b$  bits. Every subblock is split into two groups of  $N_s$  subcarriers. Every group with  $p_g$  bits is applied to create an I/Q element of a ReIQ subblock, it means that  $N_s = N_c/2B$  and  $p_g = m_t/2B$ . Like the classical IMOOFDM scheme, in every group with  $p_g$  bits of ReIQ system,  $K_s$  out of  $N_s$  subcarriers are used. The remaining inactive subcarriers are 0 padded. Nevertheless, different from the traditional IMOOFDM and IQ-IMOOFDM systems in which  $K_s$  activated subcarriers convey  $K_s$  distinguishable  $M$ -ary symbols. While the ReIQ system,  $K_s$  active subcarriers for each I/Q part deliver the same  $M$ -ary symbol. Since subblocks operate independently, the author addresses only one ReIQ subblock.

In a ReIQ subblock,  $m_b$  bits are subdivided into two parts. With I/Q component per subblock, each  $p_g$ -bits group is then split into two components. The first part of  $p_1 = \lfloor \log_2 C(K_s, N_s) \rfloor$  bits is utilized to activate  $K_s$  subcarriers, where  $C(K_s, N_s)$  presents the convolution  $K_s$  of  $N_s$ ,  $\lfloor \cdot \rfloor$  is the floor function. As a result, an index symbol is generated at the output of index modulator according to active subcarriers. The activated subcarrier index set in I/Q parts is given by  $\theta^I = \{\beta_1^I, \dots, \beta_{K_s}^I\}$ ,  $\theta^Q = \{\beta_1^Q, \dots, \beta_{K_s}^Q\}$ , where  $\beta_k^I, \beta_k^Q \in \{1, 2, \dots, N_s\}$ ,  $k = 1, 2, \dots, K_s$ . Thanks to use of repeated transmission at the transmitter, symbols for indices in I/Q dimensions are respectively represented as  $\lambda^I = [\zeta_1^I, \dots, \zeta_N^I]$ ,  $\lambda^Q = [\zeta_1^Q, \dots, \zeta_N^Q]$ , in which  $\zeta_i^I = 1$  if  $i \in \theta^I$ ,  $\zeta_i^Q = 1$  if  $i \in \theta^Q$  and  $\zeta_i^I, \zeta_i^Q = 0$  if  $i \notin \theta^I, \theta^Q$ . It means that  $\lambda \in \{0, 1\}$ . The index symbol is decided by either the combination or look up table method [1].

The remaining element of data bits in every -bit I/Q group is to establish complex symbols. Whereas activated subcarriers in the I/Q dimension of classical IQ-IMOOFDM schemes transmit different complex symbols:  $s^I = [s_{\beta_1}^I, \dots, s_{\beta_{K_s}}^I]$ ,  $s^Q = [s_{\beta_1}^Q, \dots, s_{\beta_{K_s}}^Q]$ , respectively, with  $s_{\beta_k}^I, s_{\beta_k}^Q \in \mathcal{S}$  denotes  $M$ -PAM signal constellation. In the proposed scheme, the same data symbols  $s^I, s^Q$  are transmitted over all of active subcarriers in every I/Q-branch, respectively. As a result, a transmitted signal in I component is given by  $\mathbf{x}^I = \lambda^I s^I$ . Likely, a transmitted signal in Q component is reprented by  $\mathbf{x}^Q = \lambda^Q s^Q$ . After that the transmitted signal of ReIQ system is created by combining  $\mathbf{x}^I$  and  $\mathbf{x}^Q$  given by

$$\mathbf{x}_t = \mathbf{x}^I + j\mathbf{x}^Q \quad (1)$$

Signal at the ReIQ receiver is expressed b

$$\mathbf{y}_r = \mathbf{H}\mathbf{x}_t + \mathbf{n} \quad (2)$$

where  $\mathbf{H} = \text{diag}(h_1, \dots, h_{N_s})$  is the channel matrix, and  $h_\beta$ , in which  $\beta = 1, \dots, N_s$  is random variables following Gaussian distribution that denotes the fading channel coefficient on each subcarrier  $\beta$ . AWGN noise is represented by vector  $\mathbf{n} = [n_1, \dots, n_{N_s}]^T$ . On every subcarrier,  $h_\beta$  and  $n_\beta$  have distributions  $\mathcal{CN}(0, 1)$  and  $\mathcal{CN}(0, N_0)$  in which  $N_0$  is noise variance. Assume  $\mathbb{E}\{|s^2|\} = \delta E_s$ , the signal to noise ratio (SNR) per subcarrier is determined by  $\psi = \delta E_s / N_0$ , with  $E_s$  is the transmit energy of  $M$ -ary symbol and  $\delta = N_s / K_s$  represents power allocation coefficient.

In practical system, errors can appear in the channel estimation process. The estimated channel satisfies the condition  $\tilde{h}_\beta = h_\beta + e_\beta$ , where  $e_\beta$  is error of estimated channel,  $e_\beta \sim (0, \varepsilon^2)$ ,  $\tilde{h}_\beta \sim \mathcal{CN}(0, 1 + \varepsilon^2)$ , where  $\varepsilon^2$  is the error variance.

The receive side utilizes maximum likelihood (ML) detection to simultaneously detect the activated indices and the accordent data symbols.

$$\hat{\mathbf{x}}_r = \arg \min_{\mathbf{x}_t} \left\| \mathbf{y}_r - \tilde{\mathbf{H}} \mathbf{x}_t \right\|^2. \quad (3)$$

From Eq. (3), maximum likelihood has the complexity which investigated via complex multiplications in every I/Q component  $\mathcal{O}(2^{P_1} M^{K_s})$  that enhances exponentially with modulation order.

### 3 Proposed GD Detector

In this section, the GD detector obtaining the near-optimum reliability of the ML detection is introduced. The GD divided detection process to two steps: the first one is used for detection of the indices of active subcarriers,  $\hat{\theta}^I = \{\hat{\beta}_1^I, \dots, \hat{\beta}_{K_s}^I\}$ ,  $\hat{\theta}^Q = \{\hat{\beta}_1^Q, \dots, \hat{\beta}_{K_s}^Q\}$  based on the  $K_s$  largest received energies  $|y_\beta|^2$ . Then, the second step conducts recovery for symbols corresponding to active subcarriers as follows

$$\begin{aligned} \hat{x}_{\hat{\beta}} &= \arg \min_{x_{\hat{\beta}} \in \mathcal{S}} \left\| y_{\hat{\beta}} - \tilde{h}_{\hat{\beta}} x_{\hat{\beta}} \right\|^2, \\ &= \arg \min_{x_{\hat{\beta}} \in \mathcal{S}} \left\{ \left| \tilde{h}_{\hat{\beta}} \right|^2 \left[ \left| r_{\hat{\beta}}^I - x_{\hat{\beta}}^I \right|^2 + \left| r_{\hat{\beta}}^Q - x_{\hat{\beta}}^Q \right|^2 \right] \right\}. \end{aligned} \quad (4)$$

where  $r_{\hat{\beta}}^I$  and  $r_{\hat{\beta}}^Q$  denote the I/Q parts of  $r_\beta = y_\beta / \tilde{h}_\beta$ , respectively. It can be seen that I/Q components of the ReIQ subblock can be recovered independently. Therefore, the recovery of each I/Q component is conducted as follows

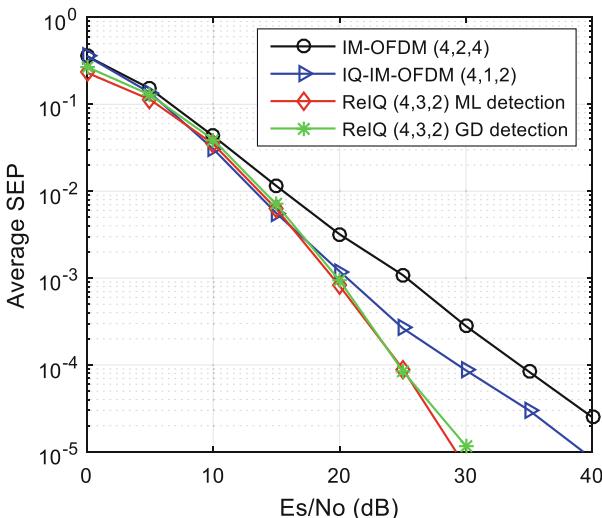
$$\hat{x}_{\hat{\beta}}^I = \arg \min_{x^I} \left| \tilde{h}_{\hat{\beta}} \right|^2 \left| r_{\hat{\beta}}^I - x_{\hat{\beta}}^I \right|^2, \quad (5)$$

$$\hat{x}_{\hat{\beta}}^Q = \arg \min_{x^Q} \left| \tilde{h}_{\hat{\beta}} \right|^2 \left| r_{\hat{\beta}}^Q - x_{\hat{\beta}}^Q \right|^2. \quad (6)$$

GD estimates active subcarrier indices without the exhausted estimation of available index combinations. This allows to decrease complexity of  $M$ -ary symbol detection. The complexity of GD detector is  $\sim \mathcal{O}(N + KM)$  which linearly increases with  $M$ .

## 4 Performance Evaluation

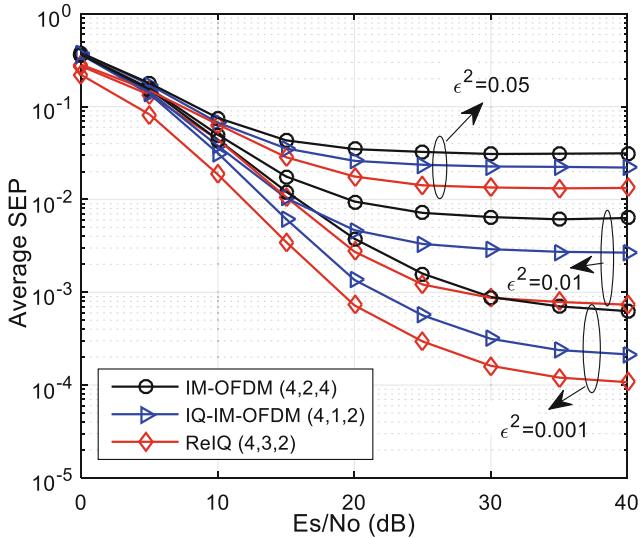
The Monte-Carlo simulation is introduced to compare the symbol error probability (SEP) of ReIQ with IMO-FDM [1] and IQ-IMO-FDM systems [2]. The system parameters are denoted by  $(N_s, K_s, M)$ . SEP of the ReIQ scheme using the proposed GD detector compared to ML and the benchmarks at them same data rate of 1.5 bps/Hz when using 4 sub-carriers,  $K_s = [1, 2, 3]$ ,  $M = [2, 4]$  is depicted in Fig. 1. Assume that CSI is known exactly at the receiver. It is clear that ReIQ achieves better performance than the benchmarks. Particularly, at  $\text{SEP}=10^{-4}$ , ReIQ obtains approximately 9.5 dB and 5 dB gains compared to IMO-FDM and IQ-IMO-FDM, respectively. This improvement in the error symbol performance is achieved because ReIQ uses the repeated transmission at the transmitter that allows the system to attain the diversity gain in frequency domain while the benchmarks without repetition code does not achieve the frequency diversity gain. From the Fig. 1, it can be observed that the GD detector provides the similar SEP to ML while the computational complexity significantly reduces. There are not many differences in SEP of ReIQ when utilizing ML as and GD detectors. GD detector is a near-optimal one, it first estimates the active indices for I and Q components independently, then detects the corresponding data symbols. Thus, GD can reduce the computational complexity especially when the values of  $N_s$  and  $K_s$  are large.



**Fig. 1.** SEP of ReIQ using proposed detector compared to benchmarks at the same data rate and perfect CSI condition.

Figure 2 presents the SEP of the ReIQ system in the comparison with the IMO-FDM [1] as well as IQ-IMO-FDM [2] schemes in the imperfect CSI condition and the same date rate of 1.5 b/s/Hz when  $N_s = 4, K_s = [1, 2, 3], M = [2, 4]$  and variance of estimated channel error  $\varepsilon^2 = [0.001, 0.01, 0.05]$ .

It can be seen that the best performance achieves at error variance  $\varepsilon^2 = 0.001$  and the worst one occurs at  $\varepsilon^2 = 0.05$ . It is clear that the increasing values of error variance



**Fig. 2.** The SEP of ReIQ and benchmarks at the same data rate and imperfect CSI condition.

make the error performance of considered systems significantly reduced. Moreover, error floors occur at the high SNR regions. The increasing values of error variance make the error floor higher.

Besides, even at imperfect CSI condition, the error performance of ReIQ still outperforms that of the benchmark systems. In particular, at the SEP of  $10^{-3}$  and  $\varepsilon^2 = 0.001$ , ReIQ attains approximately 6 dB and 1.5 dB gains in compared with IMOFDM and IQ-IMOFDM, respectively.

## 5 Conclusion

In this research, the author proposed low-complexity detector and investigated the SEP performance of the ReIQ system in the perfect and imperfect CSI conditions. The proposed GD detector can considerably reduce the computational complexity of ReIQ detection while enjoying nearly optimal performance of the ML detection. Furthermore, even in the imperfect CSI condition, ReIQ still outperforms the benchmark systems at the cost of spectral efficiency. Thank to the simple and achievable reliability, the proposals will be proper to IoT communications in which symbol error performance is preferred to the data rate. Lower complexity detectors and schemes to improve the data rate are the significantly meaningful topic for future research.

## References

1. Başar, E., Aygölü, Ü., Panayircı, E., Poor, H.V.: Orthogonal frequency division multiplexing with index modulation. *IEEE Trans. Signal Process.* **61**, 5536–5549 (2013)
2. Fan, R., Yu, Y.J., Guan, Y.L.: Improved orthogonal frequency division multiplexing with generalised index modulation. *IET Commun.* **10**, 969–974 (2016)
3. Wen, M., Basar, E., Li, Q., Zheng, B., Zhang, M.: Multiple-mode orthogonal frequency division multiplexing with index modulation. *IEEE Trans. Commun.* **65**, 3892–3906 (2017)
4. Huyen Le, T.T., Vu-Duc, N., Minh-Tuan, L., Nam, T.X.: Repeated index modulation with coordinate interleaved OFDM. In: 5th NAFOSTED Conference on Information and Computer Science (NICS), pp. 115–119 (2018)
5. Huyen Le, T.T., Nam, T.X.: Repeated index modulation for OFDM with space and frequency diversity, advanced technologies for communications (ATC). In: 2017 International Conference on IEEE, pp. 97–102 (2017).
6. Wang, L., Chen, Z., Gong, Z., Wu, M.: Space-frequency coded index modulation with linear-complexity maximum likelihood receiver in the MIMO-OFDM system. *IEEE Signal Process. Lett.* **23**, 1439–1443 (2016)



# Improving Reliability of Laser Sensor in Aviation Equipment with Anti-jamming Encoding Solution

Thi Nguyen Duc<sup>1</sup>, Trung Dang Tien<sup>2</sup>(✉) , Tuyen Tran Van<sup>3</sup>, and Giang Vu Hoang<sup>2</sup>

<sup>1</sup> General Department of Defense Industry, New Delhi, India

<sup>2</sup> Electric Power University, Hanoi, Vietnam

dangtientrung@gmail.com

<sup>3</sup> Military Technical Academy of Vietnam, Hanoi, Vietnam

**Abstract.** This paper proposes an anti-jamming solution using spread spectrum encoding technique for laser sensor of aviation equipment. Along with conventional anti-jamming methods, this solution is expected to improve the efficiency of laser sensor under active jamming, spectral interference circumstances. Simulation in Matlab-Simulink shows that at signal-to-noise ratio (Vs/Vn) of 5.5 dB, the number of error bits is zero and the ranging error is 0.015 m. Simulation results also prove that this is highly reliable solution with minor error and can be developed to apply for advanced laser sensor.

**Keywords:** Laser sensor · Spread spectrum encoding · Binary phase shift keying

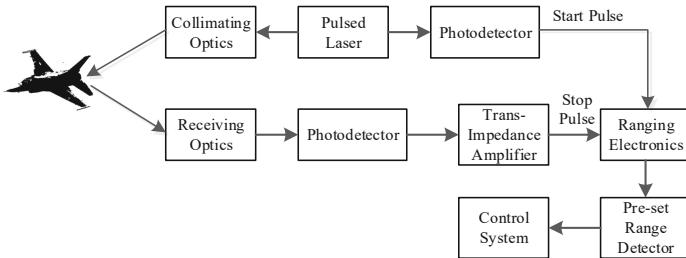
## 1 Introduction

Today, modern military aviation equipment often use laser sensors to measure distance [1, 2, 11]. The working conditions of the sensor are becoming more and more complicated, many kinds of interference affect the receiving and transmitting lines such as: distortion due to Doppler effect, transmission dust, sunlight., unnatural light sources, scattering due to the object's surface has been processed, the enemy's active waveform, ...Therefore, anti-jamming solutions for the sensors are necessary. It is possible to combine some anti-jamming solutions, where additional encoding technique is used for laser beam from the sensor's transmitter in order to help the receiver to get useful signal out among other sources of optical interferences, that plays a significant role for quality improvement of the sensor [3, 4]. Additionally, with the use of high performance microcomputer systems, we can make complex encoding techniques available to the sensor. In this paper, the authors propose an anti-jamming solution using spread spectrum encoding technique to improve quality of laser sensor.

## 2 Operating Principle of Laser Sensor

In principle, the operation of a laser sensor is basically the same as a radio sensor, except that it uses laser for target illumination, specifically: Pulsed laser from the source is

illuminated on the target; Detecting and receiving reflected laser beam from the target; Measuring the time-of-flight interval of pulsed laser from the source to target and back to receiver for distance calculation between the sensor and target [11] (Fig. 1).

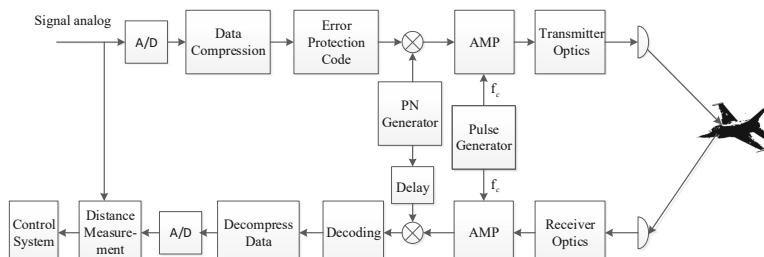


**Fig. 1.** Functional block diagram of laser sensor.

### 3 Reliability Improvement of Laser Sensor Using Spread Spectrum Encoding Technique

Currently, there are several following anti-jamming techniques used for laser sensor [4, 5, 7, 9]: Setting a threshold to firing signal; Placing a coloured glass on top of the laser receiver in accordance with wavelength of emitting laser beam; Using laser receiver and transmitter with narrow field of view; Synchronizing laser receiver and signal processing unit with laser transmitter; Choosing amplitude-frequency characteristics of signal amplifier; Gated-impulse generation at signal processing unit. Nowadays, in addition to passive noise, laser sensors are also subject to active interference, affecting working quality and accuracy.

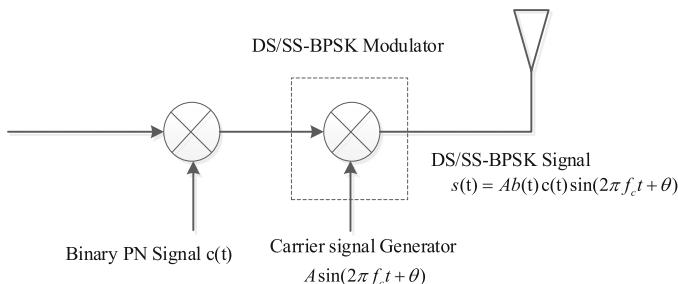
Due to reasons above, this paper introduces an anti-jamming method that called probe beam encoding of laser sensor. The encoder generates encoded signals to control the laser source. The processing unit will compare the reflected signal with the given coding rule, if the rule is correct, it will generate a signal that is sent to the device to measure the distance to the target.



**Fig. 2.** Functional block diagram of a spread spectrum encoding system for a laser sensor

Spread spectrum encoding technique is widely used in radio communication systems, especially in military applications due to its anti-jamming capability [7, 8]. DS/SS (Direct Sequence/Spread Spectrum) systems spread the spectrum by multiplying the source with pseudo-random signal. DS/SS technique is simple but provides a high security level. So we propose to apply this technique on the sensor. Each bit of data is encrypted into a new sequence of bits (with length equal to the length of pseudo-random code), so that the length of received sequence could be longer (more losses on transmission), but in return it gets difficult to jam or intercept, because it is required to have correct key to decode the signal. One of the data encoding techniques is Binary Phase Shift Keying (BPSK) [10]. Implementing DS/SS - BPSK technique, we propose a spread spectrum encoding system for a laser sensor as described in Fig. 2.

This system has two main components: receiver and transmitter. The transmitter is responsible for sending encoded signals using BPSK technique for anti-jamming purpose. The receiver is responsible for decoding collected signals in order to detect the target and active the detonator.



**Fig. 3.** DS/SS-BPSK signal generator

DS/SS-BPSK generation scheme [10] is described in Fig. 3. Laser generator modulates the signal as a function  $b(t)$ , which has values  $\pm 1$  and can be expressed as follow:

$$b(t) = \sum_{k=-\infty}^{\infty} b_k p_T(t - kT) \quad (1)$$

where  $b_k = \pm 1$  is the  $k$ -bit of data and  $T$  is its length (so, the baud rate is  $1/T$  bps). Signal  $b(t)$  is spread out by multiplying with Pseudo random-Noise (PN)  $c(t)$ , which has following form:

$$c(t) = \sum_{k=-\infty}^{\infty} c_k p_T(t - kT_c) \quad (2)$$

The resulting signal  $b(t)$ ,  $c(t)$  then is modulated with the carrier signal using BPSK technique for generating DS/SS- BPSK signal:

$$s(t) = Ab(t)c(t)\sin(2\pi f_c t + \theta) \quad (3)$$

where  $A$  is Amplitude,  $f_c$  is frequency of laser generation and  $\theta$  is initial phase.

The block diagram of DS/SS-BPSK receiver is described in Fig. 4. The purpose of the receiver is to restore  $b(t)$  from the received signal, which includes the reflected signal and parasitic noise. Due to existing delay  $\tau$  in transmission line, the received signal is express as follow:

$$s(t - \tau) + n(t) = Ab(t - \tau)c(t - \tau)\sin(2\pi f_c(t - \tau) + \theta) + n(t) \quad (4)$$

$n(t)$ : noise of the channel and front-end of receiver.

In order to restore the transmitted signal, assuming that the noise is negligible, first, signal with wide bandwidth is de-spread over a narrower band, then it is demodulated to receive signal with original bandwidth. In order to de-spread the sequence, received signal is multiplied by PN signal  $c(t - \tau)$  generated at the receiver and the result as below:

$$\begin{aligned} w(t) &= Ab(t - \tau)c^2(t - \tau)\sin(2\pi f_c t + \theta') \\ &= Ab(t - \tau)\sin(2\pi f_c t + \theta') \end{aligned} \quad (5)$$

Since  $c(t) = \pm 1$  and  $\theta' = \theta - 2\pi f_c \tau$ . This is signal with narrow bandwidth of  $2/T$ . For demodulation, we assume that receiver can detect phase  $\theta'$  and frequency  $f_c$  and the start point of every bit as well. The BPSK demodulator consists of a correlator followed with a threshold unit. To detect  $i$ -bit of data, the correlator performs below calculation:

$$\begin{aligned} z_i &= \int_{t_i}^{t_i+T} w(t) \sin(2\pi f_c t + \theta') dt = A \int_{t_i}^{t_i+T} b(t - \tau) \sin^2(2\pi f_c t + \theta') dt \\ &= \frac{A}{2} \int_{t_i}^{t_i+T} b(t - \tau)[1 - \cos(4\pi f_c t + 2\theta')] dt \end{aligned} \quad (6)$$

where,  $t_i = iT + T$  is start point of  $i$ -bit. Since  $b(t - t) = +1$  or  $-1$  in each bit, the first term in the integral has values  $T$  or  $-T$ . The second term is a double frequency component and has value approximated as 0 after integration. Therefore,  $z_i$  is equal to  $AT/2$  or  $-AT/2$ . Passing this signal through a threshold unit (or comparator) with a threshold = 0, we will receive binary 1 (logic “1”) or  $-1$  (logic “0”).

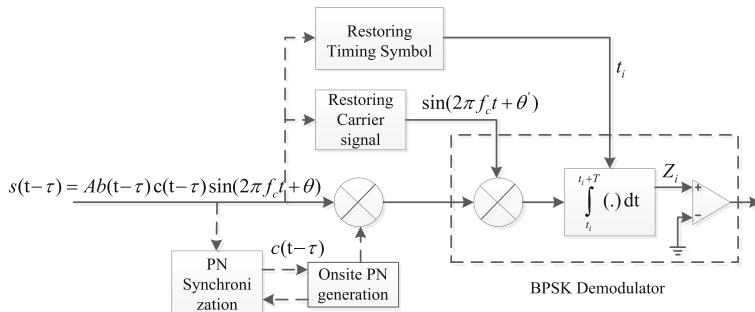


Fig. 4. DS/SS-BPSK receiver

PN Signal is sent to both transmitter and receiver. Since the receiver on the missile knows the code then it is possible to de-spread SS signal to restore the data. On the other hand, enemy's receiver has no idea about the code  $c(t)$ , therefore, it is unable to decode the data in normal conditions.

In addition, another advantage of this technique is ability of multipoint data transmission, spectrum overlapping. This means at the same time many transmitters can send the data, while 1 or more receivers receive the data. So that it is possible to implement multi-layer security for the sensors by increasing number of transmitters and each of them has its own code.

## 4 Simulation, Experiments and Results

No.	Parameters	Value	Unit
1	Frequency of gold code generator	14.4	MHz
2	Length of gold code	32	Bit
3	Frequency of transmitted data	0.452	MHz
4	Number of bits transmitted	1000	Bit
5	Signal-to-noise ratio of transmitter	4.5	dB

Pseudo-random code generator: using Gold code, 2 sets  $m = 5$  registers, length  $N = 2^m - 1 = 31$ . The simulation of transmitter-receiver system of a laser sensor using *Spread Spectrum Encoding* technique, comparision of quality of received signals in different signal-to-noise ratios (SNR) has been made (Fig. 5).

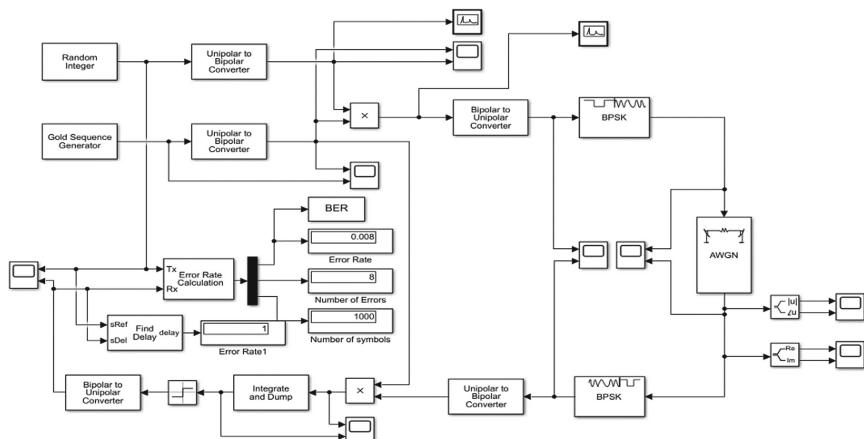
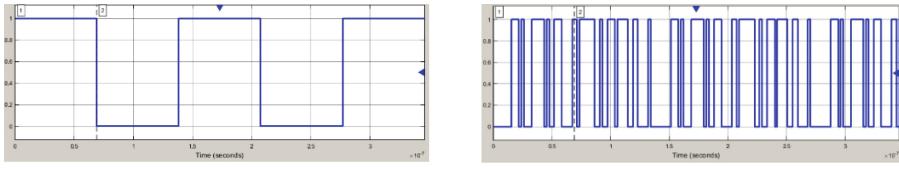


Fig. 5. Matlab-simulink diagram of single receiver-transmitter DSSS system

## Results Analysis

Input data sequence of transmitter and encoded data (Fig. 6):

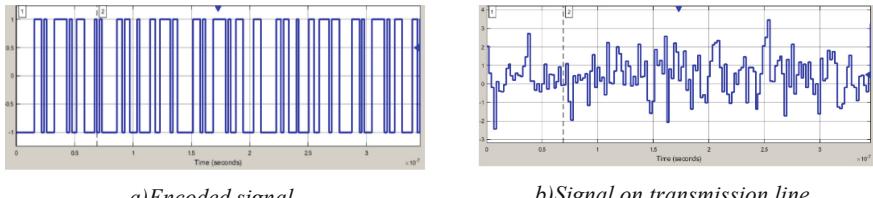


a) Data sequence

b) PN sequence (Gold code)

Fig. 6. Transmitted data and gold code

After being encoded in transmitter, signal is multiplied with pseudo-random noise on transmission line and the received signal has the form below (Figs. 7 and 8):



a) Encoded signal

b) Signal on transmission line.

Fig. 7. Encoded signal mixed with noise

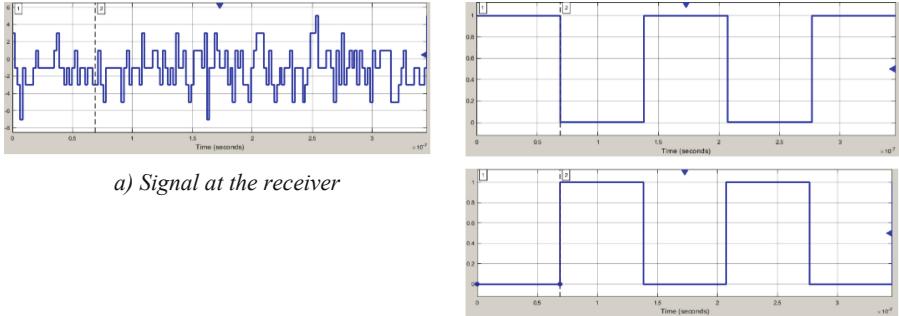


Fig. 8. Data sequence after demodulation at the receiver

The results show that delay time between transmitted and received signals is  $\Delta t_p = 68.99$  ns, processing time of FPGA is  $\Delta t_{fpga} = 30 * 62.5$  ns. Assuming that the drone is flying with speed of  $v = 3705$  km/h (1029 m/s) and in opposite direction to the target, we can obtain ranging error calculated on 1 encoded bit frame for 1 data bit that requires location:

$$\Delta d = (\Delta t_p + \Delta t_{fpga}) * 2 * v = (68.99 + 30 * 62.5) * 10^{-9} * 2 * 1029 * 1000 = 4 \text{ mm}$$

This error is negligible for today's modern flying machines. About bit error, in 1000 bits of data transmitted, there are only 7 error bits, which is equal to 0,7% of transmitted data. This is low level of errors.

## Experimental Results

Transmitter design:

- Binary data generator:
  - + Frequency of random denominator generation: 1 MHz
  - + bit '0'/'1' ratio: 60%
- Binary Gold code generator:
  - + Frequency of code generation: 31 MHz
  - + Poly chain 1: [1 0 0 1 0 1],  $(D^5 + D^2 + 1)$
  - + Poly chain 2: [1 1 1 1 0 1],  $(D^5 + D^4 + D^3 + D^2 + 1)$
  - + Length of Gold code: 31 bits
- Binary phase shift keying (BPSK)
  - + Frequency: 31 MHz (same as frequency of Gold code).

### Receiver Design:

Receiver block of modulated signals from transmitter: received signals are multiplied with carrier signal for demodulation spread spectrum signals from transmitter.

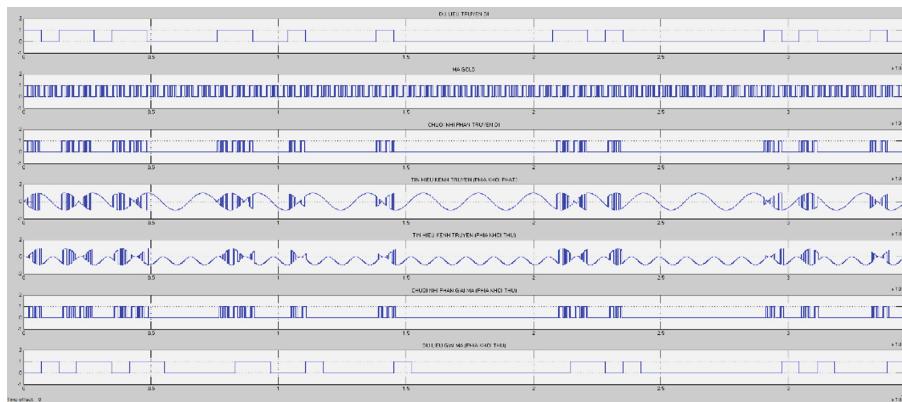
Functional and computing blocks has been made by the means of simulation and conducting experiments on Spartan 3E FPGA KIT (Hardware Co-Simulation).

- Hardware connection diagram in reality (Figs. 9 and 10):



**Fig. 9.** Experimental Model on Spartan 3E FPGA Kit

Time delay in processing is  $6.92 * 10^{-8}$  (s) or 69.2 ns, which is equivalent to simulation result in Simulink). As we can see that the implementation of this encoding algorithm with error correction on FPGA hardware is completely satisfying.



**Fig. 10.** Data Analysis using KIT FPGA

## 5 Conclusion

The application of DS/SS-BPSK *Spread Spectrum Encoding* technique in laser receiver-transmitter devices enhances anti-jamming ability and allows to collect useful information accurately. This helps to improve the quality of the laser sensors, reducing interferences under electronic countermeasures from the target. This method has been evaluated on Matlab-Simulink and proceeded to experiment on microprocessors. The simulation results show that *Spread Spectrum Encoding* technique allows significantly improve the accuracy of data transmission, the error caused by noise is negligible. Based on the results achieved, it is possible to apply DS/SS technique in laser sensors to improve reliability.

## References

1. Кренев, Г.А.: Асимметричный ответ высокоточному оружию, Воениздат, Москва (2006)
2. Лебедев, В.Н.: Авиационные боеприпасы, Москва, Издательство “ВВИА” (1979)
3. Hanshan, L.I., Xiaoqian, Z.: Laser echo characteristics and detection probability calculation on the space projectile proximity sensor. *Optik* **183**, 713–722 (2019)
4. Kaushal, H., Kaddoum, G.: Applications of lasers for tactical military operations. *IEEE Access* **5**, 20736–20753 (2017)
5. Steinvall, O.: “Effects of missiles shape and reflection on laser radar cross sections”, Optical Society of America, 2000
6. Nasser, A.: Recent advancements in proximity sensors technology. *Int. J. Eng. Res. Technol. (IJERT)* (2015)
7. Liu, J.: Advanced optical fuzing technology optical technologies for arming, safing, fuzing, and firing. In: William, J.T., Dickey, F.M., (eds.) *Proceedings of SPIE*, vol. 5871 (2016)
8. Wang, K., Chen, H.: Analysis on the characteristics of pulsed laser proximity sensor’s echo. In: *International Symposium on Photoelectronic Detection and Imaging*, vol. 8192, pp. 445–453 (2011)

9. Mau, L.T.: Development trend of carrying anti-aircraft missiles, Journal of specialized information The situation of foreign military technical development trends, no. 20, General Department of Technology, Hanoi (2002)
10. Trinh, D.Q., Hai, V.T.: Spread spectrum and application techniques. Military Technical Academy (2006)
11. Arora, A.K.: Proximity Sensors Theory and Techniques, Defence Research & Development Organisation, New Delhi (2010)



# Proposing System for Classifying Voucher Images Using Siamese Model

Phat Nguyen Huu<sup>1(✉)</sup>, Nguyet Giap Thi<sup>1</sup>, Hoang Tran Manh<sup>1</sup>, Thanh Le Thi Hai<sup>2</sup>, Hieu Nguyen Trong<sup>3</sup>, and Pha Pham Ngoc<sup>3</sup>

<sup>1</sup> School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam  
[phat.nguyễn\\_huu,hoang.tranmanh}@hust.edu.vn](mailto:{phat.nguyễn_huu,hoang.tranmanh}@hust.edu.vn),  
[Nguyet.GT202445M@sis.hust.edu.vn](mailto:Nguyet.GT202445M@sis.hust.edu.vn)

<sup>2</sup> School of Engineering Physics, Hanoi University of Science and Technology, Hanoi, Vietnam  
[thanh.lethihai@hust.edu.vn](mailto:thanh.lethihai@hust.edu.vn)

<sup>3</sup> National Institute of Patent and Technology Exploitation (NIPTECH), Hanoi, Vietnam  
[nthieu,pnpha}@most.gov.vn](mailto:{nthieu,pnpha}@most.gov.vn)

**Abstract.** Voucher image classification is a complex process that can Please confirm if the corresponding author is correctly identified. Amend if necessary. be influenced by many factors. Currently, people use their hands to classify voucher images. However, there are many problems such as processing time, efficiency, and accuracy of classification. When the data is small and their distributions are not uniform, the machine learning techniques used to classify them are not high accuracy. Therefore, the Siamese model is proposed to classify images with little data without equal between classes in this paper. The results show that the model achieves an accuracy of 99.57% for both forward and backward images when the number of epochs for training is only 100. This proves that the proposed method has a promising application in practice.

**Keywords:** Image classification · Machine learning · Object detection · Siamese model · Deep learning

## 1 Introduction

The system of sorting and calculating discount vouchers at supermarkets is mostly performed manually which takes a lot of time to classify aggregate and send them. When a large number of the voucher has different suppliers, the problem is that we will have to find a solution to save time and give the most accurate results quickly. The purpose of the study is to provide a solution to classify and summarize the voucher. The main contributions of the paper include as follow. Firstly, we design suitable datasets. Secondly, we select the Siamese model for extracting the features in each voucher and classify them according to the correct class. Finally, we will evaluate and improve them.

The rest of the paper includes four parts and is organized as follows. Section 1 will discuss the related work. In Section 2, we present the proposal system. Section 3 will evaluate the proposed system and analyze the results. In the final section, we give conclusions and future research directions.

## 2 Related Work

There are many problems for classifying images [2–9]. In computer vision, each image can be represented as a 2D array of numbers called pixels [6]. Although computers perform differently than humans, we can still guide them to recognize patterns. We must think about images differently. To help them to recognize objects, we use a type of artificial neural network (ANN) called convolutional neural network (CNN). CNNs have two main components. In the hidden layer or feature extraction part, the network will perform a series of convolution and pooling computations to detect the features. At the classification part, a class with full associations will act as a classifier of previously extracted features. This layer will give the probability of an object [3].

## 3 Developing System

### 3.1 Autoencoder and Siamese Model

Autoencoder [1] is a type of neural network that the size of output and input layers is equal. It is responsible for compressing the input into latent-space (bottleneck) of lower dimension and then reconstructing this image. Autoencoder has three parts:

1. **Encoder:** A module that compresses input data into latent space with a smaller size. This is a set of convolution blocks followed by the pooling module.
2. **Latent-space:** The module contains all the features learned by the encoder. It is an important part of the network. Therefore, the encoder-decoder structure has applications in image feature extraction and in establishing useful correlations between different inputs.
3. **Decoder:** This module decompresses the feature from latent space and reconstructs the image. The output is then compared with the real data.

Similar learning is a field of unattended. The goal is to learn several similar functions to measure or related levels of the words and return the same value. It returns a high value when the words are similar. Otherwise, vice versa are different elements. A Siamese neural network [4] contains two or more identical subnets. They have a similar model, parameter, and weight. We only need to train a subnet and use them with others. These networks are used to find the sameness of the inputs by comparing them with typical vectors.

### 3.2 Overview Proposed System

When you have a new voucher code, we just add and store it into the database. Objects of the system are vouchers for discount products. The function of the system is to classify the classes according to them. However, there is little training data when a new one is added. Therefore, we have to collect more data and retrain the model with them. It will take a lot of time as more new data types. The main purpose of the system is to properly classify the classes with little data. When there is more data, we still analyze and classify them correctly without having to re-train. To aggregate and reduce each item code, we will perform OCR to extract information on the code of each voucher. In the paper, we focus on classifying vouchers correctly. The deployed system will consist of three blocks, namely extracting image features, storing blocks extracted from information, and classifying data blocks.

### 3.3 Extracting Information Module

Data samples have a lot of properties. These attributes will reduce the training speed of the models significantly. Feature information extraction is the process used to reduce the dimension and convert it into a simpler and smaller form [7]. In the paper, using an autoencoder is the feature extracting part. However, it will remove the decoder part in which only the part from the beginning to the latent space is taken to perform the extraction for the images as shown in Fig. 1.

### 3.4 Image Classification Module

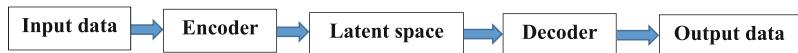
Siamese network is a special type of neural network created from two inputs followed by two identical neural networks. They simply have the same parameters. They are fed into ConvNet to create feature vectors  $h(\times 1)$  and  $h(\times 2)$  with two input images ( $x_1$  and  $x_2$ ). If two images belong to the same class, their feature vectors must be the same. Otherwise, their feature vectors will be different. Therefore, the similarity that is produced by the output Sigmoid. In the two cases, they must be different. The main purpose of this network is to calculate the probability between two input images. If two input images belong to the same class, the value will be 1. Otherwise, it will be close to 0 as shown in Fig. 2.

As mentioned above, the proposed system consists of two main models. The first model is the autoencoder part. However, we only get the latent space and the decoder part will be frozen. We then enter the image classifier to calculate the distance and put it into the sigmoid to give the result in the range (0,1). If the result is closer to 1, the two input images are of the same class. Otherwise, if the closer the result is to 0, the two images will not be of the same class. The full model is as shown in Fig. 3.

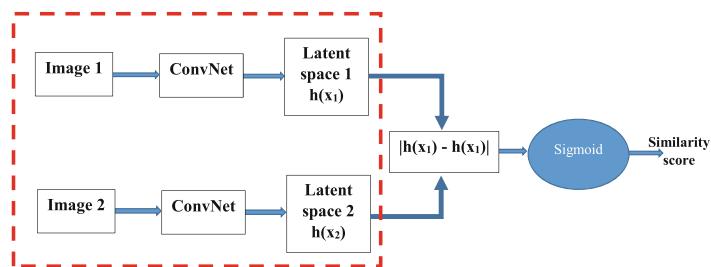
## 4 Result

### 4.1 Dataset

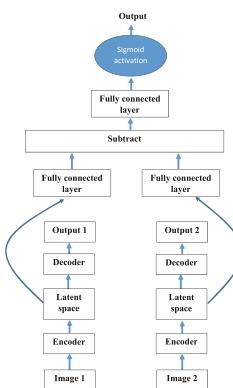
The dataset includes many folders and each of them contains many classes. Each class will have each image. The set of all images for training is 497 and 483 uneven images are divided in each class. To use color-based feature information extraction, we switch from RGB to HSV color system (H channel corresponding to 180, S, and V are 255). Therefore, it requires fast response speed after deploying to the system. We resized the input image to (256,521) to reduce its size. They still have enough information for the model to learn the features of the image. Since autoencoder is an unsupervised machine learning algorithm, we just need to feed the image into it without any labels for them. Before entering the model, the image is converted to an HSV color system and resized to size (256,512) to enhance training. The data input to the Siamese model is the pairs



**Fig. 1.** Antoencoder overview for extraction.



**Fig. 2.** Model Siamese for classifying images.



**Fig. 3.** Overall proposed system using Siamese model.

associated with their label. Pairs of images with the same class will have a label of 1, and pairs of different classes will have a label of 0. When inputting into the image classification system, the machine will not distinguish whether the input images are forward or backward. Therefore, we need to process them for the classification exactly. The data are created as follows to be able to distinguish between the front and back images of pairs of similar or different classes. We will divide them into two types, namely positive and negative pairs. The positive and negative pairs are labeled as 1 and 0, respectively.

Sigmoid\_focal\_cossentropy is introduced in [5]. It is extremely suitable when the data among classes are not balanced. This loss function generalizes cross-entropy binary by adding a parameter  $\gamma$  (gamma), known as the concentration parameter, allowing hard-to-classify examples to be penalized more heavily than easy-to-classify examples by

$$L(y, \hat{p}) = \alpha y(1 - \hat{p})^\gamma \log(\hat{p}) - (1 - y)\hat{p}^\gamma \log(1 - \hat{p}), \quad (1)$$

where  $p \in [0, 1]$  is an estimate of the probability of the positive class,  $y \in 0, 1$  is a binary class label,  $\gamma$  is the concentration parameter, and  $\alpha$  is a hyperparameter that governs the balance between precision and recall by calculating the weighted error for positive class up or down ( $\alpha = 1$  is the default). Configuration system is shown in Table 1.

**Table 1.** Configuration information is used for training.

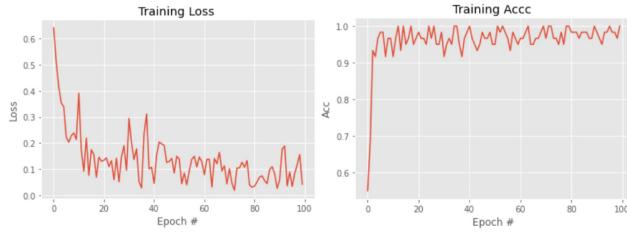
No.	Module	Description
1	Processor	Intel(R) Core(TM) i9-7920X CPU @ 2.90 GHz
2	Ram	64 GB
3	GPU1	NVIDIA Corporation GP102 [GeForce GTX 1080 Ti]
4	GPU1	NVIDIA Corporation GP102 [GeForce GTX 1080 Ti]
5	Cuda version	11.2

After generating data and selecting loss functions, the next step is to perform training. The hyperparameters are selected as  $batch\_size = 512$ , with optimizer Nadam where  $learning\_rate = 1e-4$ . Loss function is calculated by SigmoidFocalCrossEntropy(). We perform with  $epochs = 100$  and  $steps\_per\_epoch = 100$ . The hyperparameters are selected as  $batch\_size = 32$  with optimizer Nadam where  $learning\_rate = 1e-4$ . Loss function is calculated by binary\_crossentropy, with metrics, namely BinaryAccuracy, Precision, and Recall.

## 4.2 Result

To evaluate the classification results, we need to have a set of standard images for each class. To classify the image, the database (DB) will be stored by a data file with columns ID, Extract\_up, Extract\_down, height, and date. After

creating the DB, we proceed to put each image to give a feature vector ( $h_2$ ). We will calculate  $|h_2 - Extract\_up|$  or  $|h_2 - Extract\_down|$  passing through the sigmoid function that returns the highest result. We then will know that class of the input image belongs backward or forward. We will take the flag from the response as 1 and the forward image as 0. We perform with  $epochs = 200$  and  $steps\_per\_epoch = 10$ . The result is as shown in Fig. 4.



**Fig. 4.** Total loss and accuracy of model.

We give the results after classifying each column corresponding to the actual image input. The classification score is shown in Fig. 5.

No.	Name	ID	Score	Predicting image	ID image
0	6_24_a.jpg	6	0.9762002		
1	3_79_a.jpg	3	0.9709794		
2	10_132_a.jpg	10	0.96273375		
3	14_143_a.jpg	14	0.8742094		
4	6_84_a.jpg	6	0.9794533		

**Fig. 5.** Result of image classification.

## 5 Conclusion

This article proposes a Siamese model for classifying images. The results show that the model improves accuracy up to 99.6% with little data. However, there is still the problem of OCR of the code characters on each voucher and from code that will know how much each voucher will be reduced in the proposed model. Therefore, the next development direction is to improve the accuracy of the model and build a product that can be applied in practice.

**Acknowledgment.** This research is carried out in the framework of the project funded by the Ministry of Science and Technology (MOST), Vietnam under the grant 04.2020M008. The authors would like to thank the MOST for the support.

## References

1. Bank, D., Koenigstein, N., Giry, R.: Autoencoders [arXiv:2003.0599](https://arxiv.org/abs/2003.0599), pp. 1–19, March 2020
2. Christ, M.C.J., Parvathi, R.M.S.: Segmentation of medical image using clustering and watershed algorithms. *Am. J. Appl. Sci.* **8**, 1349–1352 (2011)
3. Jain, A., Mao, J., Mohiuddin, K.: Artificial neural networks: a tutorial. *Computer* **29**(3), 31–44 (1996)
4. Koch, G., Zemel, R., Salakhutdinov, R.: Siamese neural networks for one-shot image recognition **37**, 1–8 (2015)
5. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2999–3007 (2017)
6. Nebauer, C.: Evaluation of convolutional neural networks for visual recognition. *IEEE Trans. Neural Netw.* **9**(4), 685–696 (1998)
7. Poynton, C.A.: A guided tour of colour space. In: New Foundation for Video Technology: The SMPTE Advanced Television and Electronic Imaging Conference, pp. 167–180 (1995)
8. Stockman, G., Shapiro, L.G.: Computer Vision. Prentice Hall PTR, USA, 1st edn. (2001)
9. Sural, S., Qian, G., Pramanik, S.: Segmentation and histogram generation using the HSV color space for image retrieval. In: Proceedings of the International Conference on Image Processing, vol. 2, p. II (2002)



# Synthesis of Target Angle Tracking System Combining Kalman Filter Algorithm with Fuzzy Logic Controller

Nguyen Thi Dieu Linh<sup>1</sup> and Tran Ngoc Duc<sup>2</sup>

<sup>1</sup> Hanoi University of Industry, Hanoi, Vietnam

Nguyen.linh@hau.edu.vn

<sup>2</sup> Namdinh University of Technology Education, Nam Dinh, Vietnam

**Abstract.** The article proposes a method to synthesize the target angle tracking system combining the Kalman filter algorithm with a fuzzy logic controller, to improve the quality of the target phase coordinate filter. The tracking system is capable of adapting to the varied maneuvering present in the reality of the target as the evaluation process progresses the most suitable model. As a result, the evaluation quality of the target phase coordinates is advanced, the tracking error is significantly reduced. Because the tracking system does not use directly the signal balance direction as the target coordinate evaluation signal, but uses a separate line of sight coordinate evaluation filter. At the same time, the related states are evaluated by the Kalman filter algorithm in the tracking loop. The structure of the filter is simple, the evaluation error is small.

**Keywords:** Target · Maneuvering · Kalman filter · Fuzzy logic controller

## 1 Introduction

Applying the linear Kalman filter technique [4, 8], the maneuvering frequency and maneuvering intensity parameters are only selected by constants. Therefore, these parameters are not representative of every maneuvering motion of the target. The tracking error will increase when the actual movement of the target does not suitable for the hypothetical model.

Adaptive filtering algorithm with maneuvering detection based on amplification factor correction according to tracking error signal [5], the amplification factor increase when the target is maneuvering will increase the influence of observe noise on the result of evaluation. Therefore, the evaluation accuracy is not high, especially for the line of sight angular speed component. Although simple algorithms but need improvements when used [2].

Adaptive filtering algorithm with maneuvering detection based on state model parameter recognition [1]. Because the forecast matrix is continuously recognized, the state forecast will be more accurate. Because the algorithm does not increase the amplification factor, it does not increase the influence of observations noise to the state evaluation.

Therefore, the state evaluation accuracy of the algorithm is advanced. However, because the state transition matrix adaptation does not use prior information about the different maneuverability of the target, the recognition accuracy for different maneuvering cases is different [3, 6, 7].

With the application of fuzzy logic controller, the article proposes algorithm to perfect the state forecasts matrix recognition plan to respond to more diverse layer of maneuvering models than the two above options. In this algorithm, the target kinematics are hypothesized to follow  $N$  models. The fuzzy system determines the reasonableness level of each model to correct the state forecasts matrix, in order to increase the factor of the most reasonable model to the recognition model.

Therefore, the article is set out to build a target angle tracking system with high accuracy in the conditions of maneuvering and super maneuvering targets.

## 2 Algorithm to Determine Target Angle Coordinates Combining Kalman Filter with Fuzzy Logic Controller

Through the recognition of the state transition matrix  $\Phi$ , the line-of-sight angle coordinate evaluation filter is synthesized on the basis of the state model parameter recognition to perform the adaptation on the forecast phase and on the correction phase of the Kalman filter algorithm [8]. The essence of this method is to find the matrix  $\Phi$  such that the “state forecast at time  $k$ ” coincides with the “state evaluation at time  $k$ ” when using the model recognized at the previous time point. The result of recognition  $\Phi$  will be used for the next time step.

However, the change of matrix  $\Phi$  depends on the imposition of the error correlation matrix  $D_a$  and the intensity matrix  $Q_a$ . In other words, the amplification matrix  $K_a$  in the matrix recognition algorithm  $\Phi$  does not take into account the different changes of the maneuvering frequency and maneuvering intensity. When the target is maneuvering with high frequency and intensity,  $Q_a$  needs to choose large, and when the target is not maneuvering,  $Q_a$  needs to choose very small, in order to reduce the evaluation error of  $\Phi$ .

The determine domain of  $\Phi$  is not known in advance.  $\Phi$  is continuously adapted to fit the maneuvering motion of the target [1, 6]. This adaptation process only uses a priori information  $D_a$ ,  $Q_a$  and does not use prior information about the frequency of the target's maneuvers.

The article proposes an algorithm to recognize matrix  $\Phi$  to overcome the above mentioned disadvantages. The essence of the algorithm is the matrix  $\Phi$  is recognized on the basis of  $N$  matrix  $\Phi^n$  ( $n = 1, 2, \dots, N$ ) ( $\Phi^n$  are known in advance and are created from two layers of target models); where, each  $\Phi^n$  is suitable for a degree of maneuverability. Determining the influence level of each model ( $\Phi^n$ ) on the result of recognition  $\Phi$  is performed by a fuzzy logic controller.

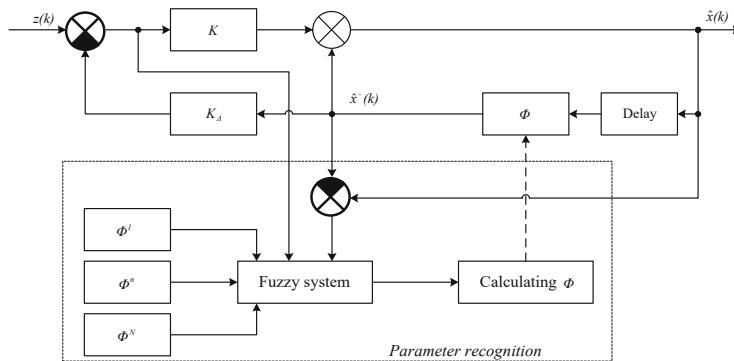
## 2.1 Structure of the Line-of-Sight Angle Coordinate Evaluation Filter Using Kalman Filter Algorithm Combined with Fuzzy Logic Controller

The change of line of sight angle coordinate  $\varphi_m$  and angular speed  $\omega_m$  are determined by the initial state space model, which has the form [1, 2]:

$$\begin{aligned}\varphi_m(k) &= \varphi_m(k-1) + \tau \omega_m(k-1), \quad \varphi_m(0) = \varphi_{m0}; \\ \omega_m(k) &= (1 - \tau \alpha_m) \omega_m(k-1) + \xi_{\omega_m}(k-1), \quad \omega_m(0) = \omega_{m0};\end{aligned}\quad (1)$$

$k$  - Discrete time;  $\tau$  - Discrete time step;  $\alpha_m$  - Characteristic coefficient for the maneuverability of the target;  $\xi_{\omega_m}$  - Central Gause white noise with known variance.

When taking into account the maneuverability of the target, model (1) is extended into two model layers.



**Fig. 1.** Structure diagram of adaptive filter using Kalman algorithm combined with fuzzy logic controller

The first layer consists of  $N_1$  models with size 2. The models differ only by the parameter  $\alpha_m^n$ . The model has the form:

$$\begin{aligned}\varphi_m(k) &= \varphi_m(k-1) + \tau \omega_m(k-1) \\ \omega_m(k) &= (1 - \tau \alpha_m^n) \omega_m(k-1) + \xi_{\omega_m}(k-1)\end{aligned}\quad (2)$$

where, the state transition matrix  $\Phi^n$ :

$$\Phi^n = \begin{bmatrix} 1 & \tau \\ 0 & 1 - \tau \alpha_m^n \end{bmatrix}; \quad n = 1, 2, \dots, N_1 \quad (3)$$

The second layer consists of  $N_2$  models with size 3, ( $N_2 = N - N_1$ ), the state space size is expanded compared to model (1). The models differ only by the parameter  $\alpha_m^n$ . The model has the form:

$$\begin{aligned}\varphi_m(k) &= \varphi_m(k-1) + \tau \omega_m(k-1) + \frac{1}{2} \tau^2 \dot{\omega}_m(k-1) \\ \omega_m(k) &= \omega_m(k-1) + \tau \dot{\omega}_m(k-1) \\ \dot{\omega}_m(k) &= (1 - \tau \alpha_m^n) \dot{\omega}_m(k-1) + \xi_{\omega_m}(k-1)\end{aligned}\quad (4)$$

The state transition matrix  $\Phi^n$ :

$$\Phi^n = \begin{bmatrix} 1 & \tau & \frac{1}{2}\tau^2 \\ 0 & 1 & \tau \\ 0 & 0 & 1 - \tau\alpha_m^n \end{bmatrix}; \quad n = N_1 + 1, \dots, N \quad (5)$$

In (3), (5) parameters  $\alpha_m^n$ , ( $n = 1, 2, \dots, N$ ) are selected in range  $\tau\alpha_m^n = 0 \div 1$ .

## 2.2 Algorithm to Evaluate the Line of Sight Angle Coordinates

Suppose, the state transition matrix  $\Phi$  is used to evaluate the line of sight angle coordinates of the form:

$$\Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \Phi_{13} \\ \Phi_{21} & \Phi_{22} & \Phi_{23} \\ \Phi_{31} & \Phi_{32} & \Phi_{33} \end{bmatrix} \quad (6)$$

The line-of-sight angle coordinate evaluation filter is performed on two “blocks” (Fig. 1):

The block evaluates the phase coordinates of the line of sight when the matrix  $\Phi$  is known; The matrix recognition block  $\Phi$ , when evaluating the known phase coordinates and according to the parameters of the hypothetical model  $\alpha_m^n$ .

- To recognition  $\Phi$ , suppose at the time  $(k-1)$  evaluates the state corresponding to the recognized model as  $\hat{\varphi}_m(k-1)$ ,  $\hat{\omega}_m(k-1)$  and  $\hat{\dot{\omega}}(k-1)$  then:

According to the recognized model, the state forecast will be:

$$\begin{bmatrix} \hat{\varphi}_m^-(k) \\ \hat{\omega}_m^-(k) \\ \hat{\dot{\omega}}_m^-(k) \end{bmatrix} = \Phi \begin{bmatrix} \hat{\varphi}_m(k-1) \\ \hat{\omega}_m(k-1) \\ \hat{\dot{\omega}}_m(k-1) \end{bmatrix} \quad (7)$$

$\Phi$  has the form (6)

The matrix  $\Phi$  is determined such that the state forecast produced by it is equal to the weighted sum of the state forecasts produced by the hypothetical models, i.e.:

$$\begin{bmatrix} \hat{\varphi}_m^-(k) \\ \hat{\omega}_m^-(k) \\ \hat{\dot{\omega}}_m^-(k) \end{bmatrix} = \sum_{n=1}^N \beta^n \begin{bmatrix} \hat{\varphi}_m^{n-}(k) \\ \hat{\omega}_m^{n-}(k) \\ \hat{\dot{\omega}}_m^{n-}(k) \end{bmatrix} \quad (8)$$

The weight  $\beta^n$  determines suitable level (probability) of the  $n^{th}$  model with the actual motion of the target, so we can choose:

$$\sum_{n=1}^N \beta^n = 1 \quad (9)$$

Replace, (3), (5) in (8), get (10): The matrix  $\Phi$  depends on the known parameters  $\alpha_m^n$  ( $n = 1, 2, \dots, N$ ) of the hypothetical models and depends on the unknown parameters  $\beta^n(k)$  ( $n = 1, 2, \dots, N$ ).

$$\Phi = \begin{bmatrix} 1 & \tau & \frac{1}{2}\tau^2 \sum_{n=N_1+1}^N \beta^n \\ 0 & 1 - \tau \sum_{n=1}^{N_1} \beta^n \alpha_m^n & \tau \sum_{n=N_1+1}^N \beta^n \\ 0 & 0 & \sum_{n=N_1+1}^N \beta^n (1 - \tau \alpha_m^n) \end{bmatrix} \quad (10)$$

The description model the change of  $\beta^n(k)$  is selected as follows:

$$\beta^n(k) = \frac{\beta^n(k-1) + \Delta\beta^n}{\sum_{i=1}^N (\beta^i(k-1) + \Delta\beta^i)} \quad (11)$$

The amount of correction  $\Delta\beta^n$  is chosen to be  $\Delta\beta^n = K_\gamma \times \gamma^n$ , for  $K_\gamma = const$ , then:

$$\beta^n(k) = \frac{\beta^n(k-1) + K_\gamma \gamma^n}{1 + K_\gamma} \quad (12)$$

Models (11), (12) show that, the fit level of the  $n^{th}$  model at time  $k$  is determined from the fit level at time  $(k-1)$  (is  $\beta^n(k-1)$ ), by adding an amount of correction  $\Delta\beta^n$ . The correction amount  $\Delta\beta^n$  is determined from parameter  $\gamma^n$ , with the amplification coefficient  $K_\gamma$  by a constant; where,  $\gamma^n$  needs to be adapted from the observation information. The denominator component of (11) is included for the purpose of normalizing  $\beta^n(k)$ , ensuring  $\sum_{n=1}^N \beta^n = 1$ .

The parameters  $\gamma^n$ , ( $n = 1, 2, \dots, N$ ), needs to be determined from the observation information. The value of  $\gamma^n$  is large when the  $n^{th}$  model is considered fit and opposite. To build an algorithm to determine the vector  $\gamma$  ( $\gamma = [\gamma^1, \gamma^2, \dots, \gamma^n, \dots, \gamma^N]^T$ ), need to consider the errors from the hypothetical models.

– Evaluate phase coordinates of line of sight when matrix  $\Phi$  is known:

+ Tracking error in hypothetical models: Suppose, at time  $(k-1)$  the received status evaluations are  $\hat{\phi}_m(k-1)$ ,  $\hat{\omega}_m(k-1)$  and  $\hat{\omega}(k-1)$ .

The tracking error of the first model layer can be determined through the tracking error of the recognized model:

$$\Delta z^n(k) = \Delta z(k) + \frac{1}{2}\tau^2 K_\Delta \left( 1 - \sum_{l=1}^{N_1} \beta^l \right) \hat{\omega}_m(k-1); \quad n = 1, 2, \dots, N_1 \quad (13)$$

The tracking error for the second model layer is:

$$\Delta z^n(k) = \Delta z(k) - \frac{1}{2} \tau^2 K_\Delta \sum_{l=1}^{N_1} \beta^l \hat{\omega}_m(k-1); \quad n = N_1 + 1, N_1 + 2, \dots, N \quad (14)$$

The tracking error of the hypothetical filters (of the layer of filters) is determined from the tracking error of the filter using the recognition model.

+ Error of state evaluation from state forecast: Error between the status evaluated at the current time  $k$  and status forecast when using the  $n^{\text{th}}$  hypothetical model is:

$$\begin{bmatrix} \Delta\varphi^n(k) \\ \Delta\omega^n(k) \\ \Delta\dot{\omega}^n(k) \end{bmatrix} = \begin{bmatrix} \hat{\varphi}_m(k) \\ \hat{\omega}_m(k) \\ \hat{\dot{\omega}}_m(k) \end{bmatrix} - \Phi^n \begin{bmatrix} \hat{\varphi}_m(k-1) \\ \hat{\omega}_m(k-1) \\ \hat{\dot{\omega}}_m(k-1) \end{bmatrix} \quad (15)$$

When using the first model layer, the second model layer, the recognition model, we have:

$$\begin{aligned} \Delta\varphi^n(k) &= \Delta\varphi(k) + \frac{1}{2} \tau^2 \sum_{l=N_1+1}^N \beta^l \hat{\omega}_m(k-1) \\ \Delta\omega^n(k) &= \Delta\omega(k) + \left( -\tau \sum_{l=1, l \neq n}^{N_1} \beta^l \alpha_m^l \hat{\omega}_m(k-1) + \tau \sum_{l=N_1+1}^N \beta^l \hat{\omega}_m(k-1) \right) \end{aligned} \quad (16)$$

$$\begin{aligned} \Delta\dot{\omega}^n(k) &= \Delta\dot{\omega}(k) + \sum_{l=N_1+1}^N \beta^l (1 - \tau \alpha_m^l) \hat{\dot{\omega}}_m(k-1) \text{ with } n = 1, 2, \dots, N_1 \\ \Delta\varphi^n(k) &= \Delta\varphi(k) - \frac{1}{2} \tau^2 \sum_{l=1}^{N_1} \beta^l \hat{\omega}_m(k-1) \text{ with } n = N_1 + 1, N_1 + 2, \dots, N \\ \Delta\omega^n(k) &= \Delta\omega(k) + \left( -\tau \sum_{l=1}^{N_1} \beta^l \alpha_m^l \hat{\omega}_m(k-1) - \tau \sum_{l=1}^{N_1} \beta^l \hat{\dot{\omega}}_m(k-1) \right) \\ \Delta\dot{\omega}^n(k) &= \Delta\dot{\omega}(k) + \sum_{l=N_1+1, l \neq n}^N \beta^l (1 - \tau \alpha_m^l) \hat{\dot{\omega}}_m(k-1) \end{aligned} \quad (17)$$

From (16), (17), found that: The errors when using the  $n^{\text{th}}$  hypothetical model are determined from the error of the recognition model.

The comments for  $\Delta\varphi^n(k)$  are similar to those for  $\Delta z^n(k)$ . Therefore, to determine  $\gamma$  either use  $\Delta z^n(k)$  or use  $\Delta\varphi^n(k)$ .

### 2.3 Evaluation Algorithm the Line of Sight Angle Coordinates Using Fuzzy Logic Controller

To determine  $\gamma^n$ , ( $n = 1, 2, \dots, N$ ), we use a fuzzy logic controller with errors  $\Delta\varphi(k)$ ,  $\Delta\omega(k)$  and  $\Delta\dot{\omega}(k)$  as input. This fuzzy system ensures to increase the efficiency of the most reasonable model to the recognition model. That is, if the  $n^{\text{th}}$  model is considered reasonable, then  $\gamma^n$  takes a large value, and  $\gamma^i$  ( $i \neq n$ ) takes a small value.

A fuzzy logic controller, the basic components include: Fuzzifying stage, composition equipment and defuzzification (fuzzy solving) stage [8].

– Fuzzing stage: The input variable of the fuzzy system includes:  $\Delta\varphi(k)$ ,  $\Delta\omega(k)$  and  $\Delta\dot{\omega}(k)$ .

+ For the input variable  $\Delta\varphi(k)$ , is fuzzed by two language variables with a selection interdependencies function of the form Gause:

$$\mu_{1i} = \exp\left(-\frac{1}{2}\left(\frac{\Delta\varphi(k) - c_{1i}}{\sigma_{1i}}\right)^2\right); i = 1, 2 \quad (18)$$

$$c_{11} = -\frac{1}{2}t^2 \sum_{n=N_1+1}^N b^n(k-1)\hat{w}_m(k-1) \quad (19)$$

$$c_{12} = \frac{1}{2}\tau^2 \sum_{n=1}^{N_1} \beta^n \hat{w}_m(k-1) \quad (20)$$

The widths of these interdependencies function are chosen equally  $\sigma_{11} = \sigma_{12} = \sigma_1$ , so that the interdependencies function  $\mu_{11}$  and  $\mu_{12}$  intersect. Choose

$$\sigma_1 = k_{\sigma_1} |c_{11} - c_{12}| \quad (21)$$

+ For the input variable  $\Delta\omega(k)$ , is fuzzed by  $(N_1 + 1)$  language variables. The selected interdependencies function has the form Gause:

$$\mu_{2i} = \exp\left(-\frac{1}{2}\left(\frac{\Delta\omega(k) - c_{2i}}{\sigma_{2i}}\right)^2\right); i = 1, 2, \dots, N_1 + 1 \quad (22)$$

$$c_{2n} = \tau \left( \sum_{l=1}^{N_1} \beta^l \alpha_m^l - \beta^n \alpha_m^n \right) \hat{w}_m(k-1) - \tau \sum_{l=N_1+1}^N \beta^l \hat{w}_m(k-1); n = 1, 2, \dots, N_1 \quad (23)$$

$$c_{2n} = \tau \sum_{l=1}^{N_1} \beta^l \alpha_m^l \hat{w}_m(k-1) + \tau \sum_{l=1}^{N_1} \beta^l \hat{w}_m(k-1); n = N_1 + 1 \quad (24)$$

The interdependencies function widths of the language variables  $L_n^2$ , ( $n = 1, 2, \dots, N_1 + 1$ ), are chosen so that consecutive interdependencies function intersect. For simplicity, can choose  $\sigma_{2n} = \sigma_2$  with every  $n = 1, 2, \dots, N_1 + 1$ .

$$\sigma_2 = k_{\sigma_2} |\max(c_{2n}) - \min(c_{2n})| / N_1 \quad (25)$$

+ For the input variable  $\Delta\dot{\omega}(k)$ , is fuzzed by  $(N_2 + 1)$  language variables, ( $N_2 = N - N_1$ ). The selected interdependencies function has the form Gause:

$$\mu_{3i} = \exp\left(-\frac{1}{2}\left(\frac{\Delta\dot{\omega}(k) - c_{3i}}{\sigma_{3i}}\right)^2\right); i = 1, 2, \dots, N_2 + 1 \quad (26)$$

$$c_{3n} = \sum_{l=N_1+1}^N \beta^l (1 - \tau \alpha_m^l) \hat{\omega}_m(k-1); n=1 \quad (27)$$

$$c_{3n} = \left( \sum_{l=N_1+1}^N \beta^l (1 - \tau \alpha_m^l) - \beta^{N_1-1+n} (1 - \tau \alpha_m^{N_1-1+n}) \right) \hat{\omega}_m(k-1) \quad (28)$$

The interdependencies function widths of the language variables  $L_n^3$ , ( $n = 1, 2, \dots, N_2 + 1$ ), are chosen so that consecutive interdependencies function intersect. For simplicity, can choose  $\sigma_{3n} = \sigma_3$  with every  $n = 1, 2, \dots, N_2 + 1$ .

$$\sigma_3 = k_{\sigma_3} |max(c_{3n}) - min(c_{3n})| / N_2 \quad (29)$$

+ Fuzzifying of the output variable  $\gamma$ :

Interdependencies function of the output variable selected form “singleton” [3].

This method is extended to the vector form output variable. That is,  $\gamma$  is fuzzed by  $N$  fuzzy set, denoted by  $A_\gamma^n$  is the  $n^{th}$  fuzzy set with interdependencies function;

$$\mu_{A_\gamma^n}(\gamma) = \begin{cases} 1 & \text{when } \gamma = \gamma_n \\ 0 & \text{when } \gamma \neq \gamma_n \end{cases} \quad (30)$$

$\gamma_n$  - Vector of size  $N$ .

- Composition equipment: Composition equipment are built through the selection of fuzzy composition law. Selecting  $N$  composition law, the  $n^{th}$  composition law determines the increasing influence of model  $n^{th}$  on the recognition model.

The fuzzy set  $A_\gamma^n$  is determined by the interdependencies function  $\mu_{A_\gamma^n}(\gamma)$ . The values of the elements of vector  $\gamma_n$  are selected so that the  $n^{th}$  element has the largest value, the remaining elements have a smaller value. For simplicity can choose equal.

$$\gamma_n = [\gamma_n^1 \dots \gamma_n^i \dots \gamma_n^N]^T \quad (31)$$

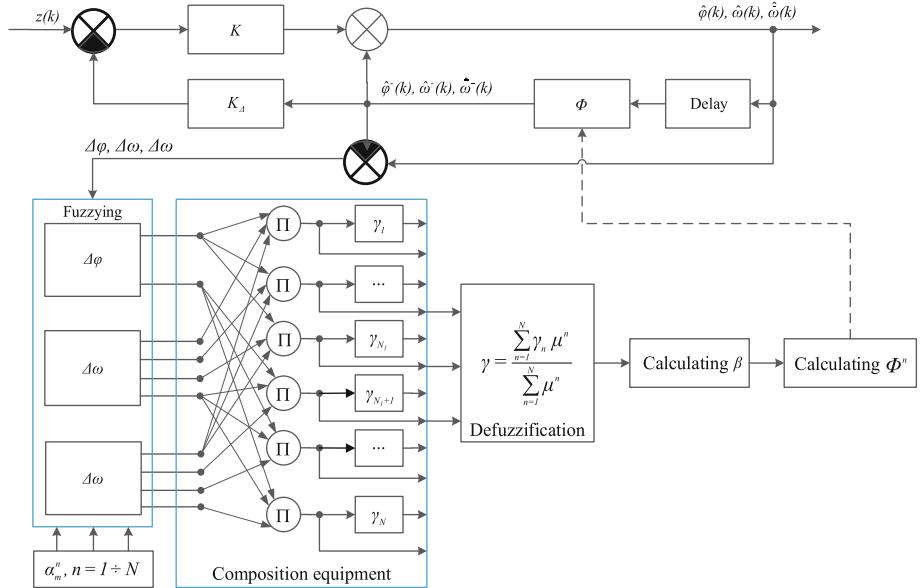
With,  $\gamma_n^i$ , ( $i = 1, 2, \dots, N$ ), determined as follows:

$$\gamma_n^i = \begin{cases} \gamma_{max} & \text{when } i = n \\ (1 - \gamma_{max}) / (N - 1) & \text{when } i \neq n \end{cases} \quad (32)$$

$\gamma_{max}$  - design parameter, its value is selected in range  $1/N < \gamma_{max} \leq 1$ .

- Defuzzification: Choosing the central defuzzification method, with the composition law as described above, the output variable  $\gamma$  is determined by (Eq. 33).

$$\gamma = \frac{\sum_{n=1}^N \gamma_n \mu^n}{\sum_{n=1}^N \mu^n} \quad (33)$$



**Fig. 2.** Structure diagram of adaptive filter using Kalman algorithm combined with fuzzy logic controller to evaluate line of sight angle coordinates

### 3 Simulation Results and Analysis

– The sample trajectory is generated from the following kinetic model:

$$\begin{aligned}\varphi_m(k) &= \varphi_m(k-1) + \tau \omega_m(k-1) \\ \omega_m(k) &= (1 - \tau \alpha_m) \omega_m(k-1) + \alpha_m \tau u \\ z(k) &= \varphi_m(k) + \xi_z \varphi(k)\end{aligned}\quad (34)$$

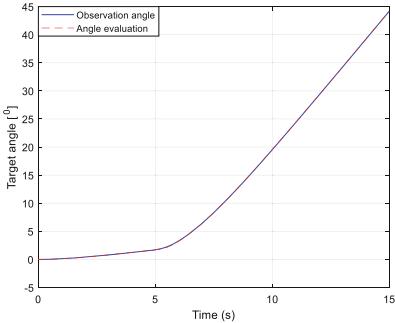
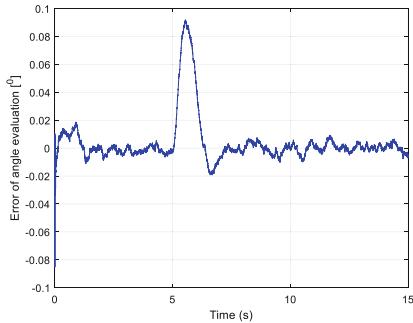
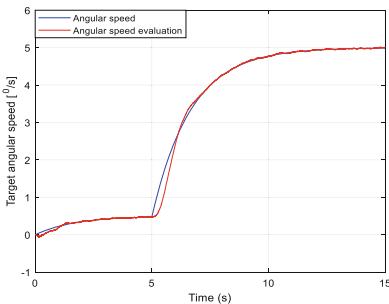
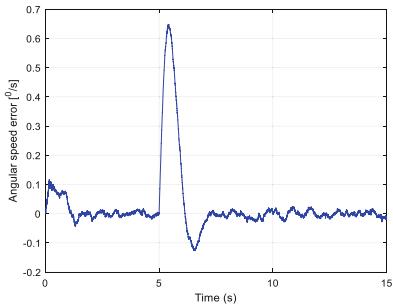
– Parameters for creating line of sight trajectory:

$$\begin{aligned}\tau &= 0,001(s), \alpha_m = 0,6(1/s), \varphi_{m0} = 0(rad), \omega_{m0} = 0(rad/s), \sigma_{\xi_z}^2 = \\ &0,05^2(o^2), u(^0/s) = \begin{cases} 0,5 & \text{when } t < 5s \\ 5 & \text{when } t \geq 5s \end{cases}, \text{simulation time } t_\Sigma = 15(s).\end{aligned}$$

– Filter parameters using fuzzy adaptive system:

- + Layer 1:  $N_1 = 4$ ;  $\alpha_m = 0,001; 0,33; 0,66; 1$ .
- + Layer 2:  $N_2 = 4$ ;  $\alpha_m = 1; 0,66; 0,33; 0,001$
- + Other parameters:  $\beta^n = 0,25$ ;  $\gamma_{max} = 0,85$ ;  $\sigma_{\omega_m} = 0,0015$ .

– The simulation results are shown by Figs. 3, 4, 5 and 6.

**Fig. 3.** Evaluation of angle coordinates  $\hat{\phi}_m$ **Fig. 4.** Evaluation error of angular coordinates**Fig. 5.** Evaluation of angular speed  $\hat{\omega}_m$ .**Fig. 6.** Error of angular speed evaluation

- When the target is maneuvering (at time 5 s), the adaptation process will change the parameters of the model, ensuring a reduction in the error of evaluation of angle coordinates and angular speed.
- In the initial stage of maneuvering, because the model is not suitable, both the error of evaluating the angle coordinates and the angular speed are large, but this error value is still small. When the model's parameters are suitable, the evaluation error is reduced.
- When using fuzzy adaptive system, the quality is improved than using state model parameter recognition adaptive system. That is, the error when using an fuzzy adaptive system is reduced even if the target is maneuvering with the change angular speed. In the early stages of the adaptation process ( $t = 5 \div 7$ s), the errors of the two methods are roughly equivalent.

## 4 Conclusion

Adaptive filtering algorithm with maneuvering detection based on using fuzzy adaptive system, controlled the reasonableness of each component model and increased the weight of the most reasonable model in the recognition model, so the recognition accuracy is improved in the case of the maneuvering targets to varying degrees. The application of the adaptive fuzzy controller has received more simple algorithm, avoiding the interactive multi-model evaluation technique are very complex.

Target angle coordinate determination system with maneuvering detection based on the application of Kalman filter algorithm combined with fuzzy logic controller with high accuracy according to both position and speed in case of the maneuvering target.

From the simulation results, can be seen that the target angle coordinate system using the Kalman filter algorithm combined with a fuzzy logic controller significantly reduces the tracking error when the target changes maneuvering form. Capable of adapting to the target's maneuvering when the evaluation process progresses to the most suitable model.

## References

1. Van Bang, N.: Synthesis of the maneuvering target coordinate determination algorithm on the basis of application of the interactive multi-model adaptive filtering technique. *East Eur. Sci. J.* **1**(67), 3/2021, 16–23 (2021)
2. Van Bang, N.: Improving the quality of the line of sight angle coordinates filter of the self-guidance missile when using the snake style maneuvering model. *Int. J. Softw. Hardw. Res. Eng.* **9**(4), 7–13 (2021)
3. Qian, G.H., Li, Y., Luo, R.J.: One maneuvering frequency and the variance adaptive filtering algorithm for maneuvering target tracking. *J. Radars* **2**(6), 258–264 (2017)
4. Simon, D.: Kalman filtering with state constraints - a survey of linear and nonlinear algorithm. *Control Theory Appl. IET* **4**(8), 1303–1318 (2010)
5. Rovisco Pais Av. Kalman and Extended Kalman Filters. Concept, Derivation and Properties, M. Isabel Ribeiro (2014)
6. Li, X.R., Jilkov, V.P.: Survey of maneuvering target tracking. *IEEE Trans. Aeros. Electron. Syst.* **41**(4), 1255–1321 (2015)
7. Ren, J., Zhang, X., Sun, J., Zeng, Q.: Interactive multi-model target maneuver tracking method based on the adaptive probability correction. In: Tan, Y., Shi, Y., Tang, Q. (eds.) ICSI 2018. LNCS, vol. 10942, pp. 235–245. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-93818-9\\_22](https://doi.org/10.1007/978-3-319-93818-9_22)
8. van Amerongen, J.: Intelligent Control, Lecture notes, University of Twente, The Netherlands, March 2014



# Design of an Adaptive Fuzzy Hierarchical Sliding-Mode Control for Uncertainties Pendubot

Van-Truong Nguyen<sup>(✉)</sup> and Quoc-Cuong Nguyen

Department of Mechatronics Engineering, Hanoi University of Industry, Hanoi, Vietnam  
nguyenvantruong@hau.edu.vn

**Abstract.** Based on the dynamic model, an adaptive fuzzy hierarchical sliding mode control (AFHSMC) for pendubot with uncertainty is presented in this study. The AFHSMC controller's settings are selected via adaptive fuzzy logic control. The Lyapunov approach is used to prove the pendubot's stability. The simulation findings reveal that the suggested control method is both dynamically and robustly effective.

**Keywords:** Pendubot · Fuzzy systems · Adaptive control · Uncertainties

## 1 Introduction

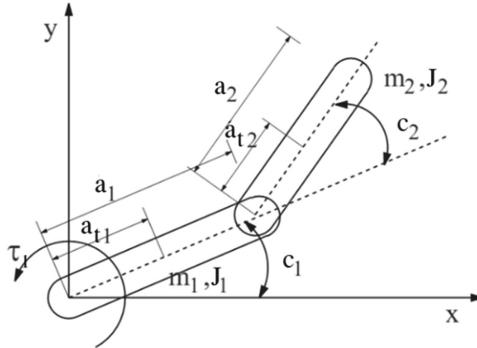
Pendubot is a mechanical system with two links spin on two joints, in there two controllable output position values, but only one control signal for the input [1]. The first link is transmitted through a direct current motor, and the second link rotates freely around the first link. The pendubot's mechanical structure is basic, yet it has a lot of nonlinearities. As a result, the pendubot is frequently utilized in research and education [2].

Various approaches for controlling pendubot are developed in the literature review, including PD [3], PID [4], sliding mode control [5], neural network control [5], etc. According to [5], sliding mode control (SMC) is a typical control approach established by dynamic model characteristics of the systems. Because the pendubot system has two links, then it is divided into two subsystems for the purpose of designing a hierarchical sliding mode controller (HSMC) [6]. However, the HSMC controller includes a sign(.) function that results in chattering around the sliding surface [6]. The fuzzy control method is very flexible and adapts to the system over time. However, the stability of the fuzzy control approach may not be guaranteed. Hence, the hybrid of fuzzy control with the adaptive or robust control methods has been proposed to a more optimal control approach.

The goal of this paper is to further develop an AFHSMC controller established by a hybrid of HSMC and fuzzy logic control (FLC). The FLC is adopted to approximate the parameter of the suggested strategy. The proposed controller's initial parameters are chosen using a genetic algorithm. The law of control is proven stable according to the Lyapunov theory.

The organization of the paper has 5 sections. Section 2 presents the dynamic model of pendubot in the standard state. Section 3 shows the design of the AFHSMC controller. Next, simulation and evaluates the stability of the suggested controller are shown in Sect. 4. Section 5 presents the conclusion.

## 2 System Dynamic of the Pendubot



**Fig. 1.** The pendubot's structure

The pendubot system's dynamic model in the standard state, with no friction, noise, or other factors, is as follows [1] (Fig. 1):

$$M(c)\ddot{c} + R(c, \dot{c})\dot{c} + p(c) = \tau \quad (1)$$

$$M(c) = \begin{bmatrix} \omega_1 + \omega_2 + 2\omega_3 \cos c_2 & \omega_2 + \omega_3 \cos c_2 \\ \omega_2 + \omega_3 \cos c_2 & \omega_2 \end{bmatrix} \quad (2)$$

$$R(c, \dot{c}) = \begin{bmatrix} -\omega_3 \sin(c_2)\dot{c}_2 & -\omega_3 \sin(c_2)\dot{c}_2 - \omega_3 \sin(c_2)\dot{c}_1 \\ \omega_3 \sin(c_2)\dot{c}_1 & 0 \end{bmatrix} \quad (3)$$

$$p(c) = \begin{bmatrix} \omega_4 g \cos c_1 + \omega_5 g \cos(c_1 + c_2) \\ \omega_5 g \cos(c_1 + c_2) \end{bmatrix} \quad (4)$$

where  $c = [c_1, c_2]^T$ ,  $c_1$  represents the angle formed by the first link with the horizontal,  $c_2$  represents the angle formed by the second and first links;  $\dot{c}$  represents the angular velocity vector;  $\ddot{c}$  represents the angular acceleration vector;  $\tau = [\tau_1, 0]^T$ ,  $\tau_1$  represents the external force activated to the first link.

The pendubot system's parameters are determined as:

$$\begin{cases} \omega_1 = m_1 a_{t1}^2 + m_2 a_1^2 + J_1 \\ \omega_2 = m_2 a_{t2}^2 + J_2 \\ \omega_3 = m_2 a_1 a_{t2} \\ \omega_4 = m_1 a_{t1} + m_2 a_1 \\ \omega_5 = m_2 a_{t2} \end{cases} \quad (5)$$

where  $m_1, m_2$  are the masses of the first link and the second link, respectively; the lengths of the first and second links are  $a_1$  and  $a_2$ , respectively;  $a_{t1}$  represents the measurement from the first joint to the first link's center of mass, the measurement between the second joint and the second link's center of mass is  $a_{t2}$ ;  $J_1$  and  $J_2$  are the first and second links' moments of inertia around their centroids.

The system's equation of state is obtained as:

$$\begin{bmatrix} \ddot{c}_1 \\ \ddot{c}_2 \end{bmatrix} = M(c)^{-1}\tau - M(c)^{-1}R(c, \dot{c})\dot{c} - M(c)^{-1}p(c) \quad (6)$$

where

$$\ddot{c}_1 = \frac{1}{\omega_1\omega_2 - \omega_3^2 \cos^2(c_2)} \begin{bmatrix} \omega_2\tau_1 + \omega_2\omega_3(\dot{c}_1 + \dot{c}_2)^2 \sin c_2 + \omega_3^2\dot{c}_1^2 \sin c_2 \cos c_2 \\ -\omega_2\omega_4g \cos c_1 + \omega_3\omega_5g \cos c_2 \cos(c_1 + c_2) \end{bmatrix} \quad (7)$$

$$\ddot{c}_2 = \frac{1}{\omega_1\omega_2 - \omega_3^2 \cos^2(c_2)} \begin{bmatrix} (-\omega_2 - \omega_3 \cos c_2)\tau_1 - \omega_3(\omega_2 + \omega_3 \cos c_2)(\dot{c}_1 + \dot{c}_2)^2 \sin c_2 \\ -\omega_5g(\omega_1 + \omega_3 \cos c_2) \cos(c_1 + c_2) \\ +\omega_4g(\omega_2 + \omega_3 \cos c_2) \cos c_1 - \omega_3\dot{c}_1^2 \sin c_2(\omega_1 + \omega_3 \cos c_2) \end{bmatrix} \quad (8)$$

### 3 Fuzzy Hierarchical Sliding Mode Control

Let  $x_1 = c_1 - \pi/2, x_2 = \dot{c}_1, x_3 = c_2, x_4 = \dot{c}_2, u = \tau_1$ . The state Eq. (6) is expressed in the following simple form [6]:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = f_1 + b_1u \\ \dot{x}_3 = x_4 \\ \dot{x}_4 = f_2 + b_2u \end{cases} \quad (9)$$

where

$$f_1 = \frac{1}{\omega_1\omega_2 - \omega_3^2 \cos^2(c_2)} \left[ \omega_2\omega_3(\dot{c}_1 + \dot{c}_2)^2 \sin c_2 + \omega_3^2\dot{c}_1^2 \sin c_2 \cos c_2 - \omega_2\omega_4g \cos c_1 + \omega_3\omega_5g \cos c_2 \cos(c_1 + c_2) \right] \quad (10)$$

$$b_1 = \frac{\omega_2}{\omega_1\omega_2 - \omega_3^2 \cos^2(c_2)} \quad (11)$$

$$f_2 = \frac{1}{\omega_1\omega_2 - \omega_3^2 \cos^2(c_2)} \left[ -\omega_3(\omega_2 + \omega_3 \cos c_2)(\dot{c}_1 + \dot{c}_2)^2 \sin c_2 + \omega_4g(\omega_2 + \omega_3 \cos c_2) \cos c_1 - \omega_5g(\omega_1 + \omega_3 \cos c_2) \cos(c_1 + c_2) - \omega_3\dot{c}_1^2 \sin c_2(\omega_1 + \omega_3 \cos c_2) \right] \quad (12)$$

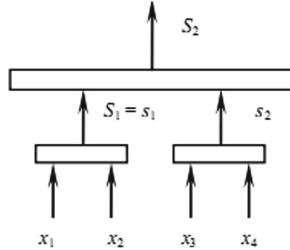
$$b_2 = \frac{-\omega_2 - \omega_3 \cos c_2}{\omega_1\omega_2 - \omega_3^2 \cos^2(c_2)} \quad (13)$$

The subsystems sliding surfaces are designed following way [6]:

$$\begin{cases} s_1 = e_1x_1 + x_2 \\ s_2 = e_2x_3 + x_4 \end{cases} \quad (14)$$

Equivalence control rules of two subsystems are obtained by derivative (9) according to time and considering the system's stability when  $\dot{s}_1 = 0$  and  $\dot{s}_2 = 0$ .

$$\begin{cases} u_{ss1} = -(e_1 x_2 + f_1)/b_1 \\ u_{ss2} = -(e_2 x_4 + f_2)/b_2 \end{cases} \quad (15)$$



**Fig. 2.** HSMC surface structure

Figure 2 shows the construction of the HSMC surfaces. The first layer surface  $S_1$  is also the slip surface of the subsystem  $s_1$ . In other words,  $S_1 = s_1$ . Then the sliding mode surface of the second layer  $S_2$  is put together from  $S_1$  and  $s_2$  as  $S_2 = \alpha S_1 + s_2$ . Combined with (9) and (14)–(15), the proposed controller law is calculated as follows:

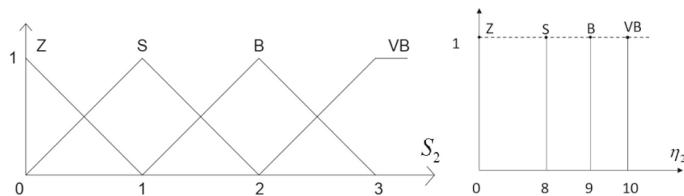
$$u = \frac{\alpha b_1 u_{ss1} + b_2 u_{ss2} - k_2 S_2 - \eta_2 \text{sign}(S_2)}{\alpha b_1 + b_2} \quad (16)$$

where  $\alpha$ ,  $k_2$  are positive constants that have been chosen to fit the system;  $b_1, b_2$  are positive constants calculated from the parameters of the pendubot system.

Parameter  $\eta_2$  is calculated from parameter  $S_2$  through fuzzy rule. Bring into play the Sugeno fuzzy model, the inference rules for the single input and output approximator are as follows:

$$\text{if } S_2^i \text{ is } A_j \text{ then } \eta_2^i \text{ is } B_j$$

The input  $S_2$  has linguistic values of Z (Zero), S (Small), B (Big), VB (Very Big) corresponding to the output variable  $\eta_2$  with corresponding values of 0, 8, 9, 10 are presented in Fig. 3.



**Fig. 3.**  $S_2$  and  $\eta_2$  of membership functions

*Theorem:* Considering the dynamic systems of pendubot in (1), the laws of the controller in (16), the subsystems' sliding surfaces in (14), we can conclude that the pendubot stability is ensured.

*Proof:* The Lyapunov function is as follows:

$$V = \frac{1}{2}S_2^2 \geq 0 \quad (17)$$

$$\begin{aligned} \dot{V} &= S_2 \dot{S}_2 = S_2(\alpha \dot{s}_1 + \dot{s}_2) \\ &= S_2(\alpha(e_1x_2 + f_1) + (e_2x_4 + f_2) + (\alpha b_1 + b_2)u) \\ &= S_2(-\alpha b_1 u_{ss1} - b_2 u_{ss2} + \alpha b_1 u_{ss1} + b_2 u_{ss2} - k_2 S_2 - \eta_2 \text{sign}(S_2)) \\ &= S_2(-k_2 S_2 - \eta_2 \text{sign}(S_2)) \end{aligned} \quad (18)$$

It is rewritten as:

$$\dot{V} = (-k_2 S_2^2 - \eta_2 |S_2|) \leq 0 \quad (19)$$

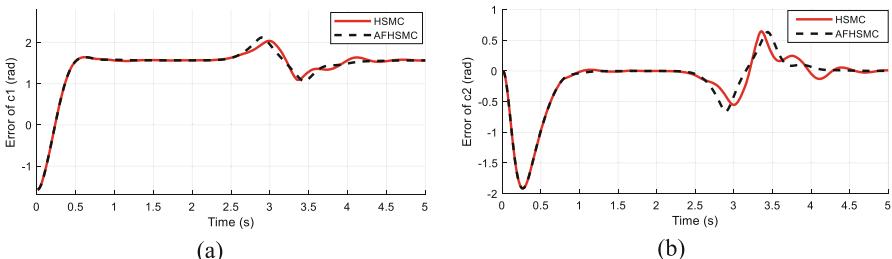
That as result, Lyapunov proves that the entire AFHSMC is stable.

## 4 Demonstrative Example

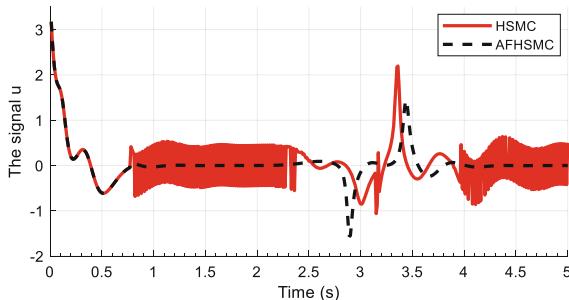
Firstly, the genetic algorithm is used to choose the parameters  $\alpha, k_2, e_1$  and  $e_2$ . Then the fuzzy rule is used to calculate  $\eta_2$  from  $S_2$ . Finally, derive the control law  $u$  using Eq. 16.

The simulation results are produced in this section to compare the proposed controller to the HSMC controller [6] to demonstrate its effectiveness. The time in all simulations is 5 s, as well as the sampling time is 0.01 s. The following uncertainty noise is applied to the proposed controller is as follows:

$$e(t) = \frac{23\pi}{180} \times \exp\left(-\frac{(t-3)^2}{0.5 \times 0.5^2}\right) (\text{rad/s}) \quad (20)$$



**Fig. 4.** The error of the joints under HSMC and AFHSMC: (a)  $c_1$  (b)  $c_2$



**Fig. 5.** The control signal  $u$  of HSMC and AFHSMC

Figure 4 shows the values of the joints' tracking error. The proposed controller is meet the swing-up and balancing requirements when equilibrium is reached. Additionally, the AFHSMC controller reaches a stable equilibrium faster than the HSMC controller [6]. When the system is subjected to noise, the AFHSMC controller is more capable of restoring the desired equilibrium than the HSMC controller. The control signals are shown in Fig. 5. It is clear to show that the chattering phenomenon of the proposed controller is eradicated.

## 5 Conclusion

This study proposes an AFHSMC controller for the dynamics model of pendubot. The developed approach is proposed in hybrid of the HSMC method and the fuzzy method to update the optimal constant value. The pendubot's stability is guaranteed using the Lyapunov theorem. The developed approach tracks to the desired trajectory well, with a fast and stable response time when the system is under uncertain noise.

## References

1. Toan, T.V., et al.: Hybrid control for swing up and balancing pendubot system: An experimental result. ICSSE, pp. 450–453 (2017)
2. Ramírez-Neria, M., et al.: Active disturbance rejection control for reference trajectory tracking tasks in the pendubot system. IEEE Access **9**, 102663–102670 (2021)
3. Nguyen, V.T., et al.: Adaptive PD networks tracking control with full-state constraints for redundant parallel manipulators. IFSA-SCIS, pp. 1–5 (2017)
4. Nguyen, V.T., et al.: Adaptive PID tracking control based radial basic function networks for a 2-DOF parallel manipulator. ICSSE, pp. 309–312 (2017)
5. Nguyen, V.T., et al.: Adaptive chattering free neural network-based sliding mode control for trajectory tracking of redundant parallel manipulators. Asian J. Control **21**(2), 908–923 (2019)
6. Qian, D., et al.: Hierarchical sliding mode control to swing up a pendubot. ACC, pp. 5254–5259 (2007)



# Adaptive Neural Network Hierarchical Sliding-Mode Control for Pendubot Based Genetic Algorithm Optimization

Van-Truong Nguyen<sup>(✉)</sup>, Xuan-Thang Vu, and Hai-Binh Giap

Department of Mechatronics Engineering, Hanoi University of Industry, Hanoi, Vietnam  
nguyenvantruong@hau.edu.vn

**Abstract.** Based on the RBF neural networks (RBFNN) approximation and genetic algorithm (GA) optimization, a new adaptive hierarchical sliding mode control (AHSMC) is developed in this research. The proposed controller utilizes RBFNN approximation to solve the problem of uncertain nonlinear functions of the pendubot. The GA is adopted to determine unknown constants of the RBFNN. The pendubot's stability is guaranteed when the Lyapunov synthesis approach is used. Finally, the simulation results in pendubot proved that the proposed approach outperforms in comparison the hierarchical sliding mode control.

**Keywords:** Pendubot · Sliding mode control · RBFNN · Genetic algorithm

## 1 Introduction

Sliding mode control (SMC) is studied and applied for nonlinear systems, including the pendubot [1, 2]. Besides, there have been many studies on variations of the conventional SMC for improving the quality of the control system, typically hierarchical SMC (HSMC) for the pendubot system [2]. However, the sign(.) function in the control law of SMC and HSMC causes the heavy chattering phenomenon [2]. In addition, the quality of them is poor when the system is subjected to large noise or the reference signal changes with time. Moreover, calculating the equivalent control component of the SMC requires full knowledge of the nominal functions of the object. In fact, it is very difficult to do that because they contain nonlinear parameters.

In order to overcome these problems, the HSMC is combined with the intelligent controllers [3–7]. In [3], fuzzy logic control compact to HSMC is proposed to control nonlinear systems with input saturation. In [4], the adaptive HSMC (AHSMC) based on fuzzy logic control is developed to track nonlinear dynamic systems. HSMC is combined with neural network is applied to nonstrict-feedback nonlinear systems in [5]. These methods give good controlling qualities. However, these controllers have some drawbacks such as requiring expert knowledge in fuzzy controllers, or the parameters of the system are not optimized.

This paper refers to an AHSMC based on RBFNN (AHSMC-RBF) approximation and GA optimization. The RBFNN is used to approximate the pendubot nonlinear functions. By using GA, the RBFNN's parameters are optimized. In addition, the proposed

controller's sign(.) function is substituted by an equivalent function to reduce chattering. The simulation results reveal that the suggested controller is capable of tracking a changing reference signal good and rejects the chattering phenomenon.

## 2 The Proposed Methodology

The Pendubot system's state variables are as follows  $z_1 = \varphi_1 - \frac{\pi}{2}$ ,  $z_2 = \dot{\varphi}_2$ ,  $z_3 = \varphi_2$ ,  $z_4 = \dot{\varphi}_2$ . With a state vector defined as  $Z = [z_1 \ z_2 \ z_3 \ z_4]^T$ , the dynamic model of the pendubot is expressed as follows [1]:

$$\begin{cases} \dot{z}_1 = z_2 \\ \dot{z}_2 = R_1(Z) + I_1(Z)u \\ \dot{z}_3 = z_4 \\ \dot{z}_4 = R_2(Z) + I_2(Z)u \end{cases} \quad (1)$$

where:

$$u = \tau_1 \quad (2)$$

$$R_1 = \frac{1}{\beta_1\beta_2 - \beta_3^2 \cos^2(\varphi_2)} \begin{bmatrix} \beta_2\beta_3(\dot{\varphi}_1 + \dot{\varphi}_2)^2 \sin \varphi_2 + \beta_3^2\dot{\varphi}_1^2 \sin \varphi_2 \cos \varphi_2 \\ -\beta_2\beta_4g \cos \varphi_1 + \beta_3\beta_5g \cos \varphi_2 \cos(\varphi_1 + \varphi_2) \end{bmatrix} \quad (3)$$

$$I_1 = \frac{\beta_2}{\beta_1\beta_2 - \beta_3^2 \cos^2(\varphi_2)} \quad (4)$$

$$R_2 = \frac{1}{\beta_1\beta_2 - \beta_3^2 \cos^2(\varphi_2)} \begin{bmatrix} -\beta_3(\beta_2 + \beta_3 \cos \varphi_2)(\dot{\varphi}_1 + \dot{\varphi}_2)^2 \sin \varphi_2 + \beta_4g(\beta_2 + \beta_3 \cos \varphi_2) \cos \varphi_1 \\ -\beta_5g(\beta_1 + \beta_3 \cos \varphi_2) \cos(\varphi_1 + \varphi_2) - \beta_3\dot{\varphi}_1^2 \sin \varphi_2(\beta_1 + \beta_3 \cos \varphi_2) \end{bmatrix} \quad (5)$$

$$I_2 = \frac{-\beta_2 - \beta_3 \cos \varphi_2}{\beta_1\beta_2 - \beta_3^2 \cos^2(\varphi_2)} \quad (6)$$

where:  $\beta_1 = v_1d_{c1}^2 + v_2r_1^2 + J_1$ ,  $\beta_2 = v_2d_{c2}^2 + J_2$ ,  $\beta_3 = v_2r_1d_{c2}$ ,  $\beta_4 = v_1l_{c1} + v_2r_1$ ,  $\beta_5 = v_2d_{c2}$ ;  $v_1$  and  $v_2$  are the link's masses.;  $r_1$  and  $r_2$  are the link's length;  $d_{c1}$  and  $d_{c2}$  are the distances between the their center of gravity and the joints;  $J_1$  and  $J_2$  are the inertia moments of each link at their center of gravity;  $g$  is the acceleration due to gravity.

The error vector of the state variables is defined as:

$$e(t) = \begin{bmatrix} z_1 - \varphi_{1d} \\ z_3 - \varphi_{2d} \end{bmatrix} = \begin{bmatrix} e_1 \\ e_3 \end{bmatrix} \quad (7)$$

where  $\varphi_{1d}$  and  $\varphi_{2d}$  are the reference angles.

Sliding surfaces of each link are proposed as:

$$\begin{cases} s_1 = c_1e_1 + e_2 \\ s_2 = c_2e_3 + e_4 \end{cases} \quad (8)$$

with  $e_2$  and  $e_4$  being the time derivatives of  $e_1$  and  $e_3$ , respectively.

According to [2], the general sliding surface of the pendubot is as:

$$S = a_1 a_2 s_1 + a_2 s_2 \quad (9)$$

In [2], the HSMC control law for pendubot is expressed as:

$$u = \frac{-(a_1 a_2 c_1 e_2 + a_1 a_2 R_1 + a_2 c_2 e_4 + a_2 R_2 - a_1 a_2 \ddot{\varphi}_{1d} - a_2 \ddot{\varphi}_{2d} + k_2 S + \eta_2 \text{sign}(S))}{(a_1 a_2 I_1 + a_2 I_2)} \quad (10)$$

where  $k_2, \eta_2, a_1, a_2, c_1, c_2$  are the positive constants.

In a real-time application, the nonlinear functions  $R_1, R_2$  are hardly determined exactly because of external uncertainty. In order to resolve it, in this paper, nonlinear functions  $R_1, R_2$  are approximated based on RBF algorithms. The outputs of the RBF neural network are determined as follow [6, 7]:

$$\begin{cases} \hat{R}_1(Z) = \hat{W}_1^T H(Z) \\ \hat{R}_2(Z) = \hat{W}_2^T H(Z) \end{cases} \quad (11)$$

where  $\hat{W}_i^T = [\hat{w}_{i1} \hat{w}_{i2} \hat{w}_{i3} \hat{w}_{i4}]$  ( $i = 1, 2$ ) is the estimation weight matrix,  $\tilde{W}_i^T = [\tilde{w}_{i1} \tilde{w}_{i2} \tilde{w}_{i3} \tilde{w}_{i4}]$  is the error weight matrix;  $H = [h_1 \ h_2 \ h_3 \ h_4]^T$  is the hidden layer matrix,  $h_j = \exp\left[\frac{-\|Z-C_j\|^2}{2d^2}\right]$  ( $j = 1, 2, 3, 4$ ) is radial basis function;  $C_j = [c_{j1} \ c_{j2} \ c_{j3} \ c_{j4}]^T$  is the center vector of each hidden nodes;  $d = [d_1 \ d_2 \ d_3 \ d_4]^T$  is a positive width;  $\hat{R}_1, \hat{R}_2$  are the function approximated by RBFNN;  $\tilde{R}_1$  and  $\tilde{R}_2$  are the error of RBFNN, respectively.

The outputs of the RBFNN are defined as:

$$\begin{cases} R_1 = \hat{R}_1 + \tilde{R}_1 \\ R_2 = \hat{R}_2 + \tilde{R}_2 \end{cases} \quad (12)$$

The ideal weight matrix is rewritten as:

$$W_i^T = \hat{W}_i^T + \tilde{W}_i^T \quad (13)$$

The weight updating rules is chosen as:

$$\begin{cases} \dot{\hat{W}}_1 = \gamma_1 a_1 a_2 S H(Z) \\ \dot{\hat{W}}_2 = \gamma_2 a_2 S H(Z) \end{cases} \quad (14)$$

where  $\gamma_1, \gamma_2 > 0$  are learning rates.

Substituting (11) to (10), the proposed controller is defined as:

$$u = \frac{-(a_1 a_2 c_1 e_2 + a_1 a_2 \hat{R}_1 + a_2 c_2 e_4 + a_2 \hat{R}_2 - a_1 a_2 \ddot{\varphi}_{1d} - a_2 \ddot{\varphi}_{2d} + k_2 S + \eta_2 (S / (\text{abs}(S) + 0.01)))}{(a_1 a_2 I_1 + a_2 I_2)} \quad (15)$$

**Theorem:** Considering the dynamic systems of pendubot in (1), the suggested controller in (15), the updated rule in (14), we can conclude that the pendubot stability is ensured.

**Proof:** The Lyapunov function:

$$V(t) = \frac{1}{2} \left( S^2 + \sum_{i=1}^2 \tilde{W}_i^T \gamma_i^{-1} \tilde{W}_i \right) \quad (16)$$

$$\dot{V}(t) = SS + \sum_{i=1}^2 \tilde{W}_i^T \gamma_i^{-1} \dot{\tilde{W}}_i \quad (17)$$

$$\begin{aligned} \dot{V}(t) = & S \left( a_1 a_2 c_1 e_2 + a_2 c_2 e_4 + a_1 a_2 \hat{R}_1 + a_2 \hat{R}_2 - a_1 a_2 \ddot{\varphi}_{1d} - a_2 \ddot{\varphi}_{2d} + (a_1 a_2 I_1 + a_2 I_2) u \right) \\ & + S \left( a_1 a_2 \left( \tilde{W}_1^T H(Z) \right) + a_2 \left( \tilde{W}_2^T H(Z) \right) \right) - \sum_{i=1}^2 \tilde{W}_i^T \gamma_i^{-1} \dot{\tilde{W}}_i \end{aligned} \quad (18)$$

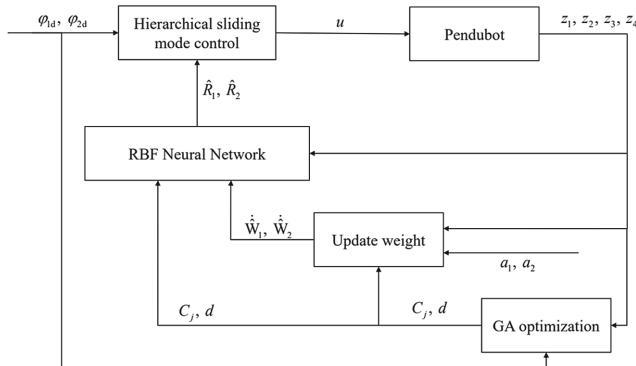
Combining (13) and (18), the result of  $\dot{V}(t)$  is obtained:

$$\dot{V}(t) = -k_2 S^2 - \eta_2 |S| + \tilde{W}_1^T \left( a_1 a_2 S H(Z) - \gamma_1^{-1} \dot{\tilde{W}}_1 \right) + \tilde{W}_2^T \left( a_2 S H(Z) - \gamma_2^{-1} \dot{\tilde{W}}_2 \right) \quad (19)$$

Replacing (14) into (19), we have:

$$\dot{V}(t) = -k_2 S^2 - \eta_2 |S| \leq 0 \quad (20)$$

From (20), the Lyapunov stability is satisfied. The overall system of AHSMC-RBF is described in Fig. 1:

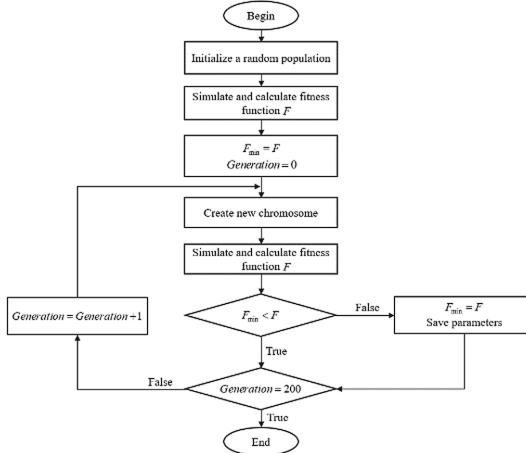


**Fig. 1.** AHSMC-RBF structure for pendubot

To be able to find the optimal parameters, the RBFNN's unknown parameters are determined using GA optimization [8]. The GA flowchart is followed Fig. 2. The fitness function is given as follows:

$$F = \sum_{i=1}^n \left[ E_1(i) \cdot E_1^T(i) + E_2(i) \cdot E_2^T(i) \right] \quad (21)$$

where  $E_1$ ,  $E_2$  are the matrixes with  $n$ -row and one column, each  $i$  element of  $E_1$  is the error between  $z_1$  and  $\varphi_{1d}$ , each  $i$  element of  $E_2$  is the error between  $z_3$  and  $\varphi_{2d}$ .



**Fig. 2.** The optimization process by GA

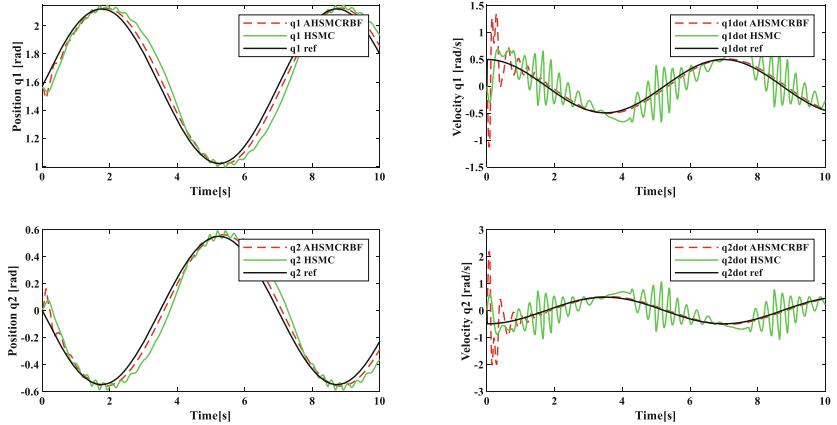
The parameters of GA optimization are chosen in Table 1.

**Table 1.** GA optimal parameters

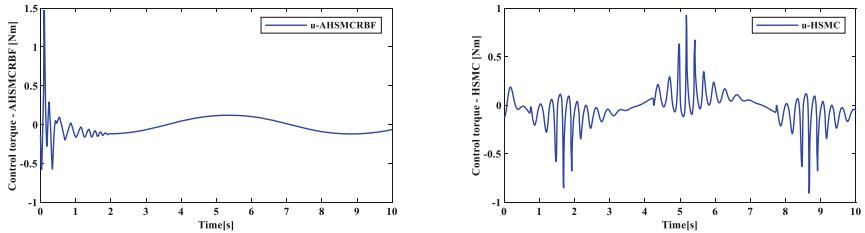
Generation	Population size	Crossover	Mutation
200	20	0.7	0.3

### 3 Simulation Results

To be able to show the efficiency performance, the HSMC [2] and the suggested controller are compared in simulation. The reference parameters of  $\varphi_1$ ,  $\varphi_2$ ,  $\dot{\varphi}_1$ ,  $\dot{\varphi}_2$  are described as  $\varphi_{1d} = \frac{\pi}{2} + 0.55 \sin(0.9t)$ ,  $\varphi_{2d} = -0.55 \sin(0.9t)$ ,  $\varphi_{1ddot} = \dot{\varphi}_{1d}$ ,  $\varphi_{2ddot} = \dot{\varphi}_{2d}$ . The HSMC control law parameters are  $a_1 = 1.740$ ,  $a_2 = 5.605$ ,  $c_1 = 9.287$ ,  $c_2 = 9.753$ ,  $k_2 = 8.195$ ,  $\eta_2 = 2.383$ ,  $I_1 = 4$ ,  $I_2 = 4$ . The initial weight matrix is  $W_1^T = W_2^T = [0 \ 0 \ 0 \ 0]$ . The parameters of RBFNN under training-based GA are:  $C_1 = [-1.5 \ -1.5 \ -1.5 \ -1.5]^T$ ,  $C_2 = [2.5 \ 2.5 \ 2.5 \ 2.5]^T$ ,  $\gamma_1 = 0.01$ ,  $\gamma_2 = 0.01$ ,  $C_3 = [6.5 \ 6.5 \ 6.5 \ 6.5]^T$ ,  $C_4 = [10.5 \ 10.5 \ 10.5 \ 10.5]^T$ ,  $D = [4 \ 4 \ 4 \ 4]^T$ .



**Fig. 3.** Responses of two links



**Fig. 4.** Control torque of AHSMC-RBF and HSMC

**Table 2.** Error of two controllers.

	RMSE_q1	RMSE_q2	MAE_q1	MAE_q2
HSMC	0.0916	0.0951	0.0733	0.0758
AHSMC-RBF	0.0412	0.0437	0.0363	0.0374

The simulation outcomes are displayed in Figs. 3 and 4. The responses of the positions and the velocities of the links are shown in Fig. 3. The RMSE and MAE comparisons are shown in Table 2. Compared with the HSMC, the results indicate that the suggested controller performs better: closer response to the reference signal, lower error-types values. The control signal is displayed in Fig. 4. The chattering phenomenon of the proposed controller has been almost eliminated by replacing the sign(.) function with the equivalent function.

## 4 Conclusion

This paper studies AHSMC based on RBFNN and GA optimization. The uncertain non-linear functions are approximated thanks to a learning neural network, and its parameters

are optimized via genetic algorithm optimization. When evaluating simulation results, it turns out that the suggested scheme outperforms HSMC. In the future, other artificial intelligence is adapted to extend to the remaining nonlinear parameters of hierarchical sliding mode control.

## References

1. Block, D.J., et al.: Mechanical design and control of the pendubot. *SAE Int. J. Commer. Veh.* **104**, 36–43 (1995)
2. Yi, J., et al.: Hierarchical sliding mode control for a class of SIMO under-actuated systems. *Control Cybern.* **37**(1), 159 (2008)
3. Zhao, X.D., et al.: Adaptive fuzzy hierarchical sliding-mode control for a class of MIMO nonlinear time-delay systems with input saturation. *IEEE Trans. Fuzzy Syst.* **25**, 1062–1077 (2017)
4. Hwang, C.L., et al.: Adaptive fuzzy hierarchical sliding-mode control for the trajectory tracking of uncertain underactuated nonlinear dynamic systems. *IEEE Trans. Fuzzy Syst.* **22**, 286–299 (2014)
5. Zhao, X.D., et al.: Adaptive neural hierarchical sliding mode control of nonstrict-feedback nonlinear systems and an application to electronic circuits. *IEEE Trans. Syst. Man Cybern. Syst.* **47**, 1394–1404 (2017)
6. Nguyen, V.T., et al.: Adaptive PD networks tracking control with full-state constraints for redundant parallel manipulators. *IFSA-SCIS*, pp. 1–5 (2017)
7. Nguyen, V.T., et al.: Adaptive PID tracking control based radial basic function networks for a 2-DOF parallel manipulator. *ICSSE*, pp. 309–312 (2017)
8. Zhou, Y.S., Lai, L.Y.: Optimal design for fuzzy controllers by genetic algorithms. *IACET*, pp. 429–435 (1995)



# Classification of Malicious Websites Using Feature Based Machine Learning Techniques

Swetha Mahendrakar ShyamRao<sup>1(✉)</sup>, Rakshit Danti<sup>1</sup>,  
Praveen Narayanappa Manjunath<sup>1</sup>, and Muneshwara Mahendrakar ShyamRao<sup>2</sup>

<sup>1</sup> Department of Information Science and Engineering, BMS Institute of Technology and Management, Bengaluru, India  
`{swethams_ise2014,1by19is121}@bmsit.in`

<sup>2</sup> Department of Computer Science and Engineering, BMS Institute of Technology and Management, Bengaluru, India  
`muneshwarams@bmsit.in`

**Abstract.** The start of the COVID-19 pandemic has revealed a huge increase in malware sites. The rising instances of phishing, spamming and malware have created a key requirement for a robust mechanism which can precisely detect and classify malicious URLs. Conventional classification methods such as blacklisting and signature-based detection are not adaptive due to enormous volume of data, innovations and technology changing over time, along with complicated features. In this work, we propose a feature-based machine learning system for the malicious URLs identification as a two-class problem and assess the performance of ML classifiers namely Random Forest classifier and Support Vector Machine. We have utilized a Kaggle dataset consisting of 72000 URLs for training our ML model. This research highlights various view-points related to the URL classification process which classifies the objective site as malicious or benign. Proposed method has wide applications in cyber security applications as well as cyber forensic investigations.

**Keywords:** Malicious URL · Classifier · Artificial intelligence (AI)

## 1 Introduction

Over the past several years, the amount and variety of online resources, such as social networking, e-Commerce websites, and video sharing pages, have increased tremendously. New means for the people to connect as well as new possibilities for other criminals, are generated as a result of these activities. The fact that Google segregates nearly 10,000 websites which look suspicious daily and list them on a “Google blacklist” could be a clear indicator and an obvious sign that hackers are exploiting these gaps. Almost every second computer in China is affected by some type of malware. China has the largest malware infection rate of 47%, then followed Turkey reported 42%, Taiwan reported with 39% and the United States with 30% [6]. Computer viruses’ statistics reports that malware attacks in 2020 have been recorded as costing a normal US company an average of \$2.4 million per year [2, 9].

The popular web page login portal is imitated by attackers to cover users and it totally appears as a genuine webpage. Once an unsuspected user visits the link, the script running behind it extracts data and makes use of it by the attacker. Malware steals personal, financial, sensitive, business, and commercial information for a variety of benefits. Malwares are used to attack companies or government organizations to collect confidential information to disrupt their work. The consequences of such web attacks are data breach, increase of computation time, reduced computer performance, system crash, replay of automatic messages or unsolicited emails without the knowledge of the owner, reducing memory space, Phishing, theft of user accounts, and financial losses by stealing credit card details. In customary methods like blacklisting technique makes it inadmissible for recognizing new malicious websites or obscure malicious patterns [1, 4].

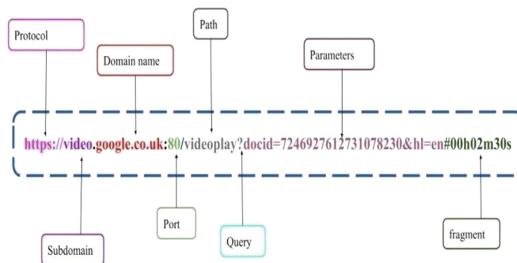
A few techniques have been created utilizing boycotting and rule-based ways to deal with phishing assaults, for example, Safe Browsing by Google, Netcraft Anti-phishing Toolbar, McAfee Site Advisor, Spoof Guard are used to tackle cyber threats but still victims are being deceived and credentials are being taken. However, there is no substantial answer for preventing such digital assaults. Identifying malicious URLs at the right time is a very great concern and order of the day. We made an attempt to develop a system that uses two-layer analysis to detect malicious URLs on the internet. The system automatically learns itself. Thinking about this, to identify phishing dangers, a feature based machine learning approach for classification of URL is proposed in this paper. From the test investigation, it is observed that the Random Forest classifier beats the other classifier models as far as accuracy and precision are concerned. The accuracy of URL classification with Random Forest classifier was seen to be 96.3%.

## 2 Literature Survey

Malware website detection has been thoroughly researched in recent years. Brij B Gupta et al. [3] proposed an extraction mechanism which acquires diverse lexical based data from URL. Lexical data obtained is then divided into separate elements, and the occurrence of an URL is arranged which contains every data important to group the URL as authentic or phishing. Faris Setiadi et al. [7] suggested a rule based approach with an intention of improving the model accuracy and response time. Several ML algorithms have been compared to see precision and response time. In this paper the authors made exhaustive comparative analysis to come up with an effective approach for identification of malicious URL's. Malicious URL Detection using Deep Learning is proposed in [11, 15] in which raw URLs are encoded using characters level embedding. It is a cutting-edge technique in NLP to address the characters in numeric organization. In [10] the author describes how to analyze and classify Malicious web sites by combining a static analysis algorithm and a runtime monitoring mechanism to extract rich features of JavaScript code.

In [14] the author highlights supervised machine learning to hinder phishing assaults. The performance is assessed with the dataset which comprises dynamic phishing assaults

and contrast it with Google Safe Browsing (GSB) which provide security in most well-known internet browsers. Riton Patgiri et al. [12] applied the convolutional neural network model to recognize the malignant URLs. The CNN model is prepared and experimented with malware webpages. This methodology reduces the load on computational gadgets. In [13] the author describes Fuzzy logic association rules which groups the authentic and phishing websites based upon the features that are changed to membership values as *Low, Medium & High*. Further the association rule are applied to classify the URLs using apriori algorithm to generate binary matrix. URL i.e., Uniform Resource Locator used to access a website. It recognizes the location on the computer network and assists in getting into the webpage. Generally URL consists of seven components such as *Protocol, Domain name, Subdomain, Parameter, Path, Top-level—domain and Query* as shown in Fig. 1.



**Fig. 1.** Sample URL

We imagine that we can stay away from the malware from the sites contaminating our PC by keeping away from choosing the URL or promotions or by declining to download the product from the malware sites. Still there are lot more ways for our PC cyber threats:

- **Driven by download:** Without our knowledge whenever we visit some unsafe URL's malware is injected into our gadget and activated.
- **Infection by Java Script:** Since the majority of the websites are developed by JS. The programmers make a malware which contaminates gadget by downloading the JS which executes naturally & begins its measure by installing the malware.
- **Malvertising:** It incorporates infusing malware-loaded advertisements into genuine web-based advertising organizations and websites.
- **Malicious Redirect:** Malicious redirect is a code embedded into a site with the purpose of diverting the webpage guest into another site.
- **Browser Hijackers:** It is a type of undesirable programming that alters an internet browser's settings without a client's authorization.

### 3 Proposed Methodology

In this paper a malware identification framework is proposed to assist the users to envision the credibility and threat of a URL. We made an attempt to develop a system which

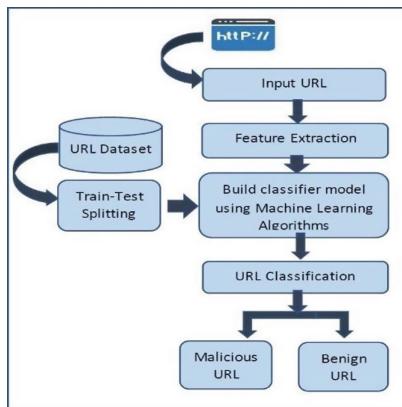
extracts the component vectors or potential features from the given URL whenever a user visits. The component vectors are pre-processed and input into a ML model which uses Random Forest classifier and SVM Classifier to approve the authenticity of a URL. The proposed system addresses the following issues as shown in Fig. 2.

**Reliability:** We expect to assemble a phishing recognition approach which produces the least number of false positives.

**Real-time:** The malicious URL must be blocked by a trustworthy phishing recognition method without giving any form of data to the hackers.

**Less Reaction Time:** We target assembling an enemy of phishing that can identify malicious URLs as right on time. The less response time will offer the hackers generally very less chance to take the credentials.

**Identification of Newly Established Phishing Sites:** We target developing a phishing identification system that outperforms the existing methodology for malware identification and can distinguish new malicious websites in the future.



**Fig. 2.** Proposed system architecture

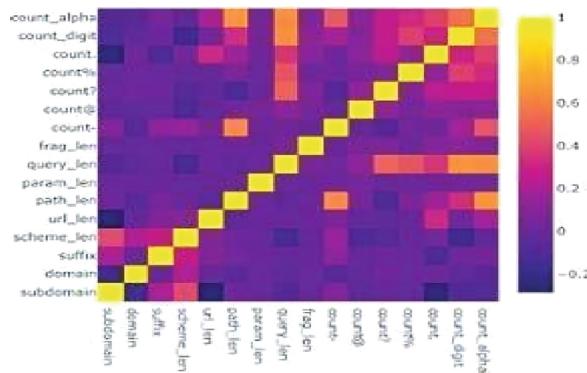
To design a malicious URL detection system, we have applied supervised machine learning models namely *Support Vector Machine* and *Random Forest*. The ML model is trained using Python 3 and the SciKit-Learn library.

### 3.1 Dataset Collection and Pre Processing

We have gathered a publicly available dataset of Kaggle consisting of 72000 URLs to build the model [5]. Out of which 40000 URLs were benign and 32000 URLs were malicious. We have split our dataset in 80:20 mix i.e., 80% of the dataset used for training and 20% of data is used for testing. The data collected is normalized and transformed to a matrix for further processes.

### 3.2 Feature Extraction

It includes the process of selecting key features that are crucial for training the model and classification. Scatter matrices and Correlation were used to analyses the strength of the relation between the features, with the purpose of extracting only the most significant ones as shown in Fig. 3.



**Fig. 3.** Correlation matrix

The potential features used in this approach are:

**Lexical Features:** To separate the lexical components we will slash the URL into different fragments. Lexical components incorporate the length of URL, underscores, No. of semicolons, equals to, question marks, zeroes, ampersands, No. of queries, parameters, Digit-to-letter ratio etc. Based on these components a dictionary is built with the presumption that the combination of these features is distinct for malicious and benign URLs.

**Host-based Components:** They are gathered through the host-name features of the URL. They give us information about the area, character, and the administration style, properties of vindictive hosts. URL lifetime is perhaps the main element since malware websites have significantly less lifetime than the legitimate web pages. Through host based components it is feasible to get data about when and who visited the domain along with IP address. Because of the trouble of getting new IP addresses these features are crucial to classify the URLs.

**Content Features:** Content features are collected by downloading the contents of the webpage. This feature helps to accomplish a higher precision rate. The HTML code of the page can be obtained to find out the total word count, average words per line, sensitive words, connections to distant contents and imperceptible objects. Usually, malware code is scrambled in HTML. JavaScript which are utilized by hackers to encode malware code which executes without consent. We examine the use of functions such as “exec ()”, “link ()”, “eval ()”, “escape ()”, “search ()” because they are used to spread malware.

**Link Popularity:** It is estimated depending on the approaching connections from different site pages, the quantity of popups and the nature of plugins. The occasions that the URL is visited can be an indicator, in light of the fact that malevolent URLs are less accessed; subsequently they have lesser link popularity. Summary of the above features are given in Table 1.

**Table 1.** Potential features used

URL component	Type of feature used
URL	Length of the URL
	Underscores, No. of semicolons, zeroes, Question marks, equals, delimiters, ampersands, dots
	Digit to letter proportion
	URL popularity
Query	Number of queries
Parameters	Number of parameters
	Length of the parameter
Protocol	Protocol used
Domain	The domain age since it is registered
	No of subdomain
	Domain name properties
Host	IP address properties
	Location
Path	Path length
	Number of sub directories

Any users enter a new URL, all the above features are extracted and check a match with the set of malicious URLs features. If a match is found then the ML model will hinder the user from using that URL. Else, the features extracted from the new URL can be utilized for training and testing purposes. Brief description of lexical features is given below.

**URL Length:** Normally, phishing URLs are lengthy. Attackers include multiple domain names, longer paths for noxious activity which make the URL extended. Number of dots, zeroes, semicolons which are comparatively more in numbers in the phishing websites to misguide the legitimacy of the website. Consider the example, [www.flipkart.com](http://www.flipkart.com) is represented as [www.flipk.art.com](http://www.flipk.art.com). For this situation, many clients don't focus on the URL and become a casualty to the assailants.

**No. of Delimiter:** The presence of delimiter such as / or + symbol in URL like www.flipkart.com, www.flip+kart.com can make a phishing URL resemble a legitimate URL. Analysis indicated that there are generally more delimiters in the phishing domain of URLs than legitimate URLs.

**Number of Subdomains:** Many benign URLs use only a single subdomain, like, “www”. Conversely, phishing URLs use multiple subdomains for social engineering purposes with an intention of abusing a recognized brand.

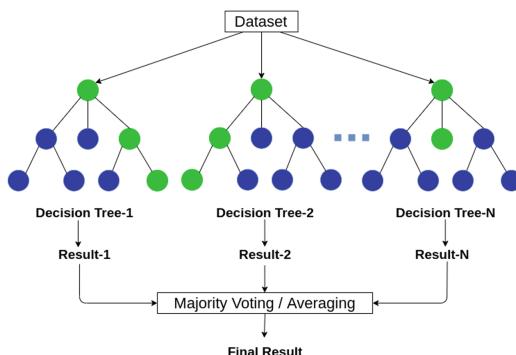
**Domain Length:** A legitimate URL normally contains short domains where we determine the domain length of every URL and store it as a feature.

**Host-Based Features:** Host based elements reveal “where” vindictive destinations are facilitated, “what their identity is” and “how” they are directed. These elements are utilized as malevolent sites facilitated in illegitimate hosting centers.

### 3.3 Model Training and Testing

The proposed model is implemented using two level classifiers namely Random Forest classifier and SVM Classifier. The system performance of the ML model which is acceptable (high classification accuracy), will be utilized in the detection stage.

**Random Forest:** It belongs to a category of ensemble of classifiers which fits various decision trees on various subsets of the data. The final approach is based on the mean value of various decision trees trained. Normally, it performs better compared to decision tree and it even deals with overfitting issue moreover. Random forest algorithm is easy to implement quick in activity and have demonstrated to be successful in variety of domains. Figure 4 show the Random Forest model.



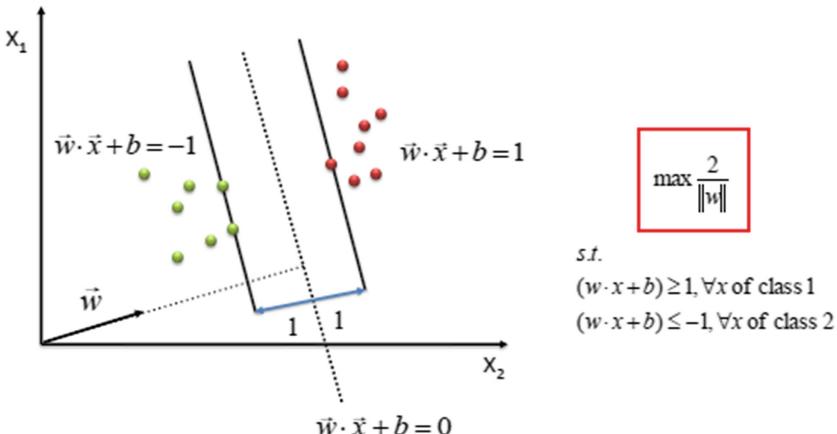
**Fig. 4.** Random forest classifier model

Random Forest model is constructed using the *Mean Squared Error (MSE)* to identify data branches from every node. Equation (1) determines the distance of every node from the predicted actual value, to determine which branch is the best decision for forest.

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (f_i - y_i)^2 \quad (1)$$

where  $N$ : no of nodes,  $f_i$ : Actual distance of node  $i$ ,  $y_i$ : Predicted distance of node  $i$ .

**Support Vector Machine:** SVMs are directed learning calculations to predict a hyperplane that classifies the information with the most extreme edge. New data is mapped by the side of the hyperplane they lie on. They are robust and extremely efficient even in situations where the quantity of training dataset is little. SVM method additionally shows more prominent capacity to generalize and high probability of producing efficient classifiers. In case of gigantic data, it is memory capable which stands out from various models since it trains on a subset of data. Figure 5 shows the model.



**Fig. 5.** SVM classifier model

Hyperplanes are represented by the Eq. (2 & 3).

$$\overrightarrow{w} \cdot \overrightarrow{x} + b = -1 \quad (2)$$

$$\overrightarrow{w} \cdot \overrightarrow{x} + b = 1 \quad (3)$$

where  $x$ : Point in the plane,  $w$ : weight factor,  $b$ : bias

Given data can be discriminated linearly by selecting 2 parallel hyperplanes that distinguishes the two classes of data with a maximum perpendicular distance. The area between these two hyperplanes is known as margin and the maximum margin hyperplane lies in the middle of them. In Fig. 5, the above right-side area of the hyper plane shows the malicious URLs and the left side area of hyper plane depicts the benign URLs.

## 4 Experimental Analysis

In this paper, we look at the assessment of measurements of all features together to choose the best outcomes and to gauge the rightness of the output. The performance measurements that we used for assessing the models are, *accuracy*, *confusion matrix*, *precision*, *F1-Score* and *recall*. We observed that the Random Forest classifier gave an accuracy of 96.28% and it outperforms other classifier methods.

### 4.1 Comparative Analysis

We have compared our experimental results with other malicious URL detection models and observed that our proposed model is comparable with other models due to its potential features and also it outperforms the other models with an accuracy of 96.3% as shown in Table 2.

**Table 2.** Comparison of proposed model with other models

Paper	Approach	Accuracy
Brij B Gupta, et. al. [3]	Logo image used to find the identity of the web page by comparing real and fake webpages	93.40%
Humam, Faris Setiadi [7]	Phishing URLs classification based on hybrid SVM and KNN approach	90.04%
R. Vinayakumar, et. al. [11]	Rule based phishing web page detection methods	86.6%
Proposed Model	Malicious website classification using Random Forest	96.3%

## 5 Conclusion

In this paper, we gave an outline of phishing assaults and a couple of existing approaches to manage malware assaults. We further depicted the constraints of the procedures and proposed our methodology in malicious URL detection. Though literature reveals enough work in this area, the proposed approach delivered a precision of 96.28% in malevolent URL recognition. The results obtained by our proposed method are compared with other popular methods with different estimations, correlations as given in Table 3. Malicious URL detection plays a major role for several cyber-security applications as well as cyber forensic investigations and clearly, we suggest that the proposed approach is certainly a promising tool for malicious URL detection in the field of cyber security.

**Table 3.** Performance metrics in percentage

Algorithm	Accuracy	Precision	Recall	F1 score
Random Forest Classifier	96.28	91.44	94.42	92.91
SVM Classifier	91.70	92.43	89.45	90.92

## References

1. Abuzuraiq, A., Alkasassbeh, M., Almseidin, M.: Intelligent methods for accurately detecting phishing websites. In: 11th International Conference on Information and Communication Systems (ICICS) (2020)
2. Eshete, B., Villafiorita, A., Woldemariam, K.: Binspect: Holistic analysis and detection of malicious web pages. in Security and Privacy in Comm. Networks. Springer, pp.149–166 (2013)
3. Gupta, B.B., Yadav, K., Razzak, I., Castiglione, A., Chang, X.: A novel approach for phishing URLs detection using lexical based machine learning in a real- time environment. Computer Communications Vol. 175, (2021)
4. Canfora, E. Medvet, F. Mercaldo, Visaggio, C.A.: Detection of malicious web pages using system calls sequences. In: Availability, Reliability, and Security in Information Systems. Springer, pp. 226–238 (2014)
5. Dataset : <https://www.kaggle.com/datasets>
6. <https://dataprot.net/statistics/malware-statistics>
7. Faris, H., Setiadi: Phishing web Page detection methods: URL and HTML features detection. In: IEEE International Conference on IOT and Intelligence System (IoTaIS) (2020)
8. Ma, J., Saul, L.K., Savage, S., Voelker, G.M.: Identifying suspicious urls: An application of large-scale online learning. In: 26th International Conference on Machine Learning. ACM, pp. 681–688 (2009)
9. Muneshwara, M.S., Swetha, M.S., Thungamani, M., Anil, G.N.: Digital genomics to build a smart franchise in real time applications. In: IEEE International Conference on Circuit, Power and Computing Technologies (ICCPCT), IEEE page no 1–4 (2017)
10. Hiremath, P.N.: A analyzing and detecting malicious JavaScript code in webpages. (2021)
11. Vinayakumar, R., Sriram, S., Soman, K.P., Alazab, M.: Malicious URL detection using Deep Learning. Computer Science IEEE (2020)
12. Patgiri, R., Biswas, A. and Nayak, S.: Deep BF: Malicious URL detection using learned bloom filter and evolutionary DeepLearning. Transactions on Cybernetics (2021)
13. Nivedha, S., Gokulan, S., Karthik, C., Gopinath, R., Gowshik, R.: Improving phishing URL detection using fuzzy association mining. Int. Jr. of Eng. Sci. 6 (2019)
14. Sinha, S., Bailey, M., Jahanian, F.: Shades of grey: On the effectiveness of reputation-based blacklists in malicious & unwanted software. In: 3rd International Conference on IEEE, pp. 57–64 (2008)
15. Sarraf, G., Swetha, M.S.: Intrusion prediction and detection with deep sequence modeling. In: Thampi, S.M., Martinez Perez, G., Ko, R., Rawat, D.B. (eds.) SSCC 2019. CCIS, vol. 1208, pp. 11–25. Springer, Singapore (2020). [https://doi.org/10.1007/978-981-15-4825-3\\_2](https://doi.org/10.1007/978-981-15-4825-3_2)



# A Design Model Network for Intelligent Web Cache Replacement in Web Proxy Caching

Truong Nguyen Xuan<sup>1</sup>(✉), Van Tran Thi<sup>2</sup>, and Lam Ho Khanh<sup>1</sup>

<sup>1</sup> Hung Yen University of Technology and Education, Hung Yen, Vietnam  
truongnx@utehy.edu.vn

<sup>2</sup> University of Labour and Social Affairs, Cau Giay, Hanoi, Vietnam

**Abstract.** Internet web caching architecture and web caching replacement policy have been used by internet service providers (ISP) and internet content providers (ICP) to save bandwidth, offload in servers and reduce time response to the user. Thereby, the internet services have improved the performance and quality. In recent years, machine learning (ML) is increasingly developing both in hardware as well as in algorithms, which could impact web caching replacement to become more intelligent. In this paper, we present a model network and apply two algorithms: the original method and an intelligent method based on the decision tree in ML of web caching replacement. The simulation results are shown in JMT software and compare the performance of the two algorithms.

**Keywords:** Web Proxy Caching · Petri net · Queueing network

## 1 Introduction

On the internet, the proxy servers of ISP like the access network, institutional network, regional network, national backbone network, and server [1, 2] and the proxy caches like access cache (AC), institutional cache (IC), regional cache (RC), central cache (CC), and origin cache (OC) are a pair to offload for the webserver and reduce the response time in the request HTTP from the client by using the web cache replacement policy and its algorithms [3]. The parameters as the hit rate (HR), byte hit rate (BHR), delay saving rate (DSR), proxy server utilization (PSU), and average response time for HTTP from client E[R] need to be computed to show the efficiently of each algorithm.

Normally, the least recently used (LRU) and least frequently used (LFU) have well-known original methods of web caching replacement policies, which could be improved the performance of network systems based on policies of the size, function, randomized [4, 5]. Although these methods get high value in HR, BHR, and DSR, they could not predict the behavior of the user in accessing the web. Therefore, the server could not select the best web cache replacement policy for the proxy cache server, and the value HR, BHR, and DSR will reduce in these cases.

To overcome the original method, machine learning (ML) was integrated into the web cache replacement policy, which are fuzzy logic, neural networks, support vector machine, naive Bayes, and decision tree. These techniques could predict the behavior

of both the user operators and the requested web object. With any algorithm in ML, the data is an important part of the system more intelligent. In this case, web data mining is considered based on a weblog file, which is created in the HTTP session of the web browser, proxy server, and application server [6]. It contains three steps: Data preprocessing (Data cleaning, user and session identification, data filtering, and data formatting), sample data (classification, clustering, association rules, sequential pattern, path analysis, association rules, sequential patterns, clusters, and classification rule), and sample data analysis (SQL query, Online analytical processing) [7, 8].

Based on data in the weblog file, Ali [9] presented the adaptive neuro-fuzzy inference system (ANFIS) to predict the accessed object. But, this method was selected the time feature, delay, and size of data by the manual. Romano [10], Cobb [11], Tian et al. [11, 12] used the back-propagation neural network (BPNN) to predict and get a decision in replacement object. Herein, neural network proxy cache replacement (NNPCR) used the feature of size object, frequency, and recency access for the training data. Some researchers [4, 13, 14] used other features as HR, BHR, time response, size of the script, CPU time, and bandwidth for the BPNN to get the decision of LRU.

For the advanced method, Koskela et al. [15] used a multilayer perceptron network classifier (MLP) which predicts object web classifier. Especially, training data could select the feature as the response of server in HTTP protocol and structure of HTML. To classification web objects cached in multilevel classes, Sajeev and Sebastian [16] used the regression based on the multinomial logistic regression (MLR) which focuses on the feature of training data as the delay, object popularity, size object, recent access, classify objects, consistency of popularity. While Benadit et al. [17] used expectation-maximization Naive Bayes classifier (EMNB) based on recency, frequency, size, sliding window length frequently (SLW-frequently), and estimate the request next time which could improve the performance of web cache due to combining with LRU and greedy dual size frequently (GDSF).

Besides, these above methods bring high performance, the decision tree (DT) method also has a good way because it predicts web objects with high priority access. Some algorithms using DT named DT-SIZE [18] and DT-Hybrid [19] have HR and BHR higher than NB, SVM, hay BPNN và ANFIS, and reduce time response. This paper will design a model network that contains the original method (HYBRID) and an intelligent method based on DT (DT-HYBRID). These models will be simulated on java modeling tools (JMT) software. and Time response, system throughput, and the number of customers in term terminals-Browsers, web proxy, CacheMiss\_and\_Full, CacheHit, LFU\_NoSatisfy, WebTrue, and CacheFree are calculated and taken for comparing to show the performance of internet web caching architecture.

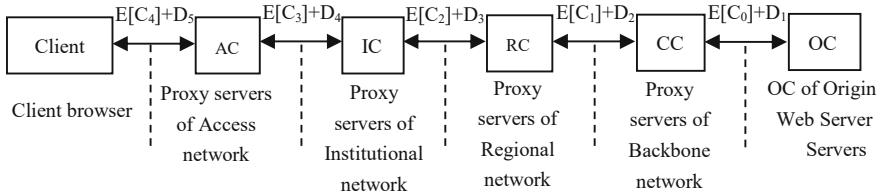
The paper is organized as follows: Firstly, the modeling network design in internet web caching architecture is presented in detail. Secondly, two algorithms named HYBRID and DT-HYBRID are applied in JMT software to evaluate the parameter effect on the network performance. Then, all results are presented and discussed. Finally, the conclusions are given in Sect. 4.

## 2 Model Network Designing in Internet Web Caching Architecture

Based on internet web caching architecture [1], a model network is designed as in Table 1 and Fig. 1 using the proxy server and its caches are in ISP and the client in LAN.

**Table 1.** The parameter of the client

Parameters	Values
Client level	5
AC	4
IC	3
RC	2
CC	1
OC	0



**Fig. 1.** Modeling design in internet web caching architecture

Consider the ISP has  $n$  network-level. The average of time response between the proxy cache server and proxy caches level  $i$  for HTTP request [1] in the client can be computed as

$$E[R_i] = (D_{nM} + D_{nREQ}) + (D_{n-1M} + D_{n-1REQ}) + \dots + (D_{i+1M} + D_{i+1REQ}) + E[C_i] + D_{i+1} + \dots + D_{n-1} + D_n \quad (1)$$

where  $D_{nM}$  is the delay of level  $n$ ;  $D_{nREQ}$  is the delay from level  $n$  to level  $n - 1$ ;  $D_i$  is the delay from level  $n$  to level  $i$ , and it depends on the web object size;  $E[C_i]$  is the web object time response to the cache of level  $i$ .

Average of time response for HTTP request of web caching and web in AC( $D_{ACH}$ ), IC( $D_{ICH}$ ), RC( $D_{RCH}$ ), CC( $D_{CCH}$ ), and OC( $D_{OCH}$ ) of LAN configuration can be expressed as:

$$D_i = E[C_i] + D_{i+1} \quad (2)$$

The HR and BHR in the ISP network for the proxy cache servers can be computed as

$$HR_{AC} = \frac{\sum_{i \in O} h_{iAC}}{\sum_{i \in O} n_i}; HR_{IC} = \frac{\sum_{i \in O} h_{iIC}}{\sum_{i \in O} n_i}; HR_{RC} = \frac{\sum_{i \in O} h_{iRC}}{\sum_{i \in O} n_i}; HR_{CC} = \frac{\sum_{i \in O} h_{iCC}}{\sum_{i \in O} n_i} \quad (3)$$

$$BHR_{AC} = \frac{\sum_{i \in O} s_i \cdot h_{iAC}}{\sum_{i \in O} s_i \cdot n_i}; BHR_{IC} = \frac{\sum_{i \in O} s_i \cdot h_{iIC}}{\sum_{i \in O} s_i \cdot n_i}; BHR_{RC} = \frac{\sum_{i \in O} s_i \cdot h_{iRC}}{\sum_{i \in O} s_i \cdot n_i}; BHR_{CC} = \frac{\sum_{i \in O} s_i \cdot h_{iCC}}{\sum_{i \in O} s_i \cdot n_i} \quad (4)$$

where,  $s_i$  is the size of web object  $i$  (byte);  $h_{iAC}$ ,  $h_{iIC}$ ,  $h_{iRC}$ ,  $h_{iCC}$  are number HTTP of correct web object  $i$  web cache to AC, IC, RC, CC;  $n_i$  is total request to web object  $i$ ,  $O$  is collective the accessed web object.

The average response of internet web caching in ISP ( $D_{IWC}$ ) and its based on the size of web object ( $D_{IWCS}$ ) are computed as

$$D_{IWC} = D_{ACH} + (1 - HR_{AC})(D_{ICH} + (1 - HR_{IC})(D_{RCH} + (1 - HR_{RC})(D_{CCH} + (1 - HR_{CC}))(D_{WSH}))) \quad (5)$$

$$D_{IWCS} = D_{ACH} + (1 - BHR_{AC})(D_{ICH} + (1 - BHR_{IC})(D_{RCH} + (1 - BHR_{RC})(D_{CCH} + (1 - BHR_{CC}))(D_{WSH}))) \quad (6)$$

The ratio for decrementing delay between the correct web object  $i$  web cache and the missing web object  $i$  web cache is defined as

$$DSR_{AC} = \frac{\sum_{i \in O} D_{ACH} \cdot h_{iAC}}{\sum_{i \in O} D_{WSH} \cdot n_{iWSH}}; DSR_{IC} = \frac{\sum_{i \in O} D_{ICH} \cdot h_{iIC}}{\sum_{i \in O} D_{WSH} \cdot n_{iWSH}}; DSR_{RC} = \frac{\sum_{i \in O} D_{RCH} \cdot h_{iRC}}{\sum_{i \in O} D_{WSH} \cdot n_{iWSH}}; DSR_{CC} = \frac{\sum_{i \in O} D_{CCH} \cdot h_{iCC}}{\sum_{i \in O} D_{WSH} \cdot n_{iWSH}} \quad (7)$$

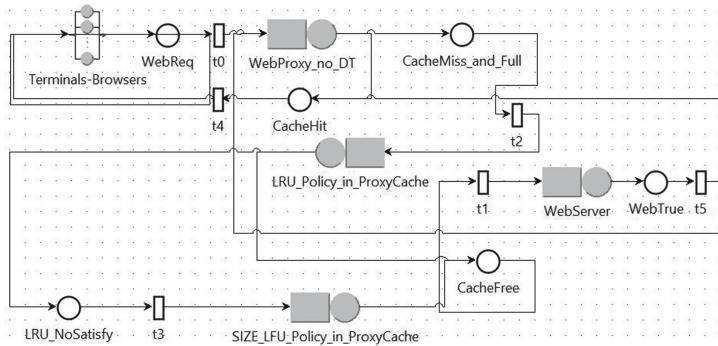
Herein, the web caching replacement policy in proxy servers depends on HR, BHR, and DSR parameters, and affects the average time response of internet web caching in ISP. For clearly, a model network will design and apply the algorithm to calculate and measure the parameters that affect the performance of the network system in Sect. 3.

### 3 Simulations and Results

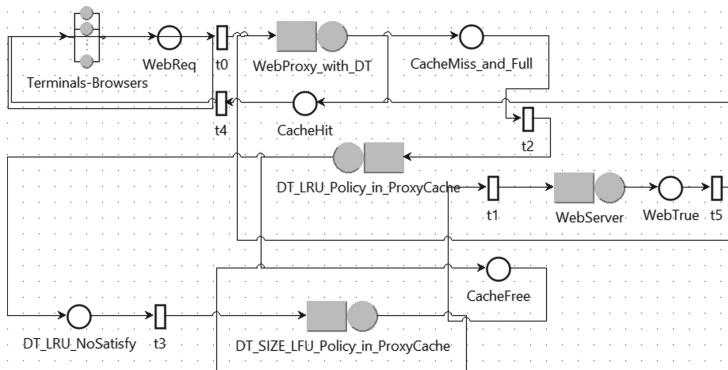
In this section, we selected two algorithms named HYBRID (using the original method based on LRU and LFU) and DT-HYBRID (using decision tree based on LRU and LFU) of the web caching replacement policy to demonstrate our computation in Sect. 2. These algorithms provide high HR values [20]. To show the value as time response, system throughput, and the number of customers in term terminals-Browsers, web proxy, Cache-Miss\_and\_Full, CacheHit, LFU\_NoSatisfy, WebTrue, and CacheFree in the network system, the JMT [21] is used to design two models (HYBRID-M and DT- HYBRID-M) as shown in Figs. 2 and 3.

Assume that all proxy caches levels of ISP are a web proxy that is attached to the web caching replacement policy. The network models contain three main queues as Terminals-Browsers, WebProxy, and WebServer connected to WebTrue, WebReq, CacheMiss\_and\_Full, CacheHit to control data flow in the network. The difference point between HYBRID-M and DT-HYBRID-M is the queue with a different parameter. The cache\_miss in DT-HYBRID-M is defined lower than HYBRID-M as Table 3 [22]. All simulations use the parameters in Table 2 and are performed within 20 s.

Using the above models and configuration parameters, the simulation in JMT is performed and measured the number of customers, response time, and system throughput to evaluate the performance of the system. All results are shown in Figs. 4, 5, 6, 7 and 8 and Table. 3.



**Fig. 2.** HYBRID model based on PetriNet and queueing network of web caching



**Fig. 3.** DT\_HYBRID model based on PetriNet and Queueing Network of web caching

Using the below models and configuration parameters, the simulation in JMT is performed and measured the number of customers, response time, and system throughput to evaluate the efficiency of the system. All results are shown in Figs. 4, 5, 6, 7 and 8 and Table 3.

**Table 2.** The nodes definition in Figs. 2 and 3

Queue, place, transition	Meaning	Time function $f(t) = \lambda e^{-\lambda t}$
Terminal-Browsers (unlimited queue M/M/4)	Number clients web requests N = 10	Mean = 10000, $\lambda = 0.0001$
$t_0$ (timed transition)	Clients send the request to the Proxy	Mean = 5, $\lambda = 0.2$

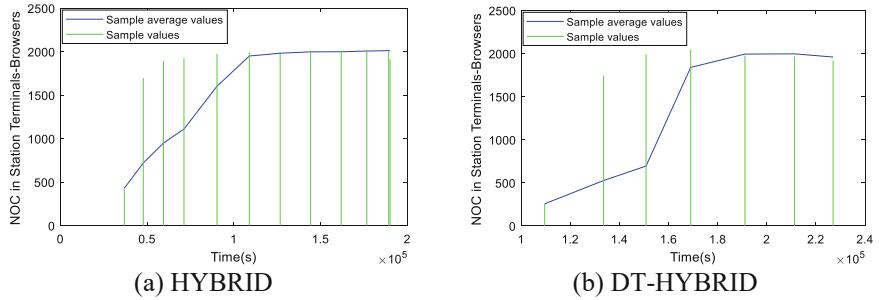
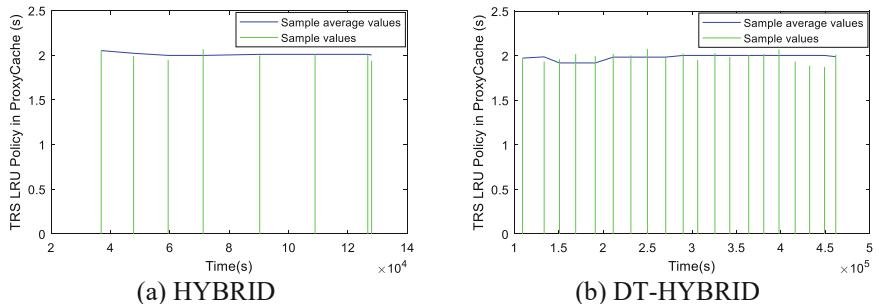
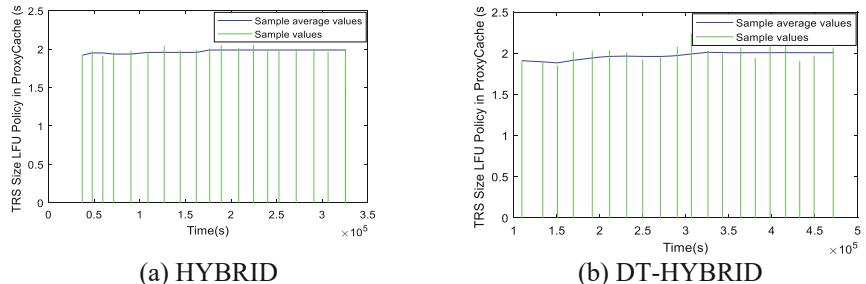
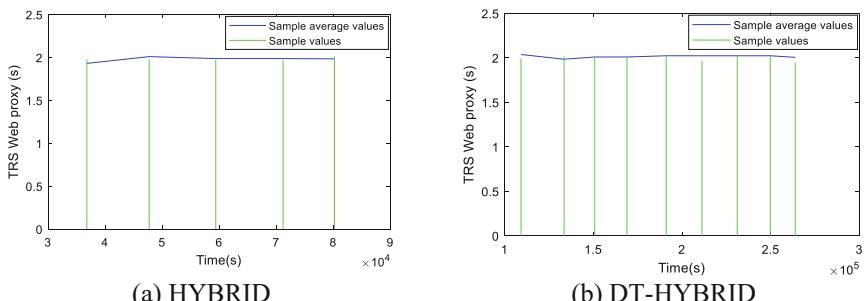
(continued)

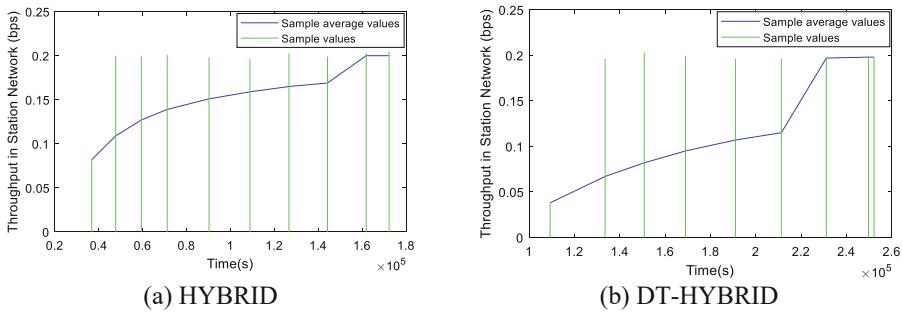
**Table 2.** (*continued*)

Queue, place, transition	Meaning	Time function $f(t) = \lambda e^{-\lambda t}$
t1 (time transition)	The proxy sends the request to webserver	Mean = 5, $\lambda = 0.2$
t2 (timed transition)	Start at policy LRU/DT-LRU in Proxy	Mean = 5, $\lambda = 0.2$
t3 (timed transition)	Start at policy SIZE-LFU/DT-SIZE-LFU	Mean = 5, $\lambda = 0.2$
t4 (timed transition)	Web Proxy server return to web contents	Mean = 5, $\lambda = 0.2$
t5 (timed transition)	Return web contents to Proxy	Mean = 5, $\lambda = 0.2$
WebServer (unlimited queue M/M/8)	Web Server receives the request	Number services: 8, Mean = 2, $\lambda = 0.5$
WebProxy_no_DT/ WebProxy_with_DT (unlimit queue M/M/4) Routing Probabilities	Proxy Server with HYBRID/DT-HYBRID	Number services: 4, Mean = 2, $\lambda = 0.5$
LRU_Policy_in_ProxyCache/ DT_LRU_Policy_in_ProxyCache (place) Routing Probabilities	Using policy LRU/DT-LRU in Proxy	Number services: 4, Mean = 2, $\lambda = 0.5$
SIZE_LFU_Policy_in_ProxyCache/ DT_SIZE_LFU_Polilcy_in_ProxyCache (unlimitqueue) Routing Probability	Using policy SIZE-LFU/DT-SIZE-LFU to: CacheFree: 1.0	Number services: 4, Mean = 2, $\lambda = 0.5$
LRU_NoSatisfy/DT_LRU_NoSatisfy (place)	No satisfy policy LRU/DT-LRU	

**Table 3.** Parameters for HYBRID-M and DT-HYBRID-M

Queue, place, transition	HYBRID	DT-HYBRID
WebProxy	Proxy Server to: CacheMiss_and_Full: 0.6 to: CacheHit: 0.4	Proxy Server to: CacheMiss_and_Full: 0.3 to: CacheHit: 0.7
Policy_in_ProxyCache	Using policy LRU in ProxyCache to: CacheFree: 0.4 to: LRU_NoSatisfy: 0.6	Using policy DT-LRU in ProxyCache to: CacheFree: 0.7 to: LRU_NoSatisfy: 0.3

**Fig. 4.** Number of Customer (NOC) in Terminals-Browsers**Fig. 5.** Time Response of Station (TRS) in LRU\_Policy\_in\_ProxyCache**Fig. 6.** Time Response of Station (TRS) in SIZE\_LFU\_Policy\_in\_ProxyCache**Fig. 7.** Time Response of Station (TRS) in WebProxy



**Fig. 8.** System throughput of station network

**Table 4.** Compared performance between HYBRID and DT- HYBRID

Parameters	HYBRID	DT-HYBRID
System Throughput (Second/N)	0.2002	0.1982
Response Time in WebProxy (s)	1.9835	1.9702
Response Time in SIZE_LFU_Policy_in_ProxyCache (s)	1.9961	1.9843
Response Time (s) in LRU_Policy_in_ProxyCache (s)	2.0034	1.9987
Number of Customer (web requests) in WebTrue	204.8635	0.7638
Number of Customer (web requests) in CacheFree	257.1057	0.7322
Number of Customer (web requests) in LFU_NoSatisfy	1.4772	0.1132
Number of Customer (web requests) in CacheHit	4.32E4	2.78E4
Number of Customer (web requests) in CacheMiss_and_Full	9659.2394	0.7324
Number of Customer (web requests) in WebProxy	0.7909	0.5662
Number of Customer (web requests) in Terminals-Browsers	2008.7415	1959.8498
Number of Customer (web requests) in WebReq	136.6407	274.2248

As result in Table 4, the DT-HYBRID has the number of customers lower than HYBRID in terms of the terminals-Browsers, web proxy, CacheMiss\_and\_Full, CacheHit, LFU\_NoSatisfy, WebTrue, and CacheFree. Especially, the Web-true in DT-HYBRID provides 0.7638 number requests which is much better than HYBRID with 204.8635 number requests. Furthermore, WebReq in DT-HYBRID-M has more responses to the client than HYBRID-M. It shows that the web-server source is offloaded by using DT-HYBRID.

Besides, the response times and the system throughput also consider evaluating the web cache replacement policies. In this simulation, the DT-HYBRID proves that it has fast response than HYBRID in terms WebProxy, SIZE\_LFU\_Policy\_in\_ProxyCache, LRU\_Policy\_in\_ProxyCache. And the system throughput in DT-HYBRID is lower than HYBRID.

Based on all results of the simulation, all features of the network system could be calculated and estimated accurately by using any web caching replacement policies. They could be used to evaluate many networks with specific parameters and compared based on mathematical expressions.

## 4 Conclusion

This paper presents the method to calculate the parameters affected to performance in internet web caching architecture. Therefore, two algorithms named HYBRID and DT-HYBRID has performed a simulation in JMT to display the accuracy value of the number of customers, response time, and system throughput to evaluate the performance in the web cache replacement policy. All results the DT-HYBRID provides higher quality than the HYBRID algorithm based on response number from JMT. Our method could succeed in both internet web caching of ISP network with  $n$  level and queueing Petri nets.

## References

1. Lam, H.K., Truong, N.X.: Performance analysis of hybrid web caching architecture. *Am. J. Networks Commun.* **4**(3), 37–43 (2015)
2. Kim, S.: Cooperative Inter-ISP traffic control scheme based on bargaining game approach. *IEEE Access* **9**, 31782–31791 (2021)
3. Sufian, R.S., Liu, T., Paul, A.: Gluon distributions and their applications to Ioffe-time distributions. *Phys. Rev. D* **103**(3), 036007 (2021)
4. Ma, T., Hao, Y., Shen, W., Tian, Y., Al-Rodhaan, M.: An improved web cache replacement algorithm based on weighting and cost. *IEEE Access* **6**, 27010–27017 (2018)
5. Yovita, L.V., Syambas, N.R.: Caching on named data network: a survey and future research. *Int. J. Electr. Comput. Eng. (IJECE)* **98**(6), 4456–4466 (2018)
6. Rais, R.N.B., Khalid, O.: Study and analysis of mobility, security, and caching issues in CCN. *Int. J. Electr. Comput. Eng. (IJECE)* **10**(2), 1438–1453 (2020)
7. Negara, R.M., Syambas, N.R.: Caching and machine learning integration methods on named data network: a survey. In: 2020 14th International Conference on Telecommunication Systems, Services, and Applications (TSSA), pp. 1–6 (2020)
8. Im, Y., Prahladan, P., Kim, T.H., Hong, Y.G., Ha, S.: SNN-cache: a practical machine learning-based caching system utilizing the inter-relationships of requests. In: 2018 52nd Annual Conference on Information Sciences and Systems (CISS), pp. 1–6 (2018)
9. Ali, W., Shamsuddin, S.M.: Intelligent client-side web caching scheme based on least recently used algorithm and neuro-fuzzy system. In: Yu, W., He, H., Zhang, N. (eds.) ISNN 2009. LNCS, vol. 5552, pp. 70–79. Springer, Heidelberg (2009). [https://doi.org/10.1007/978-3-642-01510-6\\_9](https://doi.org/10.1007/978-3-642-01510-6_9)
10. Romano, S., ElAarag, H.: A neural network proxy cache replacement strategy and its implementation in the Squid proxy server. *Neural Comput. Appl.* **20**(1), 59–78 (2011)
11. Cobb, J., ElAarag, H.: Web proxy cache replacement scheme based on back-propagation neural network. *J. Syst. Softw.* **81**(9), 1539–1558 (2008)
12. Jarukasemratana, S., Murata, T.: Web caching replacement algorithm based on web usage data. *New Gener. Comput.* **31**(4), 311–329 (2013)
13. Sulaiman, S., Shamsuddin, S.M., Forkan, F., Abraham, A.: Intelligent web caching using neurocomputing and particle swarm optimization algorithm. In: 2008 Second Asia International Conference on Modelling and Simulation (AMS), pp. 642–647 (2008)

14. Ma, T., Qu, J., Shen, W., Tian, Y., Al-Dhelaan, A., Al-Rodhaan, M.: Weighted greedy dual size frequency based caching replacement algorithm. *IEEE Access* **6**, 7214–7223 (2018)
15. Koskela, T., Heikkonen, J., Kaski, K.: Web cache optimization with nonlinear model using object features. *Comput. Networks* **43**(6), 805–817 (2003)
16. Sajeev, G.P., Sebastian, M.P.: A novel content classification scheme for web caches. *Evolving Syst.* **2**(2), 101–118 (2011)
17. Julian Benadit, P., Sagayaraj Francis, F., Muruganantham, U.: Improving the performance of a proxy cache using expectation maximization with Naive Bayes classifier. In: Jain, L.C., Behera, H.S., Mandal, J.K., Mohapatra, D.P. (eds.) Computational Intelligence in Data Mining - Volume 2. SIST, vol. 32, pp. 355–368. Springer, New Delhi (2015). [https://doi.org/10.1007/978-81-322-2208-8\\_33](https://doi.org/10.1007/978-81-322-2208-8_33)
18. Julian, B.P., Pahuja, K., Sidhu, M.S.: Enhancements to content caching using weighted greedy caching algorithm in information centric networking. *Procedia Comput. Sci.* **171**, 2435–2444 (2020)
19. Nimishan, S., Shriparan, S.: An approach to improve the performance of web proxy cache replacement using machine learning techniques. In: 2018 IEEE International Conference on Information and Automation for Sustainability (ICIAfS), pp. 1–6 (2018)
20. Mahmoud Al-Qudah, D.A., Olanrewaju, R.F., Wong Azman, A.: Enhancement web proxy cache performance using wrapper feature selection methods with NB and J48. *IOP Conf. Ser. Mater. Sci. Eng.* **260**, 012012 (2017)
21. Bertoli, M., Casale, G., Serazzi, G.: JMT: performance engineering tools for system modeling. *SIGMETRICS Perform. Eval. Rev.* **36**(4), 10–15 (2009)
22. Kumar, P.V., Reddy, V.R.: Web proxy cache replacement policies using decision tree (DT) machine learning technique for enhanced performance of web proxy. *Int. J. Recent Innov. Trends Comput. Commun.* **2**(2), 302–309 (2014)



# An Abnormal Heart Rhythm Warning System Based on a Low-Cost Two-Electrode ECG Signal Using Threshold and Neural Network Approaches

Thuong H. N. Nguyen<sup>1(✉)</sup>, Mai Hue T. Truong<sup>2</sup>, Duc M. Le<sup>2</sup>, Ngoc Huynh Tran<sup>2</sup>, Thi Hai Hau Pham<sup>2</sup>, B. Hoang Nguyen<sup>2</sup>, Tuan V. Pham<sup>2,3</sup>, and Anh Thu T. Nguyen<sup>2,3</sup>

<sup>1</sup> TAPIT Engineering Co. Ltd., Danang, Vietnam

<sup>2</sup> The University of Danang, University of Science and Technology, Danang, Vietnam

{pvtuan, ntathu}@dut.udn.vn

<sup>3</sup> L.Y.D.I.N.C Co. Ltd., Danang, Vietnam

**Abstract.** Early detection methods of abnormal heart rhythms are extremely demanded to support effective diagnosis and treatment of patient heart diseases and avoid heart stroke which is the main cause of human death. This study proposes a low-cost system to continuously record electrocardiogram signals by two electrode plates placed on the chest for the purpose of monitoring cardiac electrical signals, and detecting abnormalities in heart rhythm using the threshold-based and neural network-based approaches. A device for signal acquisition and pre-processing has been designed, fabricated and tested with normal people to create a self-built normal ECG database. The recognition results obtained from our self-built database and a published online ECG database. It shows high performance in normal and abnormal heart rhythm recognition which would be helpful for physicians to diagnose the abnormal signal of the heart in time.

**Keywords:** ECG · Abnormal heart rhythm warning · 2-electrode ECG · R-R interval

## 1 Introduction

According to WHO, cardiovascular diseases (CVDs) are the leading cause of death globally. An estimated 17.9 million people died from CVDs in 2019, accounting for 32% of all deaths globally. Of these deaths, 85% were due to heart attack and stroke [1]. In Vietnam, cardiovascular disease was the cause of 31% of all deaths in 2016 equivalent to more than 170,000 deaths. Cardiovascular manifestations often appear suddenly and often show no signs on examination. Early detection of abnormalities through heart rhythm is therefore really necessary.

Today methods used to measure heart rate include Electrocardiography, Photoplethysmography, Oscillometry and Phonocardiography. According to the manual of

exercise testing, Electrocardiography is the best way to measure and calculate instantaneously through the R-R interval [2]. However, measurements taken through such ECG devices are often encrypted and can only be decoded with software provided by the manufacturer. Recently, some manufacturers provide devices that allow users to develop applications associated with the devices with quite an expensive cost [3]. Currently, the 12-lead ECG (10-electrode ECG) and 5-lead ECG (5-electrode ECG), 3-lead ECG (3-electrode ECG) systems are considered as common ECG systems in the world. The 12-lead ECG is considered the gold standard in clinical practice [4], with ten electrodes merged on the patient's chest, arms and legs, and operated by qualified physicians in hospital due to its complexity.

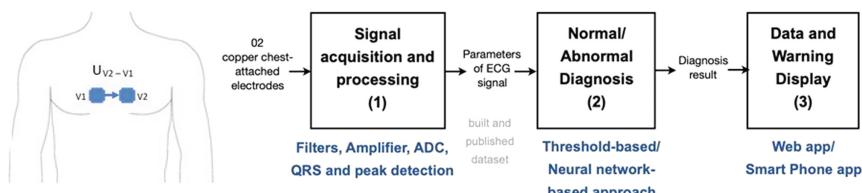
However, this 12-lead ECG device is not suitable for patients with chronic heart disease who want home care and athletes who want to monitor their health while exercising. For the same reason, this measurement method is also not suitable for soldiers, firefighters, policemen, mine rescuers, etc. Furthermore, the cost and discomfort caused by the electrodes in contact with the skin makes it impossible for patients to use 12-lead ECG measurement for daily heart rate monitoring. The same problem occurs with the 5-lead ECG and the 3-electrode ECG device.

It is therefore desirable to have two electrodes to reduce the costs and increase patient comfort. However, removal of the third electrode is challenging due to the significantly higher electromagnetic interference and lower signal-to-noise ratio in two-electrode compared to three-electrode ECG acquisition systems [5]. A number of recent studies [5, 6, 7, 8] are aimed at designing a system to measure ECG from 2 electrode plates with good results like the 3-electrode ECG device. This paper presents our proposal on designing and fabricating an low-cost electrocardiogram signal acquisition system using only two electrode plates. The synthesized ECG signal is then automatically monitored and analyzed for abnormal heart rhythm detection using threshold-based and neural network-based approaches.

## 2 Methodology and Materials

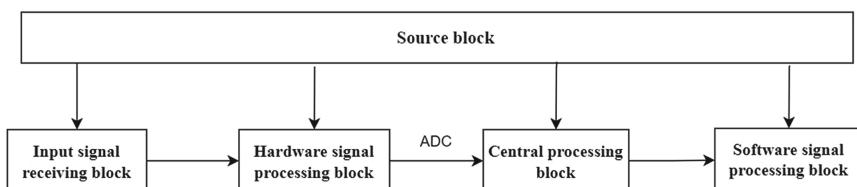
### 2.1 A Subsection Sample

In order to meet the stated demands above, an abnormal heart rhythm monitoring and warning system has been designed in this study with three blocks shown in Fig. 1.



**Fig. 1.** The electrocardiogram monitoring and warning system block diagram.

The hardware and software design process is demonstrated in Fig. 2. Particularly, the power (source) block employs a power filter composed of a series of capacitors and resistors, which aids in the reduction of source noise created by conventional power sources, which has a significant impact on the ECG signal. The ECG signal is then recorded by utilizing two electrode plates positioned on the patient's chest. In the hardware signal processing block, output of the DRL circuit (for noise cancellation) is connected to the input electrodes via high-value resistors in order to create a two-electrode ECG. The STM32 microcontroller receives the output ECG signal amplitude from the hardware signal processing block and converts it to digital signal. Finally, in the software signal processing block, the 4th order Butterworth low-pass filter with cutoff frequency  $f_c = 25$  Hz is then applied to reduce noise.



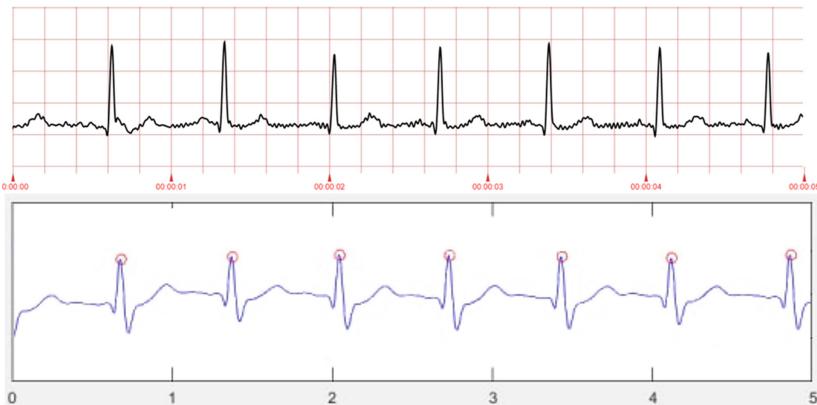
**Fig. 2.** Block diagram of the acquisition and processing of electrocardiogram signals.

The power (source) block employs a power filter composed of a series of capacitors and resistors, which aids in the reduction of source noise created by conventional power sources, which has a significant impact on the ECG signal. The ECG signal is then recorded by utilizing two electrode plates positioned on the patient's chest. In the hardware signal processing block, output of the DRL circuit (for noise cancellation) is connected to the input electrodes via high-value resistors in order to create a two-electrode ECG. The STM32 microcontroller receives the output ECG signal amplitude from the hardware signal processing block and converts it to digital signal. Finally, in the software signal processing block, the 4th order Butterworth low-pass filter with cutoff frequency  $f_c = 25$ Hz is then applied to reduce noise. The Fig. 3 describes the hardware prototype of signal acquisition and processing device.



**Fig. 3.** Hardware prototype of signal acquisition and processing device.

In order to detect the QRS complex container, the following processes are performed in sequence: obtaining the signal derivative, squaring the signal amplitude difference, and applying the threshold extraction technique. The R peaks are the positive peaks of the QRS complex. They are detected by comparing the relative magnitudes in each QRS region. In order to ensure that the output data is valid, the signal acquisition and processing unit's output signal is normalized with the standard ECG signal model. We also analyzed and compared the measured results of the device that the research team performed with the signals published at the ECG-ID Database published at PhysioNet [9]. Extracting a data sample with a duration of 5 s from the reference database (upper part of Fig. 4) and data collected from the equipment performed by our research (lower part of Fig. 4). We can recognize all the characteristic ECG waves obtained from the device, the R-R intervals are similar and within the normal heart rate range. R-R interval is a feature that we use in this study. The ECG data obtained from the device will be used for identification by threshold method and neural network method presented in the next section of this paper.



**Fig. 4.** The ECG waves obtained from the device and from the reference database.

## 2.2 Normal/Abnormal Diagnosis

**Threshold-based approach.** The heart rates which are calculated from the R-R intervals are compared with the normal human heart rates in 60–100 beats per minute. A tachycardia signal is one in which the number of beats is greater than 100 bpm, whereas a bradycardia signal is one in which the number of beats is fewer than 60 bpm. There are several factors influencing the average human heart rates are age, exercise level, etc.

**Convolutional Neural Network-based approach.** A high-accuracy ECG arrhythmia classification method based on convolutional neural network (CNN) has been proposed in [10] with quite good accuracy of heart disease classification. In this study, a CNN model has been built up of basically three types of main layers [11], where a one dimensional

signal is classified. The CNN model takes the time series data in one dimensional form wherein the data are arranged in the order of sequential time instants. The input multiplier network is a time series of a 6 s ECG signal sampled at 200 Hz. The network has an architecture of 33 convolutional layers, followed by a fully connected layer and softmax.

### 2.3 Data and Warning Display

For user-friendly heart rate signal monitoring and analysis, a web application has been designed to display raw heart rate waveform after hardware and software filtering (described in Sect. 2.1). In addition, the application displays the number of beats and a list of timing when abnormal heart rhythm is detected, more specifically, the status of the abnormality is tachycardia or bradycardia.

## 3 Results and Evaluation

### 3.1 Threshold-Based Model

**MIT-BIH Arrhythmia Database.** This database of cardiac electrical signals is divided into two abnormal data sets (“Atrial premature beat” (A) and “Premature ventricular contraction” (V)) and one normal data set. The recognition rate of 90% is obtained from the 10 abnormal A signal samples, and 97.67% obtained from the 43 abnormal V signal samples. From 2174 normal cardiac electrical signal samples, a recognition rate of 90.98% is achieved by the thresholding approach.

**TAPIT Self-built database.** TAPIT self-built data set consisting of 551 samples of normal ECG signals measured from 100 normal people using the designed 2-electrode ECG device described in Sect. 2. The results show that the normal ECG signals are recognized with Recall of 99%.

### 3.2 Neural Network-Based Model

**MIT—BIH Arrhythmia Database.** This dataset was divided into the train and test set with the ratio of 7:3 (9936 training and 4333 testing samples). The CNN model with configuration described in Sect. 2.2 has been trained and tested. As shown in Table 1, the obtained accuracy of 97.7% for two abnormal data sets and one normal data set is much higher than the one of 90.98% derived by the threshold-based approach.

**Table 1.** The training and testing results on the MIT—BIT Arrhythmia dataset.

Metrics	AUC	Accuracy	Recall	Precision	Specificity	Prevalence
Training	0.996	0.981	0.965	0.975	0.989	0.315
Testing	0.993	0.977	0.959	0.968	0.985	0.313

**Table 2.** CNN model's evaluation on TAPIT normal Self-built data set.

Metrics	AUC	Accuracy	Recall	Precision	Specificity	Prevalence
Testing	0.991	0.982	0.963	0.976	0.986	0.314

**TAPIT Self-built Database.** The trained CNN model is then tested on the normal self-built data set consisting of only normal 2-electrode ECG signals. As show in Table 2, the obtained results show accuracy of 98.2% which is quite similar to performance resulting from the threshold-based method. With the usage of the CNN model trained on the ECG signals collected from the conventional 12-electrode ECG device (MIT—BIH Arrhythmia database), this recognition result also reflects the quality of the ECG signals measured from the self-designed 2-electrode ECG device.

## 4 Conclusions and Discussion

In this paper, we first built the 2-electrode ECG device using the hardware and software co-designing process. The ECG signal measured by the device was then compared with the published one from a standardized 12-electrode ECG device for validation. Two different threshold-based (at low complexity) and neural network-based approaches have been trained to detect normal/abnormal heart rhythms. The results tested on a self-built database and standardized database show high recognition rates of normal and abnormal ECG signals. Thus this system can be potentially used in practice to early warn users and assist doctors in electrocardiograms-based heart disease diagnosis. In the future, we may develop abnormal heart rhythms equipment/systems for a variety of people, including active subjects, athletes, and so on. We also integrated some new development directions, such as adding more methods for distinguishing normal and abnormal ECG signals to the project. In addition, a more complete EDGE AI hardware will be designed in the upcoming projects to be able to train and deploy the CNN model for real-time electrical signals. However, this study only puts forward suggestions for improvement of the ECG signal devices and abnormal warning software in terms of early detection of a variety of heart disease while people still can wear the device in daily working life.

## References

1. World Health Organization: Cardiovascular diseases (CVDs), June (2021)
2. Kumar, S.: Efficient Heart Rate Monitoring. Cypress Semiconductor Corp (2011)
3. Duong, T.L., Nguyen, D.T., Nguyen, T.H., Trinh, Q.D., Nguyen, P.K., Phi N.T.: Design of the ECG signal acquisition system in real-time based on soundcard interface integrated on personal computer. VNU J. Sci. (2014)
4. David, E. Albert: Two electrode apparatus and methods for twelve lead ECG. (2016)
5. Babusiaik, B., Borik, S., Smoñdrk, M.: Two-electrode ECG for ambulatory monitoring with minimal hardware complexity, April 22, (2020)
6. Pereira, C.D.M., Mendes, P.M.: Development of a two-electrode ECG acquisition system with dynamic interference rejection, March 4 (2019)

7. Hsieh, H.Y.: Wireless potential difference electrocardiogram constituted by two electrode-pairs wearing comfort, August 20, (2020)
8. Rashkovska, A., Depolli, M., Tomašić, I., Avbelj, V. Trobec, R.: Medical-grade ECG sensor for long-term monitoring, March 18, (2020)
9. Lugovaya, T.S.: Biometric human identification based on electrocardiogram. [Master's thesis] Faculty of Computing Technologies and Informatics, Electrotechnical University "LETI", Saint-Petersburg, Russian Federation, June 2005
10. Zhang, D., Chen, Y., Chen, Y., Ye, S., Cai, W., Chen, M.: An ECG heartbeat classification method based on deep convolutional neural network. J. Healthcare Eng. vol. 2021, Article ID 7167891, 9, (2021)
11. Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., Inman, D.J.: 1D convolutional neural networks and applications: A survey. Mech. Syst. Signal Process., Volume 151, (2021)



# Edge AI Implementation for Recognizing Sounds Created by Human Activities in Smart Offices Design Concepts

Thuong H. N. Nguyen<sup>1(✉)</sup>, Quy C. Nguyen<sup>2</sup>, Viet H. H. Ngo<sup>2</sup>, Fabien Ferrero<sup>3</sup>, and Tuan V. Pham<sup>2,4</sup>

<sup>1</sup> TAPIT Engineering Co. Ltd., Da Nang, Vietnam  
nhatthuongqnn@gmail.com

<sup>2</sup> The University of Danang, University of Science and Technology, Da Nang, Vietnam  
pvtuan@tut.udn.vn

<sup>3</sup> Universite Cote d'Azur, Nice, France  
fabien.ferrero@univ-cotedazur.fr

<sup>4</sup> L.Y.D.I.N.C Co. Ltd., Da Nang, Vietnam

**Abstract.** In this study, an Edge AI solution has been proposed to recognize specific sounds created by human activities in smart office concepts. A convolutional neural network model has been developed on personal computers using collected data sets including 5 different office sounds made by humans including phone ringing, keyboard typing, door knocking, glass breaking and people talking. This trained CNN model is then deployed on an STM32 microcontroller to build standalone Edge AI recognition applications for smart offices. The evaluation results show the average recall of 95% obtained from the collected training datasets, and almost 90% obtained from the testing sounds simulated in a real office environment. The testing takes roughly 0.1 s per sample on the CNN model imported on STM32 microcontroller. This good recognition performance derived from the limited resources of memory and computational speed of the STM32F746NG MCU opens potential applications of Edge AI for human activity recognition while meeting the constraints of real time processing and other requirements.

**Keywords:** Edge AI · Convolutional Neural Network · Sound recognition · Smart office

## 1 Introduction

In recent years, there is more and more research, products and services on smart offices worldwide. Office design is critically studied to improve environment comfort and working efficiency while keeping employee health, security and safety [1]. There are a variety of concepts, designs and technologies for smart offices which have been studied [2, 3]. Automatic recognition of human behavior or activities is one of the most potential approaches to actively sense working environments for delivering next intelligent actions. Some previous studies have suggested many ways to detect human activities

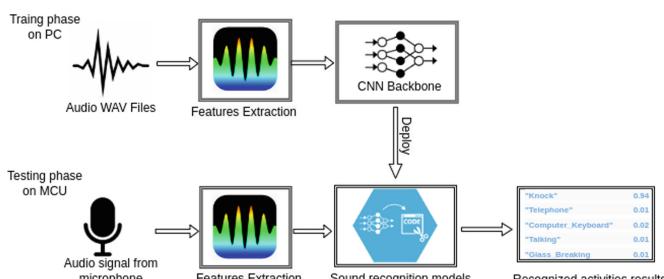
such as using cameras and motion sensors to collect data about human activities in the office. However, due to many barriers, hidden corners, people moving or low light intensity, etc. in office working environments, the sensor and vision-based approaches will face many challenges. There is another way to overcome the mentioned issues by using sound recognition to detect human activities [4]. Audio signals which consist of noises and other sounds created by human activities propagate fully in indoor spaces and convey all useful information associated with their specific behaviors in working offices.

As a result, our goal is to enable sound recognition on small, embedded edge devices, which typically have limited memory and computation capabilities due to power constraints. For mobile devices [5], smartphones, CPUs, and GPUs [6], much effort has already been put into developing smaller, more efficient CNN models. Smartphones and comparable devices, on the other hand, have compute power orders of magnitude more than the embedded devices we're targeting, necessitating considerably more radical minimization. This paper presents our proposal on developing an Edge AI platform to recognize five different sources of sounds made by human activities including phone ringing, keyboard typing, door knocking, glass breaking and people talking. A CNN model has been trained on a computer and then deployed on the Discovery kit STM32F746NG MCU using STM32 microcontroller with limited resources on memory and computing power. This novel design of human behavior related sound recognition has been tested with collected data sets and audio data simulated in real office working environments.

## 2 System Design

### 2.1 System Overview

The process of designing a sound recognition platform is presented in Fig. 1 with the following phases: The training phase consists of audio data collection, CNN training on personal computer, and implementation of the trained CNN model on the Discovery kit STM32F746NG MCU; The test phase is carried out with simulation of five different sound types in real office environments and recognition by the imported CNN model on the STM32 microcontroller.



**Fig. 1.** The overall diagram of the proposed audio recognition models.

## 2.2 Feature Extraction

Audio features could be extracted in time-domain and frequency-domain for different purposes of recognition. According to [7], frequency domain features reveal deeper patterns in the audio source when there is insufficient and complicated data which is the case we are dealing with in working offices. Frequency-domain features like Mel-Frequency Cepstral Coefficients (MFCCs) [8] which are the most common representation in traditional audio signal processing are also considered as dominant features in combining with a variety of deep learning techniques, according to extensive research by Purwins, et al. [9]. Because the Log-Mel Spectrogram was obtained in step 4 during MFCC computation, the calculation time to determine the Log-Mel Spectrogram is shorter than MFCC. As a result, the Log-Mel Spectrogram features have shown their advantages on low computation complexity which matches our research target on Edge AI implementation for sound recognition. In this study, the Log-Mel Spectrogram features have been estimated using a frame length of 1024 (64 ms), hop length of 512 (32 ms), a sample rate of 16000 Hz, and converting the frequency axis of the spectrum to a mel scale with a factor of mel = 30.

## 2.3 Dataset

The standardized audio databases for sound recognition published by FSD50K, ESC-50 [10] and FSD Kaggle [11] have been re-used for training the CNN model. Data classification of 5 different sources is presented in Table 1 with a number of audio files and corresponding samples. Non-audio portions that aren't beneficial for model training were removed. The collected database was then divided into 80%, 10%, and 10% for training, validating, and testing the model, respectively.

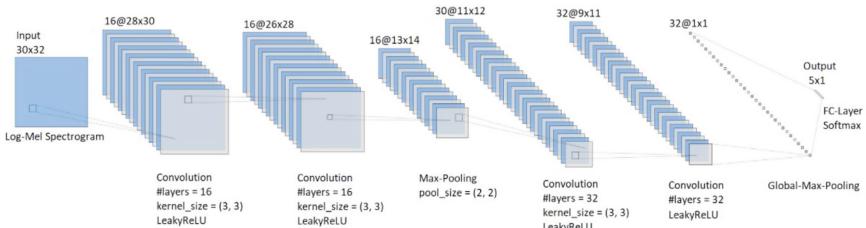
**Table 1.** The number of specific files and samples.

	Keyboard typing	Door knocking	People talking	Phone ringing	Glass breaking
Train (File/Sample)	119/7088	279/3340	120/1152	210/2224	311/984
Test (File/Sample)	26/886	39/430	48/144	49/279	68/123

## 2.4 Proposed CNN Model

Liu et al. [12] conducted a survey on deep learning models and found that the Convolutional Neural Network (CNN) outperformed other deep learning models on various applications such as computer vision, pattern recognition, speech recognition, etc. At the same time, when extracting audio features using Log-Mel Spectrogram, we can treat

audio as images for classification. For the above reasons, we have done sound recognition using a self-developed CNN in this study. For Environmental Sound Classification on the Edge, Mohaimen Uzzaman, et al. [13] presented the ACDNet Architecture, which has 12 convolution layers, each followed by a batch normalization layer and a ReLU activation layer. Because this model is too huge to execute on the MCU, we recommend a simplified model that comprises the following layers. The input is the Log-Mel Spectrogram feature; It is followed by 2 convolution layers with 16 filters and  $\text{kernel\_size} = (3, 3)$ ; Then the 2 convolution layers but with the number of filters is 32 is established; The next layer is a Global Max-Pooling layer to retain the highest activation values of each layer; Finally, it will be the output layer with the softmax activation function to determine the recognition probabilities for each type of sound. The training process with hyperparameters is as follows: Epochs = 350, batch\_size = 128. Keras Checkpoint is used to save the best model parameters during model training. Besides, in order to avoid the case that the model falls into overfitting, we also use some regularizer techniques as follows: Batch Normalization, Spatial Dropout and L2 Regularizer. The model utilizes the Adam optimization algorithm to optimize the accuracy of the model and minimize the loss of the model. Model structure is shown in Fig. 2.



**Fig. 2.** Model structure.

## 2.5 Embedding the Model into the Microcontroller

With the tool called STM32Cube.AI published in [14], minimum requirements on hardware resources for implementing the trained model have been analyzed. The minimum requirements of 58.05 KB of RAM and 64.96 KB Flash meets the hardware resources of the STM32 microcontroller consisting of 1 MB Flash and 320 KB RAM. In order to capture sounds in real environments, a normal microphone is installed and connected to the STM32's ADC peripheral. This device converts analog audio signals to digital signals with sampling frequency of 16 kHz which is similar to the audio files that have been used for training the CNN model on the computer. The MFCC feature extraction is implemented on each captured frame lasting in 64 ms with the overlap of 32 ms between two successive frames.

### 3 Result

#### 3.1 Result of Training Model on the PC

The trained CNN model has been evaluated with the training set described in Sect. 2.2. There are three criteria traditionally used for evaluating classification performance are employed in this research as: Precision ( $TP/(TP + FP)$ ), Recall ( $TP/(TP + FN)$ ) and F1-score =  $(2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$ ; where TP: true positives, FN: false negatives, TN: true negative, FP: false positives. As shown in Table 2, the recognition results reach almost 96% which are very high rates. Among five different sounds, the people talking sounds are recognized at the highest performance due to the MFCC features originally designed for voice recognition. Confusion matrix of test set is shown in Table 3.

**Table 2.** Precision, recall and F1-score derived from the training set.

	Keyboard typing	Door knocking	People talking	Phone ringing	Glass breaking	Average scores
Precision	0.965	0.930	0.977	0.971	0.937	0.956
Recall	0.962	0.940	0.986	0.962	0.937	0.957
F1-Score	0.963	0.935	0.982	0.967	0.937	0.957

**Table 3.** Confusion matrix of test set.

	Keboard typing	Door knocking	People talking	Phone ringing	Glass breaking
Keyboard typing	<b>296</b>	2	2	1	0
Door knocking	2	<b>105</b>	3	1	1
People talking	2	1	<b>399</b>	0	1
Phone ringing	1	0	1	<b>185</b>	1
Glass breaking	0	1	4	1	<b>96</b>

#### 3.2 Result of Deploying Model on the MCU

In order to build a stand-alone application, the trained CNN model has been imported onto the Discovery kit STM32F746NG MCU. With the STM32F746 ARM Cortex M7 microcontroller and the number of cycles per MACC of 6, the calculation complexity

of the MACC parameter will be 3.376.459. With this calculation complexity as well as the operating frequency at  $f_{clock} = 184.5$  MHz, the processing time of the model (from input to output) is just about 0.1 s which meets real-time processing of most applications using sound recognition:

$$t = \frac{n \times MACC}{f_{clock}} = \frac{6 \times 3.376.459}{184.5 \times 10^6} \simeq 0.1s \quad (1)$$

The imported Edge AI model is then tested with the real sounds including phone ringing, keyboard typing, door knocking, glass breaking and people talking simulated in our office. For safety and cost concerns, collected audio sounds of broken glass are played back. Each sound type was tested 100 times. The tested results are shown in Table 4.

**Table 4.** Actual test results on MCU.

	Keyboard typing	Door knocking	People talking	Phone ringing	Glass breaking	Average scores
Recall (%)	93.0	94.0	89.0	91.0	88.0	91.0

Although the recognition results are not comparable to the ones obtained from the AI model run on PC, the designed Edge AI application still shows good recognition rates of about 90%. All input sounds which result in fail recognition have been analyzed with the main causes of noisy signals.

## 4 Conclusion

The built Edge AI application running on STM32 microcontroller provides an average recognition rate of 90% for 5 different sounds (phone ringing, keyboard typing, door knocking, glass breaking and people talking) while still meeting the constraint of real time processing. In addition, by exploiting the Edge AI platform design, raw data will not be transferred to servers which keep saving bandwidth and securing data. It is obvious that energy consumption will be reduced by operating this application on the STM32 microcontroller kit. All of the above benefits will be significantly favorable when implementing this smart sound recognition application in all offices of the large buildings/companies. For future work, noise removing algorithms will be embedded in the pre-processing stage. Furthermore, there will be more types of sounds which often occur in offices to be trained for building smart sound recognition applications utilized in smart office operation. We will statistically analyze the advantage of Edge AI implementation in the future, we are going to use the whole Environmental Sound Classification datasets ESC-10 and ESC-50, rather than just a portion as it is today, to compare with the results in [3] and the ESC-50 dataset to compare whether with a dataset, our proposed model run on MCU gives better results than the model run on PC of previous studies.

**Acknowledgment.** We would like to thank L.Y.D.I.N.C Co. Ltd. for technical assistance and funding sponsor for project implementation.

## References

1. Zhang, X., Zheng, P., Peng, T., He, Q., Lee, C.K.M., Tang, R.: Promoting employee health in smart office: A survey. *Adv. Eng. Inform.* **51**, 101518 (2022)
2. Papagiannidis, S., Marikyan, D.: Smart offices: A productivity and well-being perspective. *Int. J. Inf. Manag.* **51**, 102027 (2020)
3. Röcker, C.: Services and applications for smart office environments - a survey of state-of-the-art usage scenarios. *World Academy of Science, Engineering and Technology, Open Science Index* 37. *Int. J. Humanit. Soc. Sci.* **4**(1), 51–67 (2010)
4. Jung, M., Chi, S.: Human activity classification based on sound recognition and residual convolutional neural network. *Autom. Constr.* **114**, 103177 (2020)
5. Zhang, X., Zhou, X., Lin, M., Sun, J.: Shufflenet: An extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6848–6856 (2018)
6. Wu, B., Dai, X., Zhang, P., Wang, Y., Sun, F., Wu, Y., Tian, Y., Vajda, P., Jia, Y., Keutzer, K.: Fbnet: Hardware-aware efficient convnet design via differentiable neural architecture search. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10734–10742 (2019)
7. Venkataraman, K., Rengaraj Rajamohan, H.: Emotion recognition from speech. [arXiv: 1912.10458v1](https://arxiv.org/abs/1912.10458v1) 22 Dec 2019
8. Kopparapu, S.K., Laxminarayana, M.: Choice of Mel filter bank in computing MFCC of a resampled speech. In: 10th International Conference on Information Science, Signal Processing and their Applications, pp. 121–124 (2010),
9. Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S., Sainath, T.: Deep learning for audio signal processing. *IEEE J. Sel. Top. Signal Process.* **13**(2), 206–219 (2019)
10. Piczak, K.J.: Esc: Dataset for environmental sound classification. In: Proceedings of the 23rd ACM International Conference on Multimedia, pp. 1015–1018 (2015)
11. Fonseca, E., Plakal, M., Font, F., Ellis, D.P., Favory, X., Pons, J., Serra, X.: General purpose tagging of freesound audio with audioset labels: Task description, dataset, and baseline. [arXiv: 1807.09902](https://arxiv.org/abs/1807.09902) (2018)
12. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E.: A survey of deep neural network architectures and their applications. *Neurocomputing* **234**, 11–26 (2017)
13. Mohaimen Uzzaman, M., Bergmeir, C., West, I.T., Meyer, B.: Environmental sound classification on the edge: Deep acoustic networks for extremely resource-constrained devices. [arXiv:2103.03483](https://arxiv.org/abs/2103.03483) (2021)
14. Key, D.: Working with ADC and DMA - Maker.io (2019)



# The Remarkable Enhancement in Fish Freshness Classification on Various Species Using Neural Network on Physiological Characteristics Features

Anh Thu T. Nguyen<sup>1,3(✉)</sup>, Nguyet Ha T. Tran<sup>1</sup>, Phuong Dieu T. Nguyen<sup>1</sup>, Dat Xuan Ho<sup>1</sup>, Uyen Kim Hoang<sup>1</sup>, Nguyen Ba Hoang<sup>1</sup>, Thuong H. N. Nguyen<sup>2</sup>, and Tuan V. Pham<sup>1,3</sup>

<sup>1</sup> The University of Danang, University of Science and Technology, Da Nang, Vietnam

{ntathu, pvtuan}@dut.udn.vn

<sup>2</sup> TAPIT Engineering Co.Ltd., Da Nang, Vietnam

<sup>3</sup> L.Y.D.I.N.C Co. Ltd., Da Nang, Vietnam

**Abstract.** People are becoming increasingly concerned about food quality, especially the interest in the freshness of fish. This paper presents extensive research from the previous study for detecting fresh or spoiled fishes using image processing and neural network-based classification. A large data set of fisheye photos has been collected from four types of fish. From the proposed assumptions, four low-cost physiological feature sets as 12 Intensity Slices, Minimum Intensity, Histogram, and Standard Deviation were extracted. Four neural network models have been designed and trained on four feature sets alternatively to detect the freshness of each fish type. The test results show high average recalls of 94.9%, 89.9%, and 88.4% derived from the first three feature sets, respectively. The 12 Intensity Slices feature set resulting in the highest average recall of 95% should be remarked as a common feature for freshness detection on various species. The usage of the simple neural network leads to potential development of fish freshness detecting applications on smartphones.

**Keywords:** Fish freshness classification · Fisheye feature · Neural network

## 1 Introduction

Fish demand has been increasing worldwide due to its high nutritional value fit for consumption. Consequently, many researches have clarified on different approaches for fish quality and freshness measurement over the years. The traditional method of detecting fish freshness involves a sensory examination of the fish's gills, eyes, and skin. The sensory evaluation can be conducted through the senses of smell, taste, touch, and hearing by humans [1]. Even though these indicators are visible to the naked eye, it is not easy to implement due to the need for expertise in the field. Other evaluations are the instrument methods including physical, biochemical [2], chemical [3], biosensor

[4], and more. However, disadvantages of the instrument method are the dependence on equipment, invasion, etc. In recent years, image processing is considered an effective method for fish freshness classification because of its economical, non-invasive, and flexible technique. Specifically, Issac et al. in 2016 [5] showed an effective technique that validates the fish freshness and existence of pesticides. The results of correlation achieves 92.4% compared to the reality. After that, the study published in 2018 [6] proposed an approach utilizing discriminatory features based on a relationship between the values of wavelet coefficients and the freshness levels of the stored fish. In the same year, Sengar et al. [7] showed a method for fish freshness identification with accuracy of 96.66% based on skin tissue using the saturation channel of hue saturation value color space model.

In 2020, our previous research [8] proposed a fish freshness recognition approach which relied on the threshold-based and neural network models with features extracted from fisheye images. Four assumptions were made on physiological appearance of fish eyes as getting stale, including increased background intensity of the pupil region, increased haziness amount of the whole eye area, shifting faraway from 0 value of intensity histogram, and decreased intensity variation of the iris region. These assumptions with four extracted features were well proved for the fish freshness classification on the Crucian carp fish with both recognition models. In particular, there were eight over nine models obtaining the precision of 100% on the training data and seven over nine reaching 100% on the test data. However, the research validated only for one species (Crucian carp) with limited dataset.

With the purpose of exploring the low-cost models and assumptions on various fish species, this study has been conducted on a larger dataset with four types of fish: *Siganus guttatus*, Crucian carp, *Anabas testudineus*, *Mugil cephalus*. Given the proposed assumption [8], the four features extracted from fisheye images composing 12 Intensity Slide, Minimum Intensity vector, Histogram, and Standard Deviation have been reused to evaluate the efficiency of the features, threshold-based technique, and neural network approach in fish freshness classification, and generalize the assumptions for various fish species.

## 2 Methodology and Materials

### 2.1 Data Collection

A new dataset including 37 samples of 4 fish species: *Siganus guttatus*, Crucian carp, *Anabas testudineus*, *Mugil cephalus* has been collected and labeled over several months. They are selected due to their popularity and high consumption in the middle area of Vietnam. Each species contains a different number of samples, which were randomly bought from conventional marketplaces as testing materials. Live fish samples were maintained at room temperature until death, and then we began gathering the images of their fisheye at zero hours. In order to demonstrate a normal condition of image capturing by users, an iphone 6 was used to manually capture fisheyes with the distance of 5–10 cm between the fisheye and device, at room temperature (22 °C–26 °C) and under normal room light conditions, at different time points, from 0–24 h after death, as described in Fig. 1. We've set up the method for capturing the fisheye as Fig. 1, at room temperature

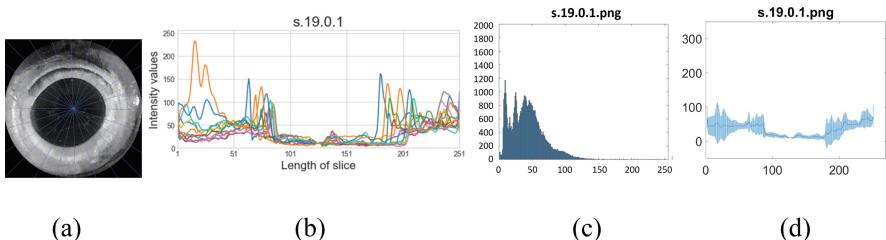
(22 °C–26 °C) and under normal room light conditions, at different time points, from 0–24 h after death. A total of 594 images were acquired. The captured photos are labeled as “Fresh” during initial hours (0 h–5 h), “Neutral” in middle hours (6–19 h), and “Stale” after about 20 h (20 h–24 h) for further analysis. This labeling is based on our previous study [8].



**Fig. 1.** Image capturing setup.

## 2.2 Pre-processing and Feature Extraction

Fisheye images of  $250 \times 250$  pixels were cropped, converted into grayscale, and segmented for pupil and iris areas. Similar to the previous research [8], the feature F0 consisting of 12 Intensity Slices were taken from the fisheye images by cropping straight lines across the image every 15°, from 0 to 180°. For feature F1, minimum values of 12 intensity slices were taken. For feature F2, the histogram of each gray fisheye image was determined to illustrate the intensity distribution of the image pixels in the range of [0, 255] at a given step size of 1. Lastly, feature F3 is the standard deviation estimated at each pixel position belonging to the 12 slices. Evolution of the features is depicted in Fig. 2.



**Fig. 2.** Grayscale image (a) and Features F0 (b), F2 (c), and F3 (d) of *Siganus guttatus*'s eye at 0 h.

## 2.3 Neural Network Training

In order to obtain the most relevant data labeling for each type of fish, the “Neutral” group was not considered, and the labeling process was refined based on self-visual inspection on fisheye images. In the total of 493 images selected for class “Fresh” and

class “Stale”, 368 fisheye images were chosen for training and the remaining 125 images were utilized for testing models, as shown in Table 1.

As studied in [8], a feed-forward neural network formed from three layers has been trained using the stochastic gradient descent algorithm. The input layer size will be the size of the feature vectors. For the purpose of low-complexity of the classifier, one hidden unit has been configured to build the neural network model with one output unit associating to two classes named Fresh Fish and Spoiled Fish. Other learning parameters have been set as in our previous research [8]. All data values were normalized before training and testing depending on the specific features.

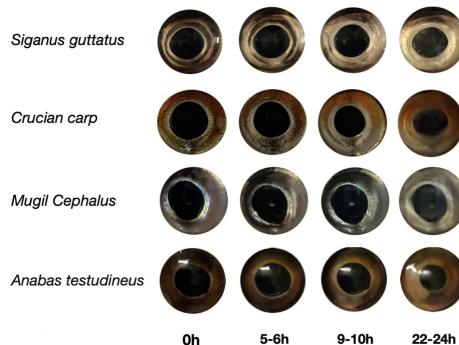
**Table 1.** Arrangement on training and test dataset for 4 fish types.

Types of fish	Sample (37)	Training (368)	Testing (125)	Sum (493)
Siganus guttatus	20	229	77	306
Mugil cephalus	2	28	10	38
Crucian carp	11	67	23	90
Anabas testudineus	4	44	15	59

### 3 Observation and Discussions

#### 3.1 Data Visual Analysis

Due to the variety of the fish species, the physiological features of fisheyes appear differently corresponding to their variation from fresh-to-stale as demonstrated in Fig. 3. In particular, Table 2 lists noticeable observed common features of the four species, and distinguishing features of each species.



**Fig. 3.** Eye images of four selected fish species at several typical time points (0, 5–6 h, 9–10 h, 22–24 h after death).

**Table 2.** Observable common and distinguishing features of the four fish species.

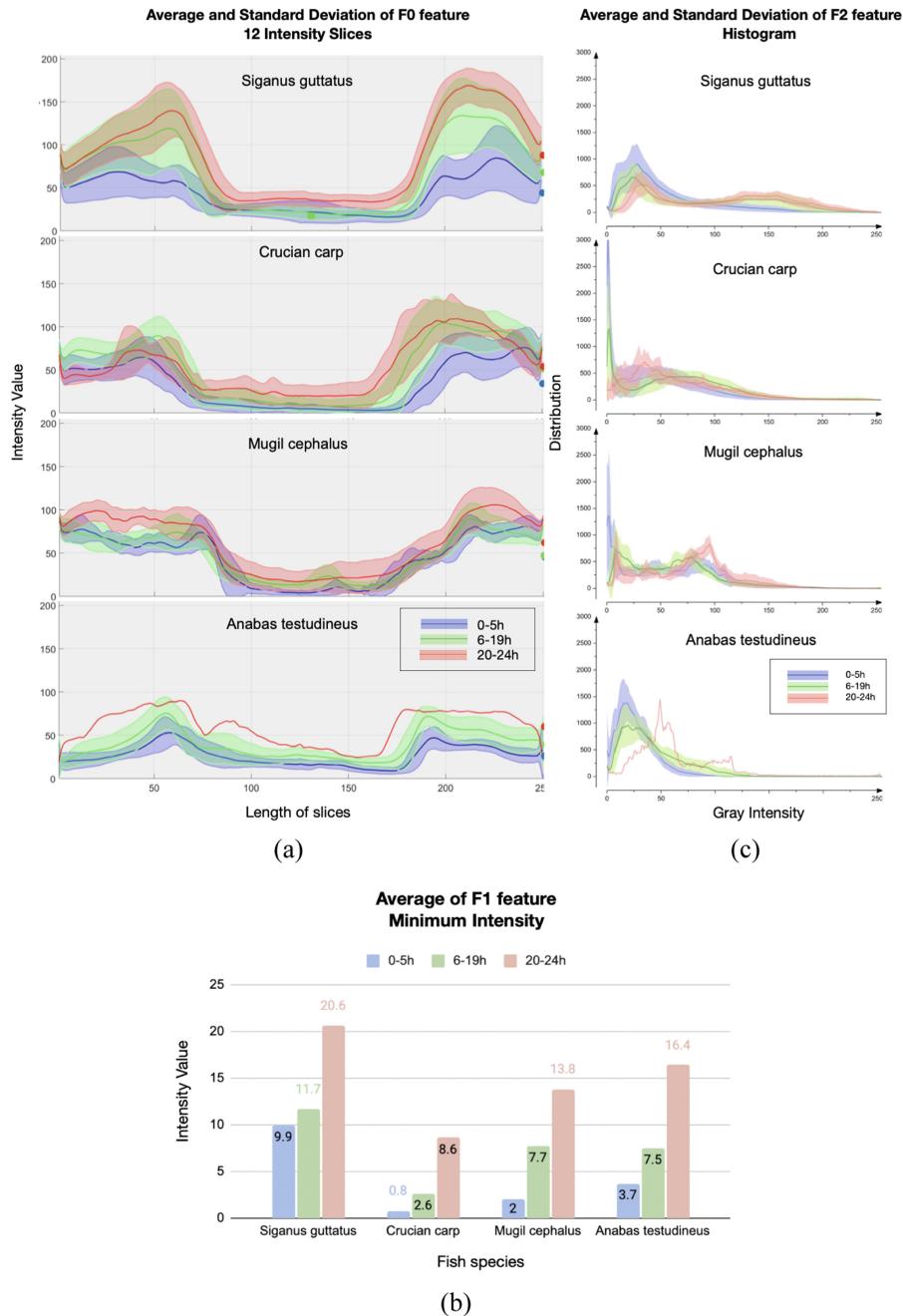
Fish types	Common features	Distinguishing features
<i>Siganus guttatus</i> (20 samples)	At initial hours, fisheyes are transparent, bright, convex, clear, shiny, and have completely dark pupils	- Iris color: was black and golden brown as being fresh and mostly white as being stale; - Pupil size: showed no change
<i>Crucian carp</i> (11 samples)	After approximately 22 h, fisheyes turn opaque, sunken, and are covered by a haze layer	- Iris color: was white and dark orange as being fresh and light orange as being stale, but after 22 h–24 h, it gets dull; - Pupil size: decreased; - Other signs: some black dots in the iris area
<i>Anabas testudineus</i> (4 samples)		- Iris color: was black and dark brown as being fresh, and light brown as being stale - Pupil size: decreased
<i>Mugil cephalus</i> (2 samples)		- Iris color: was black and dark brown; - Pupil size: showed no change; - Other signs: the brighter border between the pupil and iris, which turned dull with time

### 3.2 Observation on Extracted Features

The features F0 (12 Intensity Slice), F1 (Minimum Intensity), F2 (Histogram), and F3 (Standard Deviation) of 4 fish types were extracted and arranged in 3 groups (0–5 h, 6–19 h, 20–24 h). For each fish type, on each feature and on each group, an average vector of all the feature vectors were calculated and plotted in blue, green, and red color curves/bars (Fig. 4a–c) with surrounding corresponding red, green, blue areas for standard deviation values (Fig. 4a, c).

The plots in Fig. 4a indicates the increase of the F0 statistical values (blue to green and red) of fisheyes when they got stale, especially with *Siganus guttatus* fish. The F1 feature in Fig. 4b clearly presents the increase in the minimum values of fishey over three groups in all four species. The F2 graphs in Fig. 4c reveals a reduction in the number of low-intensity pixels but an increase in the higher intensity ones. In general, these extracted features show a good match with the proposed assumptions on increased haziness level on the whole eye area (increased F0), increased background intensity of the pupil area (increased F1), shifting faraway from 0 value of intensity histogram (F2 shift). This consolidates the common features from visual observation in Table 3.

Despite the good match with assumptions, the separations among these groups of “Fresh”, “Neutral” and “Stale” are not really well revealed in statistics of the three features F0, F1 and F2 as observed in Fig. 4. Furthermore, the F3 values (not presented



**Fig. 4.** Average values of (a) F0 features, (b) F1 feature (c) F2 feature of 4 fish species.

in Fig. 4) show random changes in the standard deviation of intensity on the fisheyes over time, which does not match with the corresponding assumption on decreased variation. This would be due to the large variety of features from different types of fisheye, which links to visually observed distinguishing features described in Table 3. These challenges make it non-effective and unsuitable to identify specific threshold values of the extracted features for the threshold-based method in fish freshness classification. The following session therefore will put in development only on neural network models and suitable extracted features for classification.

### 3.3 Classification Results and Analysis

The trained neural network models have been evaluated with the test sets described in Table 1. Three measurements that were used in previous research [8] are re-employed in this research.

The average scores (Precision, Recall and Accuracy) over four types of fish presented in Table 3 show highest classification performance at 89%, 95%, and 89%, respectively, obtained from the feature set F0 - 12 Intensity Slices. This should be recognized as the most common feature set for fish freshness detection. This finding matches well with the stated research purpose is to determine the common assumptions and common features for detecting freshness of various types of fish.

The feature F2 could be considered as the second-best common feature set with the average scores from 87% to 88%. While this feature F2 leads to very high classification performance (100% Precision, 92%–100% Recall) for Crucian carp and Anabas testudineus, it shows lower performance for Siganus guttatus and even much lower rates for Mugil cephalus. Thus this feature can be assumed as a group feature for Crucian carp and Anabas testudineus fish.

The feature F1 presents minimum values of the F0 shows relatively good classification scores but is not robust for various fish types. The feature set F3 certainly shows the lowest classification rates. Among four different types of fish, it seems that Mugil cephalus fish which have quite different characteristics of fish eyes compared to others presents more challenges to these proposed feature sets.

**Table 3.** Classification scores on the test set of 4 fish types (Siganus guttatus / Crucian carp / Mugil cephalus / Anabas testudineus).

Features	F0	F1	F2	F3
Precision (%)	93.9 / 86.7 / 85.7 / 90.0 / <b>Avg: 89.1</b>	84.5 / 93.8 / 66.7 / 100 / <b>Avg: 86.2</b>	88.7 / 100 / 60.0 / 100 / <b>Avg: 87.2</b>	66.7 / 90.0 / 60.0 / 71.4 / <b>Avg: 72.0</b>
Recall (%)	97.9 / 100 / 100 / 81.8 / <b>Avg: 94.9</b>	96.1 / 83.3 / 100 / 80.0 / <b>Avg: 89.9</b>	87.0 / 91.7 / 75.0 / 100 / <b>Ave: 88.4</b>	90.5 / 60.0 / 75.0 / 83.3 / <b>Avg: 77.2</b>
Accuracy (%)	94.8 / 91.3 / 90.0 / 80.0 / <b>Avg: 89.0</b>	85.7 / 82.6 / 80.0 / 86.7 / <b>Avg: 83.8</b>	83.1 / 95.7 / 70.0 / 100 / <b>Ave: 87.2</b>	70.1 / 69.6 / 70.0 / 80.0 / <b>Avg: 72.4</b>

## 4 Conclusion

This study has verified the physiology-based assumptions, corresponding low-cost features and neural network models for freshness classification of the previous research in the larger dataset of four species of fish. The database of fisheye images which is self-built from the four fish species: *Siganus guttatus*, Crucian carp, *Anabas testudineus*, *Mugil cephalus* fishes was described on common and distinguishing characteristics and will be published. The extracted feature sets of 12 intensity slices, minimum values, and histogram show a good match with the assumptions and visual inspection on the various fish species. Furthermore, the distinct characteristics of each type make a contribution to clarify exactly the status of fish. Specifically, the combination of low-cost 12 intensity slices and histogram features with simple neural network models clearly reflect potential for developing a user smartphone application.

**Acknowledgement.** We thank L.Y.D.I.N.C Co. Ltd. for technical assistance and funding sponsor for building FISH Database under collaboration network between UD-DUT and LYD.

## References

1. U.S. Food and Drug Administration, <https://www.fda.gov/food/buy-store-serve-safe-food/selecting-and-serving-fresh-and-frozen-seafood-safely>. Accessed 28 Oct 2021
2. Castillo-Yáñez, F.J., et al.: Postmortem biochemical changes and evaluation of the freshness in the muscle of tilapia (*O. nilo-ticus*) during the storage in ice. *J. Fish. Aquat. Sci.*, **9**(6), 435–445 (2014)
3. Andrade, S.C.S., Marsico, E.T., Godoy, R.L.O., Franco, R.M., Junior, C.A.C.: Chemical quality indices for freshness evaluation of fish. *J. Food Stud.* **3**(1), 71–87 (2014)
4. Endo, H., Wu, H.: Biosensors for the assessment of fish health: a review. *Fish. Sci.* **85**(4), 641–654 (2019). <https://doi.org/10.1007/s12562-019-01318-y>
5. Issac, A., et al.: An efficient image processing based method for gills segmentation from a digital fish image. Paper presented at the 3rd international conference on signal processing and integrated networks, pp. 645–649. IEEE (2016)
6. Arora, M., Mangipudi, P., Dutta, M.K., Burget, R.: Image processing based automatic identification of freshness in fish gill tissues. Paper presented at the international conference on advances in computing, communication control and networking, pp. 1011–1015. IEEE (2018)
7. Sengar, N., Gupta, V., Dutta, M.K., Travieso, C.M.: Image processing based method for identification of fish freshness using skin tissue. Paper presented at the 4th international conference on computational intelligence and communication technology, pp. 1–4. IEEE (2018)
8. Nguyen, A.T.T., Le, M., Vo, H.N., Hong, D.N.T., Phuoc, T.T.A., Pham, T.V.: Proposed novel fish freshness classification using effective low-cost threshold-based and neural network models on extracted image features. In: Huang, Y.-P., Wang, W.-J., Quoc, H.A., Giang, L.H., Hung, N.-L. (eds.) GTSD 2020. AISC, vol. 1284, pp. 60–71. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-62324-1\\_6](https://doi.org/10.1007/978-3-030-62324-1_6)



# Enabling the Smart City Through Agile LP-WAN Network in Da Nang City

Thao Nguyen Manh<sup>1,3,4(✉)</sup>, Lic Tran Van<sup>1,2,4</sup>, Thanh Ngo Dinh<sup>2,4</sup>,  
Thu Nguyen Thi Anh<sup>2,4</sup>, Huy Le Quoc<sup>2,4</sup>, Huy Trinh Le<sup>3</sup>, Fabien Ferrero<sup>1,4</sup>,  
and Nhan Le Than<sup>1,4</sup>

<sup>1</sup> Université Côte d'Azur, Nice, France

{manh-thao.nguyen,fabien.ferrero,nhan.le-thanh}@univ-cotedazur.fr

<sup>2</sup> University of Danang - DUT, Da Nang, Vietnam

{tvlic,ndthanh,ntathu,lqhuy}@dut.udn.vn

<sup>3</sup> University of Information Technology - UIT, Ho Chi Minh City, Vietnam  
huytl@uit.edu.vn

<sup>4</sup> Danang International Institute of Technology (DNIIT), Da Nang, Vietnam

**Abstract.** The objective of this paper is to present the Long Range Wireless Area Network (LoRaWAN), is a type of Low Power Wide Area Network (LPWAN), was deployed in Danang city, Vietnam. This paper presents the different steps from the preliminary experiment to the full LoRaWAN network deployment. The full methodology and equipment used to assess network coverage are given. An extension of the network based on relay feature is also introduced and evaluated.

**Keywords:** LPWAN · LoRa · Relay · Coverage evaluation

## 1 Introduction

In the past decade, low power wide area (LP-WAN) network has been very attractive in Vietnam thanks to its multiple advantages including: low cost, power efficiency, good coverage, deep indoor penetration and standardized protocols. These features make LP-WAN a very practical solution for Internet of Things (IoT) applications in Smart Building, Smart City, Smart Farming, Smart Campus etc. The city of Danang is the one of the largest urban area in Vietnam, with more than one million inhabitants. It is also the key economic and university center and of the central region in Vietnam. Danang has been for a long time considering the Smart City deployment among the most active and innovative cities in Vietnam. The city has invested since several years in wireless communication infrastructure including IoT and other internet network such as LoRa and Wifi [1].

---

L. T. Van—Contributed equally to this work.

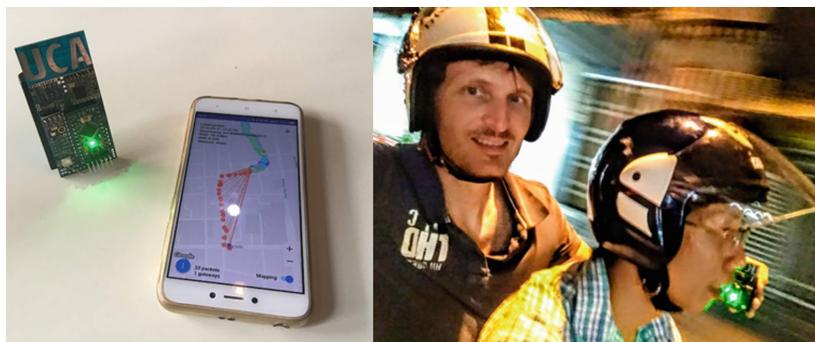
## 2 First Deployment and Coverage Testing

University Côte d'Azur and University of DaNang has started since 2016 to work on the test and deployment of LoRa technology [2]. This wireless LP-WAN technology is based on sub-GHz frequency and unlicensed bands. Concerning the ISM bands in Vietnam, since November,2020 the newest authority regulation has defined a frequency plan for Lower Power Wan Networks is in the band 920 Mhz to 923 MHz (following the Regulation of frequency management department in Vietnam, No. 1122:2020/BTTTT, on the Equipment use frequency in the Low Power Wide Area Networks (LPWAN), with frequency band operating from the 920 MHz to 923 MHz). The frequency plan is an unlicensed band between the limited uplink with frequency 915 MHz and limited downlink with frequency 925 MHz, belong the channels of the mobile networks in Vietnam. Several LoRaWAN gateways were configured and deployed using 920 Mhz-923 MHz frequency plan. We use two type of LoRaWAN gateways Rak7240 and Kerlink iBTS, were used in this experiment and designed to be installed outdoors environment. The frequency channels was configured as described in Table 1, with Radio 0 is 921.8 MHz and Radio 1 is 922.7 MHz.

**Table 1.** Frequency plan in Vietnam

Channel	0	1	2	3	4	5	6	7	8	9
Freq. (MHz)	921.4	921.6	921.8	922	922.2	922.4	922.6	922.8	922.7	923
Radio	0	0	0	0	0	1	1	1	1	1
BW. (kHz)	125	125	125	125	125	125	125	125	250	125

Gateway are connected to the network server using Ethernet and Power over Ethernet (PoE) was used to power the Gateway. The Things Network (TTN), which is the public and open source LoRaWAN network server, was configured to receive data from Gateway. In order to collect LoRaWAN coverage, TTN mapper application was used to see coverage information such as location, SNR and RSSI directly from the The Things Network (TTN). An example of practical measurement is shown on Fig. 1.



**Fig. 1.** Coverage measurement in DaNang using TTN Mapper all around the city



**Fig. 2.** Results of LoRaWAN Gateway coverage testing

In order to maximize the coverage and enable an optimal coverage of important locations, the LoRaWAN gateways location needs to be considered, following criterias:

1. Higher is the better : the Gateway location should be installed on the top of the mountains or buildings.
2. Avoid interference : Location of Gateway should far from cellular site, which Limit to install Gateway antennas near conductive objects and ensure a minimum distance about 2m between antennas.
3. Make it safe : mechanical structure should to be fixed carefully to reduce the influence of wind and storm, and should install lightning protection.

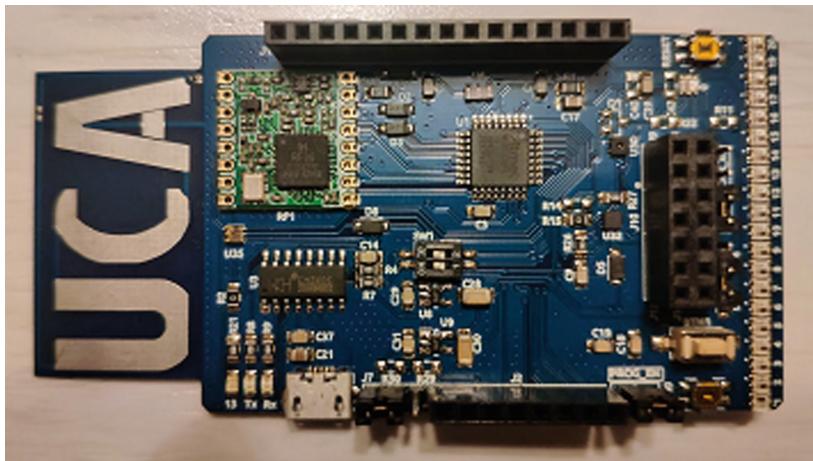
In the first step, the Free-LoRa project in Danang city, Vietnam was implemented. There are 4 gateways with different manufacturers, were installed to check the coverage and optimize the location of LoRaWAN Gateway through experimentation via TTN Mapper. The LoRaWAN Gateway location are shown in the Table 2. Coverage test were realized in the different place in Danang and a stable connection was observed for a 3km radium around the gateway. Communication up to 15 km were observed as shown in Fig. 2.

**Table 2.** Gateways location

Location	GPS-Longitude	GPS-Latitude	GPS-Attitude(m)	Height(m)
Building of DSP	108.2230598	16.07553058	90	2
Building of SPT	108.2290901	16.0786091	40	2
Building of DRT	108.231547	16.057383	45	2
Son Tra Mountain	108.273661	16.1184	812	2

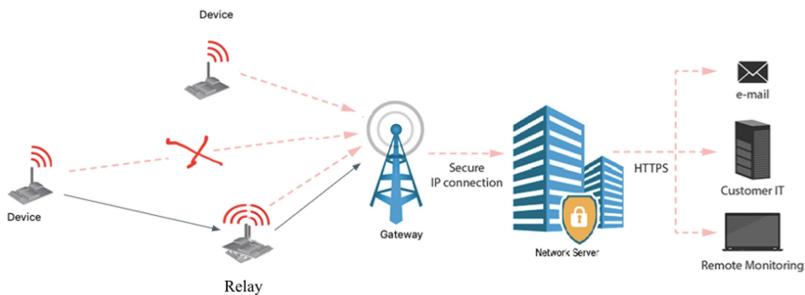
### 3 LoRaWAN End-Node Device for the Coverage Testing

To evaluate the coverage and performance of LoRaWAN network, requiring some end-nodes. The UCA board, which are open-sources development board, has been used to evaluate the LoRaWAN propagation performance since the year 2017 [5]. The antenna of this board is an on-board linear polarized antenna, were discussed in [3]. This UCA board also updated 2021 version with several environment sensor, to show the data such as accelerometer, barometer for the testing. A picture of the board is presented in Fig. 3. The board can be easily mounted with a battery to be used as an autonomous terminal.

**Fig. 3.** UCA lorawan development board version 2021

### 4 Extending the Coverage with LoRaWan Relay

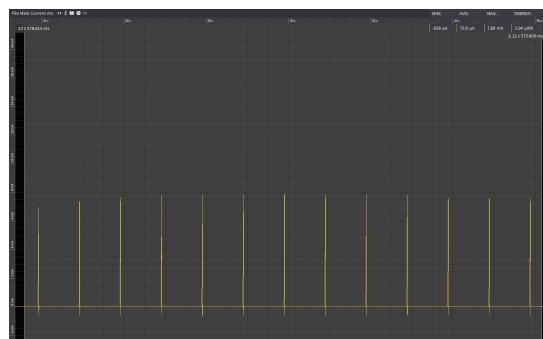
To extend the LoRaWAN network to coverage blind spots in Danang city, one or multiple devices called LoRaWAN Relay are installed between end-devices and LoRaWAN Gateways. Figure 4 demonstrates the LoRaWAN Network Architect with the Relays. The Relay repeats the same packets received from end-devices



**Fig. 4.** LoRaWAN network architecture with LoRa relay



**Fig. 5.** RF200 development board used for relay operation



**Fig. 6.** RF200 Relay's receiving power consumption

in the network to the Gateways or other Relays. Therefore, the Relays are technically transparent in the view of application managers. The minimum hardware requirements for a Relay are a micro-controller and a Lora module connecting to an appropriate-frequency antenna. Figure 5 is the RF200 Kit used as LoRaWAN Relay in LoRaWAN network infrastructure deployed in Da nang city. The RF200

Kit has dimensions of  $100\text{ mm} \times 100\text{ mm} \times 42\text{ mm}$  and is powered by a battery. Along with MCU and LoRa module, the RF200 Kit is also integrated with a GPS module and environmental sensors. These components make monitoring the network operations less complicated. The RF200 Kit has a 3D circular polarized antenna. This antenna structure allows devices to communicate in any alignments.

End devices first send a packet to Gateways to check if it is located in the blind spot of Gateways coverage. If there is no ACK from any Gateways, the end device sends the packet to relays. Meanwhile, LoRaWAN relays are always in the Channel Activity Detect mode. In this mode, the micro-controller is kept in sleep mode and only wake-up to repeat the packet if there is incoming signal. This helps the relays work efficiently with low-power consumption. In the implementation with RF200 Kit, the relay archived the power consumption of ( $1.04\text{ }\mu\text{Wh}$ ) which give around a year of battery life with a  $1200\text{ mAh}$  battery. Figure 6 show the measurement result of RF200 current consumption versus time in relay mode. As it can be seen, a short peak current areas every second to active receiver during  $10\text{ ms}$ . The autonomy of the relay can be easily extended by adding a small harvesting module like solar cell or Thermal Electric Generator (TEG).

## 5 Deployment Plan in the Future

From the estimated data of the LoRaWAN gateway coverage distance in the experiment, the next phase will focus on covering whole Danang city using the LoRaWAN network. In next step, a total from 15 to 20 gateways will be deployed in the city. Additional experiments are also on-going to investigate environmental effects in hard conditions like severe weather (heavy rain, tropical storms). Last but not least, the network will be extended to cover blind spot and deep indoor location using a sub-array of relay that will enable to provide a full and efficient coverage in the entire DaNang area.

## 6 Conclusion

Da Nang city has clearly demonstrated an efficient deployment of a LP-WAN wireless network over a large scale for a whole Danang city in Vietnam, the frequency form  $920\text{ MHz}$  to  $923\text{ MHz}$  was used as a new frequency bands. First experimental results showed that the coverage distance of LoRaWAN Gateway is about  $2\text{ km}$  to  $3\text{ km}$  in the dense urban area, and is around  $10\text{ km}$  for sub-urban areas. A simple extension and coverage of blind spot can be efficiently obtained using low-power consumption relay approach.

**Acknowledgments.** This work is in the frame of Danang Free-LoRaWAN, implemented by Danang City, Côte d'Azur University, University of Science and Technology (DUT) and Danang International Institute of Technology (DNIIT). This work was supported by The Murata Science Foundation and The University of Danang, University of Science and Technology, code number of Project: T2022-02-07 MSF.

## References

1. LoRa Technology Real World Solutions: Smart Cities. Semtech White paper (2018)
2. LoRa Alliance, White Paper, «LoRaWAN, What is it?» (November 2015)
3. Trinh, L.H., et al.: Miniature antenna for IoT devices using LoRa technology. In: 2017 International Conference on Advanced Technologies for Communications (ATC), 18 October 2017, Quy Nhon, VN, pp. 170–173 (2017)
4. Trinh, L.H., Truong, N.V., Ferrero, F.: Low cost circularly polarized antenna for IoT space applications. Electronics **9**, 1564 (2020)
5. <https://github.com/FabienFerrero/UCA21>



# An Effective Beamformer for Interference Mitigation

Luyen Tong<sup>1</sup>(✉) Cuong Nguyen<sup>1</sup>, and Duy Le<sup>2</sup>

<sup>1</sup> Hanoi University of Industry, Hanoi, Vietnam

[luyentv@hau.edu.vn](mailto:luyentv@hau.edu.vn)

<sup>2</sup> Hanoi-Amsterdam High School, Hanoi, Vietnam

**Abstract.** This paper proposes an effective beamformer for a half-wave dipole uniform linear array (DULA) based on binary bat algorithm (BBA) by controlling complex weights (both amplitudes and phases) of each element in the array. The proposed beamformer can mitigate interfering signals by imposing nulls at the direction of interfering signals while maintaining the main lobe and sidelobe levels. To demonstrate this ability, the proposal will be evaluated on several scenarios and compared to a beamformer based on binary particle swarm optimization (BPSO).

**Keywords:** Beamforming · Binary bat algorithm · Optimum array processing · Interference mitigation · Array pattern synthesis

## 1 Introduction

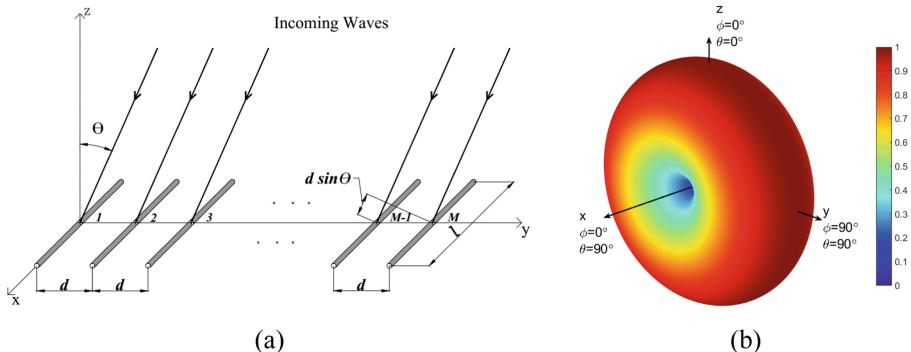
In smart antenna systems, to significantly improve the ability of interference mitigation, energy-saving and spectrum utilization in radar and wireless communications applications, array pattern synthesis has been taken into account in numerous research papers. Some pattern nulling methods including weight control, position-only control [1] and array thinning [2] have been adopted to suppress interferences with their benefits and drawbacks. In the wireless revolution, adaptive beamformers based on weight control methods for interference mitigation are of great interest [3–5].

In excitation weight-based control methods, the simplest one is the amplitude-only control because the only control parameters are weight's amplitudes [5, 6]. Nevertheless, for this control, the main lobe cannot be steered; various kinds of nulls are placed less flexibly, too [6]. The phase-only control is another simple approach due to only controlling phases. The superiority of phase controls is the ease of steering the main lobe and the utilization of existing phased arrays without incurring additional costs [5, 7]. The complex weight control, which simultaneously adjusts both weight's amplitudes and phases, has been deemed to be the best performance for array pattern synthesis, compared to the two aforementioned controls. Due to requiring a complete set of a controller, an attenuator, and a phase shifter for each element, this control may be the most complex and expensive to implement. But in exchange, its flexibility and effectiveness are the most remarkable [5, 8, 9].

Recently, metaheuristic algorithms for optimization such as bat algorithm (BA) [6–8], particle swarm optimization (PSO) [10] and genetic algorithm (GA) [10, 11], which outperform classical optimization techniques, have all been proved to be effective global optimization algorithms to obtain optimal patterns. Among those, BA is superior to the other algorithms on the different types of benchmark functions as well as multiple engineering problems [12, 13]. Adaptive beamformers utilizing BA were first introduced in [14] and successfully performed for uniform linear arrays in [6–8, 15]. The results in [6–8, 15] show that BA-based beamformers have been proved to be completely superior to GA and accelerated PSO-based ones in respect of the pattern nulling. Moreover, except for the solution in [15] utilizing BBA in [16], weight vectors of these beamformers are optimized in the real numbers while the excitation amplitude or phase of elements are commonly adjusted by digital attenuators and/or digital phase shifters, which mitigates the quantization error and the perturbation of array patterns.

## 2 Problem Formulation

In this paper, a half-wave DULA of  $M$  elements illustrated in Fig. 1 (a) with the normalized field pattern of each element displayed in Fig. 1 (b) has been employed.



**Fig. 1.** The geometry of a  $1 \times M$  DULA (a) and the normalized field pattern of a horizontally polarized half-wave dipole (b).

The array factor can be expressed as [17]:

$$AF(\theta) = \sum_{m=1}^M I_m e^{j(m-1)d k \sin(\theta)}, \quad (1)$$

where:  $I_m = a_m e^{j\delta_m}$  is the complex weight;  $\lambda$  is wavelength;  $k = 2\pi/\lambda$ ;  $d = \lambda/2$ .

The pattern of the DULA  $P(\theta)$  is expressed as [4]:

$$P(\theta) = EP(\theta)AF(\theta). \quad (2)$$

Besides, during transferring electromagnetic energy in antenna arrays, radiation characteristics, including impedance and radiation pattern, of an excited antenna element is

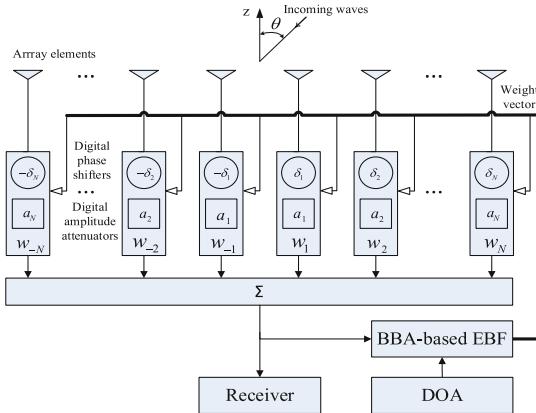
affected by the presence of the other antennas, which is known as mutual coupling (MC). This damaging effect considerably impacts array patterns, for instance, the main lobe direction, the sidelobe level (SLL), and the null depth level (NDL); therefore, it is very necessary to take account of MC effects during the design of adaptive beamformers for pattern nulling.

To characterize MC, therefore, mutual impedance, coupling matrix, S-parameter, or embedded element pattern methods are widely employed [17, 18]. This study will utilize the mutual impedance method as in [7].

### 3 The Proposal of the Beamformer

#### 3.1 The Block Diagram

The diagram of the proposed beamformer with  $M$  antenna elements ( $M = 2N$ ) based on complex weight control is depicted in Fig. 2.



**Fig. 2.** The diagram of the proposed beamformer.

To acquire better efficiency in pattern nulling, the minimum phase perturbation with an odd phase shift [19] ( $\delta_{-n} = -\delta_n$ ) and the amplitude with an even function ( $a_{-n} = a_n$ ) are leveraged in this work. Then, (1) can be reworded as:

$$AF(\theta) = 2 \sum_{n=1}^N a_n \cos(ndk \sin(\theta) + \delta_n), \quad (3)$$

According to (3) and the configuration given in Fig. 2, the number of controllers, attenuators and computational time will be cut in half.

#### 3.2 The Fitness Function

The fitness function in this proposal is developed for the receiver, but this development is similar for the transmitter.

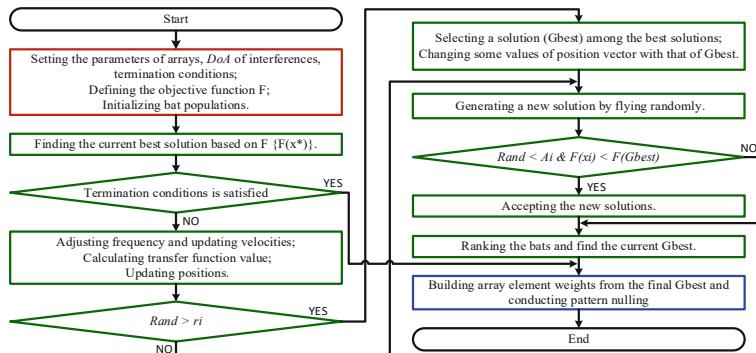
The proposed beamformer requires mitigating interferences while keeping the main lobe and sidelobes at a predefined level. This means that a problem required to solve is a constrained optimization problem. By applying the penalty method in [20], the fitness function to solve this problem can be formulated as:

$$\begin{aligned} F(\mathbf{w}, \xi) &= f(\mathbf{w}) + \xi P_e(\mathbf{w}) \\ &= \sum_{\theta=-90^\circ}^{90^\circ} [ |P_o(\theta) - P_d(\theta)|^2 ] + \xi \sum_{i=1}^{N_i} |P_o(\theta_i)|^2 \end{aligned} \quad (4)$$

where:  $P_o$  and  $P_d$  are the optimized patterns obtained by utilizing optimization algorithms, which will be BBA or BPSO and the Chebyshev method, respectively;  $\theta_i$  and  $N_i$  correspond to the direction and the number of interfering signals.

### 3.3 The Proposed Algorithm

The proposed algorithm for EBF is displayed in Fig. 3, where the termination conditions are chosen as maximum iterations in simulation scenarios apart from the case of determining the computational time presented in Sect. 4.1.



**Fig. 3.** The flowchart of the BBA-based EBF.

## 4 Numerical Results

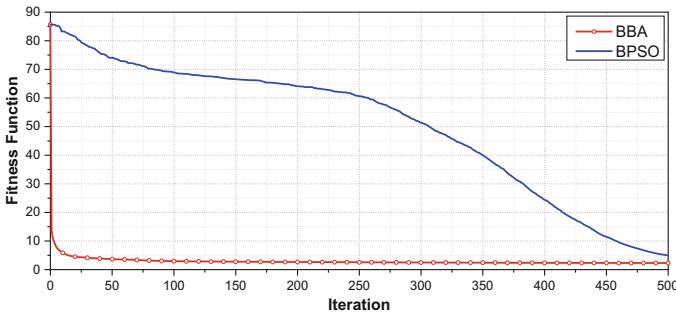
In this section, four scenarios will be implemented to evaluate the performance of the proposal for interference mitigation.

Common parameters for all scenario simulations are set as follows: the half-wave DULA includes 20( $2N$ ) elements;  $P_d$  is the pattern using the Dolph-Chebyshev method [21] with SLL =  $-30$  dB; the penalty parameter  $\xi$  in (4) is 500; the angle step is  $1^\circ$ . The proposed beamformer is optimized for 7-bit attenuators and 7-bit phase shifters where the 3 least significant bits are used for adaptive nulling. The results of all simulation scenarios are the averaged value of 200 Monte Carlo simulations. The population of BBA and BPSO are randomly initialized apart from the first solution initialized by the weights of the Chebyshev pattern.

- BBA:  $f_{min} = 0$  and  $f_{max} = 2$ ;  $A = 0.25$ ;  $r = 0.1$  [16].
- BPSO:  $C_1 = C_2 = 2$ ;  $W = [0.4, 0.9]$ ;  $V_{max} = 6$  [22].
- The transfer function is V-shaped [16, 22].

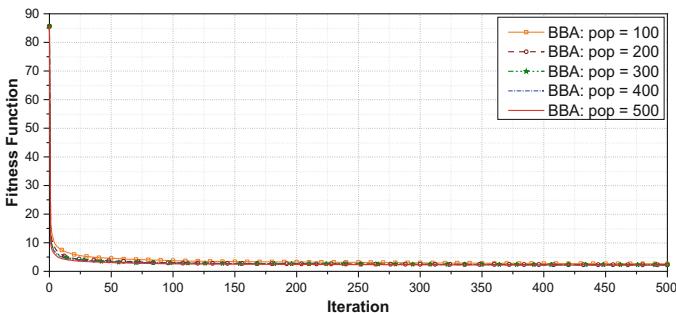
#### 4.1 Convergence Characteristics

In the first scenario, the computational time of the BBA-based EBF has been evaluated and compared to the BPSO-based one with a null placed at a sidelobe peak ( $20^\circ$ ) in the Chebyshev pattern. The simulation results of the fitness function with the population  $pop = 200$  and the iteration  $ite = 500$  are displayed in Fig. 4. The computational time of two beamformers to achieve the same value of the fitness function ( $F < 10$ ) have been **0.25** s and **16.79** s corresponding to BBA-based EBF and BPSO-based one, respectively on a PC (CPU Intel(R) Xeon(R) CPU @ 2.20 GHz, RAM 12 GB). The BBA-based EBF has clearly converged much faster than BPSO-based one.



**Fig. 4.** The comparison of fitness function between BBA-based EBF and BPSO-based one.

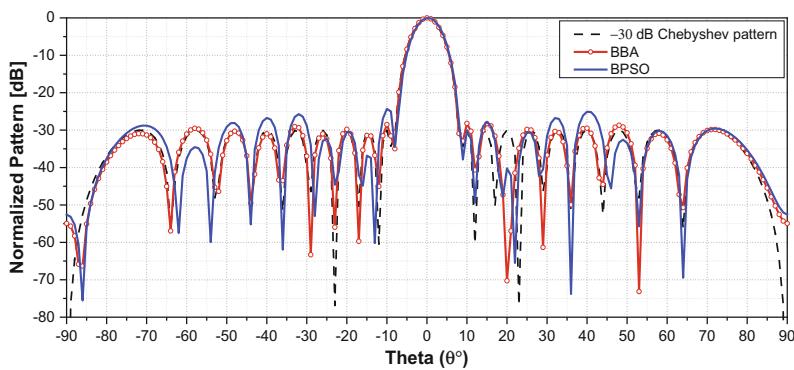
With different population sizes, moreover, the fitness function of the BBA-based EBF is illustrated in Fig. 5. The BBA-based EBF has taken 110 iterations, 50 iterations, 36 iterations, 37 iterations and 22 iterations to roughly converge ( $F \leq 3.65$ ) corresponding to  $pop = 100, 200, 300, 400$  and  $500$ , respectively. For illustrative purposes,  $pop = 200$  and  $iter = 50$  have been set for the next scenarios.



**Fig. 5.** The fitness function of BBA-based EBF with different population sizes.

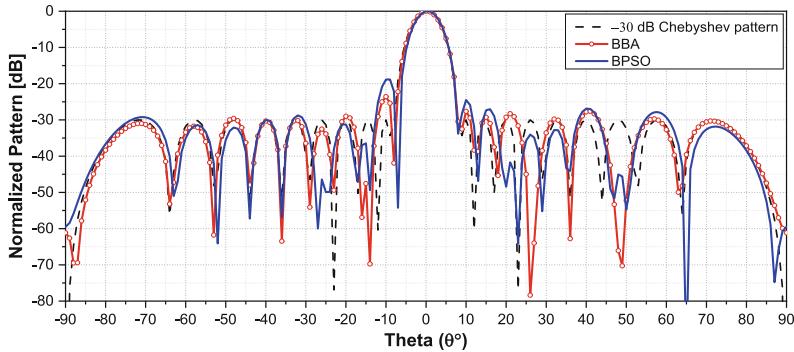
## 4.2 Effective Null-Steering Ability

This scenario presents the adaptive null-steering capability of the proposed beamformer. Firstly, the optimized pattern with a null, which can be arbitrarily set at any angle, has been considered. Figure 6 demonstrates optimized patterns with a null (at  $20^\circ$  in this test case) obtained by BBA and BPSO-based EBF. It can be seen that the optimized pattern of BBA-based EBF has preserved almost characteristics of the Chebyshev pattern such as first null beamwidth (FNBW =  $18^\circ$ ), half-power beamwidth (HPBW =  $6^\circ$ ) and SLL ( $-30$  dB) except for the maximum SLL =  $-28.19$  dB, and NDL at  $20^\circ$  =  $-70.27$  dB. Moreover, Fig. 6 also proves that the pattern optimized by BBA-based EBF has outperformed the BPSO-based one in respect to NDL, maximum SLL and maintaining sidelobes.

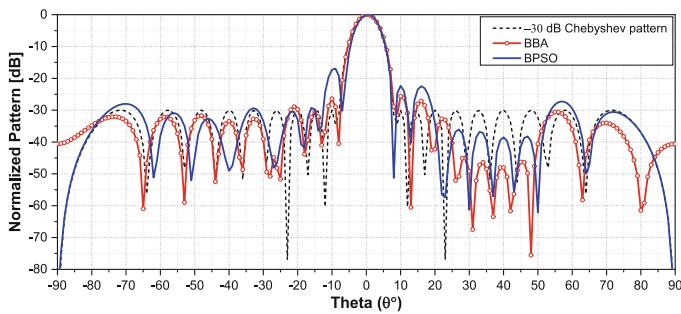


**Fig. 6.** Optimized patterns with a null at  $20^\circ$ .

Secondly, the proposal will be used to separately impose nulls at  $-14^\circ$ ,  $26^\circ$  and  $48^\circ$  corresponding to the peak of sidelobes of the Chebyshev pattern and impose a broad null at a specified sector of  $[25^\circ, 45^\circ]$ . As shown in Fig. 7 and Fig. 8, corresponding to placing multiple nulls and a broad null, the pattern optimized by BBA-based EBF has shown that HPBW and FNBW have been mostly preserved as the Chebyshev pattern, and all NDLs have been successfully imposed. Detailed results for maximum SLL and NDL have been summarized in Table 1.



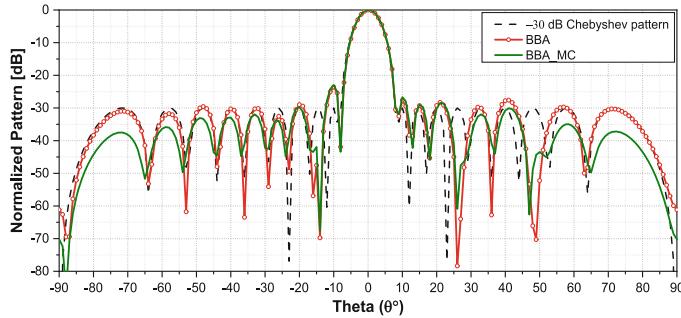
**Fig. 7.** Optimized patterns with three nulls at  $-14^\circ$ ,  $26^\circ$  and  $48^\circ$ .



**Fig. 8.** Optimized patterns with nulls placed from  $25^\circ$  to  $45^\circ$ .

#### 4.3 Optimized Patterns When Considering MC Effects

This scenario considers MC effects by using the mutual impedance method in [7] to obtain the optimized pattern depicted in Fig. 9. The results have shown that separate nulls are successfully set at the direction of interferences but NDLs are shallower. Detailed results taking account of MC for the scenario presented in Sect. 4.2 have been presented in Table 1.



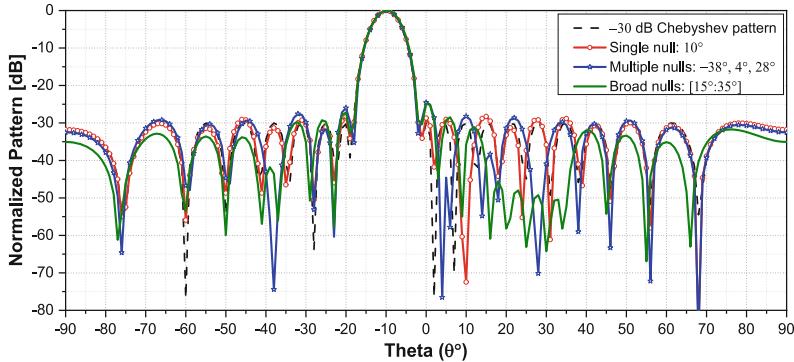
**Fig. 9.** Optimized patterns when considering MC effects (nulls at  $-14^\circ$ ,  $26^\circ$  and  $48^\circ$ ).

**Table 1.** Maximum SLL and NDL of optimized patterns are demonstrated in Fig. 6, 7, 8 and 9.

Fig.	Parameters	BPSO (dB)	BBA (dB)	
		Ideal	Ideal	MC
6	NDL at: $20^\circ$	-40.24	-70.27	<b>-62.89</b>
	Maximum SLL	-24.30	-28.19	<b>-27.05</b>
7	NDL at:	-14°	-49.40	-69.76
		26°	-34.05	-78.36
		48°	-45.05	-66.11
	Maximum SLL	-18.84	-23.57	<b>-23.00</b>
8	Maximum NDL	-61.31	-67.48	<b>-56.71</b>
	Minimum NDL	-36.08	-40.31	<b>-39.70</b>
	Maximum SLL	-16.93	-25.59	<b>-25.07</b>

#### 4.4 Optimized Patterns When Steering the Main Lobe

The efficiency of the proposal with the fixed main lobe direction has been demonstrated above. This solution, however, is also capable of steering the main lobe while placing various nulls. To do that, the main lobe is steered towards the desired direction before performing the aforementioned processes. Figure 10 shows three pattern nulling cases while steering the main lobe towards a predefine direction that is  $-10^\circ$  in this scenario. The simulation results have indicated that the proposal can perform as well as scenarios presented in Sects. 4.2.



**Fig. 10.** Pattern nulling while the main lobe is steered towards  $-10^\circ$ .

## 5 Conclusion

In this paper, a BBA-based EBF for interference mitigation of half-wave DULA has been proposed. The pattern nulling capability of the proposal has been verified via four scenarios, and the proposed beamformer has been proven that outperform the BPSO-based one in terms of execution time and adaptive nulling. For future works, the resolution of phase shifters, unknowing the direction of interferences and the direction of interferences at the main lobe should be considered.

## References

1. Guney, K., Onay, M.: Bees algorithm for null synthesizing of linear antenna arrays by controlling only the element positions. *Neural Netw. World* **17**(2), 153–169 (2007)
2. Haupt, R.L.: Adaptively thinned arrays. *IEEE Trans. Antennas Propag.* **63**(4), 1626–1632 (2015)
3. Charis, G., Showme, N.: Beamforming in wireless communication standards: a survey. *Indian J. Sci. Technol.* **10**(5), 1–5 (2017)
4. Haupt, R.L.: *Antenna Arrays: A Computational Approach*, pp. 156–176 & 484–515. Wiley (2010)
5. Steyskal, H., Shore, R., Haupt, R.: Methods for null control and their effects on the radiation pattern. *IEEE Trans. Antennas Propag.* **34**(3), 404–409 (1986)
6. Luyen, T.V., Giang, T.V.B.: Null-steering beamformer using bat algorithm. *Appl. Comput. Electromagn. Soc. J.* **33**(1), 23–29 (2018)
7. Luyen, T.V., Giang, T.V.B.: Interference suppression of ULA antennas by phase-only control using bat algorithm. *IEEE Antennas Wirel. Propag. Lett.* **16**, 3038–3042 (2017)
8. Luyen, T.V., Giang, T.V.B.: Bat algorithm based beamformer for interference suppression by controlling the complex weight. *REV J. Electron. Commun.* **7**(3–4) (2018)
9. Mouhamadou, M., Vaudon, P.: Complex weight control of array pattern nulling. *Int. J. RF Microw. Comput.-Aided Eng. Co-sponsored Center Adv. Manuf. Packag. Microw. Opt. Digit. Electron. (CAMPmode)* **17**(3), 304–310 (2007)
10. Boeringer, D.W., Werner, D.H.: Particle swarm optimization versus genetic algorithms for phased array synthesis. *IEEE Trans. Antennas Propag.* **52**(3), 771–779 (2004)

11. Khan, M.R.R., Tuzlukov, V.: Null steering beamforming for wireless communication system using genetic algorithm. In: 2011 IEEE International Conference on Microwave Technology & Computational Electromagnetics, pp. 289–292. IEEE (2011)
12. Yang, X.-S.: A new metaheuristic bat-inspired algorithm. In: González, J.R., Pelta, D.A., Cruz, C., Terrazas, G., Krasnogor, N. (eds.) Nature Inspired Cooperative Strategies for Optimization (NICSO 2010). Studies in Computational Intelligence, vol. 284. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-12538-6\\_6](https://doi.org/10.1007/978-3-642-12538-6_6)
13. Yang, X.-S.: Nature-Inspired Optimization Algorithms, pp. 141–154. Academic Press, Cambridge (2014)
14. Yao, Q., Lu, Y.: Efficient beamforming using bat algorithm. In: 2016 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO), pp. 1–2. IEEE (2016)
15. Tong, V.L., Hoang, M.K., Duong, T.H., Pham, T.Q.T., Nguyen, V.T., Truong, V.B.G.: An approach of utilizing binary bat algorithm for pattern nulling. In: Solanki, V.K., Hoang, M.K., Lu, Z., Pattnaik, P.K. (eds.) Intelligent Computing in Engineering. AISC, vol. 1125, pp. 963–973. Springer, Singapore (2020). [https://doi.org/10.1007/978-981-15-2780-7\\_101](https://doi.org/10.1007/978-981-15-2780-7_101)
16. Mirjalili, S., Mirjalili, S.M., Yang, X.-S.: Binary bat algorithm. Neural Comput. Appl. **25**(3–4), 663–681 (2013). <https://doi.org/10.1007/s00521-013-1525-5>
17. Orfanidis, S.J.: Electromagnetic Waves and Antennas, pp. 1233–1251. Rutgers University New Brunswick, NJ (2002)
18. Balanis, C.A.: Antenna Theory: Analysis and Design, 4th edn., pp. 431–480. Wiley, Hoboken (2016)
19. Shore, R.: A proof of the odd-symmetry of the phases for minimum weight perturbation phase-only null synthesis. IEEE Trans. Antennas Propag. **32**(5), 528–530 (1984)
20. Yeniyay, O.: Penalty function methods for constrained optimization with genetic algorithms. Math. Comput. Appl. **10**(1), 45–56 (2005)
21. Dolph, C.L.: A current distribution for broadside arrays which optimizes the relationship between beam width and side-lobe level. Proc. IRE **34**(6), 335–348 (1946)
22. Mirjalili, S., Lewis, A.: S-shaped versus v-shaped transfer functions for binary particle swarm optimization. Swarm Evol. Comput. **9**, 1–14 (2013)



# Support Vector Machine Used for Gender Classification

Meena Tiwari<sup>1</sup>(✉) and Ashish Mishra<sup>2</sup>

<sup>1</sup> SRK University, Bhopal, MP, India

phmeenatiwari@gmail.com

<sup>2</sup> Gyan Ganga Institute of Technology and Sciences, Jabalpur, MP, India

ashishmishra@gkits.org

**Abstract.** Biometric technology has applications in the field of example confirmation, for example single brand confirmation, face confirmation, iris confirmation, palm print confirmation, etc. Grouping by sexual orientation is a test that enables PC agencies to organize gender data and assess age based on images of unique trademarks. Therefore it becomes a staple of PC vision and is confirmed by examples due to its basic application possibilities in character confirmation, video files, robot vision and human-machine interface. In the field of vision of the PC. Highlighting the position is an essential advance in disposition of sexual orientation, because the extraction calculations performed in gender order must remove highlighting from fingerprint. The display of the unique brand confirmation will be affected by the accuracy of the sexual orientation grouping.

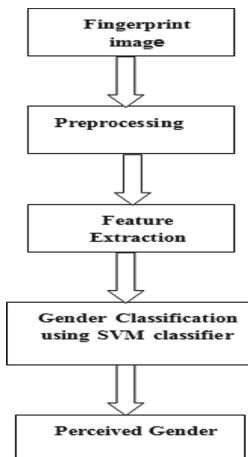
**Keywords:** Support vector machine · Fuzzy C-implies · Gender characterization · Unique mark pictures · Affiliation rule mining

## 1 Introduction

Over time, fingerprints are increasingly used in personally identifiable login and certification applications, as is actual access control and bank security. Although a variety of expanded unique trademark coordination strategies and a wide range of biometric applications are available in various ways, reliable, correct and unique trademark-based gender arrangement technology remains impossible. For single trademark confirmation analysts, characterization of sexual orientation is one of the most tested questions [2]. The extraction of highlights and the order of the samples are two important steps in grouping by gender. It seems that support vector machines are developed in it [9]. There are several applications for grouping by gender. For example, a sexual orientation substantial arrangement framework may provide a reason to use segmented data for inactive observations or to collect valuable measurements of shoppers at strip clubs. Grouping by sexual orientation can improve the presentation of the biometric framework, such as facial verification and confirmation, fingerprint confirmation of inappropriate scenes, etc.

### 1.1 Gender Classification Using SVM

When performing gender characterization through fingerprints, border widths and white lines will be determined. The disposition of sexual orientation depends on various combinations of these priorities. Highlights include edge inspection, valley count, white line count, and the ratio of edge thickness to valley thickness (RTVTR). Support vector machines are used to organize sexual orientation according to a given fingerprint. This depends on the idea of the choice plane that characterizes the choice boundary. A group of projects with unique class entry requirements are isolated by a selection plane. SVM is a non-linear classifier, often used to create incomparable permutations. The proposed SVM classifier plan for the characterization of sexual orientation is shown in Fig. 1.



**Fig. 1.** Scheme of the proposed gender classification

## 2 Related Work

Verma and Agarwal [1] use border thickness, RTVTR, and border width to identify gender. Using the SVM classifier, the completion rate for men reached 86% and that for women reached 90%.

Wang [2] use markers for border, border thickness and finger size to organize the genre. The separate key content of 57 male and 58 female subjects (age range 18 to 35 years) in Taiwan has been broken down. The proposed approach is to test the progress of the gender separation test and the multilayer perceptron (MLP) as a classifier. The best order accuracy of 86% was obtained by edge checking and finger size highlighting.

Eshak [3] use edge markup, square area, and edge thickness as unique markup highlights. In this survey, 380 Egyptian men and 372 women, aged between 20 and 30, were used. The recurring variants survey using multivariate calculations performed measurable inspections in gender groups and established an accuracy of 82%.

Sanders G [4] the number of ridges and the density of the ridges seriously affect the size of the fingertips. If males have higher number of ridges and lower density of ridges than females, then the size of the fingers is more important than the characteristics of the number of ridges and the density of ridges compared to males and females.

Ashish Mishra [5] uses a set of fingerprint data, one is female data and the second male data. This process converts the image data of each fingerprint into a sequence of digital text code and use the predefined minimum confidence and minimum condition support to minimize association rules in the filtered data set. For each text file, classification processing is done by classification and association rule extraction. After classification, the data can be classified as male or female data. This method is based on two fundamental principles of data mining.

Gnanaswamy [6] suggested using a method to classify details in time by subdividing the fingers to determine gender, thus identifying the genes used by the fingers. These components are obtained from the acceleration of the transformation of the medium, the discrete cosine transformation and the power spectral density. Add a single database of 400 data on different ages and sexual orientations. The estimated value of the digital statistical model and its comparison with the recommended conditions. They obtained results of 92.88% and 94.85% for men and women, respectively.

S. Sudha Ponarasi [7] relies on the identifiable uniqueness test strategy of the toothpaste. The subject of the survey is more than 500 people, of which 250 are men and 250 women between 1 and 90. Number of white lines, number of chains and type conditions. The vector machine was used for the most important part and got serious results.

Meena Tiwari et al. [8] in this work used four divisions: the Bassey organization, the multi-stakeholder organization, the closest neighbors, and additional vector equipment. Classification was tested in four prominent studies. These are cases that are questioned with 70% correction, 30% test, 60% preparation, 20% test, finally 60% preparation, 40% test, 10 cases. From the results it can be concluded very well that all the emergence of a common division completes representation of more than 90%. However, SVM is still the best divider proposed to be counted. Fingerprints are strong evidence of legitimacy in court.

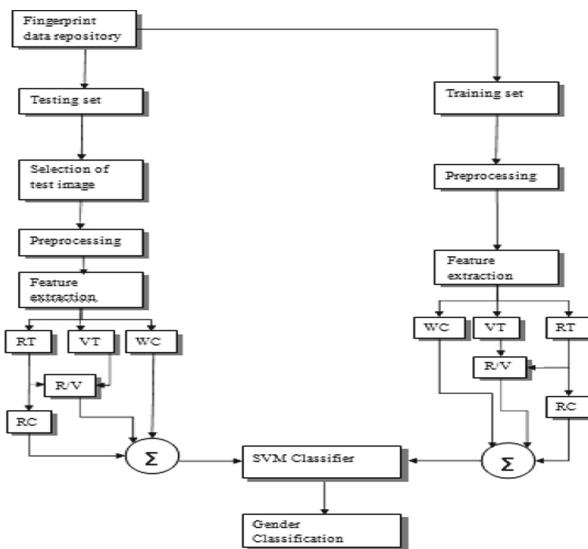
### 3 Proposed Methodology

The proposed technique uses an SVM classifier to classify the distinguishable sexual orientation of uniquely labeled images in one or the other male or female. The use of the SVM classifier to organize the sexual orientation of the uniquely marked images depends on the reflections obtained, it shows an obvious gender. To summarize the sexual orientation sequencing strategy for any uniquely marked image, a general SVM-based grouping calculation has been proposed using fingerprint highlighting (Fig. 2)..

#### 3.1 Algorithm

STEP 1: Acquire the training set and test set of fingerprint images.

STEP 2: Preprocess the image.



**Fig. 2.** Flowchart for the proposed algorithm

STEP 3: Calculate Ridge Thickness, Valley Thickness, White lines Count.  
 STEP 4: Compute RT to VT Ratio (RTVTR).

STEP 5: Input RTVTR, RC, WC to SVM Classifier.

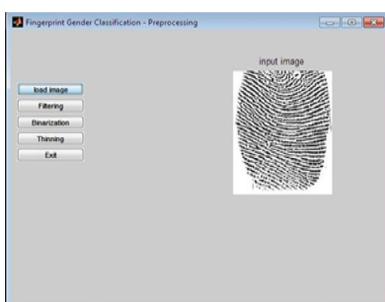
STEP 6: Classify the gender according to RTVTR, RC, WC.

STEP 7: Obtain the test set image.

STEP 8: Repeat step 2–6 for various test images until the end of the test image

## 4 Result and Discussion

In this part, the exhibition of the proposed strategy for characterization has been explored. At first twenty fingerprints are chosen. The info pictures of the fingerprints are given as

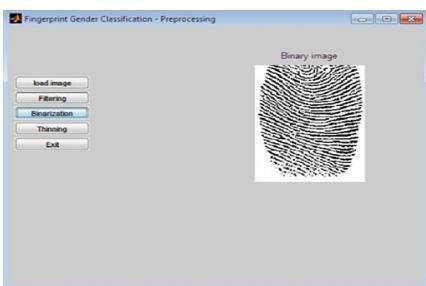


**Fig. 3.** Input image for gender classification

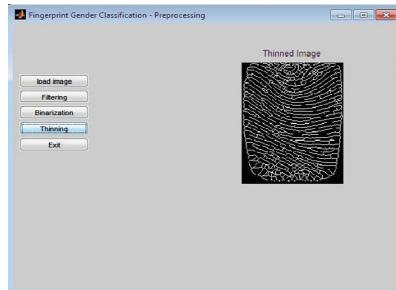


**Fig. 4.** Filtered image gender classification

information consistently as appeared in Fig. 3. Figure 4 shows the separated picture of the unique mark picture. Parallel pictures contain just two qualities either 0 or 1. It is given in Fig. 5. Figure 6 shows the diminished picture of the binarized picture. Along these lines the preprocessing strategies are completed to get the exact outcomes. In the subsequent stage, the element extraction is completed.



**Fig. 5.** Binarized image for gender classification



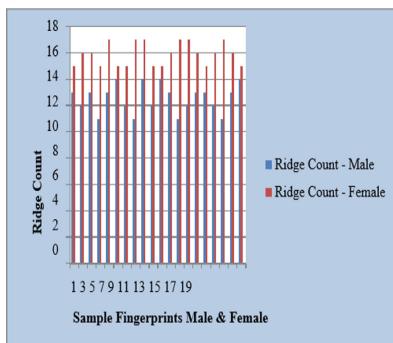
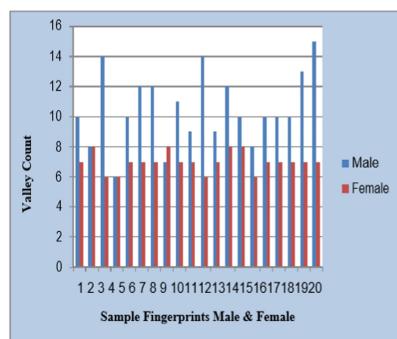
**Fig. 6.** Thinned image gender classification

Table 1 shows the border verification of randomly selected male and female fingerprints, as shown below. The related histogram is shown in Fig. 7. Table 2 shows the bottom count of randomly selected male and female fingerprints. The histogram comparison is shown in Fig. 8.

**Table 1.** Comparison of ridge count of male and female

Sl. No	Male	Female	Sl. No	Male	Female
1.	13	15	11	14	15
2.	12	16	12	13	16
3.	13	16	13	11	17
4.	11	15	14	12	17
5.	13	17	15	13	16
6.	14	15	16	13	15
7.	12	15	17	12	16
8.	11	17	18	11	17
9.	14	17	19	13	16
10.	12	15	20	14	15

Table 3 shows the relationship between edge thickness and background thickness (RTVTR) of 20 uniquely marked randomly selected male and 20 female images. Additionally, the male and female RTVTR histograms are drawn independently and are typically spoken to simplify correlation. They appear in Fig. 9, Fig. 10, and Fig. 11.

**Fig. 7.** Histogram of the ridge count**Fig. 8.** Histogram of the valley count**Table 2.** Comparison of valley count of male and female

Sl. No.	Male	Female	Sl. No.	Male	Female
1.	10	7	11.	14	6
2.	8	8	12.	9	7
3.	14	6	13.	12	8
4.	6	6	14.	10	8
5.	10	7	15.	8	6
6.	12	7	16.	10	7
7.	12	7	17.	10	7
8.	7	8	18.	10	7
9.	11	7	19.	13	7
10.	9	7	20.	15	7

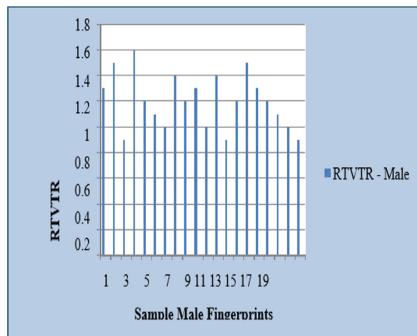
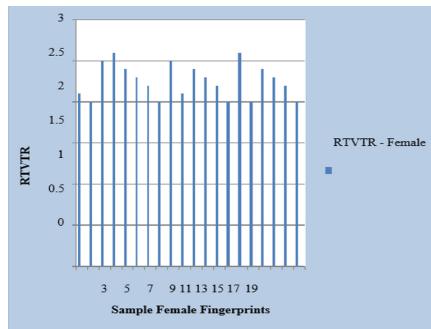
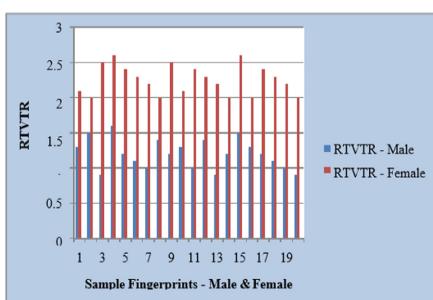
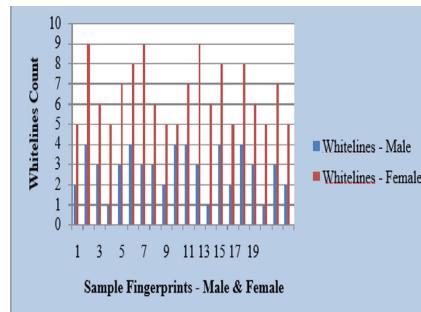
**Table 3.** Comparison of RTVTR of male and female

Sl. No.	Male	Female	Sl. No.	Male	Female
1.	1.3	2.1	11.	1.0	2.4
2.	1.5	2.0	12.	1.4	2.3
3.	0.9	2.5	13.	0.9	2.2
4.	1.6	2.6	14.	1.2	2.0
5.	1.2	2.4	15.	1.5	2.6
6.	1.1	2.3	16.	1.3	2.0

(continued)

**Table 3.** (continued)

Sl. No.	Male	Female	Sl. No.	Male	Female
7.	1.0	2.2	17.	1.2	2.4
8.	1.4	2.0	18.	1.1	2.3
9.	1.2	2.5	19.	1.0	2.2
10.	1.3	2.1	20.	0.9	2.0

**Fig. 9.** Histogram of the RTVTR (male)**Fig. 10.** Histogram of the RTVTR (female)**Fig. 11.** Histogram of the RTVTR**Fig. 12.** Histogram of the white lines count

The histogram for white lines include is outlined in Fig. 12 (Table 4).

**Table 4.** Comparison of white lines count of male and female

Sl. No.	Male	Female	Sl. No.	Male	Female
1.	2	5	11.	4	7
2.	4	9	12.	3	9
3.	3	6	13.	1	6
4.	1	5	14.	4	8
5.	3	7	15.	2	5
6.	4	8	16.	4	8
7.	3	9	17.	3	6
8.	3	6	18.	1	5
9.	2	5	19.	3	7
10.	4	5	20.	2	5

Contingent on the element estimations of the prepared SVM classifier, the test pictures are being arranged regarding the sexual orientation them groups. The last yield is given in the GUI window as shown in Fig. 13.

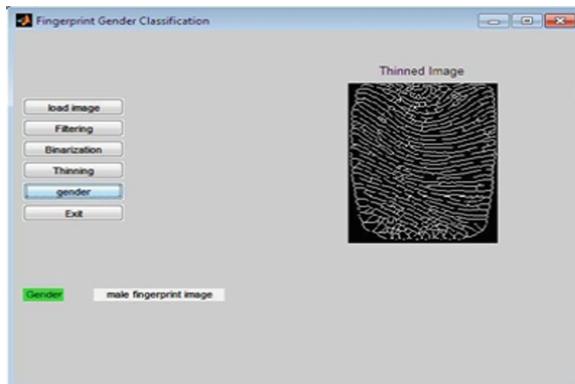
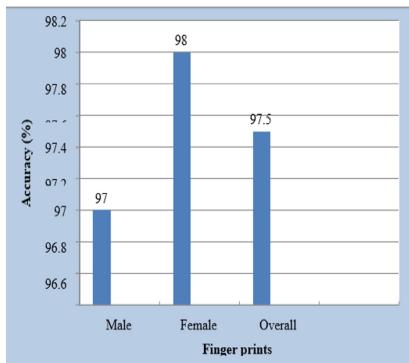
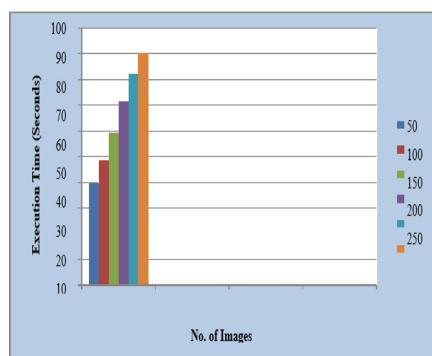
**Fig. 13.** Gender classification

Table 5 shows the exposure of gender classification based on SVM. Table 6 shows the performance metrics for this job. The figure shows a graphical survey of this correlation.

See also Fig. 14. The accuracy of Fig. 15 depends on the number of images that are accurately grouped into the total number of images in the library of tested information. Execution time is determined using the Tic and Tac instructions in the MATLAB program. It can be seen that the precision of 200 images is 97.5%; execution time is 71.35 s.

**Table 5.** Accuracy of SVM based gender classification

No. of images	Tested	Correctly classified	Accuracy (%)
Male	100	97	97
Female	100	98	98
Total	200	195	97.5

**Fig. 14.** Accuracy of SVM based gender classification**Fig. 15.** Graph execution time in sec**Table 6.** Execution time for SVM based gender classification

No. of images	50	100	150	200	250	300
Execution Time (Seconds)	39.42	48.54	59.15	71.35	82.26	90.57

## 5 Conclusion

Tests on different images found that the best method and strategy for gender distribution from fingerprints is based on SVM. Comparing presentation metrics (such as execution time and precision of the proposed strategy) and presentation metrics (such as execution time and precision), these methods are used as different methods of sexual orientation classifiers. Therefore, the SVM-based classifier is a better decision.

## References

- Verma, M., Agarwal, S.: Fingerprint based male-female classification. In: Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems, vol. 53, pp. 251–257 (2008)
- Wang, J., Lin, C., Chang, Y., Nagurka, M., Yen, C., Yen, C.: Gender determination using fingertip features. Internet J. Med. Update 3(2), 22–28 (2008)

3. Eshak, G.A., Zaher, J.F., Hasan, E.I., El-Azeem Ewis, A.A.: Sex identification from fingertip features in Egyptian population. *J. Forensic Leg. Med.* **20**(1), 46–50 (2013)
4. Sanders, G., Kadam, A.: Prepubescent children show the adult relationship between dermatoglyphic asymmetry and performance on sexually dimorphic tasks. *Cortex*. **37**(1), 91–100 (2001)
5. Mishra, A., Khare, N.: A review on gender classification using association rule mining and classification based on fingerprints. In: Fifth International Conference on Communication Systems and Network Technologies (2015)
6. Gnanasivam, P., Muttan, S.: Fingerprint gender classification using wavelet transform and singular value decomposition. *Int. J. Comput. Sci.* **9**(2), 3 (2012)
7. Sudha Ponarasi, S., Rajaram, M.: Gender classification system derived from fingerprint minutiae extraction. In: International Conference on Recent Trends in Computational Methods, Communication and Controls Proceedings Published in International Journal of Computer Applications (IJCA) (2012)
8. Tiwari, M., et al.: Development of association rule mining model for gender classification. In: 1st International Conference Computational Research and Data Analytics, (ICCRDA 2020) (2021)
9. Hsu, C.-W., Chang, C.-C., Lin, C.-J.: A practical guide to support vector classification (2003)
10. Gutta, S., Huang, J.R.J., Jonathon, P., Wechsler, H.: Mixture of experts for classification of gender, ethnic origin, and pose of human faces. *IEEE Trans. Neural Netw.* **11**(4), 948–960 (2000)
11. Brown, W., Hines, M., Fane, B., Breedlove, M.: Masculinized finger length patterns in human males and females with congenital adrenal hyperplasia. *Horm. Behav.* **42**(4), 380–386 (2002)
12. Nithin, M.D., et al.: Study of fingerprint classification and their gender distribution among South Indian population. *J. Forensic Leg. Med.* **16**(8), 460–463 (2009)
13. Kaur, R., Garg, R.K.: Determination of gender differences from fingerprint ridge density in two northern Indian populations. *Prob. Forensic Sci.* **85**, 5–10 (2011)
14. Chang, H., Zhao, X., Li, B.: Application of BP neural network in the fingerprint identification. In: International Conference on Computer Technology and Science, ICCTS 2012, IPCSIT, vol. 47 (2012)



# Relay Selection for Enhancing the Performance of Full-Duplex System with Imperfect Transceiver Hardware

Ba Cao Nguyen<sup>1,2</sup>, Le The Dung<sup>1</sup>, Tran Manh Hoang<sup>1,2</sup>, Xuan Nam Tran<sup>3</sup>,  
and Taejoon Kim<sup>1(✉)</sup>

<sup>1</sup> Chungbuk National University, Cheongju, Korea  
[nguyenbacao,tranmanhhoang}@tcu.edu.vn](mailto:{nguyenbacao,tranmanhhoang}@tcu.edu.vn), [dung.t.le@ieee.org](mailto:dung.t.le@ieee.org),  
[ktjcc@chungbuk.ac.kr](mailto:ktjcc@chungbuk.ac.kr)

<sup>2</sup> Telecommunications University, Khanh Hoa, Vietnam  
<sup>3</sup> Le Quy Don Technical University, Ha Noi, Vietnam  
[namtx@mta.edu.vn](mailto:namtx@mta.edu.vn)

**Abstract.** This paper applies the relay selection to enhance the full-duplex relay (FDR) system performance under the combined effects of two imperfect parameters, namely imperfect transceiver hardware (ITH) and imperfect self-interference cancellation (ISIC). In this circumstance, the outage and ergodic (OP & EC) expressions of the FDR system are derived over the Rayleigh fading channel for evaluation. From these expressions, it is easy to derive OP & EC expressions of FDR system with perfect transceiver hardware (PTH) and perfect self-interference cancellation (PSIC). We observe that the combined effects of ITH and ISIC are significant, especially in the high SNR region. In particular, compared with PTH and PSIC, ITH and ISIC give considerably lower performance. However, by applying relay selection, the effects of ITH and ISIC can be significantly reduced.

**Keywords:** Full-duplex relay selection · Imperfect hardware · Imperfect si cancellation · Outage probability · Ergodic capacity

## 1 Introduction

To satisfy a high capacity requirement of wireless systems in the age of the 5G and beyond networks, many solutions such as full-duplex (FD) transmission, unmanned aerial vehicles, and intelligent reflecting surfaces have been proposed and investigated [8, 15]. Among many techniques, FD transmission is still a potential candidate because its main advantage is having a two-time higher capacity than classical half-duplex (HD) transmission. However, a large amount of residual SI after all cancellation methods is challenging for deploying FD systems in practice. Despite this issue, it is expected that FD transmission can be used in the 6G of wireless networks [9].

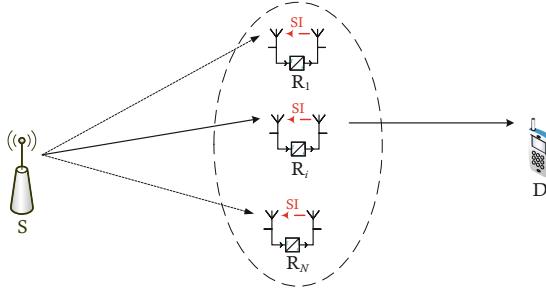
Besides high capacity, increasing the coverage is also a crucial requirement [2]. Thus, the traditional relay technique used in the 3G & 4G is still valuable for 5G & 6G wireless systems [2]. Since the relay is a low-cost and easily deployed device in practice, it can greatly improve the coverage of wireless systems. Therefore, when FD transmission is exploited, the FD relay (FDR) can receive and transmit signals simultaneously, leading to a remarkable increase in the wireless systems' capacity, coverage, and quality. Consequently, the usage of FDR in wireless systems has been widely applied for many purposes, such as in the Internet of things systems [3,13], cooperative networks [4], cognitive systems [5], and vehicle-to-vehicle systems [12].

In the literature, the FDR system performance has been evaluated under the combined effects of imperfect parameters such as imperfect SI cancellation (ISIC), imperfect CSI (ICSI), and imperfect transceiver hardware (ITH) [9]. It was proved that outage probability (OP) of the FDR systems reaches an error floor in the high SNR region even only one imperfect factor exists in the systems. Therefore, besides effectively applying various SI cancellation solutions, many methods for reducing the error floor caused by ITH and ICSI are required in the FDR systems [2]. Additionally, the usage of multi-antenna at the transmitter/receiver together with relevant techniques is still a suitable method for enhancing the performance of FDR systems [7]. Furthermore, when multiple relays are deployed, the relay selection scheme can significantly increase the performance of FDR systems [12]. However, these papers considered the cases that perfect channel state information (PCSI) and perfect transceiver hardware (PTH). In other words, using multiple relays in FDR systems with ISIC and ITH was not investigated.

The above observations motivate us to determine an FDR system with multiple relays under the effects of ISIC and ITH. Specifically, we apply the relay selection to increase the received signal power at FD relays. The main contributions of our work can be summarized as follows.

- We investigate an FDR system where multiple FD relays are deployed. Unlike previous papers, the ISIC and ITH are studied in our work. Thus, the considered FDR system is suitable in practical scenarios. Importantly, we apply relay selection to enhance the received signal power at FD relays. Thus, the performance of the FDR system is greatly improved.
- We derive the OP & EC expressions of the FDR system with a relay selection scheme over the Rayleigh fading channel under the effects of ISIC and ITH. We validate the derived expressions through computer simulations. We observe that under the combined impacts of these imperfect factors, the OP & EC of the FDR system reach constant values in the high SNR region. Furthermore, using three relays dramatically improves the FDR system performance than using only one relay. Also, the system behavior in the case of ISIC and ITH is compared with that in the case of PCSI and PTH to clarify the effects of ISIC and ITH.

The remainder of this paper is structured as follows. Section 2 presents the model where ISIC and ITH are detailedly provided. Section 3 gives mathematical



**Fig. 1.** FDR system with multiple relays.

calculations to obtain the OP and EC expressions with relay selection. Section 4 uses these derived expressions to analyze the effects of many factors such as the number of relays, ISIC level, and ITH level. Finally, Sect. 5 concludes this work.

## 2 System Model

Figure 1 describes the model of the system. Specifically, a base station (S) transmits signals to a user (D) through the support of  $N$  relays (Rs). In particular, S & D work in traditional HD mode, while Rs works in FD mode. S & D are equipped with an antenna for transmission/reception, while Rs is configured with two antennas, one for transmission and another for the reception. In practice, Rs can use only one shared antenna for both transmission and reception; however, the usage of two separate antennas as in our model can significantly increase the self-interference cancellation capability [6].

The received signal at  $R_i$  is

$$y_{R_i} = h_{SR_i}(x_S + \eta_S^t) + \eta_{R_i}^r + h_{R_iR_i}(x_{R_i} + \zeta_{R_i}^t) + \zeta_{R_i}^r + z_{R_i}, \quad (1)$$

where  $x_A$  ( $A \in \{S, R_i\}$ ) is the intended signals of A;  $\eta_S^t$  and  $\eta_{R_i}^r$  are the ITH at transmitter S and receiver  $R_i$  when transmitting/receiving  $x_S$ , respectively;  $h_{SR_i}$  is the helpful channel from S to  $R_i$ ;  $h_{R_iR_i}$  is the SI channel between two antennas of  $R_i$ ;  $\zeta_{R_i}^t$  and  $\zeta_{R_i}^r$  are the ITH at the transmitter and receiver  $R_i$  when transmitting/receiving  $x_{R_i}$ ;  $z_{R_i} \sim \mathcal{CN}(0, \sigma_{R_i}^2)$  is the Gaussian noise at  $R_i$ .

It is worth noticing that the ITH noises at transceivers are expressed as [14]  $\eta_S^t \sim \mathcal{CN}(0, (k_S^t)^2 P_S)$ ,  $\eta_{R_i}^r \sim \mathcal{CN}(0, |h_{SR_i}|^2 (k_{R_i}^r)^2 P_S)$ ,  $\zeta_{R_i}^t \sim \mathcal{CN}(0, (k_{R_i}^t)^2 P_{R_i})$ , and  $\zeta_{R_i}^r \sim \mathcal{CN}(0, |h_{R_iR_i}|^2 (k_{R_i}^r)^2 P_{R_i})$ , where  $P_A$  denotes the transmit power of A ( $A \in \{S, R_i\}$ );  $k_S^t$ ,  $k_{R_i}^r$ , and  $k_{R_i}^t$  denote the ITH levels at the transmitter S, receiver  $R_i$ , and transmitter  $R_i$ , respectively.

Using the above representations of  $\eta_S^t$ ,  $\eta_{R_i}^r$ ,  $\zeta_{R_i}^t$ , and  $\zeta_{R_i}^r$ , we rewrite (1) as

$$y_{R_i} = h_{SR_i}(x_S + \eta_S^{tr}) + h_{R_iR_i}(x_{R_i} + \zeta_{R_i}^{tr}) + z_{R_i}, \quad (2)$$

where  $\eta_S^{tr}$  and  $\zeta_{R_i}^{tr}$  are the aggregated THN noises caused by transmitter S and receiver  $R_i$  (combination of  $\eta_S^t$  and  $\eta_{R_i}^r$ ) and by transmitter  $R_i$  and receiver  $R_i$  (combination of  $\zeta_{R_i}^t$  and  $\zeta_{R_i}^r$ ), respectively.

Since the term  $h_{R_i R_i}(x_{R_i} + \zeta_{R_i}^{\text{tr}})$  is the SI induced by FD transmission including ITH noises, it has to be canceled. In particular,  $R_i$  applies all SI cancellation solutions including in propagation, analog, and digital domains. Then, the residual SI (denoted by  $I_{R_i}$ ) follows Gaussian distributed random variable, i.e.,  $I_{R_i} \sim \mathcal{CN}(0, l^2 P_{R_i})$ , where  $l^2$  denotes the residual SI at  $R_i$  [10]. Now, (2) becomes

$$y_{R_i} = h_{SR_i}(x_S + \eta_S^{\text{tr}}) + I_{R_i} + z_{R_i}. \quad (3)$$

Then, the relay handles the received signal via decoding, recoding, and forwarding it to user D. The received signal at D is

$$y_D = h_{R_i D}(x_R + \eta_{R_i}^t) + \eta_D^r + z_D, \quad (4)$$

where  $h_{R_i D}$  is the helpful channel from  $R_i$  to D;  $\eta_{R_i}^t \sim \mathcal{CN}(0, (k_{R_i}^t)^2 P_{R_i})$  and  $\eta_D^r \sim \mathcal{CN}(0, |h_{R_i D}|^2 (k_D^r)^2 P_{R_i})$  are the THN noises at the transmitter  $R_i$  and receiver D, respectively, where  $k_{R_i}^t$  and  $k_D^r$  refer to the ITH levels at  $R_i$  and D, respectively;  $x_D \sim \mathcal{CN}(0, \sigma_D^2)$  is Gaussian noise at D.

Assuming that all relays have similar conditions. In other words, we have  $P_{R_1} = P_{R_2} = \dots = P_{R_N} = P_R$ ,  $k_{R_1}^t = k_{R_2}^t = \dots = k_{R_N}^t = k_R^t$ ,  $k_{R_1}^r = k_{R_2}^r = \dots = k_{R_N}^r = k_R^r$ , and  $\sigma_{R_1}^2 = \sigma_{R_2}^2 = \dots = \sigma_{R_N}^2 = \sigma_R^2$ . Then, based on (3) and (4), the instantaneous signal-to-interference-plus-noise-and-distortion ratio SINDR at  $R_i$  ( $\gamma_{R_i}$ ) & D ( $\gamma_D$ ) are, respectively, computed as

$$\gamma_{R_i} = \frac{|h_{SR_i}|^2 P_S}{|h_{SR_i}|^2 [(k_S^t)^2 + (k_R^r)^2] P_S + l^2 P_R + \sigma_R^2} = \frac{|h_{SR_i}|^2 P_S}{|h_{SR_i}|^2 k_{SR}^2 P_S + l^2 P_R + \sigma_R^2}, \quad (5)$$

$$\gamma_D = \frac{|h_{R_i D}|^2 P_R}{|h_{R_i D}|^2 [(k_R^t)^2 + (k_D^r)^2] P_R + \sigma_D^2} = \frac{|h_{R_i D}|^2 P_R}{|h_{R_i D}|^2 k_{RD}^2 P_R + \sigma_D^2}, \quad (6)$$

where  $k_{SR}^2 = (k_S^t)^2 + (k_S^r)^2 + (k_R^r)^2$  and  $k_{RD}^2 = (k_R^t)^2 + (k_D^r)^2$  represent the aggregated ITH levels at S ( $k_S^t$ ) and  $R_i$  ( $k_R^r$ ) and at  $R_i$  ( $k_R^t$ ) and D ( $k_D^r$ ), respectively.

Since  $R_i$  uses decode-and-forward protocol, the end-to-end SINDR ( $\gamma_{e2e}$ ) is

$$\gamma_{e2e} = \min\{\gamma_{R_i}, \gamma_D\}. \quad (7)$$

### 3 Performance Analysis

We mathematically calculate the OP & EC expressions of the FDR system with multiple relays under the effects of ISIC and ITH.

#### 3.1 Outage Probability

The OP is computed as

$$\mathcal{P}_{\text{out}} = \Pr\{\log_2(1 + \gamma_{e2e}) < \mathcal{R}\} = \Pr\{< x\}, \quad (8)$$

where  $\gamma_{e2e}$  is given in (7) and  $x = 2^{\mathcal{R}} - 1$ . Herein,  $\mathcal{R}$  is required rate of the FDR system.

**Theorem 1.** *The system OP with relay selection under the effects of ISIC and ITH is given as*

$$\mathcal{P}_{\text{out}} = \begin{cases} 1 - \exp\left(-\frac{\sigma_{\text{D}}^2 x}{\Omega_{\text{RD}} P_{\text{R}}(1-k^2 x)}\right) \left[1 - \left[1 - \exp\left(-\frac{(l^2 P_{\text{R}} + \sigma_{\text{R}}^2)x}{\Omega_{\text{SR}} P_{\text{S}}(1-k^2 x)}\right)\right]^N\right], & x < 1/k^2, \\ 1, & x \geq 1/k^2. \end{cases} \quad (9)$$

where  $k^2 = \max\{k_{\text{SR}}^2, k_{\text{RD}}^2\}$ ;  $\Omega_{\text{SR}}$  and  $\Omega_{\text{RD}}$  denote the average  $\text{S} - \text{R}_i$  and  $\text{R}_i - \text{D}$  channel gains, respectively.

*Proof:* Replacing (7) into (8), we have

$$\begin{aligned} \mathcal{P}_{\text{out}} &= \Pr\{\min\{\gamma_{\text{R}_i}, \gamma_{\text{D}}\} < x\} = \Pr\{\gamma_{\text{R}_i} < x\} + \Pr\{\gamma_{\text{D}} < x\} \\ &\quad - \Pr\{\gamma_{\text{R}_i} < x\}\Pr\{\gamma_{\text{D}} < x\} = 1 - [1 - \Pr\{\gamma_{\text{R}_i} < x\}][1 - \Pr\{\gamma_{\text{D}} < x\}]. \end{aligned} \quad (10)$$

Since the relay selection scheme is applied, we have  $|h_{\text{SR}_i}|^2 = \max(|h_{\text{SR}_1}|^2, |h_{\text{SR}_2}|^2, \dots, |h_{\text{SR}_N}|^2)$ . Thus, the probability  $\Pr\{\gamma_{\text{R}_i} < x\}$  is calculated as

$$\Pr\{\gamma_{\text{R}_i} < x\} = \Pr\left\{\frac{|h_{\text{SR}_i}|^2 P_{\text{S}}}{|h_{\text{SR}_i}|^2 k_{\text{SR}}^2 P_{\text{S}} + l^2 P_{\text{R}} + \sigma_{\text{R}}^2} < x\right\} = \Pr\left\{|h_{\text{SR}_i}|^2 P_{\text{S}}(1 - k_{\text{SR}}^2 x) < (l^2 P_{\text{R}} + \sigma_{\text{R}}^2)x\right\}. \quad (11)$$

It is easy to see that in the case  $1 - k_{\text{SR}}^2 x \leq 0$  or  $x \geq 1/k_{\text{SR}}^2$ , the probability in (11) is always true. Thus, we have  $\Pr\{\gamma_{\text{R}_i} < x\} = 1$  for  $x \geq 1/k_{\text{SR}}^2$ . In the case  $1 - k_{\text{SR}}^2 x > 0$  or  $x < 1/k_{\text{SR}}^2$ , (11) is equivalent to

$$\begin{aligned} \Pr\{\gamma_{\text{R}_i} < x\} &= \Pr\left\{|h_{\text{SR}_i}|^2 < \frac{(l^2 P_{\text{R}} + \sigma_{\text{R}}^2)x}{P_{\text{S}}(1 - k_{\text{SR}}^2 x)}\right\} \\ &= \Pr\left\{\max(|h_{\text{SR}_1}|^2, |h_{\text{SR}_2}|^2, \dots, |h_{\text{SR}_N}|^2) < \frac{(l^2 P_{\text{R}} + \sigma_{\text{R}}^2)x}{P_{\text{S}}(1 - k_{\text{SR}}^2 x)}\right\}. \end{aligned} \quad (12)$$

Applying the CDF of the Rayleigh distributed channel gain [11], (12) becomes

$$\Pr\{\gamma_{\text{R}_i} < x\} = \left[1 - \exp\left(-\frac{(l^2 P_{\text{R}} + \sigma_{\text{R}}^2)x}{\Omega_{\text{SR}} P_{\text{S}}(1 - k_{\text{SR}}^2 x)}\right)\right]^N. \quad (13)$$

Combining of two above cases, we have

$$\Pr\{\gamma_{\text{R}_i} < x\} = \begin{cases} \left[1 - \exp\left(-\frac{(l^2 P_{\text{R}} + \sigma_{\text{R}}^2)x}{\Omega_{\text{SR}} P_{\text{S}}(1 - k_{\text{SR}}^2 x)}\right)\right]^N, & x < 1/k_{\text{SR}}^2 \\ 1, & x \geq 1/k_{\text{SR}}^2 \end{cases}. \quad (14)$$

Similarly, the probability  $\Pr\{\gamma_{\text{D}} < x\}$  is calculated as

$$\Pr\{\gamma_{\text{D}} < x\} = \begin{cases} 1 - \exp\left(-\frac{\sigma_{\text{D}}^2 x}{\Omega_{\text{RD}} P_{\text{R}}(1 - k_{\text{RD}}^2 x)}\right), & x < 1/k_{\text{RD}}^2 \\ 1, & x \geq 1/k_{\text{RD}}^2 \end{cases}. \quad (15)$$

Replacing (14) and (15) into (10), we obtain (9) of Theorem 1.

### 3.2 Ergodic Capacity

The system EC is

$$\mathcal{C} = \mathbb{E} \left\{ \log_2(1 + \gamma_{e2e}) \right\} = \int_0^{\infty} \log_2(1 + y) f_{\gamma_{e2e}}(y) dy, \quad (16)$$

where  $f_{\gamma_{e2e}}(y)$  is the PDF of  $\gamma_{e2e}$ . From (16), we obtain the EC as follows.

**Theorem 2.** *The system EC with relay selection under the effects of ISIC and ITH is given as*

$$\begin{aligned} \mathcal{C} &= \frac{\pi}{M \ln 2} \sum_{m=1}^M \frac{\sqrt{1 - \beta^2}}{2k^2 + 1 + \beta} \exp \left( - \frac{\sigma_D^2(1 + \beta)}{\Omega_{RD} P_R k^2 (1 - \beta)} \right) \\ &\quad \times \left[ 1 - \left[ 1 - \exp \left( - \frac{(l^2 P_R + \sigma_R^2)(1 + \beta)}{\Omega_{SR} P_S k^2 (1 - \beta)} \right) \right]^N \right], \end{aligned} \quad (17)$$

where  $\beta = \cos \left( \frac{(2m-1)\pi}{2M} \right)$ ;  $M$  is a trade-off parameter [1].

*Proof:* From (16), the EC can be computed as

$$\mathcal{C} = \frac{1}{\ln 2} \int_0^{\infty} \frac{1 - F_{\gamma_{e2e}}(y)}{1 + y} dy. \quad (18)$$

Since  $F_{\gamma_{e2e}}(y)$  is formulated by  $F_{\gamma_{e2e}}(x) = \Pr\{\gamma_{e2e} < x\}$ , we can easily derive  $F_{\gamma_{e2e}}(y)$  as

$$F_{\gamma_{e2e}}(y) = \begin{cases} 1 - \exp \left( - \frac{\sigma_D^2 y}{\Omega_{RD} P_R (1 - k^2 y)} \right) \left[ 1 - \left[ 1 - \exp \left( - \frac{(l^2 P_R + \sigma_R^2) y}{\Omega_{SR} P_S (1 - k^2 y)} \right) \right]^N \right], & y < 1/k^2, \\ 1, & y \geq 1/k^2. \end{cases} \quad (19)$$

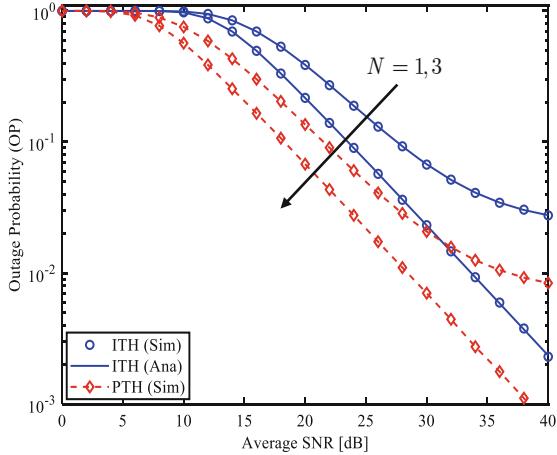
Then, substituting (19) into (18), the EC is computed as

$$\mathcal{C} = \frac{1}{\ln 2} \int_0^{1/k^2} \frac{1}{1+y} \exp \left( - \frac{\sigma_D^2 y}{\Omega_{RD} P_R (1 - k^2 y)} \right) \left[ 1 - \left[ 1 - \exp \left( - \frac{(l^2 P_R + \sigma_R^2) y}{\Omega_{SR} P_S (1 - k^2 y)} \right) \right]^N \right] dy. \quad (20)$$

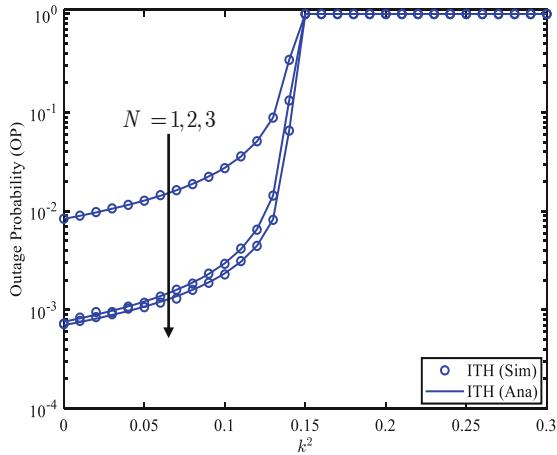
Applying [1, Eq. (25.4.30)], (20) becomes (17).

## 4 Numerical Results and Discussions

In this section, the effects of parameters on the OP and EC are examined via the above obtained expressions. Computer simulations are provided to validate our analysis. Some parameters are set as:  $\Omega_{SR} = \Omega_{RD} = 1$ ,  $\sigma_R^2 = \sigma_D^2 = \sigma^2$ ,  $P_S = P_R = P$ , and  $k_{SR}^2 = k_{RD}^2$ . Consequently,  $k^2 = \max\{k_{SR}^2, k_{RD}^2\} = k_{SR}^2 = k_{RD}^2$ . The average SNR =  $P/\sigma^2$ .

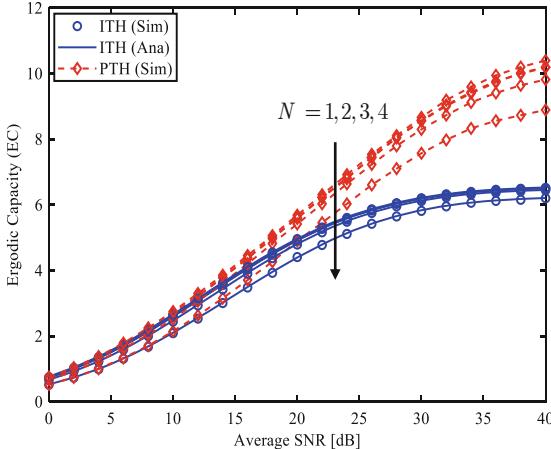


**Fig. 2.** System OP versus SNR for  $N = 1, 3$ ,  $\mathcal{R} = 3$  bpcu,  $k^2 = 0.1$ , and  $l^2 = -30$  dB.



**Fig. 3.** The effect of ITH level  $k^2$  on the system OP for  $N = 1, 2, 3$ ,  $\mathcal{R} = 3$  bpcu,  $l^2 = -30$  dB, and SNR = 40 dB.

Figure 2 depicts the system OP vs SNR for  $\mathcal{R} = 3$  bpcu,  $k^2 = 0.1$ , and  $l = -30$  dB. We investigate two cases, i.e.,  $N = 1$  and  $N = 3$  relays. We use (9) in Theorem 1 to plot the OP curves with ITH. We observe that when the number of relays increases from 1 to 3, the OPs with both ITH and PTH greatly enhance. Specifically, at SNR = 30 dB, the OPs with ITH and PTH are  $6.7 \times 10^{-2}$  and  $2 \times 10^{-2}$  corresponding to  $N = 1$ , and  $2.3 \times 10^{-2}$  and  $7 \times 10^{-3}$  corresponding to  $N = 3$ . Another observation is that the OPs when  $N = 1$  reduce slowly in the high SNR regime for both FDR systems with ITH and PTH. This result is reasonable due to the stronger effect of residual SI in higher SNR.

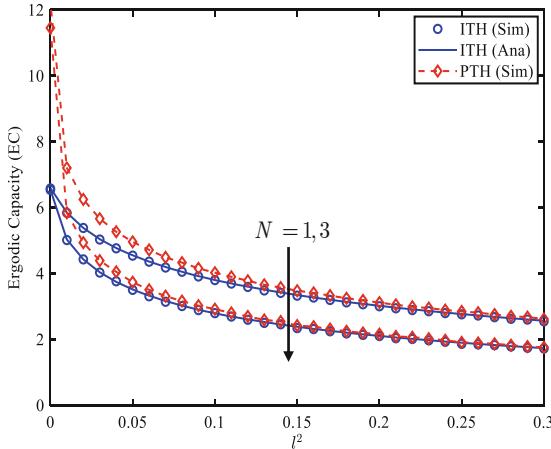


**Fig. 4.** System EC for different numbers of relays,  $k^2 = 0.01$ , and  $l^2 = -30$  dB.

Figure 3 shows the impact of ITH level  $k^2$  on the system OP for  $N = 1, 2, 3$ ,  $\mathcal{R} = 3$  bpcu, and SNR = 40 dB. Notice that when  $k^2 = 0$ , the OPs with ITH become the OPs with PTH. Specifically, when  $k^2 = 0.15$ , all OPs are equal to 1 because the second condition in (9) occurs. Thus, we should use a suitable  $\mathcal{R}$  for the FDR system with ITH to avoid the case that  $OP = 1$ . Additionally, the effect of the ITH level is significant even when  $k^2$  is small. In particular, when  $k^2 = 0$  increases from 0 (PTH) to  $k^2 = 0.1$ , the OPs increase from  $8 \times 10^{-3}$ ,  $7.1 \times 10^{-4}$ , and  $7 \times 10^{-4}$  to  $2.8 \times 10^{-2}$ ,  $3 \times 10^{-3}$ , and  $2.3 \times 10^{-3}$  corresponding to  $N = 1, 2$ , and 3. Therefore, besides applying all methods to reduce the residual SI power, various solutions for decreasing the ITH levels should be continuously proposed and applied.

Figure 4 plots the system EC for different numbers of relays, i.e.,  $N = 1, 2, 3$ , and 4. We use (17) to obtain the EC analysis curves with ITH. Similar to the OP behavior, the EC with PTH is greatly higher than that with ITH. Importantly, unlike Figs. 2 and 3, where the OP performance with PTH and ITH is greatly improved when  $N$  increases from 1 to 3, the ECs in Fig. 4 with PTH and ITH is only slightly increased for the same range of  $N$ . In other words, the relay selection scheme benefits the OP performance. Therefore, when the OP is the main requirement, we can use  $N = 3$  relays to achieve the lowest OP. Meanwhile, when the EC is the main requirement, we should use  $N = 1$  to reduce the signal processing complexity at the relay. On the other hand, the ECs of the FDR system with ITH reach the saturated ceilings in the high SNR region.

Figure 5 investigates the impact of  $l^2$  on the system EC. We should note that in the case  $l^2 = 0$  (PSIC), the ECs of the FDR system are two times in comparison with the ECs of the HDR system. When  $l^2$  increases, the ECs of the FDR system with ITH and PTH significantly reduce. Particularly, when  $l^2$  varies from 0 to 0.05, all ECs are decreased quickly, especially in the case with



**Fig. 5.** System EC versus  $l^2$  for  $N = 1, 3$ ,  $k^2 = 0.01$ , and SNR = 40 dB.

PTH. Additionally, the ECs in the cases with ITH and PTH are greatly different when  $l^2 = 0$ . In contrast, they are similar when  $l^2 = 0.3$ . As a result, the effect of ITH is significant when the residual SI level is small. In the case that the residual SI level is large enough, the effect of ITH on the EC of the FDR system can be neglected.

## 5 Conclusion

This paper provided a theoretical study on the performance of the FDR system in the cases of imperfect transceiver hardware and imperfect SI cancellation. We obtained the OP & EC expressions of the FDR system over Rayleigh fading channels. Numerical results indicated that the combined effects of ITH and ISIC are significant, especially in the high SNR region. In that context, relay selection can significantly improve the FDR system performance. Moreover, when either the ITH or ISIC level is large enough, their effects can be neglected. Therefore, besides applying all SIC solutions effectively, various methods to reduce the hardware impairments should be continuously proposed and exploited to improve the FDR system performance in practice.

**Acknowledgment.** This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) by the Ministry of Education under Grant 2020R1I1A3068305.

## References

1. Abramowitz, M., Stegun, I.A.: *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, vol. 9. Dover, New York (1972)

2. Amin, A.A., Shin, S.Y.: Capacity analysis of cooperative NOMA-OAM-MIMO based full-duplex relaying for 6G. *IEEE Wirel. Commun. Lett.* **10**(7), 1395–1399 (2021)
3. Anh, N.T., Nguyen, C.M., Duy, T.T., Hanh, T., Hai, H.D.: Reliability-security analysis for harvest-to-jam based multi-hop cluster MIMO networks using cooperative jamming methods under impact of hardware impairments. *EAI Endorsed Trans. Ind. Netw. Intell. Syst.* **8**(28), e5 (2021)
4. Do, T.N., da Costa, D.B., Duong, T.Q., An, B.: A full-duplex cooperative scheme with distributed switch-and-stay combining for NOMA networks. In: 2017 IEEE Global Communications Conference, GLOBECOM 2017, Singapore, 4–8 December 2017, pp. 1–6. IEEE (2017)
5. Doan, X.-T., Nguyen, N.-P., Yin, C., da Costa, D.B., Duong, T.Q.: Cognitive full-duplex relay networks under the peak interference power constraint of multiple primary users. *EURASIP J. Wirel. Commun. Netw.* **2017**(1), 1–10 (2017). <https://doi.org/10.1186/s13638-016-0794-y>
6. Everett, E., Sahai, A., Sabharwal, A.: Passive self-interference suppression for full-duplex infrastructure nodes. *IEEE Trans. Wirel. Commun.* **13**(2), 680–694 (2014)
7. Ha, D., Truong, V., Lee, Y.: Performance analysis for RF energy harvesting mobile edge computing networks with SIMO/MISO-NOMA schemes. *EAI Endorsed Trans. Ind. Netw. Intell. Syst.* **8**(27), e2 (2021)
8. Hoang, T.M., Nguyen, B.C., Tran, X.N., Dung, L.T.: Outage probability and ergodic capacity of user clustering and beamforming MIMO-NOMA relay system with imperfect CSI over nakagami- $m$  fading channels. *IEEE Syst. J.* **15**(2), 2398–2409 (2021)
9. Nguyen, B.C., Dung, L.T., Hoang, T.M., Tran, X.N., Kim, T.: Impacts of imperfect CSI and transceiver hardware noise on the performance of full-duplex DF relay system with multi-antenna terminals over nakagami-m fading channels. *IEEE Trans. Commun.* **69**(10), 7094–7107 (2021)
10. Nguyen, B.C., Hoang, T.M., Tran, P.T.: Improving the performance of spatial modulation full-duplex relaying system with hardware impairment using transmit antenna selection. *IEEE Access* **8**, 20191–20202 (2020)
11. Nguyen, B.C., Tran, X.N., Nguyen, L.V., Dung, L.T.: On the performance of full-duplex spatial modulation MIMO system with and without transmit antenna selection under imperfect hardware conditions. *IEEE Access* **8**, 185218–185231 (2020)
12. Nguyen, H., Nguyen, T., Vo, V., Voznák, M.: Hybrid full-duplex/half-duplex relay selection scheme with optimal power under individual power constraints and energy harvesting. *Comput. Commun.* **124**, 31–44 (2018)
13. Nguyen, N., Ngo, H.Q., Duong, T.Q., Tuan, H.D., da Costa, D.B.: Full-duplex cyber-weapon with massive arrays. *IEEE Trans. Commun.* **65**(12), 5544–5558 (2017)
14. Schenk, T.: RF Imperfections in High-Rate Wireless Systems: Impact and Digital Compensation. Springer, Heidelberg (2008). <https://doi.org/10.1007/978-1-4020-6903-1>
15. Wan, Z., Gao, Z., Gao, F., Renzo, M.D., Alouini, M.: Terahertz massive MIMO with holographic reconfigurable intelligent surfaces. *IEEE Trans. Commun.* **69**(7), 4732–4750 (2021)



# Heart Disease Prediction Using Soft Computing Methods

N. Veena<sup>1</sup> , S. Mahalakshmi<sup>1</sup> (✉) , T. Ragunthar<sup>2</sup> , Ambika Rani Subash<sup>1</sup> , and K. Vinutha<sup>1</sup>

<sup>1</sup> Department of Information Science and Engineering, BMS Institute of Technology and Management, Bengaluru, Karnataka 560064, India

{veena, maha .shanmugam, ambikasubash, vinuthak \_ise2014 }@bmsit.in

<sup>2</sup> Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamilnadu, India  
raguntht@srmist.edu.in

**Abstract.** According to WHO reports, heart related problems are the major disease because of humans' food habits and stressful living condition. Health is one of the important aspects of human life. The healthcare system indicates the physical and psychological health of the populace in clinical services and physician expertise, and also the development in drug and medicine administration. As well, it contains several considerable confronts as a consequence of enormous expenditure constraint. In this paper, we premeditated a cloud related Remote Diagnosis and Treatment System by the aid of Artificial Neural Network. According to the patient's verbally composed indication and the Brain waves, the anticipated ANN is used to discover the type and phase of disease. Additionally, we premeditated a cost-effective cloud related treatment system to generate suitable prescription based on the disease.

**Keywords:** Heart disease · ECG · Cloud-based system · ANN · SVM · KNN

## 1 Introduction

As there is a famous saying, "Health is wealth," if health is there, everything is there; otherwise, everything is of waste. In today's world, health is affected by our lifestyle, such as what we eat, stress, sleep, smoking, alcohol, and drugs. Health is also affected by how we feel about ourselves, the people around us. The society in which we live also has a significant influence on our health.

Health Sector is composed of many companies that provide services and products to medical care and health. Some of the health sectors include companies such as pharmaceutical, biotechnology, health insurance providers, and also, some companies manage clinics and hospitals.

Today much of the research is focused on the medical field. Health is a physical or mental condition, which highlights the ability to changing external and internal environments.

Human health is affected by many factors that lead to many physiological diseases, and one such disease is heart disease. Heart disease is an illness that damages the heart. Many types of heart disease affect different parts of the heart, such as Congenital Heart Disease, Arrhythmia, Coronary Artery Disease, Dilated Cardiomyopathy, Myocardial Infarction, Heart Failure, Hypertrophic Cardiomyopathy, Mitral Regurgitation, Mitral Valve Prolapse, Pulmonary Stenosis. These diseases cause significant death across the world.

WHO (World Health Organization) and CDC4 (Centres for Disease Control and Prevention) have reported that heart disease is a significant death-causing disease in many countries like Australia, Canada etc. Also, it occurs more in teenagers and early mornings because of their natural lifestyle.

Cardiovascular heart disease is the leading cause of death worldwide. According to WHO (World Health Organization) estimates 7.2 million die from this condition every year.

Many solutions have been proposed in literature for automated disease diagnosis of heart disease. Current work on disease prediction is based on snapshot data, but no methods were proposed on historic data of a patient. Continuous data analysis will lead to more effective and accurate prediction [9]. Even though disease appears sudden the reasons for it build for many months or years earlier, so by analyzing the historic of data a more accurate prediction model can be built.

This paper focus on design of methods and tools for continuous data analysis using soft computing methods.

## 2 Literature Survey

In [1] the authors have used Multiple kernel Learning (MKL) with Adaptive Neuro-Fuzzy Inference System to classify the heart disease from abnormal to normal people. They compared their system with many other system and proved that the sensitivity and specificity are high with low mean square value for KEGG Metabolic Reaction Network dataset.

In [2] the authors have discussed that cardiopulmonary exercise testing (CPET) in heart failure (HF) was performed on male candidates who are middle aged they have proved that today it can also be performed not only on male candidates but also on patients with co-morbidities, women, elderly, those with preserved ejection fraction.

In [3] authors proposed a heart disease classification system. Higher-Order Statistics and Spectra (HOS) bispectrum and cumulant features are extracted from each ECG beat. The characteristics extracted are given to Principal Component Analysis (PCA) to reduce the dimensions. Then PCA coefficients are marked using multiple ranking methods. KNN and Decision tree models are built using ranked features to get the highest accuracy. The classification cannot predict initial stages of disease.

In [4] authors implemented an expert system using decision tree and KNN classifier to predict disease. They used Pima Indians Diabetes Database to predict the diabetes. The disadvantage in this work is that it is based on snapshot data and not on continuous time series data.

In [5] authors proposed a decision support system for disease classification with more focus on feature selection. Probabilistic Principal Component Analysis (PPCA) was used for feature extraction. It extracted features with high covariance and feature dimension is also reduced. On the dimension reduced data, Radial basis function (RBF) neural network is built to predict the disease. With usage of feature selection accuracy improved, but the dataset is not continuous and it is only snapshot.

In [6] authors proposed a prediction system for coronary heart disease. It used narrative medical histories and extracted features from it for training and classification. From the unstructured text input, features are extracted using natural language processing techniques and from it PCA, mutual information filter employed to reduce feature dimension size. SVM classifier is then built to classify. The system accuracy depends on narration ability of patients and when missing information in narration, the accuracy is very low.

In [10] authors devised a heart disease classification using ECG signals. Signal Denoising is done using wavelet filters and on the filtered signal, Discrete wavelet transform (DWT) based features are extracted. The method is only for data collected at a particular snapshot.

In [11] author proposed a method based on selecting best features with random forest algorithm. After selected features are extracted, different machine learning techniques are applied to predict heart disease. They were able to achieve 75% accuracy using this method.

In [12] authors analyzed the performance of five classifiers. The classifier was Bayes's Net, SMO, KStar, MLP and J48. From their analysis Bayes Net and SMO classifiers are the optimum among the all five classifiers. Their analysis is based on data set collected at particular time and not on continuous time series data.

In [13] authors proposed a risk model for heart disease. This model can predict 1-year or more survival for Heart failure diagnosed patients. From the unstructured medical records, certain attributes are extracted using search and then multinomial Naïve Bayes (NB) is used for classification. The approach does not address missing value problem. Feature selection is not given attention.

In [14] authors analyzed heart disease prediction problem using six machine learning techniques. Thirteen distinct attributes from dataset is taken for analysis. StatLog heart disease dataset from UCI machine learning repository is taken and classifier models built on those datasets to predict heart disease. But the work does not give attention to feature set selection and continuous data.

In [15] authors used decision tree classifier to diagnose Heart Arrhythmia. Discrete wavelet transforms (DWT) are done on ECG signals to get coefficients and from its statistical features are extracted and used as features. The accuracy of the method is not satisfactory as no feature dimensionality reduction procedures are followed. Also, data is snap shot based and not on continuous based.

### 3 Methodology

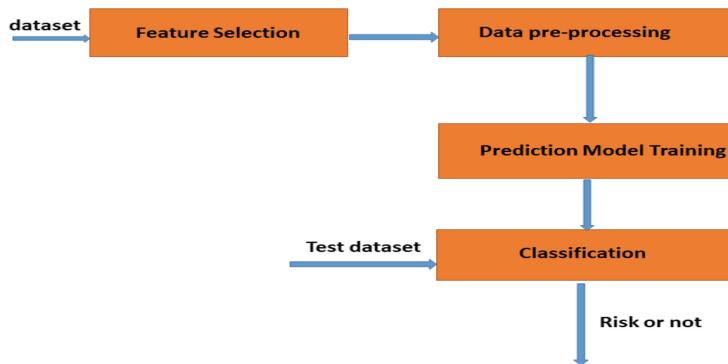
**Step 1:** Data set is given to feature selection module. Feature Selection module selects the attributes that influences prediction result.

**Step 2:** Data pre-processing module converts the feature selected dataset into a form suitable for training.

**Step 3:** Prediction Model training module constructs a classifier module which classifies any test data of patient to either risky or not risky.

**Step 4:** Classification module classifier test data of patient to risk or not by applying the classifier model [8].

We have also discussed noninvasive devices which involves EEG signals [7] instead of ECG signals. The system architecture is given below (Fig. 1):



**Fig. 1.** Architecture of proposed system

### 3.1 Feature Selection

Feature Selection is done to identify relevant attributes which influences the output. It is done mainly to reduce the training time and build a more accurate model.

The feature selection is based on calculation of symmetric uncertainty between attributes and target class. Once the symmetric uncertainty is calculated between each attribute and the target class, the attributes who symmetric uncertainty value less than a threshold value are dropped as irrelevant features and rest all features are selected for classification.

Symmetric uncertainty is calculated as

$$Su(X, Y) = 2 * \frac{MI(X, Y)}{H(X) + H(Y)} \quad (1)$$

where  $H(X)$  is the entropy of a attribute X,  $H(Y)$  is the entropy of a class variable.

$Y$  and  $MI$  is mutual information between  $X$  and  $Y$ .

The Fast Fourier Transformation (FFT) is an efficient method to calculate the Discrete Fourier Transform (DFT) and the inverse of DFT. FFT is a windowing system similar to the Wavelet Transformation. The DFT decomposes the time-series data of the patients in the sliding windows to extract frequency data for predicting the patient's condition before the day.

A normalization technique is used before FFT to scale the input such that the variables take values between 0 and 1.

The equation in 2 is for normalization.

$$D_{norm} = \frac{D_{orig} - D_{min}}{D_{max} - D_{min}} \quad (2)$$

Here,  $D_{norm}$  - normalized value,  $D_{orig}$  - original value,  $D_{min}$  and  $D_{max}$  minimum and maximum values in the dataset.

If  $x(t)$  is time series of a sliding window.

The DFT of  $x(t)$  is defined as in Eq. 3:

$$X(c^{jw}) = \sum_{t=-\infty}^{\infty} x(t)c^{-jwt} \quad (3)$$

where,

t - Discrete-time index, w - Frequency

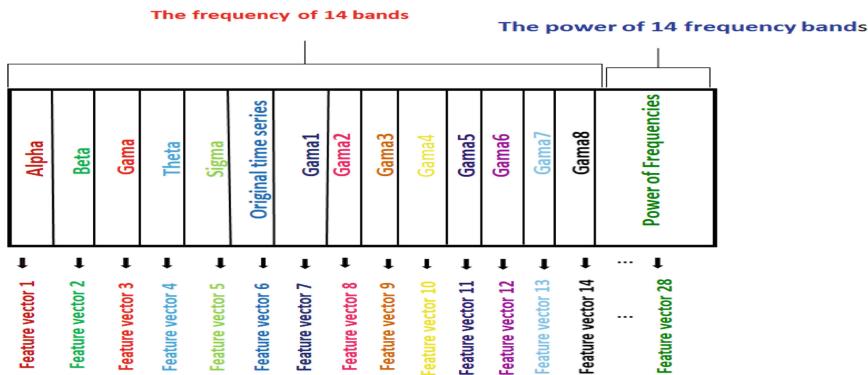
There will be T time-series input  $x(t)$ ; hence the DFT transform pair is defined as in Eq. 4

$$X(P) = \sum_{t=0}^{T-1} x(t)W_T^{tp} \leq> x(t) = \frac{1}{T} \sum_{p=0}^{T-1} X(P)W_T^{-tp} \quad (4)$$

By summarizing the equations, we need to perform  $T^2$  complex multiplications on DFT matrix W for  $x(t)$  (input time signal) with a length of T. Hence the implementation cost of factorizing FFT (W) into a matrix is less than the cost of DFT because each stage of FFT needs only T additions and  $T/2$  multiplications.

$X = \{y_1, y_2, y_3, \dots, y_n\}$  contains n data items represents the time series data. X is further divided into a set of overlapping sub-segments based on parameter k. k is the size of the sliding window of each sub-segment.

Patient's condition is predicted by analyzing the input time series by using the FFT to extract frequency information. FFT can acquire 5 frequency bands ( $\alpha, \beta, \gamma, \delta$ , and  $\theta$ ) for each sliding window as in Fig. 2. The high-frequency band is further divided into eight sub-bands as it contains majority of information regarding the sliding window; for reference, the original sliding window is also included with the extracted feature set. As an outcome, a total of 14 frequency bands are generated for each sliding window. For all 14 frequency bands power of the FFT coefficients is computed, which leads to the calculation of square of the absolute value of the Fourier coefficients. This results total of 28 frequency bands obtained for each sliding window, is shown in Fig. 2.



**Fig. 2.** Frequency Bands that can be extracted from a window

8 distinct statistical features can be obtained from each frequency band. The dataset determines the best performing features. The data may be symmetrically distributed or may have twisted distribution. Generally, the minimum and maximum are used as beat measures during symmetric distribution of time series, while mean and standard deviation is appropriate during skewed distribution.

Patient's data is collected every day over a period of time; the selected features are extracted from this data. The time-variant attributes are arranged as time series from the selected features. Smaller overlapping sliding windows of configurable size are created from the time series data then this sliding window is passed through FFT. Properties of time series data over the frequency domain is obtained by applying FFT. The following features are obtained from coefficients by applying FFT Maximum Value, Minimum Value, Mean, Standard Deviation.

The data set with selected features is now transformed as every selected time-variant feature is replaced by its statistical features.

### 3.2 Classification

Three classifier models are built on the pre-processed training data set. Following models are built.

1. Linear Support Vector Machine
2. ANN
3. K Nearest Neighbour

## 4 Result

Here symptoms are orally collected and EEG signals or ECG signals are collected from the patients. From these EEG or ECG signals we extract median, mean and mode frequency. These extracted features are used to train the Artificial Neural Network (Fig. 3).



**Fig. 3.** Collection the EEG or ECG signals plus orally collected symptoms

Results of this Artificial Neural Network is measured in terms of True Positive (TP), True negative (TN), False positive (FP), False Negative (FN) with varied training and testing data as shown below (Table 1).

**Table 1.** Performance of ANN using TP, TN, FP, FN

Training: Testing (%)	TP	TN	FP	FN
50:50	58	51	13	18
60:40	32	36	12	16
70:30	26	28	10	8
80:20	14	21	6	7

The performance of Artificial Neural Network is measured with respect to various metrics such as Accuracy, Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV) and F-Measure compared with other two models namely Support Vector Machine (SVM) and K-nearest neighbor (KNN) with varying training and testing data as shown in the figure below (Tables 2, 3 and 4):

**Table 2.** Performance of ANN using different metrics

Training: Testing (%)	Accuracy	sensitivity	specificity	PPR	NPR	F-measure
50:50	0.778571	0.763158	0.796875	0.816901	0.73913	0.779652
60:40	0.708333	0.666667	0.75	0.727273	0.692308	0.705882
70:30	0.75	0.764706	0.736842	0.722222	0.777778	0.750515
80:20	0.729167	0.666667	0.777778	0.7	0.75	0.717949

Finally, we were able to see that Artificial Neural Networks performance is better than the other two models SVM and KNN.

**Table 3.** Performance of SVM using different metrics

<b>Training: Testing (%)</b>	<b>Accuracy</b>	<b>sensitivity</b>	<b>specificity</b>	<b>PPR</b>	<b>NPR</b>	<b>F-measure</b>
<b>50:50</b>	0.742857	0.736842	0.75	0.777778	0.705882	0.743363
<b>60:40</b>	0.677083	0.64	0.717391	0.711111	0.647059	0.676489
<b>70:30</b>	0.662162	0.631579	0.694444	0.685714	0.641026	0.661521
<b>80:20</b>	0.702128	0.684211	0.714286	0.619048	0.769231	0.698925

**Table 4.** Performance of KNN using different metrics

<b>Training: Testing (%)</b>	<b>Accuracy</b>	<b>sensitivity</b>	<b>specificity</b>	<b>PPR</b>	<b>NPR</b>	<b>F-measure</b>
<b>50:50</b>	0.685714	0.689189	0.681818	0.708333	0.661765	0.685484
<b>60:40</b>	0.635417	0.6	0.673913	0.666667	0.607843	0.634812
<b>70:30</b>	0.635135	0.621622	0.648649	0.638889	0.631579	0.634848
<b>80:20</b>	0.680851	0.666667	0.689655	0.571429	0.769231	0.677966

## 5 Conclusion

An efficient heart disease diagnosis system is proposed. With the proposed method features are selected and pre-processing done on training data set. With Entropy based feature selection only essential predictor variables are selected from data set for processing. By the way of FFT based feature extraction, a way to represent continuous data for training the classifier is presented. Different classifiers were used to predict heart disease. Three solutions for classification were proposed such as Neural Network based classification, Rule based mining with PSO swarm optimization, FFT Based Feature extraction and Classification.

## References

1. Manogaran, G., Varatharajan, R., Priyan, M.K.: Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system. *Multimedia Tools Appl.* **77**(4), 4379–4399 (2018)
2. Corra, U., et al.: Role of cardiopulmonary exercise testing in clinical stratification in heart failure: A position paper from the committee on exercise physiology and training of the heart failure association of the European society of cardiology. *Eur. J. Heart Fail.* **20**(1), 3–15 (2018)
3. Acharya, U.R., et al.: “Application of higher-order spectra for the characterization of coronary artery disease using electrocardiogram signals. *Biomed. Signal Process. Control* **31**, 31–43 (2017)
4. Hashi, E.K., Zaman, M.S.U., Hasan, M.R.: An expert clinical decision support system to predict disease using classification techniques. In: International Conference on Electrical, Computer and Communication Engineering (ECCE). IEEE (2017)
5. Shah, S.M.S., et al.: Feature extraction through parallel probabilistic principal component analysis for heart disease diagnosis. *Physica A: Stat. Mech. Appl.* **482**, 796–807 (2017)
6. Buchan, K., Filannino, M., Uzuner, Ö.: Automatic prediction of coronary artery disease from clinical narratives. *J. Biomed. Inform.* **72**, 23–32 (2017)

7. Veena, N., Anitha, N.: A review of non-invasive BCI devices. *Int. J. Biomed. Eng. Technol.* **34**(3), 205 (2020)
8. Veena, N., Anitha, N.: Cloud based remote diagnoses for rural health care. *J. Adv. Res. Dyn. Control Syst.* **10**(08), 707–718 (2018)
9. Veena, N., Mahalakshmi, S., Diyya, G.A., Allada, A., Anand, M.S.: Heart disease prediction system. In: 2021 International Conference on Forensics, Analytics, Big Data, Security (FABS), pp. 1–6 (2021). <https://doi.org/10.1109/FABS52071.2021.9702552>
10. Kumar, S.U., Inbarani, H.H.: Neighborhood rough set based ECG signal classification for diagnosis of cardiac diseases. *Soft. Comput.* **21**(16), 4721–4733 (2016). <https://doi.org/10.1007/s00500-016-2080-7>
11. Mustaqeem, A., et al.: Wrapper method for feature selection to classify cardiac arrhythmia. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE (2017)
12. Sultana, M., Haider, A., Uddin, M.S.: Analysis of data mining techniques for heart disease prediction. In: 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT). IEEE (2016)
13. Saqlain, M., et al.: Identification of heart failure by using unstructured data of cardiac patients. In: 2016 45th International Conference on Parallel Processing Workshops (ICPPW). IEEE (2016)
14. Dwivedi, A.K.: Performance evaluation of different machine learning techniques for prediction of heart disease. *Neural Comput. Appl.* **29**(10), 685–693 (2016). <https://doi.org/10.1007/s00521-016-2604-1>
15. Alickovic, E., Subasi, A.: Medical decision support system for diagnosis of heart arrhythmia using DWT and random forests classifier. *J. Med. Syst.* **40**(4), 108 (2016)



# Enhancing Inflammatory Bowel Disease Diagnosis Performance Using Chi-Squared Algorithm on Metagenomic Data

HaiThanh Nguyen<sup>1(✉)</sup>, Huong Hoang Luong<sup>2</sup>, TrongThanh Tran<sup>2</sup>, NgocVan Nguyen<sup>2</sup>, and KhoiDinh Nguyen<sup>2</sup>

<sup>1</sup> College of Information and Communication Technology, Can Tho University,  
Can Tho, Vietnam  
[nthai@cit.ctu.edu.vn](mailto:nthai@cit.ctu.edu.vn)

<sup>2</sup> FPT University, Can Tho 900000, Vietnam

**Abstract.** Disease prediction is a rapidly growing research field in the age of information and technology. In the future, machinery robotics is expected to significantly support doctors in specific diagnostic areas and provide helpful medical advice. Metagenomics has been emerging in recent years and plays an essential role in personalized medicine approaches. However, data sources for metagenomic analysis are still complex and very high dimensional; hence it is necessary to provide helpful feature selection methods to apply to them. In this study, we propose to use the Chi-Squared technique to extract insights from the metagenomic features for disease diagnosis. Chi-Squared is deployed in feature selection on Metagenomic Data to extract meaningful features related to Inflammatory Bowel Disease and shows a somewhat spirited performance than other methods. As obtained results, Chi-Squared is expected to pick out valuable features that can improve the prediction performance.

**Keywords:** Feature selection · Chi-Squared · Metagenomic data · Inflammatory Bowel Disease

## 1 Introduction

In recent years, the death of patients related to the gastrointestinal tract has increased. Inflammatory Bowel Disease (IBD) is not a very dangerous disease but should be mentioned as it develops into other deadly diseases. IBD has an increased risk of developing colorectal cancer(CRC) [1], although CRC has decreased in association with IBD in recent decades (in Western countries) [2]. It is a general term used to describe disorders related to the digestive tract [3]. IBD includes Crohn's disease and ulcerative colitis [3]. Crohn's disease affects the small intestine to the large intestine in the digestive tract [4]. Ulcerative colitis affects the large intestine and rectum [4,5].

The dangers of IBD cause much harm to the body if not detected in time. Common symptoms: malnutrition with resulting weight loss, colorectal cancer (CRC), fistulas or ulcers, intestinal rupture or perforation, bowel obstruction [6]. To elucidate, we took data on CRC in symptoms of IBD. CRC is the most common cancer in both men and women and the 3rd most common cancer in humans by 2020 in the United States. Several colorectal cancer cases by 2020: 104,610 new cases [7]. The lifetime risk of colorectal cancer is 1 in 23 (4.4%) for men and 1 in 25 (4.1%) for women. They expected it to cause 53,200 deaths by 2020 and Under-55 mortality increased by 1% per year from 2008 to 2017 [7]. Triggers: Age, Race or ethnicity, Family history, Cigarette smoking, Smoking may help prevent ulcerative colitis, Non-Steroidal anti-inflammatory medications [8].

Personalized medicine is the new dose of medicine to develop health care based on genetic and pharmacological information. Several approaches to Personalized medicine are the patient's gene is. This approach reduces financial costs (drug costs). Personalized medicine is an integral part of the treatment of disease and offers a more accurate predictive value. Therefore, doctors diagnose and treat patients more accurately and make drugs and therapy more effective [9]. Personalized medicine is to tailor medical treatment to the individual characteristics of each patient. Therefore, we should find a way that is appropriate for each patient. Gene approaches make it viable to treat patients. In short, Personalized medicine is an extension of traditional methods and challenges as well as changes in how to identify and organize health problems [10]. Metagenomics is a tool to analyze DNA from the physical environment. This method is used to analyze genes from the environment collected [11]. Thereby, we see a similarity in using metagenomic to research personalized medicine.

This study proposes using the Chi-Squared algorithm to select features and predict Inflammatory bowel disease on the selected features.

The remaining parts of this work are divided into four main sections; in the next section (Sect. 2), we present the main points of our work. First, the methodology Sect. 3 will introduce the method used in this article. Then, in the experiments Sect. 4, we present the results in detail with some comparisons. Finally, we conclude and summarize the main points in this work in Sect. 5.

## 2 Related Work

Today, Deep Learning(DL) has successfully applied to images, waves, etc. When sample size ( $N$ ) is greater than the number of features ( $d$ ), DL outperforms ML. In the opposite case,  $d$  is greater than  $N$ , then using DL leads to severe overfitting. This article shows how to apply CNN to data with no original visual structure. These datasets are gradually available as many models have used this information for decision making [12] (healthcare personalization [25], diagnosis and prognosis [13], pharmacology. Study [14] etc.). The most successful techniques for omics data sets are RF and sparse regression. This paper uses DL directly onto six measurement datasets to reflect the abundance of bacteria species and their appearance in the gut of both sick and healthy patients. This

technology works especially well in image classification and focuses on CNN use. The goal is to demonstrate that the DL technique is the best tool for predicting the metagenomics context. Methods include Fig. 1-visualization of image-based representations. The division can be done on a linear or logarithmic scale. The features are then displayed as images in one of two ways, namely phylogenetic-sorting (PLG) or t-SNE [15]. t-SNE is very effective in representing the height of points in a small space (2D). Fill-up [17] and t-SNE: Fill-ups [17] are formed by arranging values into a matrix from left to right, ordered by row. The bottom left square is set to 0 (white). It uses distinct colors on the logarithmic scale to represent abundance. (Black/White for Presence/Absence). Architecture configurations of CNNs: Inspired by the VGG grid [16]. In Short, the authors proposed the MET2IMG method to predict the patient's disease using measurement data. They use two main methods to construct synthetic images: fill-up [17] and t-SNE. Fill-up [17] using small and simple images. The T-SNE feature is embedded in 2-dimensional space using classic embedding in ML [18].

### 3 Methodology

This section discusses selecting features using the Chi-squared algorithm on metagenomic data. The experiments we have run include Feature Selection using Chi-Squared and Feature Selection randomly for later comparison. The comparison is done by the specific amount of features and based initially on datasets archived in the research of Fungal microbiota dysbiosis in IBD [20].

#### 3.1 Chi-Squared Algorithm

First, we need to discover about Chi-Square distribution, the random variable X represented by the sum of squares of the standard normal variables Z (Eq. 1).

$$X^2 = \sum Z_i^2 \quad (1)$$

where  $Z_1, Z_2, \dots$  etc. are standard normal variables. Next is the Degrees of freedom, known as the maximum number of logically independent values, defined by the total number of observations minus the number of independent constraints (Eq. 2).

$$Df = N - 1 \quad (2)$$

where Df: degree of freedom, N: sample size The formula for Chi-Square is exhibited in Eq. 3.

$$X_c^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (3)$$

where c : degree of freedom, O : observed value(s), E : expected value(s).

### 3.2 Feature Selection Using Chi-Squared Method

We use Feature Selection using Chi-Squared in this study. Chi-Squared or Chi-Squared Test is Statistically Significant [19] so that it can help a lot with features selection. We experiment on the dataset mentioned in Datasets Description Sect. 4.1. The dataset included six sets of data. Each of them has several rows as subjects which can be patients or healthy subjects. Columns of the dataset are the code of bacteria and their Metagenomic Data values with subjects. We divide the data into two groups, one for feature selection. After successfully implementing Chi-Squared, we pick out ten high-value chi-squared. The other group for the train uses ten high-value chi-squared picks in the first group and then tests to get the result.

### 3.3 Feature Selection Randomly

We randomly select the number of features in the same dataset corresponding to the number of features we choose to use Chi-Squared for later comparison. The process is mainly the same as the previous method. Which is on the same dataset, and we divide the data into two groups, one as the group for selecting ten random features. The other group for the train used ten random features picked in the first group and then tested.

### 3.4 Disease Prediction Using Random Forest Classifiers

For the two above methods, feature selection uses Chi-Squared and random selection for comparison. We apply Random Forest Classifier for train and test to get the results. The classifier provides the relative importance of different features during its process, which can help feature selection. However, we have already selected our perspective features. Therefore, the next thing we will do is use Random Forest Classifier to train and test on the considered datasets. As mentioned above in Sects. 3.2 and 3.3, the dataset consists of 6 sets of data. We divided them into two groups, with the first group selecting features. So we will use the second group to make the prediction. Ten selected features will be used to train and test these sets in the second group.

## 4 The Experiments

This section presents the dataset used in the experiments and some comparisons to evaluate the efficiency of the proposed method on IBD classification tasks.

### 4.1 Dataset Description

Data obtained from 235 IBD patients and 38 healthy subjects (HS) were determined using the 16S and ITS2 sequences [20]. The six considered datasets include as followings:

- HS\_UCr - Healthy (HS) and Ulcerative Colitis (UC) in remission
- HS\_iCDr - Healthy (HS) and ileal Crohn's disease (iCD) in remission
- HS\_UCf - Healthy (HS) and Ulcerative Colitis (UC) in flare
- HS\_iCDF - Healthy (HS) ileal Crohn's disease (iCD) in flare
- HS\_CDr - Healthy (HS) Crohn's disease (CD) in remission
- HS\_CDf - Healthy (HS) Crohn's disease (CD) in flare

To predict whether an individual is identified as an IBD patient or not. We divide one set for selecting features (UCr datasets). The selected features' efficient features are evaluated on other datasets, including iCDr, UCf, iCDF, CDr, and CDf. Information about these sets is given in the Table 1.

**Table 1.** Datasets information detail

Information	Dataset					
	UCr	iCDr	UCf	iCDF	CDr	CDf
The number of features	238	258	251	248	258	260
The number of patients	44	59	41	44	77	60
The number of healthy subjects	38	38	38	3 8	38	38
Total subjects	82	97	79	82	115	98

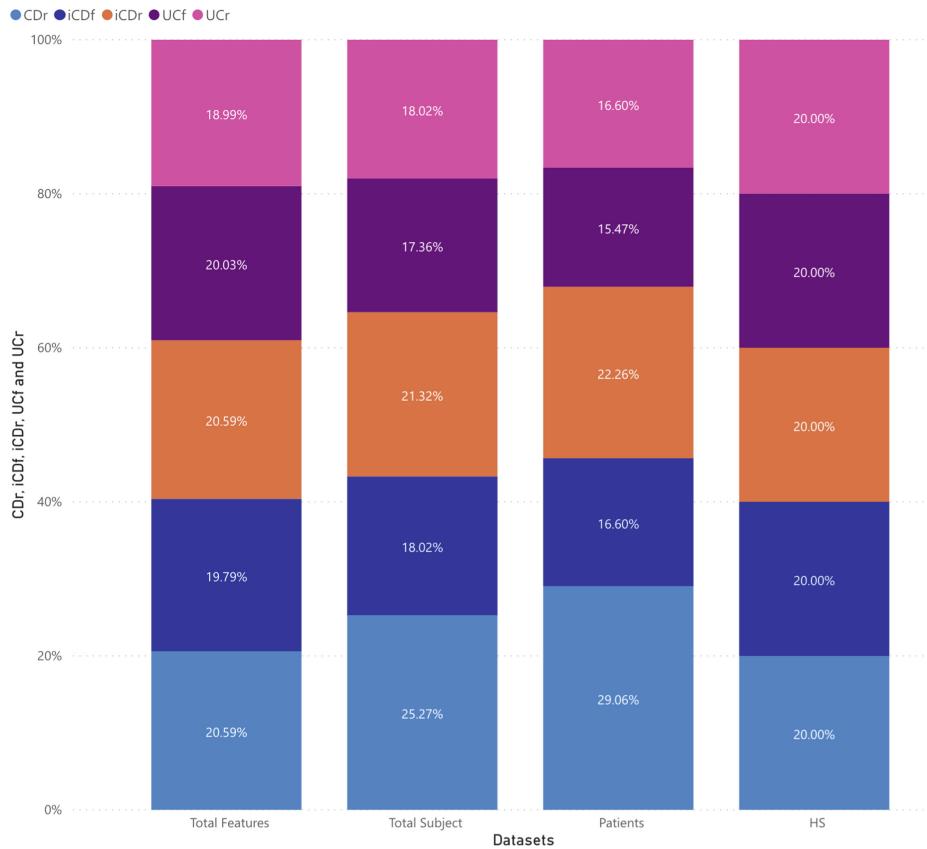
## 4.2 Data Division and Metrics for Evaluation

We have divided these datasets into two groups. The first one is for training, and the other aims to test. The selected group for feature selection later on for the training. In the other group, five others set later on for tests. To derive our results, we used the following three scoring metrics, including Accuracy (ACC), Area Under the Curve (AUC), Matthews correlation coefficient (MCC), which are commended in [21] to evaluate analysis in computational biology tasks. The performance was measured by an average value of metrics on a 5-Fold cross-validation repeated ten times. The charts also illustrate standard deviation in error bars.

## 4.3 Prediction Result with All Features

Before we get into feature selection using specific methods, we first select all features from a dataset then test on other datasets. To see how much we could improve on if we use feature selection or not exactly make any improvement.

In Table 2, we select all features in dataset CDr to train and later test on other datasets. We also illustrate some experimental results in the chart in Fig. 2. The maximum value is 0.85789 on metric ACC when trained by CDr and tested on CDf. Furthermore, the mean reaches its maximum prediction value of 0.804 on metric ACC, and the mean fall into its minimum prediction value is 0.620 on metric AUC.



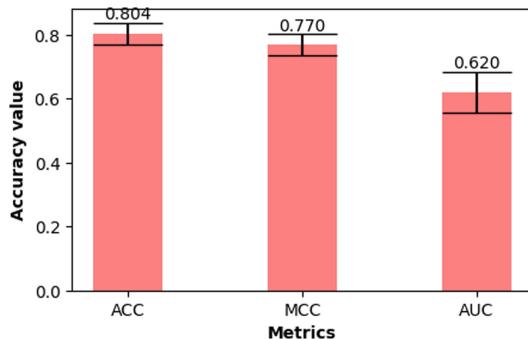
**Fig. 1.** Distribution of datasets through 4 criteria (total feature, total subject, patients and healthy subject)

**Table 2.** Results with metrics corresponding ACC, MCC and AUC.

Scenario	Train data	Metrics	Dataset				
			CDF	iCDF	iCDr	UCf	UCr
All features	CDr	ACC	0.85789	0.81996	0.79077	0.78270	0.76994
		MCC	0.71830	0.65018	0.59107	0.57894	0.56247
		AUC	0.81786	0.78815	0.75556	0.74784	0.73969

#### 4.4 Using Chi-Squared to Select Features

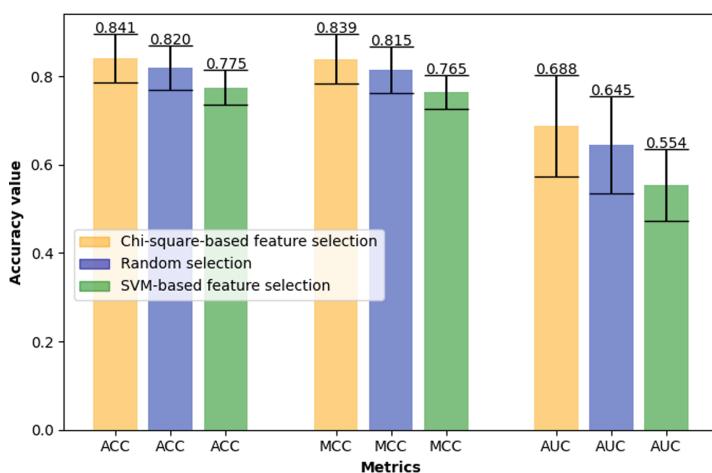
The selection was made on a dataset, and selected features were tested on other datasets. This study investigated the top ten features with Chi-squared and SVM and compared them to random selection. We change to another



**Fig. 2.** The bars plot with error bars when all features in train set is selected (trained by CDr dataset and test on others).

**Table 3.** Results with metrics corresponding ACC, MCC and AUC.

Scenario	Train data	Metrics	Dataset				
			CDf	iCDf	iCDr	UCf	UCr
Chi-Squared	CDr	ACC	0.87574	0.88912	0.82453	0.86508	0.75294
		MCC	0.75063	0.79502	0.63776	0.74607	0.51159
		AUC	0.87571	0.89099	0.81468	0.86389	0.75159
Random	CDr	ACC	0.84324	0.86376	0.79516	0.85364	0.74230
		MCC	0.68135	0.75522	0.57094	0.72374	0.49221
		AUC	0.83595	0.86598	0.78034	0.85086	0.74061
svm	CDr	ACC	0.81579	0.81471	0.76263	0.72417	0.75588
		MCC	0.61555	0.65766	0.49748	0.46034	0.53710
		AUC	0.79286	0.81270	0.72841	0.72798	0.76171

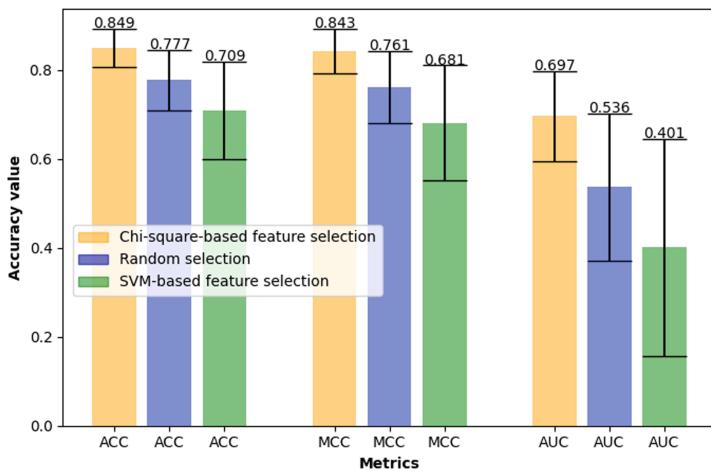


**Fig. 3.** The bars plot with error bars between random and feature selection using the Chi-Squared method (trained by CDr dataset and test on others).

In Table 3, the feature selection was made on the dataset CDr, the efficiency of selected features was evaluated on other datasets, including iCDf, iCDr, Cdf, UCr, and UCf. We also illustrate some experimental results in the chart in Fig. 3. In the prediction result, which used the Chi-Square method, we obtained all values higher than the random and SVM method. The disparity between the use of the Chi-squared method with the random and SVM method is moderately skewed towards the Chi-squared.

**Table 4.** Results with metrics corresponding ACC, MCC and AUC

Scenario	Train data	Metrics	Dataset				
			Cdf	iCDf	iCDr	UCf	UCr
Chi-Squared	Cdf	ACC	0.81652	0.90669	0.82363	0.88483	0.81397
		MCC	0.59380	0.81851	0.63263	0.79143	0.64688
		AUC	0.79345	0.90629	0.81358	0.88496	0.81506
Random	Cdf	ACC	0.73609	0.86469	0.76190	0.82595	0.69566
		MCC	0.36549	0.73609	0.50048	0.67836	0.40132
		AUC	0.67801	0.86413	0.74379	0.82422	0.69584
SVM	Cdf	ACC	0.69565	0.77941	0.78316	0.76083	0.52353
		MCC	0.23682	0.59736	0.57338	0.54409	0.05166
		AUC	0.55857	0.78948	0.77186	0.76131	0.52163



**Fig. 4.** The bars plot with error bars between random and feature selection using the Chi-Squared method (trained by Cdf dataset and test on others).

As we can see in Table 4 when using the Chi-Square method, all achieved high values. On the other hand, the rest of the sets, which give no high results,

proved that it is not as good as the Chi-Squared method for feature selection when using the random method. Besides, the values at the SVM method are not higher than the random method. Therefore, the chart in Fig. 4 and Table 4 show that stabilities can be achieved when using selection features and prevent results from imbalanced we can see in randomly chosen features and the SVM method. In comparison with Table 2, we can see that the set of features obtained by Chi-square outperforms even for the whole original set of features.

## 5 Conclusion

This study proposed a method based on the Chi-Squared technique to filter features to better results in Inflammatory Bowel Disease prediction with classic machine learning and allow a faster process to support disease diagnosis. The results have shown that our prediction has statistically more accurately predicted which subject IBD has than these selected features. Furthermore, Chi-Squared used for feature selection on Metagenomic Data can pick out the features more related to the sickness. Additionally, based on that premise, improving disease diagnosis technologies in the future might be an exciting area. We also publish the code and other results at <https://github.com/fptuni/ibd-chi-squared-v2>.

## References

1. (ASCRS) American Society of Colon & Rectal Surgeons: The Colon: What it is, What it Does and Why it is Important: Overview Cancer. <https://facsrs.org/patients/diseases-and-conditions/a-z/the-colon-what-it-is,-what-it-does>. Accessed June 2021
2. Kim, E.R., Chang, D.K.: Colorectal cancer in inflammatory bowel disease: the risk, pathogenesis, prevention, and diagnosis. World J. Gastroenterol. **20**(29), 9872–9881 (2014). <https://doi.org/10.3748/wjg.v20.i29.9872>
3. Centers for Disease Control and Prevention: Inflammatory bowel disease (IBD). <https://www.cdc.gov/ibd/what-is-IBD.htm>. Accessed June 2021
4. NIH-U.S. National Library of Medicine: Crohn's Disease. <https://medlineplus.gov/crohnsdisease.html>. Accessed June 2021
5. Innerbody: Rectum. [https://www.innerbody.com/image\\_digeov/dige14-new3.html](https://www.innerbody.com/image_digeov/dige14-new3.html). Accessed June 2021
6. Inflammatory Bowel Disease: Types, Causes, and Risk Factors (2020). <https://www.healthline.com/health/inflammatory-bowel-disease>. Accessed June 2021
7. American Cancer Society: Key Statistics for Colorectal Cancer. <https://www.cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html>. Accessed June 2021
8. Inflammatory bowel disease (IBD) - Symptoms and causes (2020). <https://www.mayoclinic.org/diseases-conditions/inflammatory-bowel-disease/symptoms-causes/syc-20353315>. Accessed June 2021
9. Cutter, G., Liu, Y.: Personalized medicine: the return of the house call? (December 2012). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3613206/>
10. Opinion: What Is Personalized Medicine And Why Is It So Important? — BioSpace (2020). <https://www.biospace.com/article/opinion-what-is-personalized-medicine-and-why-is-it-so-important-/>

11. Metagenomics - an overview — ScienceDirect Topics (2020). <https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/metagenomics>
12. Ginsburg, G.S., Willard, H.F.: Genomic and personalized medicine: foundations and applications. *Transl. Res.* **154**, 277–287 (2009)
13. Pasolli, E., Truong, D.T., Malik, F., Waldron, L., Segata, N.: Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. *PLOS Comput. Biol.* **12**(7), e1004977 (2016). Edited by Jonathan A. Eisen
14. Voora, D., McLeod, H.L., Eby, C., Gage, B.F.: The pharmacogenetics of coumarin therapy. *Pharmacogenomics* **6**, 503–513 (2005)
15. Laurens van der Maaten, G.H.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 8 (2008)
16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR abs/1409.1556 (2014). <http://arxiv.org/abs/1409.1556>
17. Nguyen, T.H., et al.: Disease classification in metagenomics with 2D embeddings and deep learning. In: The Annual French Conference in Machine Learning (CAp 2018), Rouen, France (June 2018). <https://arxiv.org/abs/1806.09046>
18. Nguyen, T., Chevaleyre, Y., Prifti, E., Sokolovska, N., Zucker, J.: Deep learning for metagenomic data: using 2D embeddings and convolutional neural networks (November 2020). <https://arxiv.org/abs/1712.00244>. Accessed 01 Dec 2017
19. Sharpe, D.: Chi-square test is statistically significant: now what? *Pract. Assess. Res. Eval.* **20**, 8 (2015)
20. Sokol, H., et al.: Fungal microbiota dysbiosis in IBD. *Gut* **66**(6), 1039–1048 (2016). <https://doi.org/10.1136/gutjnl-2015-310746>
21. Chicco, D.: Ten quick tips for machine learning in computational biology. *BioData Min.* **10**(1), 1–17 (2017). <https://doi.org/10.1186/s13040-017-0155-3>

# Author Index

## A

- Ahmed, Rashid, 253  
Aleksanyan, Grayr K., 83  
Anh Ngoc, Le, 369  
Anh Tuan, Nguyen, 369  
Anh, Dao Nam, 92  
Anh, Gian Quoc, 484  
Anh, Ho Trong, 246  
Anh, Nguyen Quynh, 92  
Anh, Thu Nguyen Thi, 623

## B

- Bui, Ngoc Dung, 33, 39  
Bui, Trung Ninh, 484

## C

- Cao, Van Luyen, 221  
Cheong, Siew Ann, 33  
Chu, Duc-Tuan, 230  
ChuDuc, Hoang, 83  
Cuong, Nguyen Manh, 83

## D

- Dang, Xuan Tho, 511  
Dang-Minh, Tuan, 287, 418, 445  
Danti, Rakshith, 581  
Dao, Nam Anh, 511  
Dao, Phuong Nam, 71  
Dao, Quy-Thinh, 260, 333  
Dien, Tran Thanh, 410  
Dinh, Thanh Ngo, 623  
Dinh, Van-Vuong, 260

## Do, Trong Tan, 71

- Do, Van-Chieu, 394  
Duc, Thi Nguyen, 541  
Duc, Tran Ngoc, 557  
Duc, Vu Minh, 141  
Dung, Le The, 650  
Dung, Nguyen Tuan, 11  
Duong, Minh-Duc, 260  
Duong, Thi Thanh Tu, 309  
Duong, Tran Binh, 53  
Duong, Van Tu, 110, 130

## E

- El-Rabaie, Sayed, 484

## F

- Ferrero, Fabien, 608, 623

## G

- Giap, Hai-Binh, 574

## H

- Ha, Le Hai, 246  
Hai, Hong Thi Chu, 519  
Hai, Thanh Le Thi, 45, 550  
Ho, Dat Xuan, 615  
Hoan, Pham Minh, 168  
Hoang, Anh-Tu, 260  
Hoang, Giang Vu, 541  
Hoang, Minh Duc, 309  
Hoang, Nguyen Ba, 615  
Hoang, Quang-Trung, 484

Hoang, Tran Manh, 650  
 Hoang, Uyen Kim, 615  
 Hoang, Xuan Tung, 39  
 Hoang-Nhu, Dong, 287, 418, 445  
 Ho-Dac, Hung, 121  
 Hua-Thi-Hoang, Yen, 76  
 Huong, Pham Thi Viet, 11  
 Huong, Trieu Thu, 519  
 Huu, Cuong Thieu, 351  
 Huu, Phat Nguyen, 45, 550  
 Huy, Hoang Quang, 11  
 Huynh, Tuan Van, 76  
 Huynh, Viet Quoc, 76

**J**

Jin, Wenquan, 253

**K**

Khan, Anam Nawaz, 253  
 Khanh, Lam Ho, 591  
 Khanh, Pham Duy, 11  
 Khoi, Huy, 1  
 Kien, Nguyen Phan, 83, 379  
 Kim, Dohyeun, 253  
 Kim, Taejoon, 650  
 Kim-Van, Kien, 287

**L**

Lam, Sinh Cong, 476  
 Le – Nguyen Thi, Thuy, 386  
 Le Dinh, Anh Huy, 298  
 Le Dinh, Thang, 194  
 Le Thanh, Nhan, 455  
 Le, Anh Ngoc, 253  
 Le, Duc M., 601  
 Le, Duc Minh, 22  
 Le, Duy Tung, 71  
 Le, Duy, 630  
 Le, Huy Trinh, 623  
 Le, Nguyen Thi, 210  
 Le, Nhan Dong, 237  
 Le, Thai Minh Phan, 437  
 Le, Viet Nam, 221  
 Linh, Nguyen Thi Dieu, 557  
 Long, Hoang Phi, 246  
 Luong, Huong Hoang, 279, 669  
 Luu, Anh Khoa Lanh, 110  
 Ly, Ngoc-Thuan, 260

**M**

Ma, Xuan-Quang, 99  
 Mahalakshmi, S., 660  
 Mahendrakar ShyamRao, Muneshwara, 581

Mahendrakar ShyamRao, Swetha, 581  
 Mai, Truc Thuy Nguyen, 410  
 Manh, Hoang Tran, 550  
 Manh, Thao Nguyen, 623  
 Minh, Nghia Pham, 351  
 Minh, Nguyen Canh, 484  
 Minh, Quan Nguyen, 158  
 Minh, Tuan Dang, 427, 465  
 Minh, Tuoi Tran Thi, 427, 465  
 Mishra, Ashish, 640  
 My, Than Tra, 379

**N**

Narayappa Manjunath, Praveen, 581  
 Ngan, Nguyen Thi, 1  
 Ngo, Khanh Hieu, 298  
 Ngo, Minh Tan, 526  
 Ngo, Truong-Giang, 394  
 Ngo, Viet H. H., 608  
 Ngoc Thang, Tran, 360  
 Ngoc, Can Nguyen, 158  
 Ngoc, Nguyen Thi Thuy, 320  
 Ngoc, Pha Pham, 550  
 Ngoc, Thien Pham, 45  
 Ngoc, Tuan Nguyen, 185, 194, 202  
 Nguyen, Anh-Tu, 495  
 Nguyen, Anh Thu T., 601, 615  
 Nguyen, B. Hoang, 601  
 Nguyen, Ba Cao, 650  
 Nguyen, Binh T. H., 39  
 Nguyen, Cuong, 630  
 Nguyen, Danh Huy, 62  
 Nguyen, Dinh Cong, 221, 237  
 Nguyen, Hai Thanh, 279, 402, 410, 437, 669  
 Nguyen, Hai Vinh Cuong, 121  
 Nguyen, Hoang Long, 221  
 Nguyen, Ho-Si-Hung, 151  
 Nguyen, Huy Hung, 110, 130  
 Nguyen, Khanh Kim, 437  
 Nguyen, Khoi Dinh, 669  
 Nguyen, Manh Cuong, 62  
 Nguyen, Minh Quan, 178  
 Nguyen, Ngoc Hai, 526  
 Nguyen, Ngoc-Kien, 99  
 Nguyen, Nhu Toan, 62  
 Nguyen, Phuong Dieu T., 615  
 Nguyen, Quoc-Cuong, 568  
 Nguyen, Quy C., 608  
 Nguyen, Tan Tien, 110, 130  
 Nguyen, The Bao, 121  
 Nguyen, The Cuong, 237  
 Nguyen, Thi Thu Nga, 309

- Nguyen, Thi-Xuan-Huong, 394  
Nguyen, Thu Anh Thi, 437  
Nguyen, Thuong H. N., 601, 608, 615  
Nguyen, Trong-The, 394  
Nguyen, Trung Tan, 535  
Nguyen, Tung Lam, 62  
Nguyen, Tuong Quan, 237  
Nguyen, Van Quan, 237  
Nguyen, Van-Anh, 62, 333  
Nguyen, Van-Tinh, 99  
Nguyen, Van-Truong, 230, 568, 574  
Nguyen, Viet Dung, 158, 178  
Nguyen, Xuan Bach, 178  
Nguyen-Duy, Anh, 418  
Nguyen-Xuan, Ha, 287, 418, 445  
Nhung, Luu Thi Hong, 210  
Ninh, Chi Khanh, 455  
Ninh, Duy Khanh, 455
- P**  
Pham, Diem Ngoc Thi, 437  
Pham, Duc-Binh, 333  
Pham, Duc-Duong, 151  
Pham, Duy-Thi, 333  
Pham, Linh Thuy Thi, 410  
Pham, Thi Hai Hau, 601  
Pham, Tuan V., 601, 608, 615  
Phan, Kha Duy, 455  
Phan, Nhi Yen K., 402  
Phan, Thuan Nguyet, 386  
Phu, Dang Nguyen, 53  
Phuong, Han Minh, 168  
Prakash, Kolla Bhanu, 484
- Q**  
Quoc, Dung Duong, 351  
Quoc, Huy Le, 623
- R**  
Ragunthar, T., 660  
Rizwan, Atif, 253
- S**  
Subash, Ambika Rani, 660
- T**  
Than, Nhan Le, 623  
Thang, Tran Ngoc, 141, 246  
Thanh, Dat Tran, 185, 194  
Thi Ngoc Anh, Nguyen, 360, 369  
Thi Thanh Chau, Do, 369  
Thi Thanh Tuoi, Tran, 360  
Thi, Nguyet Giap, 550
- Thi, Van Tran, 591  
Thuy, Do Thi, 320  
Tien, Dong Nguyen, 427, 465  
Tien, Trung Dang, 185, 202, 541  
Tiep, Vuong Tri, 83  
Tiwari, Meena, 640  
Tong, Luyen, 630  
Tran, Chi Le Hoang, 279, 437  
Tran, Dien Thanh, 437  
Tran, Dinh-Khoa, 151  
Tran, Doan Minh, 221  
Tran, Duc-Quynh, 504  
Tran, Duc-Tan, 53, 476, 484  
Tran, Duy Hung, 526  
Tran, Hai Quan, 298  
Tran, Ngoc Huynh, 601  
Tran, Nguyet Ha T., 615  
Tran, Quang Khoi, 298  
Tran, Quoc-Sang, 270  
Tran, Thi Dung, 33  
Tran, Thi Men, 178  
Tran, Thi-Hoang-Giang, 151  
Tran, Tri Duc, 130  
Tran, Trong Thanh, 669  
Tran, Van Huu, 121  
Tran, Xuan Nam, 650  
Trinh, Anh Tuan, 298  
Trinh, Nguyen Nhat, 210  
Trinh, Thanh, 342  
Trong, Hieu Nguyen, 550  
Trung, Ngo Chi, 320  
Trung, Nguyen Thanh, 379  
Truong, Cong-Doan, 504, 526  
Truong, Cong Duan, 526  
Truong, Lu Tien, 298  
Truong, Mai Hue T., 601  
Tuan Khang, Truong, 360  
Tuan, Nguyen Trung, 168  
Tuan, Tran Anh, 246  
Tuan, Vu Duc, 53  
Tung, Doan Trung, 168
- U**  
Uyen, Nguyen Duc, 53
- V**  
Van Huynh, Tuan, 386  
Van Nguyen, Ngoc, 669  
Van, Bang Le, 158  
Van, Duong Pham, 465  
Van, Ky Le, 351  
Van, Lic Tran, 623  
Van, Tuyen Tran, 194  
Van, Tuyen Tran, 202, 541  
Veena, N., 660

Vinutha, K., 660  
Vo, Trinh Kieu Ngoc, 410  
Vo, Van Len, 121  
Vu, Cong-Thanh, 495  
Vu, Ha Thanh, 22  
Vu, Loi Le, 427, 465

Vu, Tran Anh, 11, 379  
Vu, Xuan-Thang, 574  
**X**  
Xuan, Tinh Tran, 185, 194, 202  
Xuan, Truong Nguyen, 591