SAIGON TECHNOLOGY UNIVERSITY
AIOT LAB VN

Kha Tu Huynh • Cao Hao Thi • Tan Duy Le •
Thinh Vo Xuan • Khanh Pham • Vi Chi Thanh

PROCEEDINGS

THE 2024 CONFERENCE ON APPLICATIONS OF TECHNOLOGY, AUTOMATION, AND CIVIL ENGINEERING

Ho Chi Minh City, November 2024

NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT

SAIGON TECHNOLOGY UNIVERSITY
AIOT LAB VN

Kha Tu Huynh • Cao Hao Thi • Tan Duy Le •
Thinh Vo Xuan • Khanh Pham • Vi Chi Thanh

PROCEEDINGS

THE 2024 CONFERENCE ON APPLICATIONS OF TECHNOLOGY, AUTOMATION, AND CIVIL ENGINEERING

Ho Chi Minh City, November 2024

NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT

Editors: **Kha Tu Huynh, Cao Hao Thi, Tan Duy Le, Thinh Vo Xuan, Khanh Pham, Vi Chi Thanh**

ATAC 2024: Applications of Technology, Automation, and Civil Engineering

The 2024 Conference on Applications of Technology, Automation, and Civil Engineering, ATAC 2024
Ho Chi Minh City, Vietnam, 23 November, 2024
Proceedings

Editors

1. Kha Tu Huynh, International University – VNU-HCM, AIoT Lab VN, Ho Chi Minh City, Vietnam
2. Cao Hao Thi, Saigon Technology University, Ho Chi Minh City, Vietnam
3. Tan Duy Le, International University – VNU-HCM, AIoT Lab VN, Ho Chi Minh City, Vietnam
4. Thinh Vo Xuan, Saigon Technology University, Ho Chi Minh City, Vietnam
5. Khanh Pham, International University – VNU-HCM, AIoT Lab VN, Ho Chi Minh City, Vietnam
6. Vi Chi Thanh, International University – VNU-HCM, AIoT Lab VN, Ho Chi Minh City, Vietnam

Prefaces

These proceedings, presented in a single volume, contain the papers from the 2024 ATAC International Conference, held on November 23, 2024, at Saigon Technology University (STU), Vietnam. The papers in this volume reflect the breadth and diversity of discussions during the conference, encompassing topics such as Artificial Intelligence (AI) and Machine Learning (ML), Deep Learning, the Internet of Things (IoT) and Smart Systems, Big Data, Robotics, Intelligent Equipment, and Automation, Sustainable Machine Design, Civil Engineering, Construction Management, Information Technology, Electronics and Telecommunications Engineering, and Mechatronics Engineering.

ATAC 2024 marks the inaugural edition of this conference, and we are excited to welcome researchers and experts from various fields to exchange knowledge and ideas. The goal of the conference is to provide a platform for researchers, practitioners, and industry experts to discuss cutting-edge technologies and their applications across a wide range of domains.

This year, ATAC 2024 received a total of 50 submissions from authors worldwide, and after a rigorous evaluation process, 24 papers were accepted for presentation. Each submission underwent a fair and thorough review process, with each paper being reviewed by two independent reviewers.

We are honored to have three distinguished keynote speakers who shared their valuable insights at the conference:

- **Assoc. Prof. Razvan Beuran**, an expert in Information Technology, presented on *“Applications of AI/ML to Cybersecurity.”*
- **Dr. Huynh Vo Trung Dung**, an expert in Electronics Engineering, spoke on *“The Synergy of Blockchain and AI in 6G Networks: Opportunities, Challenges, and Prospective Pathways.”*
- **Ms. Ta Thuy Trang**, an expert in Civil Engineering, delivered a talk on *“Machine Learning Algorithms for Predicting Compressive Strength of Normal and High-Early Strength Concrete: A Case Study in Binh Thuan, Vietnam.”*

We would like to express our deepest gratitude to the Program Committee, external reviewers, sponsors, and all the participants who contributed to making ATAC 2024 a success. Special thanks also go to the organizing committee for their unwavering efforts in organizing this event.

We hope that ATAC 2024 serves as a significant platform for sharing ideas and fostering collaborations in advanced technology fields, and we look forward to seeing continued

contributions from researchers, scientists, academics, and industry professionals at future editions of ATAC.

November 2024
ATAC 2024 Organizing Committee

Cao Hao Thi

Kha Tu Huynh

Tan Duy Le

Thinh Vo Xuan

Organization

Program Committee

Vi Chi Thanh, International University – VNU-HCM, AIoT Lab VN
Khanh Pham, International University – VNU-HCM, AIoT Lab VN
Ho Thu Hien, Ho Chi Minh City University of Science - VNU-HCM
Le Song Thanh Quynh, Ho Chi Minh City University of Technology - VNU-HCM
Nguyen Duc Duy, Ho Chi Minh City University of Technology - VNU-HCM
Vo Dang Khoa, Ho Chi Minh City Open University
Le Khanh Dien, Saigon Technology University
Bui Cong Thanh, Saigon Technology University
Tang Van To, Saigon Technology University
Lương An Vinh, Saigon Technology University
Thien Khai Tran, HCMC University of Foreign Languages - Information Technology
Thanh Cao Tien, HCMC University of Foreign Languages - Information Technology
Hung Do Phi, HCMC University of Foreign Languages - Information Technology
Nguyen Huu Bao Ngan, Dong-A University, Busan, South Korea
Khiem Nguyen Tran Gia, Ulsan National Institute of Science and Technology, Ulsan, Korea
Nguyen Quang Phu, International University – VNU-HCM, AIoT Lab VN
Nguyen Van Dung, Hung Vuong University of Ho Chi Minh City
Ha Minh Tuan, Ho Chi Minh City University of Technology (HUTECH)
Thien Trang Ly, Tien Giang University
Huynh Phuong Thanh Nguyen, Japan Advanced Institute of Science and Technology (JAIST), Japan
Duc Dang Khoi Nguyen, Auburn University, USA
Huu Sy Le, Auburn University, USA
Nguyen Chi Trung, University of Tennessee, Knoxville, USA
Nguyen Hong Quan, Japan Advanced Institute of Science and Technology (JAIST), Japan
Binh Nguyen Le Nguyen, University of Huddersfield, United Kingdom

Conference Organizing Committee

Honorable Chairs

Tran Manh Ha, Information Technology Park - VNU-HCM
Tran Ngoc Thinh, Ho Chi Minh City University of Technology, VNU-HCM
Nguyen Tan Viet Tuyen, University of Southampton, UK
Park Sangyeong, Texas A&M University, USA
Bao An Mai Hoang, AI4IA Universal Knowledge Technology Joint Stock Company

General Chairs

Cao Hao Thi, Saigon Technology University
Kha Tu Huynh, International University – VNU-HCM, AIoT Lab VN
Tan Duy Le, International University – VNU-HCM, AIoT Lab VN
Thinh Vo Xuan, Saigon Technology University

Financial Chairs

Nguyen Ngoc Mai Oanh, AIoT Lab VN, AI4AI

Publication Chairs

Tan Duy Le, International University – VNU-HCM, AIoT Lab VN
Thinh Vo Xuan, Saigon Technology University
Duc Dat Pham, AIoT Lab VN

Secretary

Duc Dat Pham, AIoT Lab VN
Ho Thi Thanh Tuyen, Saigon Technology University
Nam Anh Dang Nguyen, AIoT Lab VN

Additional Reviewers

1. **Vi Chi Thanh,** International University – VNU-HCM, AIoT Lab VN
2. **Mai Hoang Bao An,** International University – VNU-HCM, AIoT Lab VN
3. **Huynh Vo Trung Dung,** International University – VNU-HCM, AIoT Lab VN
4. **Nguyen Quang Phu,** International University – VNU-HCM, AIoT Lab VN
5. **Dang Tam Nhan,** International University – VNU-HCM, AIoT Lab VN
6. **Nguyen Trung Nghia,** International University – VNU-HCM, AIoT Lab VN
7. **Le Thi Ngoc Hanh,** International University – VNU-HCM, AIoT Lab VN
8. **Vuong Quoc Bao,** International University – VNU-HCM, AIoT Lab VN
9. **Nguyen Duc Duy,** Ho Chi Minh City University of Technology - VNU-HCM
10. **Vo Dang Khoa,** Ho Chi Minh City Open University
11. **Le Thanh Tung,** Ho Chi Minh City University of Science – VNUHCM
12. **Do Pham Minh Thu,** University of Technology Sydney
13. **Le Nguyen Binh Nguyen,** University of Huddersfield
14. **Nguyen Duc Dang Khoi,** Auburn University
15. **Huynh Ngoc Thai Anh,** Can Tho University
16. **Le Huu Sy,** Auburn University
17. **Do Phi Hung,** HCMC University of Foreign Languages – Information Technology (HUFLIT)
18. **Nguyen Tan Viet Tuyen,** University of Southampton
19. **Ha Minh Tuan,** Ho Chi Minh City University of Technology (HUTECH)
20. **Ho Thu Hien,** Ho Chi Minh City University of Science – VNUHCM
21. **Nguyen Tien Gia Khiem,** Ulsan National Institute of Science and Technology, UNIST
22. **Ngan Nguyen,** Dong-A University

Table of contents (24 papers)

The Impact of Artificial Intelligence and Machine Learning on the Online Learning Process in Higher Education	02-07
Application of YOLOv10 in Real-Time Fist Gesture Recognition for Controlling Electromechanical Devices	08-12
Developing a Convolutional Neural Network (CNN) Model for Face Recognition in Education and School Security	13-18
Enhanced Warrant Pricing: A Machine Learning Approach vs. Black-Scholes in Vietnam	19-24
Building a Self-Driving Model Using Yolov8 and Raspberry Pi 4	25-30
Computer Vision and AI: Applications and Trends	31-34
ACNE10M: Developing a Smart Clinical Decision Support System for Enhanced Acne Care	35-40
EconVNNewsBot: Leveraging Question Answering System with Retrieval-Augmented Generation and Chain-of-Thought Reasoning for Economic News in Vietnam	41-46
Development of a CNN-Based Application for Identifying Common Flower Species in Vietnam	47-52

Table of contents (24 papers)

ASPS - AI-Driven Smart Parking Solutions with Integrated Web and Mobile Applications	53-58
EVALLOS: An Effective Solution for CLO and SLO Assessment, Supporting Enhanced Teaching Quality	59-64
Tactiwave: AI-Driven Tactile Communication for Deaf-Blind Support	65-70
NexMIND: AI-Driven Multimodal Integration for Neurological Diagnostics and Health Monitoring	71-76
SYNERGY: Smart System for Home Device Control and User Interaction	77-82
Multimodal Skin Lesion Classification With Convolutional Attention Mechanism	83-88
Efficient Image Retrieval and Activity Recognition in Vietnamese Daily Life using AI	89-94
AI Applications for Students in University	95-100
Multi-Sector Data Integration Solution for Ho Chi Minh City Based on the Data Integration Platform	101-106
Digital Control Strategy for Step-Down Energy Conversion Systems	108-112

Table of contents (24 papers)

Multi-Node Fingerprint Attendance System with Centralized Feature Storage	113-116
Design and Manufacturing of a Control System for a Pneumatic Robotic Arm for Training Purposes	117-121
Overview of natural ventilation solutions and the potential of applying solar chimneys to enhance energy efficiency and thermal environment in buildings in Ho Chi Minh City, Vietnam	123-128
Deterministic factors governing the load-bearing capacity of concrete-filled steel tube	129-132
A Research on the Compressive Resistance of Concrete Using Sea Sand and Mixed with Eco-CSSB Additive Compared to Concrete Using Sea Sand Without Additives and Traditional Concrete (in Ben Tre Province)	133-137

Section 1:
Information Technology (IT)

Description: This field leverages cutting-edge technologies such as Artificial Intelligence, Deep Learning, IoT, and Big Data to develop intelligent systems, optimize data management, and enable smart, interconnected solutions across industries.

The Impact of Artificial Intelligence and Machine Learning on the Online Learning Process in Higher Education

Nguyen Ngoc Bich
Marketing Faculty
University of Finance and Marketing
Ho Chi Minh City, Viet Nam
nnbich@ufm.edu.vn

***Abstract*—Artificial Intelligence (AI) and Machine Learning (ML) have become transformative technologies in the realm of higher education, particularly within the domain of online learning. This study explores the impact of AI on online education in Vietnam, focusing on its current status, challenges, and the ethical concerns it presents. By utilizing qualitative research methods and conducting a comprehensive literature review, the study reveals that AI has the potential to significantly personalize learning experiences, enhance the effectiveness of teaching, aid in classroom management, and provide more efficient assessments of student outcomes. Nevertheless, the adoption of AI is not without its hurdles, as it faces several technical challenges, including issues related to security and data privacy, ethical dilemmas in AI-driven decision-making processes, and the risk of over-reliance on technology. This analysis offers valuable insights into both the benefits and challenges of integrating AI into online education in Vietnam, thereby contributing to the formulation of future policies and strategies to optimize its application.**

Keywords—intelligent interaction, learning analytics, learning processes, learning support systems, personalized learning

I. Introduction

The rapid advancement of technology has had a transformative effect on various sectors, with education being one of the most profoundly impacted. In recent years, artificial intelligence (AI) and machine learning (ML) have emerged as powerful tools for enhancing online learning environments [1]. These technologies hold immense potential in personalizing learning experiences, improving teaching effectiveness, and optimizing student outcomes, particularly as the world witnessed a surge in online learning due to the COVID-19 pandemic [2]. However, despite the promising advantages, the integration of AI and ML in online education presents significant challenges, including concerns around data privacy and ethical considerations [3].

Previous studies, such as those by [4] has underscored the role of AI in personalizing education, while Holmes et al. (2019) discuss the ethical implications of AI in education [5] . This study aims to assess the impact of AI and ML on the online learning process in higher education, focusing particularly on the Vietnamese context. The research will address the current level of AI and ML integration in online learning, the challenges encountered, and the ethical issues involved in this process.

Understanding these effects is vital for developing effective and equitable online learning environments, especially in the post-pandemic era. The structure of the paper includes a review of the relevant literature, an explanation of the research methodology, a discussion of the findings, and a presentation of conclusions and recommendations. By examining the role of AI and ML in online learning, this study seeks to offer actionable insights to improve the quality and fairness of higher education.

This research holds both theoretical and practical significance. Theoretically, it contributes to the growing body of knowledge on AI applications in education, providing an in-depth analysis of the advancements and challenges posed by AI technology [6]. Practically, it offers valuable insights into the current situation in Vietnam, proposing policy recommendations and solutions for enhancing AI applications in online learning. These recommendations will help higher education institutions optimize AI usage while aiding educators and policymakers in understanding both the benefits and limitations of this technology, ultimately leading to more effective strategies for improving educational quality [7].

II. Literature Review

Numerous studies have examined the transformative effects of AI and machine learning on online learning, particularly in enhancing personalized learning experiences and improving student outcomes. For instance, Kizilcec, Piech, and Schneider (2013) indicated that AI can be used to personalize students' learning experiences in massive open online courses (MOOCs) [8]. by analyzing students' data and adjusting the content to match individual abilities and needs. This personalized approach has been shown to improve learning performance and increase engagement levels [8].

Virtual assistants and chatbots have also become key components in many online learning platforms. Winkler and Söllner (2018) found that virtual assistants offer timely learning support by answering student inquiries and providing relevant learning materials, effectively reducing the administrative burden on instructors and allowing them to focus on higher-priority tasks [9].

In addition to personalization, AI is widely used to analyze learning data and predict student performance. Brooks, Erickson, and Greer (2014) demonstrated that machine learning models could predict learning outcomes with high accuracy by analyzing variables such as learning history, engagement levels, and test results. These predictions allow educational

institutions to intervene early, supporting struggling students more effectively [10].

AI has also proven useful in automating grading processes. Burstein, Chodorow, and Leacock (2004) developed the e-rater automated scoring system, which employs machine learning algorithms to grade essays based on structural, grammatical, and vocabulary criteria. Their research showed that the system is highly accurate and serves as a valuable tool for educators in assessing student writing. [11].

Furthermore, AI can significantly enhance collaborative learning environments. The study by Rosé et al. (2008) developed an AI-based system to support collaborative learning, leveraging natural language processing to analyze student interactions and provide real-time feedback, thus improving the quality and effectiveness of group discussions [12].

III. Theorical Framework

A. Artificial Intelligence (AI)

Artificial Intelligence (AI) is a branch of computer science that focuses on developing systems capable of executing tasks that typically require human intelligence, such as perception, reasoning, learning, and adaptation. As defined by Russell and Norvig (2021), AI involves "the study of intelligent agents, where an intelligent agent is a system that perceives its environment and takes actions that maximize its chances of achieving its goals." This highlights the core purpose of AI: to create systems that can autonomously make decisions based on their understanding of the environment [13].

AI is generally divided into two categories: narrow AI (also referred to as weak AI) and general AI (strong AI). Narrow AI refers to systems designed to perform specific tasks, such as voice recognition or predictive analytics in healthcare. In contrast, general AI aims to replicate the full range of human cognitive abilities, enabling it to perform any intellectual task that a human can undertake. While current AI applications in fields like education are mostly limited to narrow AI, there is significant potential for advancements that could lead to general AI applications in the future [14].

B. Machine Learning (ML)

Machine Learning (ML) is a subset of artificial intelligence that focuses on creating algorithms and models enabling computers to learn from data, adapt, and improve performance over time without the need for explicit task-specific programming. As Mitchell (1997) describes it, machine learning is "the study of computer algorithms and models that allow systems to perform specific tasks by identifying patterns and making inferences from data" [15]. This ability to adapt based on data is a key distinction of ML compared to traditional programming.

Machine learning can be classified into three main types:

- **Supervised Learning**: In this approach, models learn from labeled datasets to make predictions about new, unseen data. For example, in the field of education, supervised learning could be applied to predict student grades based on variables such as study time, attendance, and participation [16].
- **Unsupervised Learning**: In contrast, unsupervised learning involves models analyzing unlabeled data to uncover hidden structures or patterns. A typical educational application might involve clustering students into groups based on similar learning behaviors or performance metrics [17].
- **Semi-supervised Learning:** This method utilizes a combination of labeled and unlabeled data to improve the learning process, which is particularly useful when acquiring labeled data is costly or time-consuming. This approach allows the model to make inferences using a small amount of labeled data while learning from a larger set of unlabeled data [18].

C. Applications of AI and Machine Learning in Higher Education

Artificial intelligence (AI) and machine learning (ML) are increasingly being integrated into higher education to enhance students' learning experiences and improve the efficiency of teaching. These technologies are applied in various ways, including personalized learning systems, virtual assistants, predictive analytics for decision-making, and automated assessment tools [19].

IV. Method

The author adopts a qualitative research approach to examine the effects of Artificial Intelligence (AI) and Machine Learning (ML) on online learning in higher education. This methodology emphasizes the collection of non-numeric data, such as observations and content analysis [20] drawing from sources such as books, scholarly articles, research papers, and other relevant documents. The author not only analyzes theoretical perspectives but also offers practical insights and best practices applicable across various contexts: FPT University, National Economics University, and VinGroup's Edtech initiatives, as well as educational systems in other countries. By focusing on real-world applications, the research provides a comprehensive understanding of how AI and ML can enhance learning outcomes and improve educational efficiency on a global scale.

V. Analysis of the Impact and Potential of AI in Online Learning

A. Personalized Learning

Personalized learning refers to educational approaches designed to cater to the distinct needs, preferences, and interests of individual students. Through AI, it becomes possible to create adaptive learning environments that adjust in real-time, based on data from student interactions and performance.

One of the core methods by which AI facilitates personalized learning is through adaptive learning technologies. These systems rely on algorithms to analyze student behavior, such as time spent on tasks, quiz outcomes, and interaction patterns, to identify individual strengths and weaknesses. For instance, platforms like Knewton utilize machine learning algorithms to deliver customized learning content to each student, enabling them to master topics more

effectively [21]. This tailored approach ensures that students receive the support they need at the precise moment, improving learning outcomes and efficiency.

In addition to adaptive technologies, AI also enhances personalized learning through Intelligent Tutoring Systems (ITS), which provide individualized tutoring based on the learner's current level of understanding. For example, Carnegie Learning's Math Tutor employs AI to deliver personalized feedback and hints, adjusting to the learner's pace and needs [22]. This one-on-one interaction allows students to progress at their own pace, fostering a deeper comprehension of the material.

AI-driven analytics tools further aid educators in supporting personalized learning by offering insights into students' learning patterns. Predictive analytics can identify students who may be at risk of falling behind and recommend timely interventions [23]. For example, AI algorithms can predict which students might struggle with upcoming material based on their past performance and engagement, allowing educators to provide additional resources proactively.

Beyond academic support, AI can also personalize non-academic aspects of learning. Virtual assistants and chatbots, such as Georgia Tech's Jill Watson, can respond to student inquiries, offer study tips, and assist with time management, creating a more responsive and supportive learning environment [24]. These AI tools ensure that students' unique needs are addressed, making the learning experience more engaging and effective.

B. Enhancing Teaching Effectiveness

One of the key contributions of AI in education is its ability to automate administrative tasks, enabling educators to dedicate more time to teaching and lesson planning rather than paperwork. AI-powered systems streamline processes such as grading, attendance tracking, and assignment management, reducing teachers' administrative burden and allowing them to engage more directly with students [25].

In addition, AI provides educators with advanced analytics tools that generate valuable insights into student performance and learning behaviors. These tools analyze large datasets to uncover patterns that traditional assessment methods might overlook. For instance, AI-enabled learning management systems (LMS) can track student engagement, participation, and academic performance in real-time, allowing educators to identify students who may require additional support [23]. This data-driven approach enables teachers to tailor their instructional strategies to better meet individual student needs, enhancing overall teaching effectiveness.

AI also plays a pivotal role in personalized learning, directly contributing to improved teaching outcomes. By leveraging AI to deliver customized learning experiences, educators can ensure that each student receives appropriate levels of support and challenge. Intelligent Tutoring Systems (ITS), for example, adapt to the unique learning pace and style of each student, providing personalized feedback and resources, which facilitates a deeper understanding of complex concepts. Additionally, this allows educators to monitor student progress more closely and adjust their teaching methods accordingly [22].

Moreover, AI-powered virtual assistants and chatbots serve as invaluable resources for both students and educators. These tools can address common queries, offer study materials, and even provide tutoring in specific subjects. Their ability to deliver instant support keeps students motivated and engaged, while simultaneously freeing up educators to focus on more intricate instructional tasks [24].

C. Supporting Classroom Management and Assessing Learning Outcomes

AI-driven systems significantly enhance classroom management by automating routine tasks such as attendance tracking, scheduling, and communication with students and parents. For example, AI-powered attendance systems using facial recognition technology can automatically record student attendance, saving valuable time for educators and allowing them to focus on instructional tasks [25].

Beyond administrative functions, AI contributes to more effective assessment of learning outcomes by offering advanced analytics that assess student performance in real-time. Traditional assessments, which often rely on periodic exams and quizzes, may not provide a full picture of a student's learning trajectory. AI, however, continuously analyzes various data points—such as online interactions, assignments, and participation—to deliver a comprehensive evaluation of academic progress [23].

A notable application of AI in assessment is automated grading. These systems use natural language processing and machine learning algorithms to grade essays, assignments, and exams with high accuracy and consistency. Platforms like GradeScope employ AI to assess both handwritten and typed responses, providing detailed, timely feedback while reducing the grading burden on educators [26]. This efficiency enables teachers to allocate more time to personalized instructional support.

Moreover, AI-powered predictive analytics can identify students at risk of falling behind by analyzing engagement levels, performance trends, and other data patterns. This allows educators to intervene early by offering additional support, such as tutoring or modified instructional strategies, before these issues become critical [27].

AI also plays a crucial role in formative assessments by offering real-time feedback. Intelligent Tutoring Systems (ITS) provide instant hints and corrections, enabling students to learn from mistakes and improve their understanding of the material. This continuous feedback loop fosters active learning and leads to better academic outcomes [22].

In summary, AI enhances classroom management and assessment by automating administrative tasks, offering real-time performance analytics, facilitating automated grading, and enabling predictive and formative assessments. These capabilities help educators create more efficient, responsive, and supportive learning environments.

D. Examples and Case Studies

1) Personalized Learning Platform Knewton:

Knewton is a leading example of how AI can be used to personalize learning. The platform leverages machine learning algorithms to analyze student learning data and adapt the content to meet individual needs and capabilities. Research indicates that students using Knewton perform better and report higher levels of satisfaction compared to those utilizing traditional learning methods [28]

2) Virtual Assistant Jill Watson at Georgia Tech

Jill Watson, developed at Georgia Tech, is an AI-powered virtual assistant designed to address frequently asked questions from students enrolled in online courses. This virtual assistant has significantly reduced the workload for instructors by providing timely responses, while also enhancing the learning experience by ensuring students receive quick and reliable support [24].

3) Automated Assessment System e-rater

E-rater is an automated assessment tool that uses machine learning algorithms to evaluate essays based on factors such as structure, vocabulary, and grammar. Widely employed in both exams and online courses, e-rater has been instrumental in reducing instructor workload and ensuring objective and consistent grading [11].

VI. Analysis of Online Learning in Universities

A. Current State of Online Education in Vietnam

In response to the COVID-19 pandemic, online education in Vietnam has grown rapidly and become a vital component of the education system. According to a report by the Vietnamese Ministry of Education and Training (2020), over 90% of universities and colleges have adopted online learning methods to ensure the continuity of education during disruptions [29]. Platforms such as Zoom, Microsoft Teams, and Google Classroom have become widely used, enabling students to remotely access lectures and educational resources.

Moreover, the integration of Artificial Intelligence (AI) in Vietnam's educational landscape is gaining momentum, particularly in areas such as teaching support and learning management. Several universities have begun incorporating AI technologies into their online learning systems. For example, FPT University has successfully implemented an AI-driven system to personalize student learning experiences, thereby optimizing the learning process and enhancing teaching effectiveness (Nguyen et al., 2021). This shift towards AI-driven education reflects a broader trend in modernizing Vietnam's educational infrastructure.

B. Analysis of specific universities in Viet Nam

Exploring the application of AI in education often requires evaluating its effectiveness in real-world contexts and assessing its impact on teaching practices [1]. In this regard, the author examines specific case studies in Vietnam.

1) FPT University:

FPT University has been a leader in integrating AI into education in Vietnam. The university has developed an AI-based system that personalizes the learning experience for students by using machine learning algorithms to analyze individual learning data. This system adapts content delivery to match each student's abilities and needs. Research has shown that this AI-driven approach has led to improvements in both academic performance and student satisfaction with the learning process [30].

2) National Economics University (NEU):

NEU has also embraced AI through its implementation of an online learning platform designed to support both teaching and classroom management. This platform not only facilitates online education but also includes AI features that analyze student performance and provide suggestions for improvement. According to Hoang et al. (2021) the use of AI at NEU has significantly reduced the time spent on classroom management tasks while simultaneously improving the quality of instruction [31].

3) VinGroup's Edtech Project:

VinGroup has made substantial investments in educational technology, focusing on the development of AI solutions aimed at enhancing the quality of online education. Its Edtech project employs AI to personalize learning content, support educators in delivering instruction, and provide automated assessments. The positive outcomes from this project have contributed to raising the overall quality of education in Vietnam [32].

C. Analysis in the world

Holmes and Tuomi (2022) emphasize that AI's applications in education contribute significantly to personalized learning and the overall efficiency of educational systems. Sethi et al. (2020) further illustrate this by examining specific cases across different countries. In the United States, a study on an AI-powered adaptive learning platform was found to enhance student engagement and performance in an introductory physics course. Specifically, the research indicated that students utilizing the AI tutoring system achieved double the learning gains compared to those in traditional classroom settings. The authors of the study believe that the system's ability to provide personalized feedback and allow students to adjust their learning pace are key advantages over conventional teaching methods. [33]



Fig. 1. A comparison of mean post-test performance between students taught with the active lecture and students taught with the AI tutor.

Similarly, in China, research at Beijing Normal University on AI-driven intelligent tutoring systems revealed substantial improvements in mathematics performance [34, 35]. Meanwhile, in Norway, the University of Bergen demonstrated through a chatbot study that personalized feedback increased both student engagement and satisfaction [36]. In the United Kingdom, the University of Nottingham's research on AI-driven grading systems found that these systems not only streamlined the grading process but also reduced the workload on educators [37]. Additionally, a study at the University of Toronto in Canada found that AI-powered tools for analyzing student writing improved the quality of submissions and enhanced teacher productivity [38]. These findings collectively indicate that AI can effectively address both educational challenges and cultural complexities, while also shaping the future of personalized, efficient learning environments

VII. Challenges and Ethics in the Use of AI

A. Technical and Infrastructure Challenges

One of the key challenges in integrating AI into the online learning process is the considerable demand for technical infrastructure. AI systems rely on vast amounts of data to function effectively, which requires robust storage and processing capabilities. Educational institutions must invest in high-performance servers, scalable storage solutions, and reliable high-speed internet to support these AI-driven systems. However, many universities and schools in Vietnam continue to struggle with outdated equipment, insufficient bandwidth, and limited access to advanced technologies, making AI implementation more challenging [39].

B. Security and Privacy Issues

Security and privacy remain pressing concerns when incorporating AI into education. The collection and analysis of student data, including academic records and behavioral information, pose significant privacy risks if not managed appropriately. Without stringent cybersecurity measures, personal data can be vulnerable to breaches, making AI systems potential targets for cyber-attacks. Educational institutions must therefore adopt comprehensive data protection strategies, such as encryption, access controls, and regular security audits, to safeguard student information and maintain trust [40].

C. Ethics in AI Decision-Making

While AI has immense potential, its decision-making capabilities raise several ethical issues. AI systems are trained using large datasets, which may inadvertently contain biases related to gender, race, or socioeconomic status. These biases can be replicated or even exacerbated in AI outputs, resulting in unfair or discriminatory outcomes, such as biased grading or unequal access to resources [41]. It is essential to develop and implement measures to identify and mitigate such biases. This can be achieved by using diverse training datasets, regularly auditing AI outputs, and involving ethicists in AI development.

D. Over-Reliance on Technology

Excessive dependence on AI in education poses another challenge. While AI enhances learning and teaching processes, over-reliance on these technologies can inhibit the development of critical thinking and creativity in both students and educators. Students may become overly dependent on AI for answers, neglecting independent problem-solving and analysis [42]. Similarly, educators might rely too much on AI for administrative tasks and assessments, losing touch with students' individual needs. Institutions must strike a balance, ensuring that AI complements human judgment rather than replacing it entirely.

VIII. Conclusions and Policy Implications

The integration of AI and ML into higher education has had a profound impact on improving personalized learning experiences, enhancing teaching effectiveness, and supporting classroom management. These technologies create interactive learning environments, provide instant feedback, and facilitate more flexible learning (Chen et al., 2020). The results suggest that AI can significantly improve educational quality, increase student engagement, and enhance learning outcomes [6].

To fully realize the potential of AI, universities must invest in robust infrastructure, including data storage and processing systems, and ensure that AI tools are customized to meet each institution's specific needs. In terms of privacy and security, strict data protection measures must be implemented to safeguard student and faculty information. Educational policymakers should develop clear frameworks to support AI adoption, while training programs should be established to raise awareness among faculty and students about the benefits and responsible use of AI in online education.

This study does acknowledge certain limitations. The research primarily focuses on a few universities in Vietnam, which may not fully represent the country's higher education system. Additionally, the study relies solely on qualitative methods; future research should incorporate quantitative approaches to identify more specific trends and impacts of AI in education (Creswell, 2013). Expanding future research to include more diverse institutions and analyzing long-term student learning outcomes will provide a more comprehensive understanding of AI's effectiveness in education. Moreover, gathering insights from both students and faculty will offer a fuller picture of how AI is shaping online learning experiences.

Personalized learning systems utilize AI to tailor educational content and teaching methods to the unique needs and abilities of each student, thereby promoting more effective learning [43]. In addition, virtual assistants, such as chatbots, provide timely support to students by answering frequently asked questions and guiding them through learning materials, which significantly enhances student engagement and reduces the administrative workload for educators [25].

References

[1] Dogan, M.E., T. Goru Dogan, and A. Bozkurt, *The use of artificial intelligence (AI) in online learning and distance education processes: A systematic review of empirical studies.* Applied Sciences, 2023. 15(3).

[2] WorldBank, *Vietnam: Digital transformation and the future of work.* World Bank. , 2021.

[3] Akgun, S. and C. Greenhow, *Artificial intelligence in education: Addressing ethical challenges in K-12 settings.* AI and Ethics, 2022. 2(3): p. 431-440.

[4] Castro, G.P.B., et al., *Harnessing AI for Education 4.0: Drivers of Personalized Learning.* Electronic Journal of e-Learning, 2024. 25(2): p. 01-14.

[5] Holmes, W., M. Bialik, and C. Fadel, *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning.* Center for Curriculum Redesign, 2019.

[6] Zawacki-Richter, O., et al., *Systematic review of research on artificial intelligence applications in higher education – where are the educators?* . International Journal of Educational Technology in Higher Education, 2019. 16(1): p. 39.

[7] Selwyn, N., *Should robots replace teachers? AI and the future of education.* Learning, Media and Technology, 2019. 44(2): p. 134-147.

[8] Kizilcec, R.F., C. Piech, and E. Schneider, *Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses.* . Proceedings of the Third International Conference on Learning Analytics and Knowledge, 2013: p. 170-179.

[9] Winkler, R. and M. Söllner, *Unleashing the potential of chatbots in education: A state-of-the-art analysis.* IEEE Transactions on Learning Technologies, 2018. 11(4): p. 495-507.

[10] Brooks, C., G. Erickson, and J. Greer, *Modelling and quantifying the behaviours of students in lecture capture environments.* IEEE Transactions on Learning Technologies, 2014. 7(4): p. 385-397.

[11] Burstein, J., M. Chodorow, and C. Leacock, *Automated essay evaluation: The criterion online writing service.* AI Magazine, 2004. 25(3): p. 27-36.

[12] Rosé, C.P., et al., *Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning.* International Journal of Computer-Supported Collaborative Learning, 2008. 3(3): p. 237-271.

[13] Russell, S. and P. Norvig, *Artificial intelligence: A modern approach.* Pearson, 2021.

[14] Nilsson, N.J., *The quest for artificial intelligence.* Cambridge University Press, 2010.

[15] Mitchell, T.M., *Machine learning.* McGraw-Hill, 1997.

[16] Zhu, X., *Semi-supervised learning literature survey: Computer Science Technical Report.* University of Wisconsin-Madison, 2008.

[17] Kotsiantis, S.B., *Supervised machine learning: A review of classification techniques.* Emerging Artificial Intelligence Applications in Computer Engineering, 2007. 160: p. 3-24.

[18] Zhu, X.J., *Machine teaching: An inverse problem to machine learning and an approach toward optimal education.* Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015: p. 4083-4087.

[19] Baker, R.S. and P.S. Inventado, *Educational data mining and learning analytics in learning analytics.* Springer., 2014: p. 61-75.

[20] Creswell, J.W., *Qualitative inquiry and research design: Choosing among five approaches.* Thousand Oaks, CA: SAGE Publications, 2013.

[21] Colchester, K., et al., *A survey of artificial intelligence techniques employed for adaptive educational systems within e-learning platforms.* Journal of Artificial Intelligence and Soft Computing Research, 2017. 7(`): p. 47-64.

[22] VanLehn, K., *The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems.* Educational Psychologist, 2011. 46(4): p. 197-221.

[23] Siemens, G. and P. Long, *Penetrating the Fog: Analytics in Learning and Education.* EDUCAUSE Review, 2011. 46(5): p. 31-40.

[24] Goel, A. and L. Polepeddi, *Jill Watson: A virtual teaching assistant for online education.* Proceedings of the 1st International Conference on AI in Education, 2018: p. 561-564.

[25] Luckin, R., et al., *Intelligence unleashed: An argument for AI in education.* Pearson Education, 2016.

[26] Singh, A., et al., *Gradescope: a fast, flexible, and fair system for scalable assessment of handwritten work.* In Proceedings of the fourth acm conference on learning@ scale, 2017: p. 81-88.

[27] Piech, C., et al., *Modeling how students learn to program.* Proceedings of the 44th ACM Technical Symposium on Computer Science Education, 2013: p. 153-160.

[28] Knewton, *Knewton: The world's leading adaptive learning technology provider.* 2013.

[29] Training, M.o.E.a., *Summary report for the 2019-2020 academic year and directions for the 2020-2021 academic year.* 2020.

[30] Ngo, T.T.A., *The perception by university students of the use of ChatGPT in education.* International Journal of Emerging Technologies in Learning (Online), 2023. 18(17).

[31] Hoang, Q.M., T.A. Nguyen, and D.T. Tran, *Applying AI in the management of learning and teaching in Vietnamese universities.* International Journal of Artificial Intelligence in Education,, 2021. 32(1): p. 45-59.

[32] Cuong, D.X., N.T. Huong, and V.Q. Anh, *The Use of Google Solutions in Teaching and Learning: A Case of Vietnam.* Educational technology, 2022.

[33] Manning, A.J., *Professor tailored AI tutor to physics course. Engagement doubled.* The Harvard Gazette, 2024.

[34] Sethi, K., et al., *Impact of social and cultural challenges in education using AI. In Revolutionizing Education in the Age of AI and Machine Learning* IGI Global, 2020: p. 130-151.

[35] Holmes, W. and I. Tuomi, *State of the art and practice in AI in education.* European Journal of Education, 2022. 57: p. 542-570.

[36] YILDIZ DURAK, H., *Conversational agent-based guidance: examining the effect of chatbot usage frequency and satisfaction on visual design self-efficacy, engagement, satisfaction, and learner autonomy.* Educ Inf Technol, 2023. 28: p. 471–488.

[37] Nottingham, *AI at Nottingham.* University of Nottingham, 2024.

[38] Xue, Y., *Towards automated writing evaluation: A comprehensive review with bibliometric, scientometric, and meta-analytic approaches.* Educ Inf Technol 2024.

[39] Tran, K. and T. Nguyen, *Preliminary research on the social attitudes toward AI's involvement in Christian education in Vietnam: Promoting AI technology for religious education.* Religions 2021. 12(3): p. 208.

[40] Bostrom, N. and E. Yudkowsky, *The ethics of artificial intelligence.* In Cambridge Handbook of Artificial Intelligence, 2018: p. 316-334.

[41] O'Neil, C., *Weapons of math destruction: How big data increases inequality and threatens democracy.* New York: Crown Publishing Group, 2017.

[42] Carr, N., *The shallows: What the Internet is doing to our brains. New York:.* Norton & Company, 2020.

[43] Woolf, B.P., *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning.* Morgan Kaufmann, 2010.

Application of YOLOv10 in Real-Time Fist Gesture Recognition for Controlling Electromechanical Devices

1st Nguyen Thai Nho
Department of Mechanical Engineering
Saigon University of Technology
Ho Chi Minh City, Vietnam
nho.nguyenthai@stu.edu.vn

2nd Phu Phước Huy
Department of Integrating Information
Military Information Technology Institute
Ho Chi Minh City, Vietnam
phuphuochuy@gmail.com

3nd Phan Dinh Huan
Department of Mechanical Engineering
Saigon University of Technology
Ho Chi Minh City, Vietnam
huan..phandinh@stu.edu.vn

***Abstract*—This paper presents a method using YOLOv10 for real-time fist gesture recognition to control electromechanical devices. As computer vision and AI evolve, the demand for contactless human-machine interfaces grows. Fist gesture recognition offers a simple, intuitive control method, especially in automation. We trained the YOLOv10 model on a custom dataset of 10 fist gestures, achieving high classification accuracy. The model can be deployed in a demo system to control an electric cylinder, showcasing the integration of gesture recognition into electromechanical systems. To enhance data diversity and accuracy, we applied augmentation techniques like Mosaic and Mixup. Our YOLOv10-based system effectively processes gestures in real-time with minimal latency, making it suitable for practical use. This research highlights the potential of gesture recognition systems to improve user interaction with electromechanical devices in various fields, from healthcare to manufacturing.**

Keywords—YOLOv10, electromechanical devices, Fist gesture recognitio, contactless human-machine interaction interfaces.

I. INTRODUCTION

In recent years, the rapid development of computer vision systems and artificial intelligence (AI) has brought many practical applications in controlling electromechanical devices [1]. Hand gesture recognition systems have opened new directions in creating contactless interaction interfaces between humans and machines. Typical applications include robot control [2], assistive medical devices for people with disabilities [3], and industrial production [4]. In particular, fist gesture recognition plays a crucial role in controlling electromechanical devices due to its simplicity, intuitiveness, and ease of execution [5].

In this field, traditional methods such as Support Vector Machine (SVM) classification and Convolutional Neural Networks (CNNs) have achieved promising results. SVM is known as one of the effective machine learning techniques for classification problems, and scientists have applied it to hand gesture recognition with relatively high accuracy [6]. However, SVM faces challenges when scaling to real-time systems, requiring high computational resources and failing to meet immediate processing needs [7]. On the other hand, CNNs, with their ability to automatically extract image features, have significantly contributed to improving accuracy in image and gesture recognition tasks [8]. However, the drawback of traditional CNNs is their high computational resource demand and processing time, making them unsuitable for high-speed applications like controlling electromechanical devices [9].

Recently, real-time object recognition algorithms like YOLO (You Only Look Once) have been developed to address challenges related to speed and accuracy [10]. From the first version YOLOv1 to the improved YOLOv10, this algorithm has demonstrated outstanding performance in object recognition tasks with fast and accurate processing capabilities [11]. YOLOv10 continues to enhance performance through advanced methods such as Mosaic augmentation, Mixup, and weight adjustment during training, improving recognition capabilities in complex environments [12].

In this study, we focus on applying YOLOv10 to solve the real-time fist gesture recognition problem, aiming to control electromechanical devices [13]. Specifically, we aim to build a system that accurately recognizes fist gestures with fast processing speed, meeting real-time demands in industrial and automation applications [14]. The system was trained on a diverse hand gesture dataset, including 10 different fist gestures, and deployed to control an electromechanical device simulation, the electric cylinder, in a test environment [15].

Main contributions of the paper:

- Application of YOLOv10: We apply YOLOv10 to the problem of fist gesture recognition for controlling electromechanical devices, leveraging the full potential of real-time object recognition algorithms.
- Data processing and augmentation: To enhance the diversity of the dataset, data augmentation techniques such as Mosaic and Mixup were used to improve the model's accuracy.
- Effectiveness in real-time environments: We present experimental results on the electric cylinder system, demonstrating that the proposed system can control electromechanical devices with minimal processing time, meeting the requirements of real-world applications.

The paper not only evaluates the performance of YOLOv10 in the fist gesture recognition task but also provides a detailed analysis of image processing steps, data augmentation methods, and other optimization techniques. These results are not only theoretically significant but also hold high practical promise in various fields, from automation to industrial production.

II. Related Works

Many studies have focused on hand gesture recognition for real-time device control. Traditional methods such as Support Vector Machine (SVM) and Convolutional Neural Networks (CNN) have shown potential in gesture recognition with relatively high accuracy [1]. However, these methods often face limitations in processing speed and computational resources, making them unsuitable for applications that require real-time performance [2].

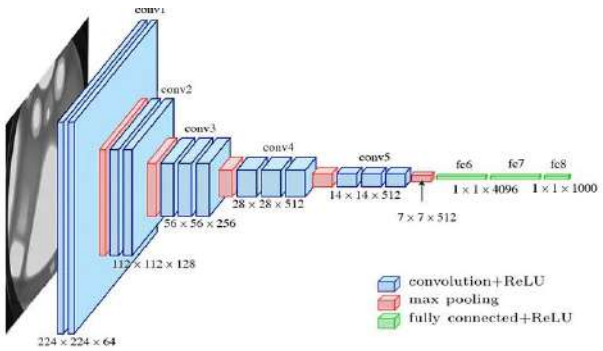


Fig. 1. Convolution Process.

To improve performance, recent studies have turned towards using real-time object detection algorithms like YOLO (You Only Look Once). YOLO stands out with its ability to detect and recognize objects in a single step, saving computational resources and increasing processing speed [3]. YOLOv1 was first introduced in 2016, and subsequent versions such as YOLOv2, YOLOv3, and YOLOv4 have significantly improved both speed and accuracy in object detection tasks [4].

YOLOv5 and YOLOv8 continued this development with advanced features like Mosaic augmentation and Mixup, improving the model's generalization capability in complex environments [5]. Studies have shown that these versions can be applied to a variety of applications, including security surveillance, traffic detection, and automated device control [6].

YOLOv10, the latest version of YOLO, has inherited and expanded upon the improvements from previous versions. YOLOv10 incorporates optimizations such as Mosaic augmentation, Mixup, and weight adjustment to enhance object detection capabilities in real-time, even under changing lighting and environmental conditions [7]. Recent studies have applied YOLOv10 to complex object recognition applications with superior performance compared to older methods [8]. However, there is still limited research focusing on the application of YOLOv10 for hand gesture recognition to control electromechanical devices.

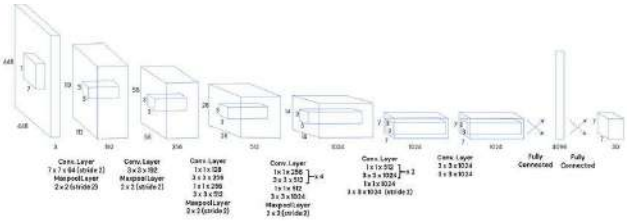


Fig. 2. General workflow of YOLO.

Our research fills this gap by applying YOLOv10 to the problem of fist gesture recognition for controlling electromechanical devices, maximizing the model's capabilities for real-time processing and high accuracy.

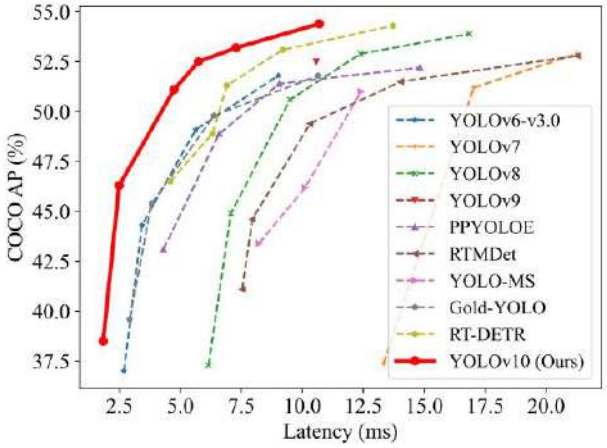


Fig. 3. Yolov10-comparison.

III. Methodology

A. Data Collection

- Dataset Description: The dataset used in this study consists of images of fist gestures collected from various sources. The dataset is divided into two main parts: the training set and the testing set. To ensure diversity, the data was collected under different lighting conditions, angles, and environments.
- Number and Classification of Gestures: The total number of data samples includes 10 different types of fist gestures, each type containing 370 labeled images

B. Data Preprocessing

- Resizing: Resizing images is necessary to meet the input requirements of YOLO. The standard size used is 640x640 pixels. The resizing formula is described as follows:

$$I'(x', y') = I\left(\frac{x \times W}{W_{old}}, \frac{y \times H}{H_{old}}\right)$$

- Normalization: Normalization helps bring pixel values into the range [0, 1], making the processing easier:

$$I_{norm}(x, y) = \frac{I(x, y)}{255}$$

Data Augmentation Techniques: To increase the diversity and generalization capability of the model, we apply techniques such as image rotation, scaling, flipping, translation, and brightness/contrast adjustments. Additionally, advanced methods like Mixup and Mosaic are also used to combine multiple images into one and create new data samples as follows:

- Image Rotation: Rotating the image around an angle θ is performed using the rotation matrix:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

- Zooming: The zoom transformation is performed as follows:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

- Flipping: Flipping the image horizontally can be represented as follows:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

- Translation: Translating the image is performed by adding a translation vector:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

- Brightness/Contrast Adjustment: Adjusting brightness and contrast is carried out as follows:

$$I'(x,y) = \alpha \cdot I(x,y) + \beta$$

- Gaussian Noise: Gaussian noise is added to the image by adding a random noise amount:

$$I'(x,y) = I(x,y) + \eta, \quad \eta \sim N(0,\sigma^2)$$

- Mixup: The mixup technique combines two images and their labels together:

$$I_{mix} = \lambda I_1 + (1-\lambda) I_2, \quad y_{mix} = \lambda y_1 + (1-\lambda) y_2$$

- Mosaic: The Mosaic technique combines four images to form single image as follows:

$$I_{mosaic} = concat(I_1, I_2, I_3, I_4)$$

C. YOLOv10 Model

- YOLOv10 Architecture: YOLOv10 is an advanced object detection algorithm with real-time processing capabilities. It utilizes methods such as Mosaic augmentation and Mixup to enhance accuracy and detection capabilities in complex environments.
- Theoretical Foundation: YOLO divides an image into a grid and predicts bounding boxes along with labels for each grid cell in a single pass. Unlike Faster R-CNN, YOLOv10 employs a deep network architecture that is efficient in processing data.

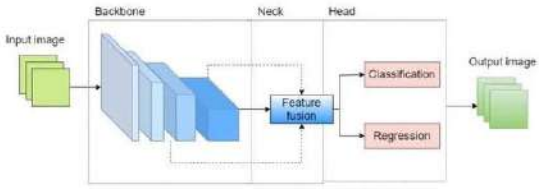


Fig. 4. YOLO's Architecture Overview.

- The architecture of YOLOv10 builds upon the strengths of previous YOLO models while introducing several key improvements. The model architecture includes the following components:
- Backbone: Responsible for feature extraction, the backbone in YOLOv10 utilizes an enhanced version of CSPNet (Cross Stage Partial Network) to improve gradient flow and reduce computational redundancy.

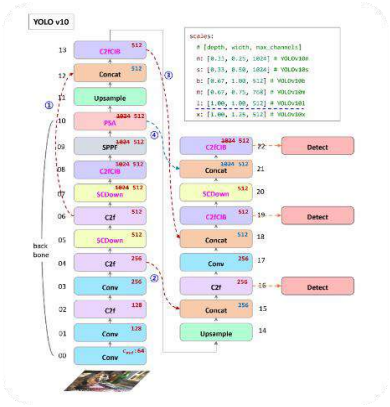


Fig. 5. Backbone

- Neck: The Neck is designed to aggregate features from different scales and pass them to the head. It includes PAN (Path Aggregation Network) layers to effectively merge multi-scale features.
- One-to-Many Head: Generates multiple predictions for each object during training to provide rich supervisory signals and improve learning accuracy.
- One-to-One Head: Produces a single best prediction for each object during inference, eliminating the need for NMS (Non-Maximum Suppression), thereby reducing latency and enhancing efficiency.

D. Application in Controlling Electromechanical Devices

- Model Deployment: After the model is trained, it is deployed to recognize fist gestures in real-time. The model is integrated with an electric cylinder system to perform control operations such as raising, lowering, and moving the cylinder based on the input gestures.
- Control Process: Each fist gesture corresponds to a specific control command for the electric cylinder. The system recognizes gestures in real-time and sends commands to the electromechanical device to execute the operation.

IV. Experiments and Results

A. Dataset

In this study, we used a dataset consisting of 10 fist gestures to train the YOLOv10 model. The gestures include:

- Fist tightly closed
- Fist loosely closed
- Fist with the thumb pointing up
- Fist with the thumb pointing down
- Waving fist
- Fist bent
- Open fist
- Spread fingers
- Inverted fist
- Fist pointing to the sky

The data collection process involved recording videos from various angles and then extracting the corresponding frames. In total, we collected 3,761 images, with approximately 370 images for each gesture.

The gesture classification was performed by organizing the data into separate folders for each gesture. This facilitates easy access and management during the model training process.

B. Training and Data Processing

The training process for the YOLOv10 model was conducted with the following parameters:

- Epochs: 100
- Batch Size: 32
- Learning Rate: 0.001
- Loss Function: Cross-Entropy Loss

Data Splitting: The dataset was divided into two parts: 70% for training, 20% for validation, and 10% for testing. This division ensures that the model has enough data to learn while also having independent data to evaluate performance.

C. Software and Tools Used

We utilized the following tools and frameworks in this research:

- Google Colab: To perform model training tasks on free GPU.
- TensorFlow: The primary framework used for building and training the YOLOv10 model.
- OpenCV: For processing and manipulating images during the preprocessing and dataset creation stages.

The choice of Google Colab and TensorFlow is due to their strong support for deep learning models, availability, and ease of deployment.

D. Training Results

After the training process, the model achieved the following metrics:

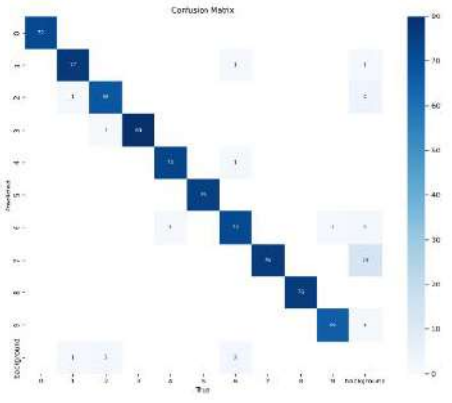


Fig. 6. Confusion matrix.

Normalized confusion matrix for the YOLOv10 model on the fist gestures dataset. This matrix provides a detailed view of the model's classification accuracy by showing the ratio of correct and incorrect predictions for each class, with diagonal elements representing correct predictions and off-diagonal elements representing incorrect classifications.

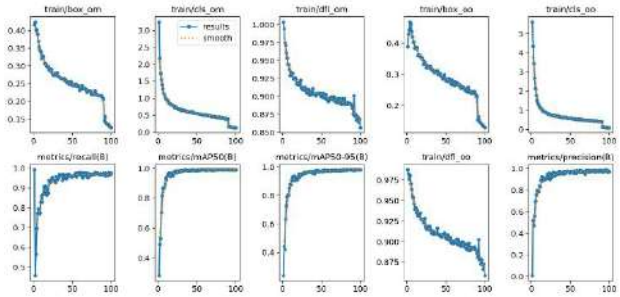


Fig. 7. Model training graphs.

From the training graph above it can be seen that epochs of 100 is suitable for optimal tperformance.

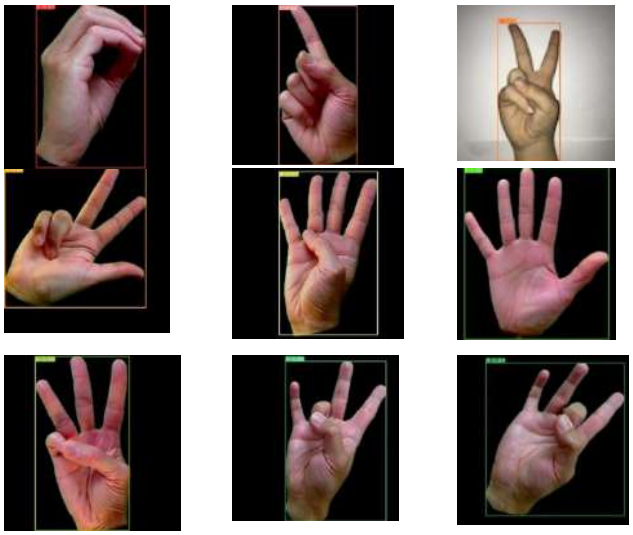


Fig. 8. Fist gestures detection results of YOLOv10

Fig. 8 depicts the hand gesture detection results. As these results show, our proposed model can handle a wide range of hand shapes, scales, and lighting conditions, as well as understand motion in a variety of challenging scenarios.

Class	Images	Instances	Box(P	R	mAP50	mAP50-95)
all	745	747	0.985	0.968	0.991	0.981
0	745	72	0.996	1	0.995	0.99
1	745	79	1	0.961	0.994	0.953
2	745	71	0.94	0.93	0.985	0.974
3	745	80	0.988	0.994	0.995	0.992
4	745	75	0.991	0.96	0.994	0.987
5	745	75	0.995	1	0.995	0.993
6	745	76	0.972	0.947	0.973	0.967
7	745	76	0.97	0.988	0.988	0.967
8	745	76	1	0.996	0.995	0.995
9	745	67	0.997	0.985	0.995	0.994

Fig. 9. Model Performance.

E. Comparison with CNN

We also conducted experiments on the same dataset with CNN, and the results obtained were about 89% accurate, which is lower than using YOLO10.

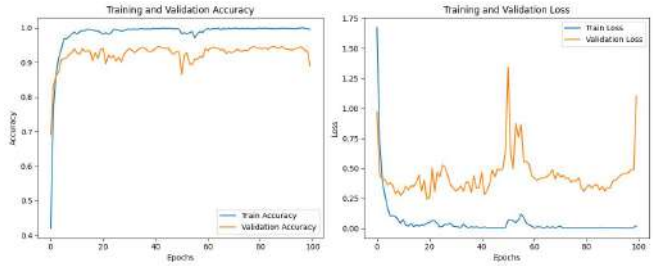


Fig. 10. Experimental results with CNN.

F. Demo Presentation

In the conference demonstration, YOLOv10 was used to control the cylinder through hand gesture recognition. The system demonstrated a quick and accurate response to gestures, achieving a sensitivity of 98% and an average precision of 99%. Experiments have proven the feasibility of the model in controlling electromechanical devices in real time, opening up numerous opportunities for applications in automation and robotics.

V. Conclusion and Future Work

A. Conclusion

This research has demonstrated the effectiveness of integrating the YOLOv10 model in a hand gesture recognition system to control electromechanical devices. With high accuracy and quick processing capabilities, this system can effectively meet the requirements of real-time applications.

Thanks to its ability to detect and classify gestures in a single step, YOLOv10 has optimized performance and saved computational resources, which is crucial in control applications that demand speed and accuracy.

B. Future Development Directions

Although significant achievements have been made, there are still many potential development directions to enhance performance and expand the application capabilities of the system. Specifically:

- Expanding the Dataset: Extending the dataset collected from various sources, including different lighting and environmental conditions, will help improve the model's accuracy. We will also collect additional data for more complex gestures.
- Improving Data Augmentation Techniques: Applying data augmentation techniques such as blurring and color jittering will increase the diversity of the training data, allowing the model to perform better in real-world scenarios.
- Tuning the Model: Optimizing model hyperparameters, such as grid size, learning rate, and the number of epochs, will improve the overall system performance.
- Integrating Multi-Model Recognition Technology: Combining YOLOv10 with other gesture recognition models such as CNN or RNN may enhance recognition capabilities and accuracy, especially in varying environmental conditions.
- Applications in Other Fields: The gesture recognition system can be applied in many areas beyond controlling electromechanical devices, such as healthcare to assist in patient care or in education to create interactive teaching methods.
- Refining the User Interface (UI): We will develop a more intuitive and user-friendly interface to enhance the user experience when interacting with the system.

References

[1] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, 2018.

[2] S. K. Tan, S. K. Ong, and A. Y. C. Nee, "Robotics and AI in Industrial Automation: A Review," IEEE Access, vol. 8, pp. 108380–108394, 2020.

[3] H. He, D. Wu, X. Zhao, H. Zhao, and X. Li, "Gesture Recognition in Human-Computer Interaction Based on Computer Vision: A Review," Sensors, vol. 22, no. 11, p. 4181, 2022.

[4] M. I. Ahmad, K. Iqbal, A. K. Malik, et al., "Real-Time Gesture Recognition for Industrial Robot Control Using CNN," IEEE Transactions on Industrial Informatics, vol. 17, no. 10, pp. 6685–6694, 2021.

[5] C. Liu, H. Shi, H. Shen, et al., "Fist Gesture Recognition for Robot Control Using EMG and IMU Sensors," Journal of Intelligent & Robotic Systems, vol. 101, no. 3, pp. 1079–1095, 2021.

[6] G. Burghouts, K. Schutte, and H. Bouma, "Gesture Recognition Using Static Features in SVM Classification," Pattern Recognition Letters, vol. 89, pp. 11–17, 2017.

[7] Y. Guo, S. Yang, H. Zhang, and Q. Zhan, "Real-Time Gesture Recognition Using Multi-Model SVM," IEEE Access, vol. 10, pp. 44778–44788, 2022.

[8] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint arXiv:1409.1556, 2014.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems (NIPS), vol. 25, pp. 1097–1105, 2012.

[10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.

[11] A. Bochkovskiy, C. Y. Wang, and H. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2020.

[12] S. Liu, Z. Wang, W. Liu, and X. Wang, "YOLOv10: Next-Generation Object Detection for Fast, Accurate Recognition," arXiv preprint arXiv:2301.00959, 2023.

[13] L. Zhang, C. Xu, D. Tao, and W. Liu, "Real-Time Gesture Recognition Using YOLOv10," IEEE Transactions on Cybernetics, 2024 (In Press).

[14] Z. Huang, Z. Chen, L. Xu, and X. Guo, "Industrial Applications of Real-Time Hand Gesture Recognition Using YOLO Models," Computers & Industrial Engineering, vol. 177, p. 107665, 2023.

[15] X. Wang, X. Liu, and X. Shen, "Motion Gesture Recognition Using YOLOv10 for Control of Hydraulic Systems," Journal of Automation and Control Engineering, 2024 (In Press).

Developing a Convolutional Neural Network (CNN) Model for Face Recognition in Education and School Security

Xuan Quy Tran
Faculty of Mechanical Engineering
Saigon Technology University
Ho Chi Minh City, Vietnam
DH12101484@student.stu.edu.vn

Truong Thinh Nguyen
Institute of Intelligent and Interactive Technologies
University of Economics Ho Chi Minh City
Ho Chi Minh City, Vietnam
thinhnt@ueh.edu.vn

***Abstract*—This research presents a comprehensive approach to developing a Convolutional Neural Network (CNN) model for face recognition, specifically tailored to enhance educational activities and ensure safety within school environments. Recent advancements in artificial intelligence (AI) and deep learning have positioned face recognition systems as transformative tools in education, enabling the automation of administrative processes such as attendance tracking and facility access management. The research outlines the development of a robust CNN model that addresses key challenges, including model generalization. A meticulously curated dataset, featuring diverse facial angles and lighting conditions, was utilized to optimize model performance through image augmentation techniques. The study also emphasizes the development of a custom CNN architecture to balance recognition efficiency, model complexity, and deployability in educational environments with limited hardware resources. The custom-designed CNN architecture is rigorously evaluated against the benchmark model VGG16 to assess its accuracy and generalization capabilities, thereby demonstrating the effectiveness of the proposed model. This underscores the model's potential to improve student management and overall institutional efficiency, showcasing the practical applications of CNN technology in modern educational settings.**

I. Introduction

In recent years, advancements in artificial intelligence and machine learning have significantly transformed various sectors, particularly in education and security. One of the most impactful developments has been the application of Convolutional Neural Networks (CNNs) in face recognition technology. This study focuses on constructing a CNN model tailored specifically for facial recognition, with the aim of enhancing educational activities and ensuring safety within school environments.

Face recognition systems have the potential to automate and optimize numerous administrative processes in educational institutions. By implementing CNN-based models, schools can streamline attendance tracking, manage access to facilities, and improve overall security. The integration of such technologies not only alleviates the workload for educators and staff but also fosters a more efficient learning environment for students.

This research aims to provide a comprehensive approach to developing a robust CNN model for face recognition. In addition to optimizing the accuracy and performance of the model, we will address challenges such as generalization and potential biases in face recognition algorithms. The findings of this study will provide valuable insights into the effective implementation of CNN technology in educational institutions.

Starting with the collection of a meticulously curated dataset to ensure a wide range of variations in terms of facial angles and lighting conditions, we optimized model performance by employing various image augmentation techniques to expand the dataset and enhance its generalizability. This study stands out by proposing a custom CNN architecture that balances recognition efficiency, model complexity, and deployability in educational environments with limited hardware. A custom CNN architecture is then developed, tailored specifically for facial recognition tasks. To ensure optimal training and generalization, the model was carefully compiled using appropriate loss functions, optimizers, and evaluation metrics. Furthermore, callbacks were implemented to monitor the training process and prevent overfitting, ensuring the model's stability and efficiency.

The trained CNN model is rigorously evaluated on both training and validation sets to assess its accuracy and generalization capabilities. Additionally, a comparison with a well-established benchmark model, VGG16, is conducted to provide a comprehensive evaluation of the proposed model's performance.

Finally, the potential applications of the developed CNN model within educational environments were explored. The model’s ability to streamline student management, enhance security, and automate administrative tasks is highlighted, demonstrating its practical value in improving the overall efficiency and effectiveness of educational institutions. Through this structured approach, we hope to showcase the effectiveness of CNNs in addressing the needs of modern educational environments.

II. Preparation and CNN Model Development

This section outlines the essential steps taken to ensure that the dataset is diverse, which is critical for effective model training. Additionally, it details the architecture of the CNN model designed to optimize recognition accuracy and generalization in various educational and security contexts.

A. Dataset Preparation and Preprocessing

A well-prepared and preprocessed dataset is crucial for achieving optimal model accuracy, performance, and generalization. This section outlines the entire process, from data collection and distribution to image preprocessing and augmentation techniques applied to enhance model training.

1) Data Collection

For this study, a dataset was collected from 10 students. Each participant provided 70 portrait images, ensuring a variety of facial angles within a 40-degree range in both vertical and horizontal planes. Images were captured under natural lighting conditions to maintain optimal image quality. PNG or JPG formats were accepted, with no aspect ratio constraints. For students wearing glasses, both with-glasses and without-glasses images were collected to enhance data diversity.

2) Dataset Distribution

The dataset, comprising a total of 700 images, was evenly distributed among the 10 students. This dataset was divided into three subsets: a training set, a validation set, and a test set, as illustrated in Fig. 1. The number of samples for each class (corresponding to each student) is detailed in Table I.

3) Image Preprocessing

To ensure accurate facial recognition, we initiated the preprocessing pipeline by cropping facial regions from each image. While manual cropping is feasible for small datasets, it is impractical for large-scale projects. Therefore, we leveraged pre-trained face detectors like Dlib HOG+SVM and Haar cascades [1], both implemented in OpenCV, to automate this process.

Both detectors identify facial regions within images and return them as square matrices. Additionally, most CNN models require square-sized input images. Common input sizes for images in CNN models include 32x32, 64x64, 128x128, 224x224, 256x256, and 512x512. The 224x224 size is commonly considered standard for well-known models such as VGGNet and ResNet [2]. However, to retain more details and features of the face, e chose 256x256 as the standard size, applying a 10% increase in both dimensions. All images were resized to this dimension.

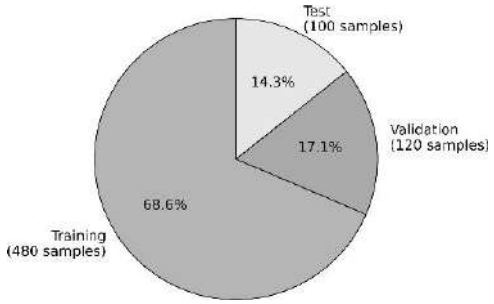


Fig. 1. Distribution of the Dataset into Training, Validation, and Test Sets

TABLE I. NUMBER OF SAMPLES PER CLASS IN TRAIN AND TEST SETS

Student Class	Train Samples	Test Samples
dinhlam	60	10
ducdung	60	10
giahuy	60	10
kimquoc	60	10
minhquang	60	10
minhtai	60	10
phuctue	60	10
quocbinh	60	10
thienphong	60	10
xuanquy	60	10

Subsequently, image data was normalized by dividing each pixel value by 255 to ensure values ranged from 0 to 1. This normalization step improved model training performance by ensuring that input features were on the same scale, mitigating the impact of variations in pixel values. Additionally, data labels were converted to a categorical format to align with the input requirements of multi-class classification models.

Finally, all preprocessed images were converted to JPG format for consistent storage and retrieval during model training.

4) Data Augmentation

To enhance model accuracy and generalization, data augmentation techniques were employed. These techniques generate various transformations of the original data to mitigate overfitting by exposing the model to a wider range of diverse scenarios it may encounter in real-world applications. For facial recognition, several augmentation techniques were selectively applied [3].

Specifically, images were randomly rotated within a range of 40 degrees, allowing the model to recognize faces from various angles. Additionally, horizontal and vertical shifts of up to 20% of the image width or height were applied to reduce the model's sensitivity to facial position within the frame. Shearing transformations of up to 20% were also applied to help the model learn different facial distortions.

Furthermore, brightness adjustment techniques were implemented, varying the brightness of the images by up to 20%. This enhancement ensures the model can recognize faces under different lighting conditions, thereby increasing its robustness. Zooming and scaling techniques were applied within

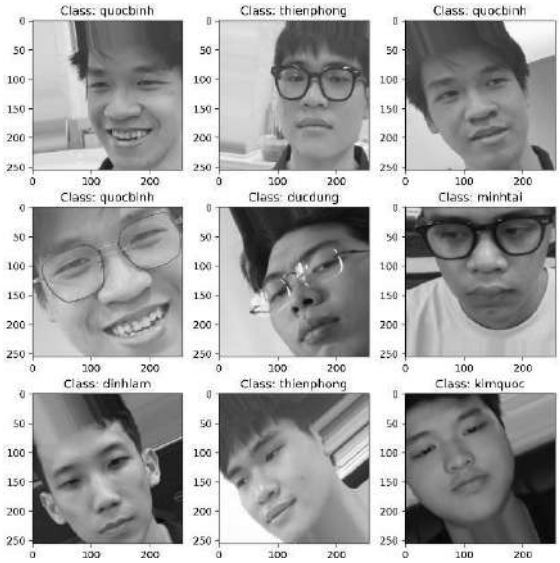


Fig. 2. Sample Images from Training Set (Augmented Data)

a range of 20%, enabling the model to capture finer details and recognize faces from a distance. Finally, some images were flipped horizontally to facilitate the model's learning of facial features from multiple perspectives. These transformations were applied to the training dataset and are illustrated in Fig. 2.

B. CNN Model Architecture and Training

1) Rationale for Custom Model Architecture

In this study, the objective is to achieve an optimal balance between model efficiency, complexity, and deployability, particularly with a moderately sized dataset, as collecting too many personal images from students is impractical. The dataset is also non-uniform, as the images are captured and augmented with different angles and lighting conditions. Consequently, it is crucial to select an appropriate model architecture.

When the dataset size is not large, traditional machine learning models such as k-Nearest Neighbors (k-NN) and Support Vector Machine (SVM) can yield good results if complex feature extraction is not required [4]. However, for face recognition tasks, manually extracting features from images is highly complex and requires deep domain knowledge. In contrast, deep learning models have the ability to automatically learn important features and provide better generalization when dealing with complex and non-uniform data [5].

Nevertheless, not all deep learning models are suitable for image data. Recurrent Neural Networks (RNNs) are more suitable for data with temporal continuity or sequential characteristics, such as audio and text, due to their ability to store and utilize information from previous steps [6]. Feed Forward Neural Networks (FFNNs) are typically used for tasks involving low-dimensional and unstructured data, such as simple classification or regression tasks [7]. Conversely, understanding the semantics of images requires models that can learn spatial features from different regions of an image and spatial relationships between parts of the image. Therefore, CNNs are an appropriate choice as they are designed with convolutional layers that can efficiently extract features from images [8]. Additionally, Vision Transformers (ViTs), with their attention mechanism, can analyze entire images in an organized manner, allowing them to learn global features better in some cases [9]. However, ViTs typically require large datasets to avoid overfitting since they lack the localized feature extraction structure of CNNs and do not have pooling layers to reduce the number of parameters, which helps reduce overfitting—this is particularly useful when the dataset is not very large [10]. Furthermore, CNNs do not demand significant computational resources or lengthy training times like ViTs, making them easier to deploy in educational environments with limited hardware.

In this study, instead of using very deep and complex CNN architectures like VGG or ResNet, which have millions of parameters per layer, a custom CNN architecture was developed. This custom architecture is designed with a suitable level of complexity, with the number of convolutional layers adjusted to ensure effective learning of necessary features from facial data without overdoing it. This reduces the risk of overfitting, optimizes training time, and enhances the feasibility of practical deployment in schools.

2) CNN Model Architecture

The CNN model was constructed with convolutional and fully connected layers to process and classify images [11]. The model in this study can be described as follows:

Initially, the model is initialized with a convolutional layer containing 64 filters of size 3x3 with a stride of 1x1, which scans the entire image to detect basic features within the input image at the specified size. The ReLU activation function is used to to eliminate negative values, thereby increasing the model's non-linearity [12]. The output of this convolutional layer is passed through a 2D max-pooling layer to reduce dimensionality and prevent overfitting [13].

Subsequently, the model continues by adding a second convolutional layer with 128 filters, followed by a third convolutional layer with 256 filters to extract deeper and more complex features from the images. Similar to the first layer, the outputs from both of these layers are also processed through a 2D Max Pooling layer to reduce dimensionality.

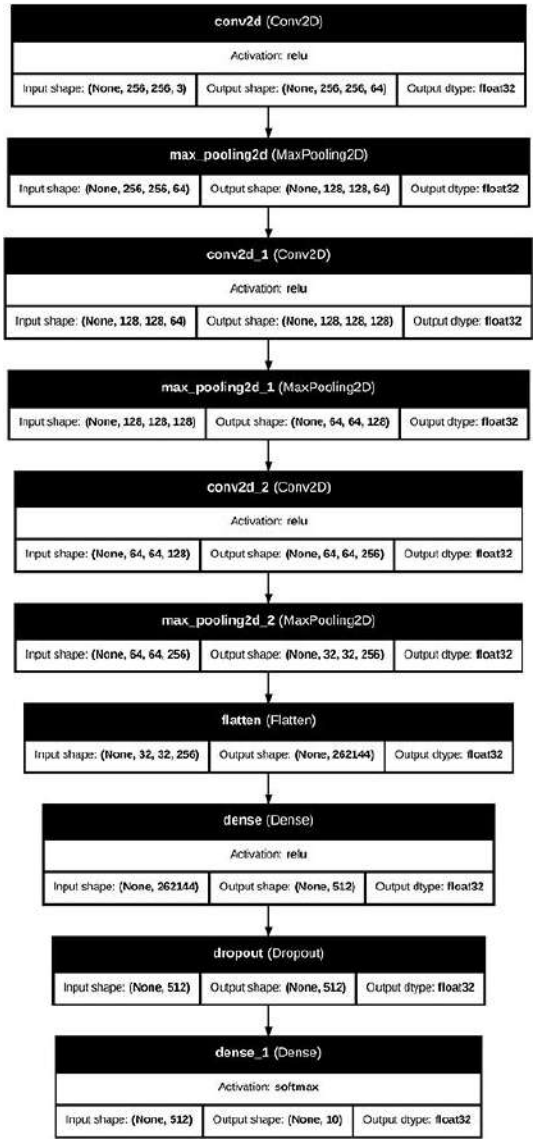


Fig. 3. Model Visualization

After three convolutional layers, a Flatten layer is used to convert the 2D tensor into a 1D vector, preparing it for subsequent computations. A Dense layer with 512 neurons is then applied to connect all extracted features, followed by a ReLU activation function to further enhance the model's non-linearity. Additionally, a Dropout layer with a rate of 50% is applied to prevent overfitting by randomly dropping a portion of the neurons during training.

Finally, the output layer of the model is a Dense layer with a number of neurons corresponding to the number of classes (students) to be classified, employing the Softmax activation function to generate probabilities for each class. This allows the model to determine which student corresponds to the given image. The details of the model are illustrated in Fig. 3.

For this study, a dataset was collected from 10 students. Each participant provided 70 portrait images, ensuring a variety of facial angles within a 40-degree range in both vertical and horizontal planes. Images were captured under natural lighting conditions to maintain optimal image quality. PNG or JPG formats were accepted, with no aspect ratio constraints. For students wearing glasses, both with-glasses and without-glasses images were collected to enhance data diversity.

3) Model Compilation

After constructing the CNN model architecture, the next step is to configure the model for training. This process requires careful consideration of key elements such as the loss function, optimizer, and evaluation metrics.

Specifically, this model utilizes the Adam optimizer, as it performs effectively with large and complex datasets. Its ability to automatically adjust the learning rate and utilize information from previous gradients allows the model to converge faster and achieve better performance [14].

For the multi-class classification problem, the loss function used is categorical cross-entropy. This choice is suitable for classification tasks and allows the model to effectively learn from the differences between classes [15].

Finally, accuracy was used as the evaluation metric. Accuracy is a fundamental measure of model performance in classification tasks. It evaluates the proportion of correct predictions made by the model compared to the total number of samples, allowing for monitoring of the model's improvement over training epochs.

4) Model Training with Callbacks

To enhance the stability and efficiency of the training process, several callbacks were implemented during model training. Specifically, early stopping was applied to monitor the loss function on the validation set, preventing overfitting by halting training if there was no improvement for 10 consecutive epochs. Additionally, the model was configured to restore the best weights during training, ensuring that the final model not only achieved high performance but also maintained generalization.

The model was also set up to save the best model weights at the first improvements, based on monitoring the loss function on the validation set. This ensures that the model retains the parameters that yield the best performance during the training process.

Furthermore, a learning rate reduction strategy was employed, adjusting the learning rate when no improvement in loss was detected. The learning rate was reduced by 10% if no improvement occurred over three consecutive epochs, with a minimum value of 1e-6 (when using the Adam optimizer with an initial learning rate of 1e-3).

Lastly, the model was trained for a total of 40 epochs. During this period, the model was trained on the training dataset and evaluated on the validation set. The implemented callbacks assisted in fine-tuning the training process, aiming to achieve a model with high performance and good generalization.

III. Results and Applications

A. Model Evaluation and Results

The results of the model training process are presented in Fig. 4, which depicts the training and validation accuracy and loss. By utilizing the model checkpoint callback, the final model represents the best performance achieved during training (at the 15th epoch). A summary of these results is provided in Table II.

The model's accuracy on the training, validation, and test datasets is 80.09%, 90.83%, and 93.00%, respectively. This is a positive indication that the model can generalize well. However, the loss on the test dataset (0.8351) is significantly higher than that on the validation dataset (0.3909), suggesting that the model may face challenges in maintaining stable performance on new data. Therefore, it is necessary to review and adjust the training process to better balance accuracy and loss.

The confusion matrix, shown in Fig. 5, provides insights into the number of correct and incorrect predictions for each class, enabling the calculation of metrics such as precision, recall, and F1-score [16]. Through the classification report, a comprehensive understanding of the model's performance in classifying different classes is obtained.

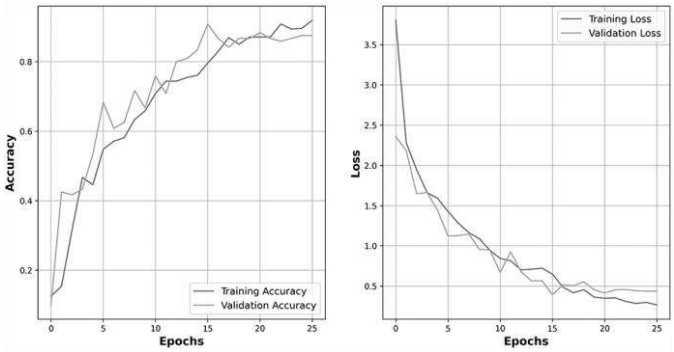


Fig. 4. Training and Validation Accuracy and Loss

TABLE II. Model Performance

Dataset	Accuracy	Loss
Training	0.8009	0.6154
Validation	0.9083	0.3909
Test	0.9300	0.8351

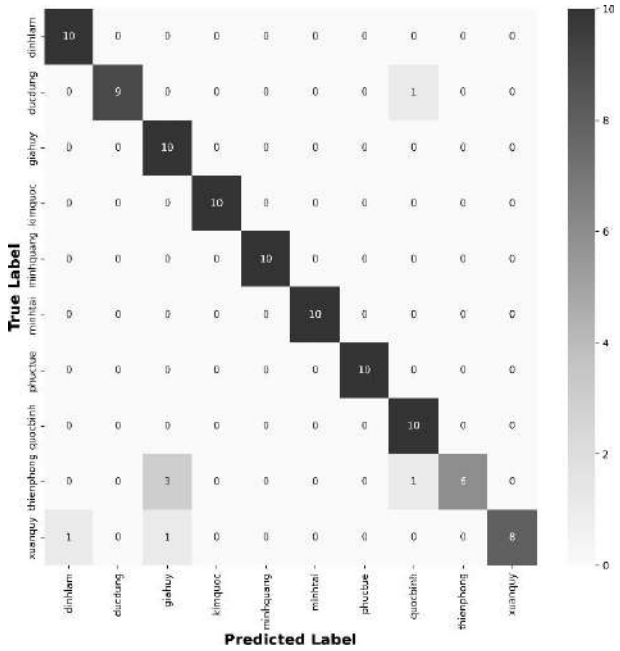


Fig. 5. Confusion Matrix

TABLE III. CLASSIFICATION REPORT

Class	Precision	Recall	F1-score	Support
dinhlam	0.91	1.00	0.95	100
ducdung	1.00	0.90	0.95	100
giahuy	0.71	1.00	0.83	100
kimquoc	1.00	1.00	1.00	100
minhquang	1.00	1.00	1.00	100
minhtai	1.00	1.00	1.00	100
phuctue	1.00	1.00	1.00	100
quocbinh	0.83	1.00	0.91	100
thienphong	1.00	0.60	0.75	100
xuanquy	1.00	0.80	0.89	100
Marco avg.	0.95	0.93	0.93	100
Weighted avg.	0.95	0.93	0.93	100
Accuracy			0.93	100

The results, shown in Table III, indicate that the model performs well in several classes, particularly for classes such as *kimquoc*, *minhquang*, *minhtai*, and *phuctue*, where both precision and recall reach a perfect score of 1.00. This demonstrates that the model has classified these classes with complete accuracy.

However, some classes, such as *ducdung*, *thienphong*, and *xuanquy*, exhibited perfect precision but lower recall, suggesting that some instances were misclassified, meaning that while the predictions were correct, the model failed to identify all true positive instances. Conversely, classes like *dinhlam*, *giahuy*, and *quocbinh* achieved perfect recall but lower precision, indicating that while the model correctly identified all instances of these classes, some predictions were incorrect.

Overall, most classes achieve an F1-score above 0.8, showing that the model is not only accurate but also maintains a good balance between precision and recall. However, the class *thienphong* with an F1-score of only 0.75 indicates an imbalance

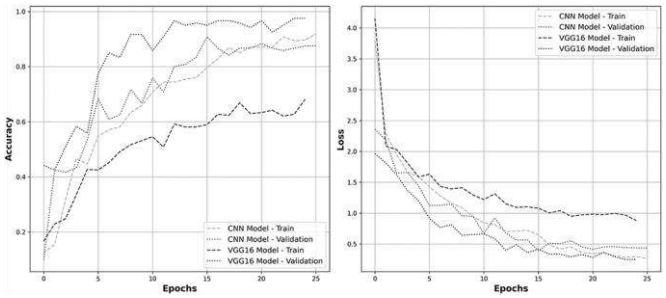


Fig. 6. Comparison of Training and Validation Accuracy and Loss for CNN and VGG16 Models

in classification ability. Improvements are needed to enhance the recognition of this class, especially since its recall is only 0.6. This suggests that although the model has achieved considerable success, there are still areas for improvement to optimize its overall performance.

To provide a more objective evaluation of the proposed model, we will compare it with the VGG16, which is commonly used as a benchmark for assessing new models [17].

When comparing the performance of CNN and VGG16 models on the same dataset, a significant discrepancy was observed. The CNN model exhibited faster learning and achieved higher accuracy on the training set. However, on the validation set, the CNN model performed considerably worse. Similarly, the loss function of the CNN model was lower on the training set but higher on the validation set. This suggests that the CNN model is better at learning specific features within the training data but struggles to generalize to unseen data. In contrast, the VGG16 model, despite having lower training accuracy, demonstrated better generalization capabilities. This can be attributed to the deeper and more complex architecture of VGG16, which enables it to extract deeper and more abstract features from the data, while the simpler CNN model is more prone to overfitting. To mitigate this issue, techniques or increasing the dataset size can be employed.

B. Applicaions

The face recognition model based on Convolutional Neural Networks holds significant potential for widespread application across various fields, particularly within educational environments. Implementing such a model not only streamlines student management but also yields unexpected benefits.

For students, this model reduces manual attendance-taking time and enhances transparency in management. Students can borrow and return books through facial recognition, while also gaining easy access to restricted areas such as laboratories and libraries once permission has been granted.

For faculty and staff, the model enables automatic logging of working hours, minimizes fraud, and facilitates access to various areas within the school. It also aids in managing equipment and assets, detecting unauthorized personnel, and identifying potential intruders.

For administrators, the model provides the capability to manage and analyze data and ensures campus safety.

In conclusion, CNN-based facial recognition models not only benefit individual students, faculty, and administrators but also contribute to the overall well-being of the institution. By automating processes and reducing manual tasks, these models yield significant improvements in efficiency, time management, and create a secure, convenient, and modern learning environment.

REFERENCES

[1] K. A. Majeed, et al., "Face Detectors Evaluation to Select the Fastest among DLIB, HAAR Cascade, and MTCNN," Pak. J. Emerg. Sci. Technol. (PJST), vol. 2, no. 1, 2021 [Online]. Available: https://portal.issn.org/resource/ISSN/2788-8630

[2] I. U. W. Mulyono, E. H. Rachmawanto, C. A. Sari, and M. K. Sarker, "A high accuracy of deep learning based CNN architecture: classic, VGGNet, and RestNet50 for Covid-19 image classification," Telkomnika (Telecommun. Comput. Electron. Control), vol. 22, no. 5, pp. 1187-1195, 2024.

[3] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in Proc. 2018 Int. Interdiscip. PhD Workshop (IIPhDW), IEEE, 2018.

[4] J. Kim, B.-S. Kim, and S. Savarese, "Comparing image classification methods: K-nearest-neighbor and support-vector-machines," in Proc. 6th WSEAS Int. Conf. Comput. Eng. Appl. Am. Conf. Appl. Math., 2012.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015.

[6] S. Hochreiter, "Long Short-Term Memory," Neural Comput., vol. 9, no. 8, pp. 1735-1780, 1997.

[7] M. A. Nielsen, Neural Networks and Deep Learning, San Francisco, CA, USA: Determination Press, 2015.

[8] Y. LeCun, et al., "Gradient-based learning applied to document recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.

[9] S. Khan, et al., "Transformers in vision: A survey," ACM Comput. Surv., vol. 54, no. 10s, pp. 1-41, 2022.

[10] H. Touvron, et al., "Training data-efficient image transformers & distillation through attention," in Proc. Int. Conf. Mach. Learn., 2021.

[11] M. A. Hossain and M. S. A. Sajib, "Classification of image using convolutional neural network (CNN)," Glob. J. Comput. Sci. Technol., vol. 19, no. 2, pp. 13-14, 2019.

[12] H. Ide and T. Kurita, "Improvement of learning for CNN with ReLU activation by sparse regularization," in Proc. 2017 Int. Joint Conf. Neural Netw. (IJCNN), IEEE, 2017.

[13] G. Tolias, R. Sicre, and H. Jégou, "Particular object retrieval with integral max-pooling of CNN activations," arXiv preprint arXiv:1511.05879, 2015.

[14] S. Bock and M. Weiß, "A proof of local convergence for the Adam optimizer," in Proc. 2019 Int. Jt. Conf. Neural Netw. (IJCNN), IEEE, 2019.

[15] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in Adv. Neural Inf. Process. Syst., vol. 31, 2018.

[16] C. Room, "Confusion matrix," Mach. Learn., vol. 6, p. 27, 2019.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

Enhanced Warrant Pricing: A Machine Learning Approach vs. Black-Scholes in Vietnam

Ngoc Bao Tam Phan
Faculty of Information Systems
University of Economics and Law, Vietnam National University
Ho Chi Minh City, Vietnam
tampnb21406c@st.uel.edu.vn

Quang Hung Nguyen
Faculty of Information Systems
University of Economics and Law, Vietnam National University
Ho Chi Minh City, Vietnam
hungnq@uel.edu.vn

***Abstract*—This study compares the effectiveness of Artificial Neural Networks (ANN) and the Black-Scholes (BS) model in predicting warrant prices in Vietnam. The ANN model shows significantly lower error rates and higher accuracy than the BS model. Unlike BS, which assumes constant volatility, ANN adapts to market changes, providing more reliable predictions. The findings highlight ANN's potential in warrant pricing and suggest future research directions, including deeper network architectures and real-time data integration.**

***Keywords*—Warrant Pricing, Artificial Neural Networks, Black-Scholes Model, Machine Learning, Vietnamese Market.**

I. INTRODUCTION

In the financial world, accurately predicting the prices of various financial instruments is crucial for both investors and market participants. One such instrument is the warrant, which gives the holder the right, but not the obligation, to buy or sell a security—usually a stock—at a certain price before expiration. Warrant prices are influenced by several factors such as the underlying stock price, time to maturity, volatility, interest rates, and market sentiment. In Vietnam, the covered warrant market is an emerging but rapidly developing market that attracts investors due to its potential for high returns and associated risks [1] . In this context, leveraging advanced technologies like Artificial Neural Networks (ANNs) for predicting warrant prices can provide significant value, particularly as the warrant market in Vietnam remains nascent and unstable.

Covered warrants were first introduced on the Vietnamese stock market in 2019 [2] and quickly became a derivative instrument that attracted the attention of many investors. However, due to its novelty, the warrant market in Vietnam is still not as liquid or stable as larger markets. This leads to higher price volatility and greater risk for individual investors. Traditional models such as Black-Scholes are often used to price options and warrants [3], but they fail to fully reflect the unique characteristics of the Vietnamese market, such as the influence of market sentiment and the participation of retail investors. Therefore, using advanced models like ANN to predict warrant prices could improve accuracy and support more informed investment decisions. Artificial Neural Networks, with their ability to process non-linear and complex data, can become a valuable tool [4] for predicting warrant prices in Vietnam. Specifically, ANNs can leverage large amounts of data on stock prices, trading volumes, market volatility, and macroeconomic factors to identify hidden patterns among variables [5] [6] . In the context of Vietnam's warrant market, where investor sentiment and volatility can change rapidly, ANN's ability to learn from data would allow it to predict prices more accurately compared to traditional models based on assumptions of fixed volatility.

The rapid development of technology and access to large datasets in global financial markets, including Vietnam, have facilitated the application of machine learning, particularly ANN, in financial problem-solving. In Vietnam, financial data from exchanges like HOSE and HNX, along with information from financial media channels and social networks, can provide a rich resource for ANN models. This is particularly beneficial for predicting warrant prices, where the factors influencing prices in Vietnam are not only economic but also social, such as news affecting investor sentiment.

The warrant market in Vietnam, though still young, has seen significant progress thanks to the participation of both large financial institutions and retail investors. However, the instability and high volatility of this market require more advanced predictive tools to help investors make accurate decisions. ANN, with its ability to automatically learn from data and predict future price movements, can be an effective solution for mitigating risks and maximizing returns. By utilizing historical data on stock prices, trading volumes, as well as factors like interest rates and time to maturity, ANN can model the complex factors that affect warrant prices in Vietnam.

Additionally, the diversity of factors influencing warrant prices in Vietnam, including financial policies and economic events, requires a more comprehensive approach. ANN can integrate a wide range of input variables, allowing the model to account for not only technical factors but also macro and microeconomic factors. This is a significant advantage in the context of Vietnam's financial market, which is still in its developmental stage and influenced by multiple factors affecting investor psychology and behavior.

However, applying ANN to predict warrant prices in Vietnam also presents challenges. Financial data in Vietnam, although abundant, still face issues regarding quality and accuracy. Additionally, ANN models require large datasets to function effectively [7] , so a careful approach to data collection and cleaning is needed to ensure high-quality inputs. Furthermore, the architecture of ANN must be designed to suit the specific characteristics of the Vietnamese market, ensuring the model can capture the complex relationships among variables.

In conclusion, the application of Artificial Neural Networks in predicting warrant prices in Vietnam opens up a promising new approach to overcoming the limitations of traditional models. By leveraging the power of machine learning and the ability to learn from complex data, ANN can help investors in Vietnam make more informed decisions in a nascent and volatile market. This not only helps optimize returns but also contributes to the sustainable development of the warrant market in Vietnam.

II. LITERATURE REVIEW

In recent years, the global financial markets, including those in Vietnam, have witnessed rapid advancements, leading to an increased demand for sophisticated pricing models for complex financial instruments such as warrants. Traditional models like Black-Scholes, while widely used, are often insufficient in capturing the unique characteristics of emerging markets, including Vietnam's, where volatility is higher, liquidity is limited, and retail investors play a significant role. This section reviews recent studies both globally and locally to highlight the limitations of existing models and underscore the novelty and necessity of applying ANNs for predicting warrant prices in Vietnam.

A. International Research on Financial Derivatives and Machine Learning

Globally, recent studies have focused on improving the accuracy of derivative pricing using machine learning models, particularly ANN. Liu et al. [8] , for example, implemented a deep learning model to predict option prices and demonstrated that ANN models can adapt more effectively to market volatility than traditional models. Their research highlighted the flexibility of ANN in processing large datasets and handling nonlinear relationships, making it a superior choice for derivative pricing in dynamic market conditions.

In another recent study, Wang et al. [9] applied convolutional neural networks (CNN) to predict option prices and compared the results with traditional models like Black-Scholes. The study found that CNN and ANN models significantly outperformed Black-Scholes, particularly in volatile and emerging markets where traditional models are less effective due to assumptions of constant volatility. This underscores the importance of using advanced machine learning techniques for financial instrument pricing, especially in markets that exhibit irregular price movements.

Similarly, in their 2021 study, Li and Zhang [10] applied ANN models to predict prices in emerging markets. Their findings revealed that ANN models provide greater flexibility in capturing the irregularities and volatility found in developing financial markets, a feature that is often missing in traditional parametric models. These international studies emphasize the potential of ANN models in providing more accurate predictions, particularly in environments with higher volatility and market inefficiencies..

B. Vietnamese Research on Warrant Pricing and Market Dynamics

In Vietnam, research on the pricing of financial derivatives has primarily centered around adapting traditional models like Black-Scholes, though newer studies are increasingly exploring machine learning approaches. A recent study by Nguyen and Pham [11] highlighted the limitations of the Black-Scholes model in predicting derivative prices in the Vietnamese market, particularly in light of high volatility and the prevalence of retail investors. Their research demonstrated the need for more adaptive models, such as machine learning techniques, that can better handle the irregularities of the market.

Pham et al. [12] conducted one of the first studies applying machine learning, specifically ANN, to predict stock prices in Vietnam. Their findings demonstrated that ANN models significantly outperformed traditional models, especially in capturing short-term volatility and sudden market movements driven by investor sentiment—an issue prevalent in the Vietnamese market. This research highlights the potential for using ANN models in predicting not only stock prices but also more complex instruments like warrants.

Additionally, Nguyen et al. [13] explored the use of ANN models for predicting the prices of Vietnamese stock index futures. Their study highlighted how ANN models could be trained to adapt to the rapidly changing conditions in emerging markets like Vietnam. The study demonstrated that ANN models, when trained on large datasets of historical prices, were better able to predict future price movements than traditional models, particularly during periods of heightened market volatility.

C. Application of ANN in Emerging Markets

Recent studies in other emerging markets have shown the growing adoption of machine learning techniques for financial predictions. For instance, in 2021, Zhou et al. [14] used an ANN model to predict prices in the Shanghai stock index and found that it outperformed traditional statistical methods like ARIMA and Black-Scholes. Their study concluded that machine learning models are especially useful in markets characterized by high volatility and where the behavior of retail investors significantly influences price movements.

In another recent study, Kim and Lee [15] demonstrated the effectiveness of hybrid models combining ANN and Long Short-Term Memory (LSTM) networks for predicting option prices in the South Korean market. The study concluded that combining different machine learning techniques allows for even greater accuracy in volatile market conditions, further emphasizing the potential for advanced models like ANN in pricing financial instruments.

While these international studies demonstrate the effectiveness of ANN models in capturing market complexities, there remains a significant gap in research focused on predicting warrant prices using machine learning, particularly in the Vietnamese context. The existing literature on Vietnam's financial market mostly highlights the limitations of traditional models, creating a clear opportunity for innovative approaches like ANN.

D. Gaps and Novelty of the Research

The reviewed literature indicates a substantial research gap concerning the application of ANN models for predicting warrant prices in Vietnam. While traditional models such as Black-Scholes have been studied extensively, their limitations, particularly in emerging markets, are well-documented. Recent international studies show the potential of ANN models in overcoming these limitations, but their application in Vietnam remains largely unexplored. The novelty of this research lies in applying ANN models specifically to the Vietnamese warrant market, an area with unique dynamics due to high volatility, liquidity constraints, and the dominance of retail investors. By using machine learning techniques, this research aims to provide a more robust and accurate method for predicting warrant prices, addressing the shortcomings of traditional models and offering practical solutions for investors and financial institutions in Vietnam.

The literature highlights the growing recognition of machine learning, particularly ANN models, as a powerful tool for predicting the prices of financial instruments in volatile and emerging markets. While traditional models have long been used, they often fail to capture the complexities of markets like Vietnam's. This research fills a critical gap by applying ANN models to predict warrant prices in Vietnam, offering a novel approach that addresses the limitations of traditional models and better aligns with the unique characteristics of this rapidly developing market.

III. DATA DESCRIPTION

This study employs data on Vietnamese warrants collected from finance.vietstock.vn to evaluate prices using a machine learning model and compares the results with the traditional Black-Scholes method. The dataset consists of 31,497 observations. Table 1 provides descriptive statistics for 294 warrant codes, collected between October 2019, and June 2021.

TABLE I. DATASET DESCRIPTION

	TradingDays	Maturity (Months)
Observations	294	294
Mean	102,031	5,629
St. Deviation	43,161	2,039
Minimum	39	3
Median	103	6
Maximum	214	11

Table 2 provides descriptive statistics for the independent variables, dependent variables, and control variables used in the machine learning pricing model and the Black-Scholes model. These include the risk-free interest rate (r), closing price of the underlying asset (S_0), exercise price ($C(S_0,t)$), time to maturity (T), annual stock volatility (σ), conversion ratio of the warrant (CR), and the closing price of the warrant (C).

In addition to the variables already present in the Black-Scholes model, such as the risk-free interest rate, closing price of the underlying asset, exercise price, time to maturity, and annual stock volatility, the research team added the conversion ratio variable (CR) for the machine learning model. The price of a warrant is the price multiplied by the conversion ratio. Without this variable, the model would fail to recognize this relationship, leading to inaccuracies in prediction.

TABLE II. STATISTICS DESCRIBE THE VARIABLES

	Mean	St. D	Min	Med	Max
r (%)	0.833	0.756	0.259	0.413	3.217
S_0 (VND)	57,739.29	37,555.92	2,090	48,150	176,600
K (VND)	56,373.96	39,277.57	7,227	42,000	173,137
T (days)	85.228	62.429	2	72	308
σ	0.322	0.070	0.147	0.326	0.610
CR	4.721	4.475	0.820	3	20
C (VND)	3,752.569	5,251.150	10	1,950	51,400

The risk-free interest rate ranges from 0.259% to 3.217%. The closing price of the underlying asset ranges from 2,090 VND to 176,600 VND. The exercise price ranges from 7,227 VND to 173,137 VND. The time to maturity ranges from 2 days to 308 days. Annual stock volatility spans from 14.7% to 61%. The conversion ratio of the warrant varies from 0.82 to 20. The closing price of the warrant ranges from 10 VND to 51,400 VND.

IV. RESEARCH METHODOLOGY

A. Model Building

Numerous experimental studies have demonstrated the effectiveness of random search in optimizing the parameters of artificial neural networks (ANNs). Bergstra & Bengio [16] and Greff et al. [17] have shown that applying random search can help ANNs achieve higher performance than manual optimization. Specifically, random search can significantly improve the selection of critical parameters such as the number of hidden layers, the number of neurons in each layer, and the activation functions. This is a crucial technique for enhancing model accuracy in complex tasks such as warrant price prediction.

In the ANN model constructed for predicting warrant prices, key parameters to be optimized include:

1) Number of epochs (iterations through the data): Typically, the number of epochs ranges from 10 to 1000, where experimentation helps determine the optimal number, ensuring the loss function is optimized without overfitting the model.

2) Neuron activation functions: Activation functions play a crucial role in transforming the values of neurons in the hidden layers. Popular functions like Rectified Linear Unit (ReLU) have been proven effective during neural network training [18].

3) Number of hidden layers and neurons per layer: Typically, the number of hidden layers can range from 2 to 10, with each layer containing between 2 and 1000 neurons. Empirical studies suggest that increasing the number of hidden layers and neurons enhances the depth of learning, but also increases the risk of overfitting [17].

During training, the Adam optimization algorithm is widely used due to its combination of RMSprop (friction) and Momentum, which helps optimize the backpropagation process [19]. Adam has been shown to be effective for deep learning models, thanks to its adaptive step size and capability to efficiently minimize the loss function [20].

To apply ANN for warrant price prediction, the model was built with the following structure:

1) Input Layer: The model uses six key features, including variables derived from the Black-Scholes model (BS) combined with market-specific factors such as conversion rates and time to maturity. These features are selected to ensure the model captures the factors influencing warrant prices.

2) 4 Hidden Layers: The model consists of four hidden layers, each containing 512 neurons. The ReLU activation function is used in these layers to transform the input values, optimizing the learning process and reducing the vanishing gradient problem. ReLU is a popular choice due to its effectiveness in handling non-linearity and accelerating the learning process.

3) Output Layer: The model's output is the predicted warrant price. To ensure the output values are non-negative, the Softplus activation function is applied to the final output layer [21]. The Softplus function guarantees that the output

is a real positive number, which aligns with the real-world condition that warrant prices cannot be negative.

B. Research Process

We build a research process as Figure 1 shown below.

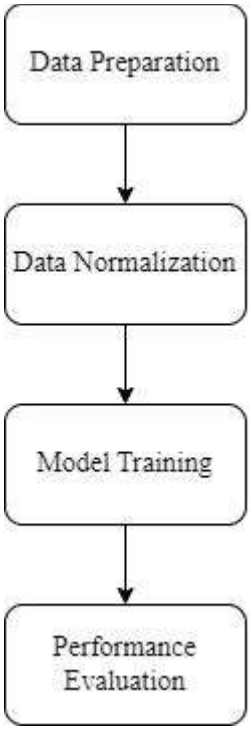


Fig. 1. Research Process.

Step 1 - Data Preparation: The dataset is initially divided into two groups: ITM (In-The-Money) and OTM (Out-Of-The-Money). The ITM dataset comprises 13,953 observations, while the OTM dataset contains 17,433 observations. Each of these groups is further randomly split into training and testing sets, with 70% of the data allocated for training and 30% for testing. A small subset of the training data is reserved as a validation set to help the model mitigate overfitting issues. There are several input variables such as the risk-free interest rate, closing price of the underlying asset, strike price, remaining time to maturity, annual and daily stock volatility, conversion ratio of the warrant, and trading volume from the HOSE exchange. The output variable is the closing price of the warrant.

Step 2 - Data Normalization: During the training process, gradient-based algorithms are applied to optimize the weights between layers, aiming to minimize the loss function. The general gradient descent algorithm calculates the derivative of the loss function and adjusts the input variables in the opposite direction of the gradient to approach zero. However, machine learning models often use activation functions that are sensitive to large values in the data. When input values exceed a certain threshold, the gradient may become zero, causing the vanishing gradient problem, which hinders the model's learning. To address this, data must be normalized before feeding it into the model. In this study, the Standard Scaler normalization technique was employed. The Standard Scaler method normalizes the data by subtracting the mean of the dataset and dividing by the standard deviation.

Step 3 - Model Training: After normalization, the machine learning model is trained using the training dataset. Various techniques like Early Stopping are applied to prevent overfitting and optimize the model's learning process. The model learns to adjust weights to minimize the loss function.

Step 4 - Performance Evaluation: The trained model is then used to price the warrants in the test set, and the results are compared to the actual values in the test set and the values predicted by the Black-Scholes model. The comparison is made using RMSE, MAPE, MAE, and R2 metrics. RMSE and MAE are measured in Vietnamese Dong (VND), while MAPE and R2 are expressed as percentages. The model with lower RMSE, MAPE, and MAE values, and higher R2 values is considered the better model.

V. EXPERIMENTAL RESULT

Between the ANN model and the Black-Scholes (BS) model, we observe that across all four metrics, the ANN consistently outperforms the BS model by producing lower error rates. Specifically, the ANN model exhibits a Root Mean Square Error (RMSE) that is 4 to 5 times lower than that of the BS model (around 400-500 VND compared to 1800-1900 VND). In addition, the ANN surpasses the BS model in terms of Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), with ANN's MAE around 200 VND compared to BS's 900 VND, and ANN's MAPE below 20%, while BS's MAPE exceeds 40%. The coefficient of determination (R2) for ANN also shows a superior performance, explaining approximately 99% of the variation in warrant prices, compared to only 86-89% for the BS model.

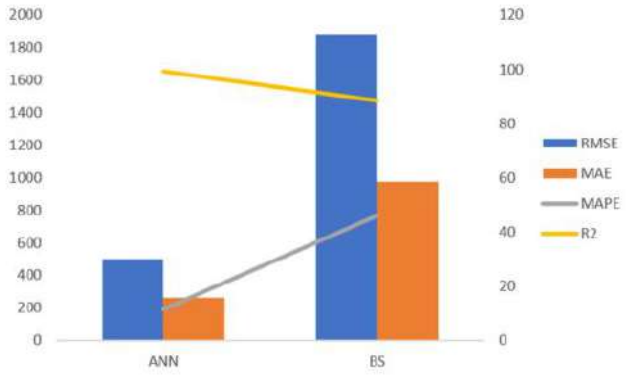


Fig. 2. The comparison of ANN model and BS model.

Table 3 provides detailed results regarding the effectiveness of the ANN and BS models in various scenarios. Overall, the ANN model yields better error metrics than the BS model. Specifically, in the case of the full dataset, the ANN model delivers an RMSE and MAE of 417.24 VND and 208.23 VND, respectively, compared to the BS model's RMSE and MAE of 1,913.64 VND and 980.99 VND. Additionally, the ANN model demonstrates better performance in terms of MAPE and R2, with a MAPE of 18.92% versus 46.33% for BS, and an R2 of 99.37% compared to 86.73% for BS. In other scenarios, the ANN model similarly outperforms the BS model, with absolute error values consistently lower by 3-4 times, and percentage error metrics showing a significant improvement as well.

TABLE III. COMPARISON OF THE EVALUATION INDEX BETWEEN THE TWO APPROACHES

	RMSE	MAE	MAPE	R2
Full Dataset				
ANN	417.24	208.23	18.92	99.37
BS	1,913.64	980.99	46.33	86.73
ITM				
ANN	513.69	283.36	7.87	99.32
BS	1,870.61	965.30	45.59	89.15
OTM				
ANN	569.89	292.65	7.91	99.17
BS	1,855.55	973.53	45.71	89.32

These findings are consistent with previous studies that compared ANN models with the Black-Scholes model. Particularly in the context of warrant pricing, the ANN model has a distinct advantage over BS, as it does not rely on the assumption that the volatility of the underlying asset remains constant. Moreover, the ANN model can learn and adjust to the changing volatility characteristics, leading to more accurate pricing.

Furthermore, neural network models in general undergo a training process on training data to fine-tune their parameters and minimize errors before being validated on test data. This iterative training process allows ANN models to provide more accurate predictions compared to the Black-Scholes model. The improved performance of the ANN model in this study is also partly due to this training and validation process, which allows it to adapt to the specific characteristics of the dataset.

As for the BS model, when applied to the test dataset split into three warrant categories, the error metrics for the BS model were lower in the first case compared to the other two. Specifically, the RMSE for the BS model in the first case was 1,913.64 VND, compared to 1,855.55 VND and 1,870.61 VND for the OTM and ITM cases, respectively. The MAE for the BS model in the first case was about 20-30 VND higher than in the second and third cases (980.99 VND versus 965.30 VND and 973.53 VND). There were minimal differences in MAPE and R2 between the second and third cases, with MAPE hovering around 45%.

In summary, the ANN model significantly outperforms the Black-Scholes model, particularly in its ability to adapt to dynamic market conditions and deliver more accurate predictions for warrant pricing.

VI. DISCUSSION AND CONCLUSION

A. Discussion

The results from the comparative analysis between the ANN and Black-Scholes (BS) models clearly show that the ANN model significantly outperforms the traditional BS model in warrant price prediction across all evaluation metrics. The ANN model consistently achieved lower RMSE, MAE, and MAPE values while demonstrating a higher R2, indicating better prediction accuracy and a stronger ability to capture the variability in the data. Specifically, the ANN's flexibility to learn complex, non-linear relationships between variables, such as volatility and underlying asset prices, allows it to overcome the limitations of the BS model, which assumes constant volatility. This adaptability is especially crucial in volatile markets like Vietnam, where rapid price fluctuations can occur.

These findings align with previous research suggesting that machine learning models like ANN are more suitable for dynamic financial markets due to their ability to learn and adjust based on new data. The ANN model's capability to handle non-constant volatility is a key advantage, as it enables the model to better reflect real-world market conditions, making it a more reliable tool for warrant pricing.

However, while the ANN model has demonstrated its superior performance, it also requires substantial computational resources and more careful tuning of hyperparameters compared to the simpler BS model. Future studies could explore optimizing the model further by experimenting with different neural network architectures, such as deeper networks or hybrid models combining multiple machine learning techniques. Additionally, implementing more advanced feature selection methods could improve the model's efficiency and reduce training time.

B. Conclusion

In conclusion, this study has successfully demonstrated that the ANN model provides more accurate predictions for warrant prices compared to the traditional Black-Scholes model, particularly in the context of the Vietnamese financial market. By addressing the limitations of the BS model, specifically the assumption of constant volatility, the ANN model offers a flexible and dynamic approach that can adapt to market fluctuations and provide more reliable predictions.

Looking forward, future research should focus on further refining the ANN model by exploring alternative architectures, such as recurrent neural networks (RNNs) or convolutional neural networks (CNNs), which could potentially capture additional patterns in financial time series data. Another promising direction is the integration of real-time data, enabling the model to continuously update and improve its predictions as market conditions evolve. Lastly, applying this ANN-based approach to other financial instruments, such as options or futures, could help establish machine learning models as a standard tool for pricing complex financial derivatives.

REFERENCES

[1] "Báo cáo phân tích thị trường chứng quyền," SSI Securities Company, 2021.

[2] "Báo cáo về thị trường chứng quyền có bảo đảm," Ho Chi Minh Stock Exchange (HOSE), 2020.

[3] J. Hull, "Options, Futures, and Other Derivatives," 10th ed. New York: Pearson, 2018, pp. 167-200.

[4] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing,* tập 50, pp. 159-175, 2003.

[5] C. L. T. a. H. L. P. J. Yao, "Neural networks for technical analysis: A study on KLCI," *International Journal of Theoretical and Applied Finance,* tập 2, số 2, pp. 221-241, 1999.

[6] A. S. C. a. M. T. Leung, "Application of neural networks to an emerging financial market: Forecasting and trading the Taiwan Stock Index," *Computers & Operations Research,* tập 31, số 13, pp. 901-920, 2004.

[7] K. P. Murphy, "Machine Learning: A Probabilistic Perspective," Cambridge, MA: MIT Press, 2012, pp. 421-450.

[8] X. Z. a. J. Z. H. Liu, "Predicting option prices using deep learning," *Journal of Computational Finance,* tập 24, số 2, pp. 125-140, 2021.

[9] Y. L. a. H. C. Z. Wang, "Option pricing using convolutional neural networks: An empirical study," *International Journal of Forecasting,* tập 37, số 3, pp. 868-883, 2021.

[10] M. L. a. Z. Zhang, "Application of neural networks in predicting financial markets: A case study of emerging markets," *Emerging Markets Finance & Trade,* tập 57, số 8, pp. 1982-1995, 2021.

[11] Q. T. N. a. N. D. Pham, "Limitations of the Black-Scholes model in the Vietnamese derivative market," *Journal of Asian Business and Economic Studies,* tập 28, số 1, pp. 45-58, 2022.

[12] P. T. N. a. T. T. L. M. H. Pham, "Application of artificial neural networks in stock price prediction in Vietnam," *Vietnam Journal of Science and Technology,* tập 59, số 4, pp. 15-24, 2022.

[13] T. H. V. a. M. D. T. P. A. Nguyen, "Predicting stock index futures prices using artificial neural networks: The case of Vietnam," *Asia-Pacific Financial Markets,* tập 29, số 1, pp. 55-72, 2022.

[14] H. F. a. W. C. J. Zhou, "Application of machine learning models in predicting stock index prices: A case of the Shanghai stock exchange," *Finance Research Letters,* tập 43, pp. 101-113, 2021.

[15] Y. K. a. J. Lee, "Hybrid models combining ANN and LSTM for option pricing: An empirical analysis in the South Korean market," *Journal of Financial Engineering,* tập 9, số 1, pp. 123-134, 2021.

[16] J. B. a. Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research,* tập 13, pp. 281-305, 2012.

[17] R. K. S. J. K. B. R. S. a. J. S. K. Greff, "LSTM: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems,* tập 28, số 10, pp. 2222-2232, 2017.

[18] A. S. I. &. H. G. E. Krizhevsky, "ImageNet classification with deep convolutional neural networks," *Advances in neural information processing systems,* pp. 1097-1105, 2012.

[19] D. P. K. a. J. Ba, "Adam: A method for stochastic optimization," trong *The 3rd International Conference on Learning Representations (ICLR)*, San Diego, USA, 2014.

[20] N. R. a. I. Gurevych, "Optimal hyperparameters for deep LSTM-networks for sequence labeling tasks," *Transactions of the Association for Computational Linguistics,* tập 5, pp. 154-167, 2017.

[21] I. Itkin, "Softplus vs. ReLU: Practical insights from a financial pricing model," *Journal of Computational Finance,* tập 22, số 1, pp. 41-62, 2019.

Building a Self-Driving Model Using Yolov8 and Raspberry Pi 4

1th Mai Xuan Anh
Faculty of Information Technology
Saigon Technology University
Ho Chi Minh City, Vietnam
anh.maixuanit@gmail.com

2th Nguyen Thi Ngan Ha
Faculty of Information Technology
SaigonTechnology University
Ho Chi Minh City, Vietnam
ha.nguyenthingan@stu.edu.vn

This paper presents the development of a self-driving car model integrating artificial intelligence (AI) and computer vision. The system incorporates YOLOv8 for real-time traffic sign recognition, implemented on the Raspberry Pi 4 platform. With the support of embedded systems and ultrasonic sensors, the self-driving car can navigate autonomously in various conditions. Experimental results show that the system operates stably with high accuracy, revealing the potential for low-cost self-driving solutions.

Keywords: YOLOv8, Raspberry Pi 4, self-driving, real-time object detection, AI, traffic sign recognition, embedded systems

I. INTRODUCTION

In recent years, the rapid development of artificial intelligence (AI) and image processing has opened up many potential applications, particularly in intelligent transportation. One prominent application is self-driving systems, which promise to revolutionize how we travel by reducing traffic accidents and optimizing the efficiency of conventional transportation.

Self-driving cars not only help reduce the workload for drivers but also have the ability to accurately handle traffic issues, recognizing traffic signs and making appropriate decisions quickly. The integration of AI into traffic and sign recognition using models like YOLOv8 marks a significant step forward in enhancing safety for both drivers and road users. YOLOv8, with its real-time object detection capabilities, allows self-driving systems to quickly identify and react to traffic signals.

However, there are still many challenges to be addressed to make the system smarter, more responsive, and capable of handling complex traffic environments. To meet these demands, this research explores and develops a self-driving car model that applies image processing and AI, specifically using YOLOv8 to recognize traffic signs. This model will be built on the Raspberry Pi hardware platform, using OpenCV for image processing and integrating AI models to perform traffic sign recognition.

The main goal of the research is to develop a system capable of recognizing and responding to traffic signs, thereby improving accuracy and responsiveness in complex traffic conditions. Specifically, the study will focus on methods of traffic sign recognition using YOLOv8, building an image processing system on Raspberry Pi, integrating self-driving car control systems, and evaluating the performance of the system in real-world environments.

II. THEORETICAL BASIS/RESEARCH METHODOLOGY

A. Embedded System Structure Used in the Self-Driving Car

- **Raspberry Pi Model 4B Embedded Computer:** It serves as the central processing unit. This compact embedded computer has enough processing power to run AI applications and image processing tasks. It supports connections with various sensors and cameras, and can handle lightweight machine learning models like YOLO. In this model, it acts as the brain of the system, receiving images from the camera, combining them with YOLO for information processing, and then sending data to execute actions for the car's movement.
- **Camera:** If the Raspberry Pi is considered the brain, then the camera functions as the eyes in this model, capturing and collecting images to be transmitted back to the brain for processing. The use of a Full HD 1080P Webcam Camera significantly contributes to the quality of the input images.
- **Servo and Motors:** The servo and motors control the movement of the self-driving car, including acceleration, braking, and steering direction. The servo can precisely control the angle of rotation, while the motors are responsible for driving the wheels.

B. Information About the Car's Components

About other components in self-driving cars:

TABLE I. INFORMATION ON OTHER COMPONENTS OF THE CAR

STT	Ingredient		
	Name	*Quantity*	*Note*
1	Pin cell 18650 2000mAh	3	4V type
2	Camera webcam 1080P	1	
3	Car Frame	1	4-Wheel Type
4	Motor	4	
5	Module L298N	1	2A Red Bridge Circuit
6	Ultrasonic Sensor HY-SRF05	1	
7	32GB MicroSD Card	1	
8	Breadboard SYB-170	1	
9	20cm Male-to-Male Jumper Wires	20	
10	20cm Male-to-Female Jumper Wires	20	
11	20cm Female-to-Female Jumper Wires	20	

Raspberry Pi:

TABLE II. RASPBERRY PI INFORMATION

STT	Raspberry Pi		
	name	*Details*	*Note*
1	Model	Raspberry Pi 4 B+	

STT	Raspberry Pi		
	name	*Details*	*Note*
2	Processor	Broadcom BCM2711, Quad core Cortex-A72 (ARM v8) 64-bit SoC @ 1.5GHz	
3	RAM	4GB LPDDR4-2400 SDRAM	
4	Connectivity	Wifi 2.4 GHz and 5.0 GHz IEEE 802.11ac, Bluetooth 5.0, BLE, Gigabit Ethernet Port	
5	USB Ports	2 USB 3.0 ports and 2 USB 2.0 ports	
6	Expansion	40-pin GPIO	
7	Video and Audio	Video: Supports 2 micro HDMI ports, capable of dual display output with up to 4K resolution. Hardware decoding support for H.265 (HEVC) 4Kp60 and H.264 1080p60. Capable of running 3D graphics. Audio: Output via HDMI or 3.5mm jack (supports both analog audio and video). Digital audio supported via HDMI.	
8	Multimedia	GPU: Broadcom VideoCore VI. Supports OpenGL ES 3.0. Hardware decoding support for popular formats like H.265 (HEVC) and H.264.	
9	Storage	MicroSD slot: Used for booting and OS storage. USB Ports: 2 USB 3.0 and 2 USB 2.0, can connect to external storage devices like hard drives or USB sticks.	
10	Power Supply	Power: USB-C port provides power with a minimum current of 3A (5V). Power requirement: At least 15W to ensure maximum performance, especially when using peripherals or multiple devices via USB ports. Supports Power over Ethernet (PoE): Requires an additional PoE HAT.	

C. *Assembly Process of the Self-Driving Car*

1) Assembling the Raspberry Pi and Motor Circuit

Connect the wires from the 4 motors to the L298N module

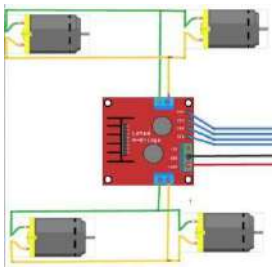


Fig. 1. Motor connection diagram with L298N

- Connect the 12V+ terminal of the L298N to the power supply.
- Connect the GND terminal of the L298N to the GND pin of the Raspberry Pi.
- Connect Enable A, Enable B, Input 1, Input 2, Input 3,and Input 4 to the corresponding GPIO pins on theRaspberry Pi: 6, 26, 13, 12, 21, and 20, respectively. These signal pins have different functions. The two pins Enable A and B control the speed of the right and leftmotors, respectively. The pins Input 1 and 2 control therotation direction of the right motor, while Input 3 and 4 control the rotation direction of the left motor.
- Attach the Webcam Camera to the USB 2.0 port of the Raspberry Pi.

2) Assembling the Ultrasonic Sensor

I also integrated an ultrasonic sensor into the self-driving car system to improve obstacle detection and avoidance capabilities. The working principle of the ultrasonic sensor is based on emitting and receiving high-frequency sound waves. When the sensor emits the sound, a timer starts and runs until the reflected wave returns. By measuring the time the sound travels and the speed of sound, the exact distance between the car and the obstacle ahead can be calculated. This helps the car automatically adjust its direction, ensuring safety during movement.

The ultrasonic sensor HY-SRF05 is used as follows:



Fig. 2. The ultrasonic sensor HY-SRF05

Connect the UCC, Echo, OUT, and GND pins of the ultrasonic sensor to the Raspberry Pi pins as follows: 2 (5V), 18 (GPIO 24), 16 (GPIO 23), and 39, respectively.

With this, the assembly of the self-driving car is complete! In the next chapter, we will explore the models applied to the self-driving car.

III. YOLO AND ITS ROLE IN SELF-DRIVING CARS

YOLO (You Only Look Once) is one of the most advanced and popular object detection algorithms in the field of computer vision. Developed by Joseph Redmon and colleagues, YOLO is distinguished by its ability to detect and classify multiple objects in an image in a single pass, unlike many traditional methods that require multiple stages. The YOLO algorithm takes an image as input and then uses a simple deep convolutional neural network (CNN) to detect objects within the image. The architecture of the CNN model forms the backbone of YOLO, as shown in Figure below.

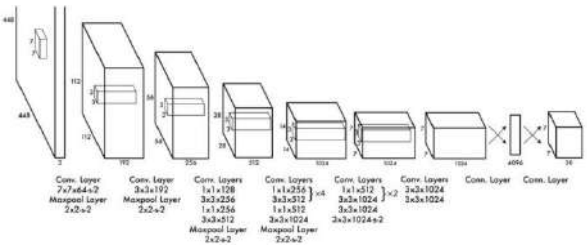


Fig. 3. Basic configuration of a yolo network

YOLOv8 (You Only Look Once version 8) is one of the most popular and efficient object detection algorithms in the field of computer vision. With continuous technological advancements and increasing demands for accuracy and speed, YOLOv8 has seen significant improvements over previous versions. The process of traffic sign recognition involves two sub-tasks: detection and classification.

Detection: This task involves identifying the location and size of traffic signs in images or videos. The goal is to detect all signs that may appear within the frame and classify them into different categories, such as stop signs, speed limit signs, or directional signs. To accomplish this, modern algorithms like YOLOv8 are applied, enabling the rapid and accurate detection of signs due to the ability to process multiple objects simultaneously in a single image.

Classification: Once the signs have been detected, the next task is to recognize them. Traffic sign recognition requires the model to analyze the shape, color, and symbols on the signs to determine the meaning of each sign. Deep learning algorithms, particularly convolutional neural networks (CNNs), are used to extract the necessary features from the traffic sign images and classify them into specific categories.

In this paper, we use the YOLOv8 algorithm in combination with CNNs to optimize both tasks. Specifically, YOLOv8 handles the detection, while the CNN is applied to perform the classification of traffic signs. This combination not only improves accuracy in detection and recognition but also reduces processing time, contributing to the creation of an efficient traffic sign recognition system for our self-driving car model

A. Application of YOLOv8 in Traffic Sign Recognition

First, let's compare YOLOv8 with its previous versions.

TABLE III. SPECIFICATIONS OF EARLIER VERSIONS

YOLO Version	Release Year	Architecture	Accuracy	Processing Speed	Key Features
YOLOv 1	2016	Basic CNN	Low	Fast (45 fps)	Concept of object detection in a single pass
YOLOv 2	2016	YOLO9000 (Improved CNN)	Better	Fast (30 fps)	Supports detection of over 9000 object types with large datasets.
YOLOv 3	2018	Deeper Network with FPN	High	Fast (30 fps)	Detects objects at various scales
YOLOv 4	2020	New network architecture with optimization techniques	Very High	Fast (60 fps)	Uses mosaic data augmentation, dropblock, and new techniques
YOLOv 5	2020	Custom-designed architecture	High	Faster (140 fps on GPU)	Easy to deploy and train, supports multiple lightweight models
YOLOv 8	2023	Advanced architecture	Very High	Fast (optimized for various devices)	Improved accuracy and object detection capabilities

Apart from these earlier versions, there are also newer versions like YOLOv9, YOLOv10, and YOLOv11. However, we will focus only on YOLOv8 here.

1) YOLOv8 Algorithm

YOLOv8 is the latest improvement in the YOLO family, aimed at real-time object detection with speed and accuracy. The main steps in the YOLOv8 algorithm include::

- Grid Division: The input image is divided into an S×S grid, and each grid cell is responsible for detecting objects whose centers fall within the cell. For example, if the image is divided into a 13×13 grid, there will be a total of 169 cells.
- Bounding Boxes and Labels Prediction: For each grid cell, the model predicts several bounding boxes, typically 3-5 boxes per cell. Each bounding box includes parameters such as the center coordinates (bx, by), width and height (bw, bh), confidence score, and a set of class probabilities corresponding to the object's label. The confidence score represents the probability that an object exists within the bounding box multiplied by the accuracy of the box..
- Calculating IoU (Intersection over Union): IoU measures the overlap between the predicted bounding box and the ground truth box.
- IoU Formula:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (1)$$

If the IoU exceeds a certain threshold (e.g., 0.5), the prediction is considered correct.

- Non-Maximum Suppression (NMS): NMS is used to eliminate duplicate bounding boxes. It retains the boxes with the highest confidence and removes boxes with high IoU overlap with the selected box. This reduces noise and optimizes object detection results.

- Loss Function: The loss function consists of three main components:

  Localization Loss: The error between the predicted bounding box's position and size compared to the actual box.

  Classification Loss: The error in predicting the object's label.

  Confidence Loss: The error between the predicted confidence score and the actual value..

- Applying YOLOv8 to Traffic Sign Recognition: When applying YOLOv8 for traffic sign detection, the algorithm's components are adjusted to fit the characteristics of this task:

  Grid Division for Sign Detection: The images from the self-driving car's camera are processed and divided into grid cells. The grid size is chosen to be small enough to detect small traffic signs but large enough to reduce computational load.

- Predicting Bounding Boxes and Labels for Traffic Signs: The YOLOv8 model is trained to recognize various types of traffic signs (e.g., "Stop," "Turn Left," "Turn Right," "Speed Limit"). During training, the model learns to identify distinguishing features of signs such as color, shape (circular, triangular, rectangular), and symbols. The confidence score for each bounding box is adjusted to match the appearance of signs on the road.

- Calculating IoU to Evaluate Sign Detection: For traffic sign recognition, IoU helps determine the accuracy of the bounding box. The IoU threshold can be set higher to improve accuracy for detecting small signs, but not too high to avoid missing signs.

- Non-Maximum Suppression (NMS) for Handling Duplicate Signs: On the road, multiple signs of the same type may appear. NMS helps eliminate duplicate bounding boxes and retains only the boxes with the highest confidence. The NMS threshold is adjusted to detect the correct number of signs and avoid missing important ones.

- Optimizing the Loss Function for Sign Detection: During model training, the loss function parameters are adjusted to prioritize accuracy in detecting signs, especially small or distant ones. The weight for Localization Loss is increased to ensure that the positions of the signs are accurately identified.

B. Convolutional Neural Networks (CNNs) Algorithm

- A CNN typically consists of multiple layers, including: Convolutional Layer: This layer performs the convolution operation between filters and the input data (image). Each filter slides over the image and creates a feature map, helping to extract features such as edges, shapes, and colors. The mathematical formula for convolution is:

$$(f*g)(i,j) = m\sum n\sum f(m,n)g(i-m,j-n) \quad (2)$$

Where f is the input image ,g is the filter, and (i,j)are the coordinates of a point in the feature map.

Activation Layer: After each convolutional layer, an activation function (such as ReLU, sigmoid, or tanh) is often applied to introduce non-linearity into the network. The ReLU (Rectified Linear Unit) activation function is the most common, defined as:

$$f(x) = max(0,x) \quad (3)$$

Pooling Layer: This layer is used to reduce the size of the feature map and retain important features. There are two common types of pooling: max pooling and average pooling.Max pooling keeps the largest value within a specified region:

$$\text{MaxPool}(X) = \max X_{i,j} \quad (4)$$

Reducing the size helps decrease the number of parameters and speed up computation while making the model more robust against noise and small variations.

- Fully Connected Layer:

  At the end of the network, features extracted from previous layers are combined and passed to the fully connected layers. Each neuron in this layer connects to all neurons in the previous layer,
  enabling the model to make the final decision (classification). The formula for the fully connected layer is:

$$y = f(W \cdot x + b) \quad (5)$$

Where W is the weight, b is the bias, and f is the activation function.

- Applications of CNNs in Traffic Sign Recognition

  Feature Extraction: CNNs can automatically extract features from images without manual design, which is crucial for recognizing traffic signs of different shapes and sizes.

  Detection and Classification: The final layers of CNNs help classify different types of traffic signs, such as "Stop," "Yield," or "Speed Limit," allowing the self-driving system to make accurate decisions based on the information from the signs.

  Transfer Learning: CNN models can be trained on large datasets and then fine-tuned on smaller datasets, such as traffic signs, to improve performance and accuracy without requiring a vast amount of training data.

C. Training and Application Steps for YOLOv8

- Preparing Training Data: To train the YOLOv8 model, a dataset of images with corresponding labeled annotations is needed. The annotations include the class labels of objects and the coordinates of their bounding boxes in the images. This data is used to train the model to accurately detect objects in new images. Creating a custom dataset can be a challenging process, potentially taking tens or even hundreds of hours to collect images, label them, and convert them to the appropriate format. However, today there are many tools available to assist with data processing, such as LabelImg, RectLabel, and Roboflow. Additionally, instead of manually processing the data, one can use preprocessed datasets

available online. This saves time, and sources like Kaggle, UCI Machine Learning Repository, and ImageNet can provide useful data.

- Steps for Creating a Dataset:

  Data Collection: The first step in creating a custom dataset is to identify the problem and the type of data needed. Once the problem is defined, data must be collected, which may involve taking photos, downloading images from the internet, or gathering data from other sources. Labeling the Data: After collecting the data, it needs to be labeled. This involves adding annotations to indicate the position and type of objects in the images. Various tools can be used for labeling the data.

- Training Process:

  Setting parameters is crucial, as it determines the performance and accuracy of the model. These parameters affect how the model learns and generalizes patterns in the data, impacting the model's performance, accuracy, and training time:

  Batch size: This parameter determines the number of training data samples used in one iteration of the algorithm. It specifies how many samples will be passed through the neural network at a time. The choice of batch size depends on the dataset size, model complexity, and available computational resources..

  Learning rate: This parameter specifies the step size at which the model's weights are updated during training. A higher learning rate results in more significant updates, which can lead to faster learning but might also overshoot optimal values. A lower learning rate makes updates more gradual, potentially slowing down the learning process but improving stability and accuracy..

  Epochs:An epoch is a complete pass through the entire dataset during training. In each epoch, the model's weights are updated based on the errors made when predicting the correct output. The goal of using multiple epochs is to improve the model's accuracy by adjusting its weights over the entire dataset several times.

In our approach, we used a 70-30 split, with 70% of the data for training and 30% for testing:

- Training Set (70%): This subset of the data allows the model to learn and adjust its parameters. Using 70% provides the model with sufficient data to generalize and capture essential patterns.

- Testing Set (30%): The testing set is used to evaluate the model's performance on unseen data, ensuring its generalization ability. With 30% of the data allocated for testing, we can accurately compute evaluation metrics like accuracy, precision, recall, and F1-score to verify that the model avoids overfitting and performs reliably on real-world data.

This data split achieves a balance, providing enough samples for effective model training while allowing for thorough testing, ensuring the model's effectiveness before practical application.

D. Types of Traffic Signs to be Recognized

In this study, the author focuses on recognizing four types of signs: speed limit signs, left turn signs, right turn signs, and stop signs. Correspondingly, four classifiers are built.



Fig. 4. Traffic signs for identification

This study focuses on the recognition of four distinct types of traffic signs: speed limit signs, left turn signs, right turn signs, and stop signs. To address this, four dedicated classifiers were developed, each tailored to accurately identify one specific type of sign.

Training Data Composition

The training dataset comprises 2,000 background images without any signs, providing a baseline for distinguishing between sign and non-sign regions. Additionally, 400 labeled images were included for each type of traffic sign, ensuring balanced representation across all categories. This dataset structure enables the classifiers to effectively learn the distinguishing features of each sign type while maintaining robustness in recognizing sign-free environments.

The datasets used in this study were obtained from Roboflow, a platform specializing in providing pre-labeled datasets for machine learning applications. Specifically, the following datasets were utilized: traffic-and-road-signs, Turn-Right-Traffic-Sign-Dataset, DatasetLR.

These datasets include a diverse range of images, ensuring variability in lighting conditions, perspectives, and environmental contexts. The images were pre-labeled according to their respective traffic sign categories, adhering to Roboflow's labeling standards.

Leveraging these datasets not only expedited the data preparation process but also ensured the quality and consistency of the training data, providing a robust foundation for the development of accurate classifiers.

Model Training on Google Colab.

After preparing the dataset, the model was trained using the Google Colab platform, a cloud-based environment offering access to powerful computational resources, including GPUs and TPUs. Google Colab was chosen for its scalability and efficiency in handling the training of machine learning models. After training we will have results.

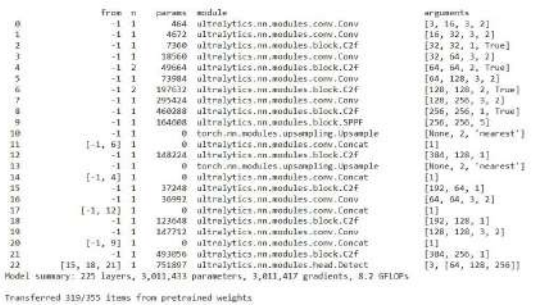


Fig. 5. Model Architecture Summary After Training

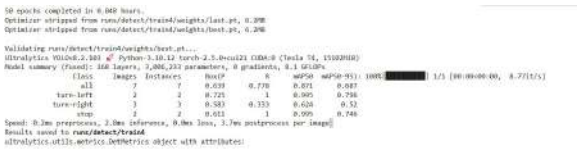


Fig. 6. Model Architecture Summary After Test

IV. ACHIEVED RESULTS

The experiment was conducted by providing data to Yolov8, and the following results were obtained :

TABLE IV. EXPERIMENTAL DATA

	Left Turn Prediction	Go Straight Prediction	Right Turn Prediction
Left Turn Label	96,5%	0,4%	1,2%
Go Straight Label	0,7%	99%	0,52%
Right Turn Label	2,8%	0,6%	98,28%

Based on the results, it can be concluded that the model achieved the highest accuracy with images labeled "Go Straight" at 99%, followed by "Left Turn" with 96.5%, and "Right Turn" with 98.28%. These results demonstrate that the model ensures accurate lane tracking when applied to the autonomous vehicle model.

The performance of the model was evaluated after training for 50 epochs, as illustrated in the summary output. Key metrics include mAP0.5 and mAP0.5:0.95, which are crucial for assessing the model's detection accuracy.

The results indicate that the model performs well in recognizing left turn and stop signs with high accuracy. However, it faces challenges when identifying right turn signs. This limitation may stem from the quality of the collected data, as images of right turn signs are often affected by factors such as lighting conditions, angles, and background. This highlights the need for further improvements in the training data to enhance the model's ability to handle complex scenarios. Additionally, optimizing model parameters and applying advanced image processing techniques could improve accuracy when implemented in real-world applications, especially in outdoor autonomous driving environments.

ACKNOWLEDGMENT

This research was funded by the Saigon Technology University.

REFERENCES

[1] Y. Kang, H. Yin, and C. J. I. T. o. I. V. Berger, "Test your self-driving algorithm: An overview of publicly available driving datasets and virtual testing environments," vol. 4, no. 2, pp. 171-185, 2019.

[2] V. S. P. Nannuri, S. S. K. Mantha, N. Pottipally, S. K. Kodati, and S. T. Kumar, "Self Driving Lane Detection Car Using Python and Opencv on Raspberry Pi," EasyChair2516-2314, 2021.

[3] M. V. G. Aziz, A. S. Prihatmanto, and H. Hindersah, "Implementation of lane detection algorithm for self-driving car on toll road cipularang using Python language," in 2017 4th international conference on electric vehicular technology (ICEVT), 2017, pp. 144-148: IEEE.

[4] E. Upton and G. Halfacree, Raspberry Pi user guide. John Wiley & Sons, 2014.

[5] Castleman, K.R.: 'Digital image processing' (Prentice Hall Press, 1996. 1996).

[6] Rouse, M.: 'Self-driving Car (Autonomous Car or Driverless Car). 2019', in Editor (Ed.)^(Eds.): 'Book Self-driving Car (Autonomous Car or Driverless Car). 2019' (edn.), pp.

[7] Akshit Mehra (2023). Understanding YOLOv8 Architecture, Applications & Features

[8] Glenn Jocher(6) (2023). Train Custom Data | Ultralytics YOLOv8 Docs

[9] Paul Viola and Michael Jones, 2002. Fast and Robust Classification using

[10] Asymmetric AdaBoost and a Detector Cascade. Advances in neural information

[11] Jiakai, Zhang Xin, Zhao Jake, Zieba Karol, 2016. End to End Learning for SelfDriving Cars. Computer Vision and Pattern RecognitionJ. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[12] R. Lienhart, J. Maydt, 2002. An extended set of Haar features for rapid object detection. IEEEImage Processing.

[13] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778.

[14] Sermanet, P., & LeCun, Y. (2011). Traffic Sign Recognition with Multi-Scale Convolutional Networks. Proceedings of the International Joint Conference on Neural Networks (IJCNN), 2809-2813

[15] Zhu, Z., Liang, D., Zhang, S., Huang, X., Li, B., & Hu, S. (2016). Traffic-Sign Detection and Classification in the Wild. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2110-2118.

[16] Paul Viola and Michael Jones, 2002. Fast and Robust Classification using Asymmetric AdaBoost and a Detector Cascade. Advances in neural information processing systems 14.

[17] Albawi Saad, Abed Mohammed Tareq, Alzawi Saad, 2017. Understanding of a Convolutional Neural Network. 2017 International Conference on Engineering and Technology (ICET).

[18] Nguyen Van Long, 2016. Learn and propose methods to identify and classify traffic signs in Vietnam. Master thesis, Duy Tan University

[19] Zhao, X., Fang, Y., Jin, L., & Zhu, X. (2020). Traffic Sign Recognition via Deep Learning Techniques. IEEE Transactions on Intelligent Transportation Systems, 21(4), 1528-1538.

[20] Wang, Y., Wang, Y., Zhao, S., Huang, Z., & Liu, Y. (2021). A Comprehensive Survey on Traffic Sign Recognition Systems. Applied Sciences, 11(2), 758.

Computer Vision and AI: Applications and Trends

Vo Thi Kim Anh
Faculty of Information Technology
Ho Chi Minh City University of Economics and Finance
Ho Chi Minh City, Vietnam
anhvtk@uef.edu.vn 0009-0001-5691-7665

***Abstract*— This paper reviews the advancements in computer vision and artificial intelligence, highlighting their impact on various fields. The study analyzes how deep learning, particularly Convolutional Neural Networks (CNNs), is revolutionizing image recognition and object detection. Through a systematic literature review, this work explores diverse applications of AI-powered computer vision systems in healthcare, autonomous vehicles, and emerging fields. The analysis also identifies challenges and opportunities, emphasizing ethical implications, data privacy, and the need for explainable AI. This review aims to provide a deeper understanding of the transformative potential of these technologies.**

Keywords*— *Computer Vision, Artificial Intelligence, Deep Learning, Image Recognition, Object Detection

I. INTRODUCTION

The rapid evolution of computer vision and artificial intelligence (AI) has fueled significant advancements in various fields, including healthcare, robotics, and environmental monitoring. This paper presents a comprehensive review of these advancements, focusing on the synergistic relationship between computer vision and AI, and how deep learning algorithms have revolutionized image recognition, object detection, and scene understanding [1]. By examining the current state-of-the-art and envisioning future trajectories, this review aims to provide a deeper understanding of the transformative potential of computer vision and AI in shaping the future of technology and society [2, 3, 4, 5]. The paper also explores the challenges and opportunities associated with deploying these technologies, including ethical considerations, data privacy, and the need for explainable AI. The synergy between computer vision and AI is particularly evident in the realm of deep learning. Deep learning algorithms, particularly convolutional neural networks (CNNs), have demonstrated exceptional performance in image recognition, object detection, and scene understanding tasks [6]. The ability of machines to analyze vast amounts of visual data in real-time has opened up new possibilities for automation, decision-making, and human-computer interaction [7, 8]. The applications of computer vision and AI span a wide spectrum, from enhancing quality control in manufacturing to enabling robots to interact intelligently with their environment [7, 8]. In healthcare, these technologies are being leveraged for medical image analysis, disease diagnosis, and surgical assistance [5]. The automotive industry is witnessing a paradigm shift with the development of autonomous vehicles, where computer vision and AI play a critical role in perception, decision-making, and control [3]. Environmental monitoring is another area where computer vision and AI are making significant contributions, enabling the monitoring of air and water quality, deforestation, and wildlife conservation efforts [9]. The potential of computer vision and AI extends beyond these domains, with emerging applications in areas such as smart agriculture, fashion, and even the metaverse [10, 11, 12]. Computer vision and AI are increasingly vital in shaping the future of technology. Advances in areas like image processing, pattern recognition, and deep learning are driving innovation and tackling real-world problems [13]. As these technologies evolve, we can expect groundbreaking applications and advancements that will profoundly impact both technology and society.

II. METHODOLOGY

This research employs a systematic literature review to analyze the current state and future trends of computer vision and AI. A comprehensive search was conducted across multiple databases, using keywords such as "computer vision," "artificial intelligence," "deep learning," "image recognition," and "object detection." To evaluate the identified AI/computer vision applications, a novel framework is proposed, encompassing the following key dimensions: Accuracy and Performance, Generalization and Robustness, Interpretability and Explainability, Ethical and Societal Considerations, and Scalability and Efficiency (Figure 1). This framework will guide the analysis and discussion of each application's strengths and weaknesses. Articles published between 2023 and 2024 were reviewed. This process resulted in a final selection of 30 papers for in-depth analysis.

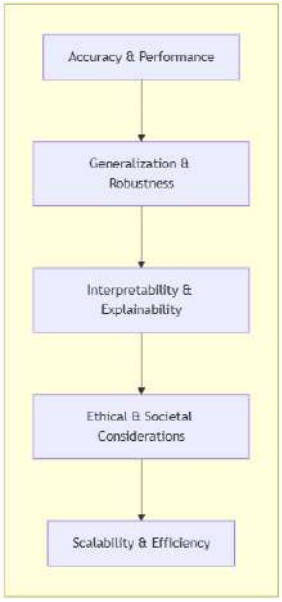


Figure 1 Framework for Evaluating AI/Computer Vision Applications

III. LITERATURE REVIEW

The literature review reveals the transformative impact of computer vision and AI across diverse sectors. In the realm of healthcare, the fusion of these technologies has revolutionized medical image analysis, enabling the detection and diagnosis of diseases with unprecedented accuracy [5]. The automation of tasks such as tumor segmentation and the identification of abnormalities in X-rays and MRIs has not only expedited diagnosis but also improved the precision of treatment plans [5]. Furthermore, AI-powered computer vision systems are assisting surgeons in real-time, providing critical visual information and enhancing surgical precision [5]. CNNs show promise in medical image analysis, but face challenges in generalizability and interpretability. Further research is needed to develop robust and interpretable models.

To assess the effectiveness of these AI-powered computer vision systems in healthcare, the proposed framework can be applied as follows:

1. Accuracy and Performance: Evaluate the accuracy of CNNs in tumor segmentation and other medical image analysis tasks.
2. Generalization and Robustness: Assess the ability of these models to generalize to different types of tumors and imaging modalities.
3. Interpretability and Explainability: Employ XAI techniques to explain the reasoning process of CNNs in medical diagnosis.
4. Ethical and Societal Considerations: Address potential biases in medical image analysis and ensure data privacy and security.
5. Scalability and Efficiency: Evaluate the computational resources required for real-time surgical assistance using AI/computer vision systems.

In the field of robotics, computer vision and AI have empowered robots to perceive and interact with their environment in a more sophisticated manner [7]. The ability to recognize and manipulate objects, navigate complex terrains, and collaborate with humans has expanded the scope of robotic applications across industries [7]. From manufacturing and logistics to healthcare and service sectors, robots equipped with computer vision and AI are augmenting human capabilities and driving efficiency gains. However, this integration also presents significant technical challenges. For example, current algorithms for object recognition often struggle with variations in lighting, viewpoint, and occlusion. This limitation can hinder the reliable operation of robots in real-world environments where such variations are common. A concrete example of this challenge is the difficulty of robotic grasping in cluttered scenes, where the accurate identification and localization of target objects can be severely affected by occlusions and background clutter. The automotive industry is undergoing a profound transformation with the advent of autonomous vehicles, where computer vision and AI are indispensable [3]. The ability of these vehicles to perceive their surroundings, make informed decisions, and navigate safely hinges on the seamless integration of these technologies [3]. The potential of autonomous vehicles to reduce accidents, improve traffic flow, and enhance accessibility is immense, and computer vision and AI are at the forefront of this revolution. Environmental monitoring has also benefited significantly from the integration of computer vision and AI [9]. The ability to analyze vast amounts of visual data from satellites, drones, and other sources has enabled scientists and researchers to monitor deforestation, track wildlife populations, and assess air and water quality with greater precision and efficiency [9]. The insights gleaned from these analyses are instrumental in informing conservation efforts, mitigating environmental risks, and promoting sustainable practices. The fusion of computer vision and AI has also found applications in emerging fields such as smart agriculture, where these technologies are being used to monitor crop health, optimize irrigation, and detect pests and diseases [10]. In the fashion industry, computer vision and AI are enabling personalized recommendations, virtual try-ons, and trend forecasting [0]. The metaverse, a virtual shared space where users can interact with a computer-generated environment and other users, is also leveraging computer vision and AI to create immersive and interactive experiences [11]. These diverse applications highlight the transformative potential of computer vision and AI in shaping the future of technology and society. The ability of machines to see, understand, and interact with the world in a manner akin to humans has opened up new frontiers of innovation and progress. However, it is crucial to acknowledge and address the challenges associated with these technologies to ensure their responsible and ethical deployment.

This review highlights diverse computer vision and AI applications across various domains, ranging from healthcare and robotics to environmental monitoring and emerging fields. However, contrasting approaches and research gaps exist within these applications. For instance, healthcare studies vary in their focus on specific medical image analysis tasks, the types of deep learning algorithms employed, and the level of performance achieved. Similarly, robotics research diverges in its exploration of computer vision techniques. These observations suggest potential avenues for future research, such as developing standardized evaluation metrics, exploring the integration of multiple computer vision techniques, and addressing limitations of existing algorithms.

IV. FINDINGS AND DISCUSSION

The literature review underscores the transformative potential of computer vision and AI across diverse domains. The integration of these technologies has led to significant advancements, revolutionizing industries and reshaping human-computer interaction. The ability of machines to perceive, analyze, and understand visual information has opened up new frontiers of innovation and progress. The analysis of the 30 articles reveals several key findings and overarching themes:

- **Deep Learning's Pervasive Impact:** Deep Learning's Pervasive Impact: Deep learning,

especially CNNs, has greatly enhanced computer vision capabilities [4, 6, 8, 14, 28]. These algorithms excel at extracting features from visual data, improving accuracy in image recognition and object detection. This has led to advancements in healthcare [5], robotics [7], and autonomous vehicles [3], where deep learning aids in perception and decision-making. CNN-based models outperformed traditional machine learning algorithms in terms of accuracy. However, more complex architectures required more processing power, highlighting trade-offs between accuracy, efficiency, and scalability. While some studies advocate for complex CNN architectures, others emphasize model interpretability and efficiency..

- **The Rise of Interdisciplinary Applications:** Computer vision and AI applications span diverse domains, blurring traditional disciplines [21, 10, 11]. In agriculture, these technologies monitor crop health and optimize irrigation [10, 28]. AI is transforming fashion through personalized recommendations and virtual try-ons [11]. The metaverse is further expanding the horizons of computer vision and AI, creating immersive experiences that blend the physical and digital worlds [12, 20, 22].
- **Navigating Challenges and Opportunities:** While the potential of computer vision and AI is immense, the literature also underscores the challenges that need to be addressed. Data privacy and security concerns, especially when dealing with sensitive information, remain a critical issue [9, 23]. The need for explainable AI, ensuring transparency and accountability in decision-making processes, is also gaining prominence [18]. Additionally, ethical considerations, such as bias in algorithms and the potential for misuse, demand careful attention [9]. The successful deployment of computer vision and AI systems hinges on addressing these challenges proactively and responsibly.
- **Emerging Trends and Future Directions:** The increasing availability of large-scale datasets and advancements in computing power are fueling the development of more sophisticated AI models [4, 16]. Integrating computer vision with other technologies like robotics, IoT, and blockchain opens new avenues for innovation [24, 7, 22]. For instance, the integration of computer vision with blockchain technology can enhance the security and transparency of data management in various applications, such as supply chain management and healthcare. Sustainability and environmental monitoring are also gaining momentum, with computer vision and AI playing a crucial role in addressing global challenges [24, 25, 9]. However, it is essential to acknowledge the inherent challenges. AI-based computer vision systems are often limited by the 'human frame,' struggling with tasks requiring nuanced understanding beyond their training data [10]. Additionally, the effectiveness of AI solutions relies heavily on data availability, quality, and security [9]. Acknowledging these limitations is crucial for setting realistic expectations and ensuring responsible implementation. Computer vision and AI are driving technological innovation across various fields. However, responsible use requires addressing challenges related to data privacy and ethical considerations. By navigating these complexities, we can harness their potential to create a more intelligent and sustainable world.

Computer vision and AI are driving technological innovation across various fields. However, responsible use requires addressing challenges related to data privacy and ethical considerations. By navigating these complexities, we can harness their potential to create a more intelligent and sustainable world.

V. CONCLUSION AND FUTURE DIRECTIONS

The convergence of computer vision and AI has ushered in a transformative era, revolutionizing industries and reshaping the way we interact with the world. In healthcare, AI-powered computer vision systems are enabling faster and more accurate diagnoses, leading to improved patient outcomes. In the automotive industry, these technologies are paving the way for autonomous vehicles, promising to reduce accidents and enhance accessibility. These advancements underscore the potential of AI and computer vision to address global challenges and drive innovation across various sectors.

To fully realize this potential, future research should prioritize addressing the ethical implications, ensuring data privacy, and developing explainable AI. This necessitates a multi-faceted approach, focusing on:

- **Advancements in Algorithms:** Investing in research to advance algorithms, with an emphasis on explainable AI, robustness, and real-time processing.
- **Ethical and Regulatory Frameworks:** Developing ethical guidelines and regulations to govern AI usage, addressing bias, transparency, and accountability.
- **Data Sharing and Collaboration:** Promoting data sharing for the development of diverse datasets to train robust AI models.

By proactively addressing these challenges and fostering collaboration between researchers, practitioners, and policymakers, we can harness the transformative power of computer vision and AI to create a more intelligent, efficient, and sustainable world.

VI. Ethical and Societal Considerations

While the advancements in AI and computer vision offer transformative potential, it is essential to address the ethical and societal implications that accompany these technologies. One critical concern is the potential for algorithmic bias, where AI models trained on biased data may perpetuate or even amplify existing societal inequalities. Additionally, the increasing automation capabilities of AI and computer vision raise questions about the future of employment and the need for workforce adaptation.

It is crucial to proactively address these concerns by promoting responsible AI development, ensuring transparency and accountability in algorithmic decision-making, and fostering public dialogue on the ethical and societal implications of these technologies.

References

[1]. Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A review of convolutional neural networks in computer vision. Artificial Intelligence Review, 57, 57-99. https://doi.org/10.1007/s10462-024-10721-6

[2]. Gill, S. S., Wu, H., Patros, P., Ottaviani, C., Arora, P., Casamayor-Pujol, V., Haunschild, D., Parlikad, A. K., \u00c7etinkaya, O., Lutfiyya, H., Stankovski, V., Li, R., Ding, Y., Qadir, J., Abraham, A., Ghosh, S., Song, H., Sakellariou, R., Rana, O., & Buyya, R. (2024). Modern computing: Vision and challenges. Telematics and Informatics Reports, 13, 100116. https://doi.org/10.1016/j.teler.2024.100116

[3]. Arafat, M. Y., Alam, M., & Moh, S. (2023). Vision-based navigation techniques for unmanned aerial vehicles: Review and challenges. Drones, 7(2), 89. https://doi.org/10.3390/drones7020089

[4]. Okatta, C., Ajayi, F., & Olawale, O. (2024). Leveraging HR analytics for strategic decision making: Opportunities and challenges. International Journal of Management & Entrepreneurship Research, 6, 1304-1325. https://doi.org/10.51594/ijmer.v6i4.1060

[5]. Tan, T., Thirunavukarasu, A., Jin, L., Lim, J., Poh, S., Teo, L., Ang, M., Chan, P., Ong, J., Turner, A., Karlström, J., Wong, T. Y., Stern, J., & Ting, D. (2023). Artificial intelligence and digital health in global eye health: Opportunities and challenges. The Lancet Global Health, 11(11), E1432-E1443. https://doi.org/10.1016/S2214-109X(23)00323-6

[6]. Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A review of convolutional neural networks in computer vision. Artificial Intelligence Review, 57, 57-99. https://doi.org/10.1007/s10462-024-10721-6

[7]. Che, C., Zheng, H., Huang, Z., Jiang, W., & Liu, B. (2024). Intelligent robotic control system based on computer vision technology. Applied and Computational Engineering, 64, 150-155. https://doi.org/10.54254/2755-2721/64/20241373

[8]. Ettalibi, A., Elouadi, A., & Mansour, A. (2024). AI and computer vision-based real-time quality control: A review of industrial applications. Procedia Computer Science, 231, 212-220. https://doi.org/10.1016/j.procs.2023.12.195

[9]. Schirpke, U., Ghermandi, A., Sinclair, M., Van Berkel, D., Fox, N., Vargas, L., & Willemen, L. (2023). Emerging technologies for assessing ecosystem services: A synthesis of opportunities and challenges. Ecosystem Services, 63, 101558. https://doi.org/10.1016/j.ecoser.2023.101558

[10]. Wang, Huajia & Gu, Jinan & Wang, Mengni. (2023). A review on the application of computer vision and machine learning in the tea industry. Frontiers in Sustainable Food Systems. 7. 10.3389/fsufs.2023.1172543

[11]. Chrimes, C., & Boardman, R. (2023). The opportunities and challenges of the metaverse for fashion brands. In Title of the book (pp. xx-xx). https://doi.org/10.1007/978-3-031-33302-6_20

[12]. Yuan, J., Zhang, L., & Kim, C.-S. (2023). Multimodal interaction of MU plant landscape design in marine urban based on computer vision technology. Plants, 12(7), 1431. https://doi.org/10.3390/plants12071431

[13]. Lee, H., Chatterjee, I., & Cho, G. (2023). A systematic review of computer vision and AI in parking space allocation in a seaport. Applied Sciences, 13(18), 10254. https://doi.org/10.3390/app131810254

[14]. Kozlov, E., & Gibadullin, R. (2024). Prerequisites for developing the computer vision system for drowning detection. E3S Web of Conferences, 474, 02031. https://doi.org/10.1051/e3sconf/202447402031

[15]. Yang, L., Kumar, R., Kaur, R., Babbar, A., Makhanshahi, G., Singh, A., Bhowmik, A., & Alawadi, A. (2024). Exploring the role of computer vision in product design and development: A comprehensive review. International Journal on Interactive Design and Manufacturing (IJIDeM), 1-48. https://doi.org/10.1007/s12008-024-01765-7

[16]. Iqbal, U., Riaz, M. Z. B., Barth\u00e9lemy, J., Perez, P., & Idrees, M. B. (2023). The last two decades of computer vision technologies in water resource management: A bibliometric analysis. Water and Environment Journal, 37(3), 10.1111/wej.12845.

[17]. Wang, C.-X., You, X., Gao, X., Zhu, X., Li, Z., Zhang, C., Wang, H., Huang, Y., Chen, Y., Haas, H., Thompson, J., Larsson, E., Di Renzo, M., Tong, W., Zhu, P., Shen, X., Poor, H. V., & Hanzo, L. (2023). On the road to 6G: Visions, requirements, key technologies and testbeds. IEEE Communications Surveys & Tutorials, PP, 1-1. https://doi.org/10.1109/COMST.2023.3249835

[18]. Gill, S. S., & Kaur, R. (2023). ChatGPT: Vision and challenges. Journal Name, Volume (Issue), page numbers. https://doi.org/10.1016/j.iotcps.2023.05.004

[19]. Liu, B., Yu, L., Che, C., Lin, Q., Hu, H., & Zhao, X. (2024). Integration and performance analysis of artificial intelligence and computer vision based on deep learning algorithms. Applied and Computational Engineering, 64, 44-50. https://doi.org/10.54254/2755-2721/64/20241374

[20]. Matsuzaka, Y., & Yashiro, R. (2023). AI-based computer vision techniques and expert systems. AI, 4(1), 289-302. https://doi.org/10.3390/ai4010013

[21]. Turchet, L. (2023). Musical metaverse: Vision, opportunities, and challenges. Personal and Ubiquitous Computing, 27. https://doi.org/10.1007/s00779-023-01708-1

[22]. Banaeian Far, S., & Hosseini Bamakan, S. M. (2023). NFT-based identity management in metaverses: Challenges and opportunities. SN Applied Sciences, 5, 260. https://doi.org/10.1007/s42452-023-05487-5

[23]. Tan, T., Thirunavukarasu, A., Jin, L., Lim, J., Poh, S., Teo, L., Ang, M., Chan, P., Ong, J., Turner, A., Karlstr\u00f6m, J., Wong, T. Y., Stern, J., & Ting, D. (2023). Artificial intelligence and digital health in global eye health: Opportunities and challenges. The Lancet Global Health, 11(11), E1432-E1443. https://doi.org/10.1016/S2214-109X(23)00323-6

[24]. Lv, Y. (2023). Transitioning to sustainable energy: Opportunities, challenges, and the potential of blockchain technology. Frontiers in Energy Research, 11. https://doi.org/10.3389/fenrg.2023.1258044

[25]. Nwankwo, C., Ebirim, W., Ninduwezuor-Ehiobu, N., Ani, E., Olu-lawal, K., & Ugwuanyi, E. (2024). Integrating sustainability into HVAC project management: Challenges and opportunities. Engineering Science & Technology Journal, 5(3), 873-887. https://doi.org/10.51594/estj.v5i3.943

[26]. Ettalibi, A., Elouadi, A., & Mansour, A. (2024). AI and computer vision-based real-time quality control: A review of industrial applications. Procedia Computer Science, 231, 212-220. https://doi.org/10.1016/j.procs.2023.12.195

[27]. Yang, B., Yang, S., Li, H., Lyu, Z., Cheng, X., & Wang, F. (2023). Computer vision technology for monitoring indoor and outdoor environments and HVAC equipment: A review. Sensors, 23(13), 6186. https://doi.org/10.3390/s23136186

[28]. Md Akbar, Jalal Uddin & Kamarulzaman, Syafiq & Muzahid, Abu & Rahman, Md Arafatur & Uddin, Mueen. (2024). A Comprehensive Review on Deep Learning Assisted Computer Vision Techniques for Smart Greenhouse Agriculture. IEEE Access. PP. 1-1. 10.1109/ACCESS.2024.3349418.

[29]. Okatta, C., Ajayi, F., & Olawale, O. (2024). Leveraging HR analytics for strategic decision making: Opportunities and challenges. International Journal of Management & Entrepreneurship Research, 6, 1304-1325. https://doi.org/10.51594/ijmer.v6i4.1060

[30]. Matsuzaka, Y., & Yashiro, R. (2023). AI-based computer vision techniques and expert systems. AI, 4(1), 289-302. https://doi.org/10.3390/ai4010013

ACNE10M: Developing a Smart Clinical Decision Support System for Enhanced Acne Care

1st Pham Le Duc Thinh
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
pldthinh.private@gmail.com

2nd Do Anh Kiet
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
dokiet2602@gmail.com

3rd Nguyen Phuc Khang
International University, VNU-HCM
Ho Chi Minh City, Vietnam
khang2073@gmail.com
(Corresponding author)

4th Nguyen Thi Minh Phuong
Hanoi Medical University - HMU
Hanoi, Vietnam
nguyenthiminhphuong3568@gmail.com

5th Nguyen Thuc Anh
Hanoi Medical University - HMU
Hanoi, Vietnam
nguyenthucanhbg02@gmail.com

***Abstract*—Unpredictable health crises like the COVID-19 pandemic have intensified the demand for accessible and efficient healthcare solutions. This study presents a mobile healthcare application dedicated to acne care, integrating telemedicine with personalized health management. The application employs a Clinical Decision Support System (CDSS) utilizing an AI model trained on a curated dataset of annotated acne images to enhance diagnostic precision and treatment outcomes. Built with React, Firebase, and Tailwind CSS, the system architecture ensures scalability, stability, and a user-centric experience. Results demonstrate the application's ability to accurately classify acne types and provide real-time feedback, facilitating personalized treatment tracking. The success of the system is attributed to its three-tier architecture and secure cloud infrastructure, ensuring efficient performance and reliable data management. Despite promising outcomes, further real-world testing is necessary to refine features and enhance user engagement. Future developments will focus on expanding functionality, incorporating advanced analytics, and fostering a collaborative community to elevate dermatological care.**

***Index Terms*—Acne, AI, Clinical Decision Support, Treatment Monitoring, UI/UX, Mobile Healthcare Application.**

I. Introduction

The COVID-19 pandemic has underscored the urgent need for accessible healthcare solutions, revealing significant deficiencies in medical infrastructure, especially in underserved regions [1]. This crisis highlights the necessity of leveraging technology to bridge the gap between patients and healthcare providers, ensuring equitable access to quality care.

Many healthcare apps today lack user-friendly interfaces, comprehensive features, and personalized care, particularly in dermatology. High consultation fees, limited specialist access, and time constraints often delay treatment, worsening conditions like acne. This highlights a critical need for accessible and effective dermatological solutions [2].

To address these issues, we present a user-friendly mobile application dedicated to acne care. The application integrates telemedicine services with certified medical professionals, providing convenient access to quality treatments and trusted skincare products. It also offers a statistical analysis system for monitoring health status and treatment progress via intuitive health charts, simplifying medical history tracking to support effective care.

Our key contributions include:

- **Development of a User-Friendly Mobile Platform**: Enabling effective acne care management through an intuitive application designed for broad adoption.
- **Integration of Telemedicine Services**: Enhancing access to professional care by connecting users with certified medical professionals.
- **Implementation of a Treatment Monitoring System**: Facilitating effective care through a statistical analysis tool that tracks health status and treatment outcomes.

By combining user-centric design with comprehensive support features, our application aims to overcome limitations of existing solutions, offering a targeted and effective approach to acne care. This work advances dermatological treatment through technology and encourages proactive personal health management.

II. Background and Related Work

A. Background

Acne is a widespread dermatological condition affecting millions of individuals worldwide, significantly impacting quality of life and psychological well-being. Despite its prevalence, effective and personalized acne care remains inaccessible to many due to high treatment costs, limited access to dermatologists, and geographical barriers [5], [6]. The advent of mobile healthcare applications offers a promising avenue to bridge these gaps by providing accessible tools for individuals to manage their skin health.

Clinical Decision Support Systems (CDSS) have been increasingly integrated into healthcare software to assist clinicians in diagnosis and patient management by leveraging

personalized data [3]. In dermatology, CDSS can enhance treatment quality and reduce diagnostic errors, leading to improved patient outcomes and experiences [4]. However, many existing applications lack the necessary specificity and user engagement features crucial for effectively managing chronic conditions like acne.

B. Related Research and Work

In this field, several software applications have offered solutions to support patients and doctors in diagnosing diseases, such as:

- **Ada Health**
  Ada Health is an AI-driven application that helps users identify potential health conditions by asking symptom-based questions [7]. While it offers general diagnostic suggestions, it lacks specificity in dermatological conditions such as acne. The app does not provide personalized treatment plans, ongoing treatment tracking, or community support features, which are essential for effective acne management.
- **DXplain**
  DXplain is a web-based CDSS designed primarily for healthcare professionals, providing diagnostic assistance by processing clinical data [8]. Although it generates potential diagnoses and explanations, it is not tailored for patient use, lacks a focus on dermatological conditions, and does not offer mobile accessibility or patient engagement features.

After assessing existing solutions, it is evident that harnessing their strengths and addressing their weaknesses is essential. This project seeks to create an integrated acne care solution.

Our software integrates a CDSS with features that assist patients and doctors in making accurate diagnoses and treatment decisions. Key features include:

- **An AI model for acne recognition**: An AI model analyzes image input to make predictions from multiple angles and lighting conditions. It classifies various lesions and offers diagnoses, treatment plans, and medication suggestions for multiple types, including Acne Scars, Blackheads, Cystic Acne, Flat Warts, Folliculitis, Keloids, Milia, Papules, Pustules, Acne Conglobata, Whiteheads, and Syringomas.
- **A classification system**: This system collects input data and supports disease classification based on a dataset curated by dermatology specialists. The system then provides a general diagnosis, improvement methods, and personalized treatment plans.
- **A tracking algorithm**: Designed to calculate treatment progress, this algorithm displays the progress in chart form, helping users easily monitor their treatment journey.
- **A treatment diary**: This feature allows users to review treatment information and history, enhancing the overall effectiveness of the treatment process.
- **Doctor consultation and booking**: The system offers users the ability to ask questions to a team of doctors and book appointments with treatment hospitals.
- **Personalized product recommendations**: Skincare products tailored to the patient's skin condition are suggested based on the personalized treatment plan.
- **Community support**: Users can engage with a community for mutual support, sharing treatment experiences, as well as receiving news on dermatology and skincare tips.

By incorporating these features, our application aims to overcome the limitations of existing software while delivering a more targeted and effective solution for acne care.

III. Methodology

To initiate this research, it is crucial to adopt a systematic and well-organized approach for planning each stage of the medical application development process. Each phase serves as a critical foundation for the subsequent steps, forming a sequential framework that ensures continuous progress. Failure to adequately execute any phase could significantly undermine the integrity and efficacy of the final product. This highlights the necessity of thorough planning and precise execution throughout the entire development cycle.

A. Use Case Diagram for System Design

The use case diagram for the "Disease Diagnosis System" serves as a high-level representation of the application's core workflows, including patient authentication, data collection, and diagnostic evaluation. This design emphasizes the roles of patients, healthcare administrators, and doctors, whose coordinated interactions are crucial for enhancing system accuracy, user safety, and operational efficiency.

The framework effectively aligns system functions with healthcare standards, enhancing both usability and reliability. It offers a clear roadmap for feature integration, ensuring data integrity and security, robust performance, and adherence to regulatory requirements.

Figure 1 illustrates these relationships, providing a reference for both system functionality and the procedural alignment necessary to support patient-centered care within a mobile healthcare context.

B. Crafting Product Interfaces

The product interface is a key factor influencing the overall success of an application. A product, regardless of its high quality, is unlikely to attract users if it lacks a sense of approachability and friendliness. To build an effective user interface (UI), it is essential to ensure that the user experience (UX) is also optimized. An aesthetically pleasing design that does not function effectively can create a negative impression and hinder users in their interactions with the application.

During the design process, Figma serves as our primary tool, offering a wide range of necessary features. The initial sketching phase of the application is conducted with a focus on accessibility, enabling patients to achieve their health goals easily. The navigation functions are designed to align with

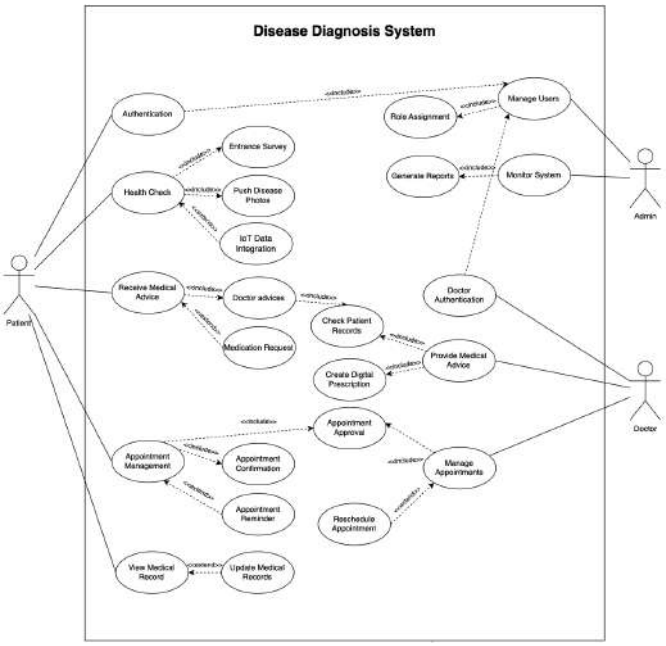


Fig. 1. The Disease Diagnosis System Use Case Diagram.



Fig. 2. UI Design Foundations Crafted with Figma.

the patients' actions, facilitating tasks such as information retrieval, disease diagnosis, and the convenient and efficient storage of medical records. Figure 2 illustrates the initial design pages of the application.

C. ERD Model Design for Data Storage

Efficient management of medical data in a mobile healthcare application requires a well-structured database. The Entity-Relationship Diagram (ERD) (Figure 3) is designed to optimize data storage, retrieval, and integrity while ensuring secure handling of patient, physician, and medical records. It also supports efficient test tracking and appointment management.

1) Core Entities:

a) PATIENT: Stores essential patient details, including **patient_id (Primary Key)**, name, gender, phone, and state. This entity connects to other records for comprehensive healthcare management.

b) PHYSICIAN: Includes physician details such as **physician_id (Primary Key)**, name, specialty, join date, and experience. Links to appointments, test orders, and diagnosis records.

c) APPOINTMENT: Tracks appointments with **appointment_id (Primary Key)**, date, type, and summary, linking patients and physicians.

d) DIAGNOSIS_RECORD: Records diagnoses with **diagnosis_id (Primary Key)**, linked to **appointment_id (Foreign Key)**, and optional medical images.

e) TEST_ORDER: Manages test requests with **test_order_id (Primary Key)**, linked to **patient_id** and **physician_id (Foreign Keys)**.

f) SCREENING_TEST: Stores test details with **test_id (Primary Key)**, test date, type, and links to **diagnosis_id** and **test_order_id (Foreign Keys)**.

g) ENTRANCE_SURVEY: Captures patient health surveys with **survey_id (Primary Key)** and links to **patient_id (Foreign Key)**.

2) Entity Relationships:

- **Patients** can have multiple appointments, submit health surveys, and be linked to test orders and results.
- **Physicians** manage appointments, test orders, and diagnosis records.

3) Data Optimization: The ERD follows normalization principles to reduce redundancy, enhance data integrity, and improve query performance through indexed primary and foreign keys. Logical entity division ensures secure, efficient, and scalable healthcare data management.

D. Software Design According to the WRSPM Process

1) Software Development Process: This software development process consists of five sequential stages designed to guide the project in the right development direction, meet user needs and expectations, and enhance the project's development capabilities.

- **W - Wants:** Identifying the desires, expectations, and objectives for the software development alongside stakeholders, including customers, users, and experts. This collaborative effort provides a comprehensive overview and development direction for the project.
- **R - Requirements:** Following the identification of wants, the project's requirements and implementation plans are clarified. The initial project requirements include:
  - Software construction techniques
  - Expertise required from the team of specialist doctors
  - Requirements for data accuracy and testing
  - A detailed implementation plan by stages
  - Resource requirements for project development
  - User experience requirements for the product
- **S - Specification:** Creating technical documentation that details the software construction process. The documentation encompasses information on system architecture,

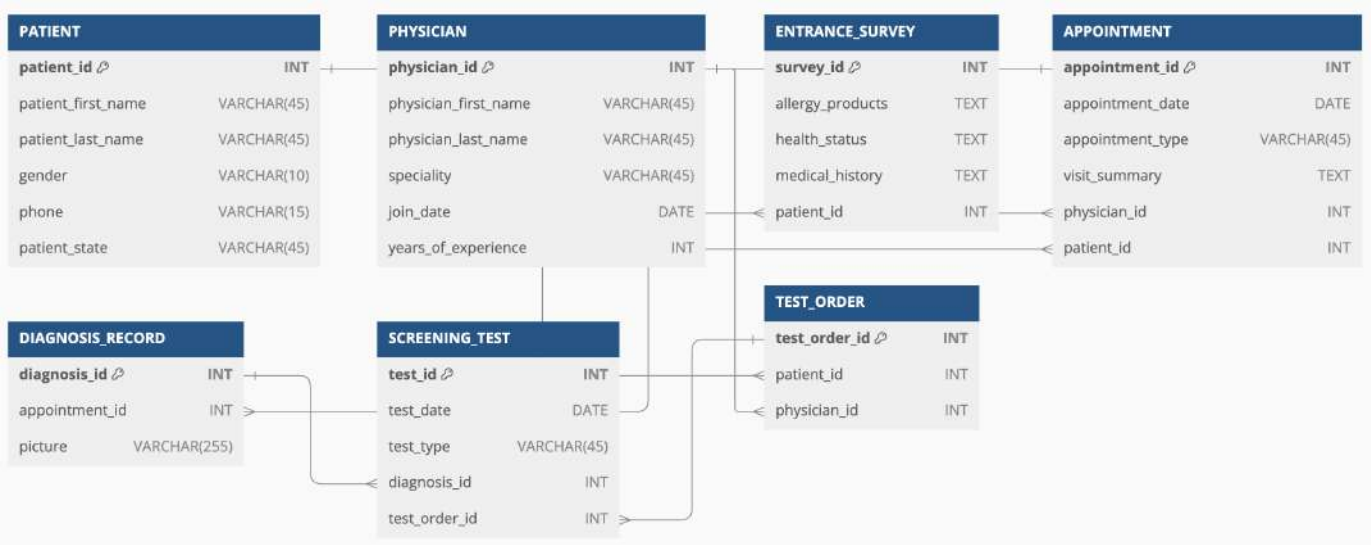


Fig. 3. Entity-Relationship Diagram (ERD) for the Clinical Decision Support System.

database structure, user interface, and software components.

- **P - Programming:** Developers start building the software according to the technical specification document, ensuring alignment with the project's development goals. They work closely with specialist doctors to improve the product's functionality and support debugging and testing efforts.
- **M - Maintenance:** After deployment, the project enters the maintenance phase, focusing on bug fixes, upgrades, and feature improvements to meet user needs. A beta version of the product is widely released to explore, investigate, and resolve potential issues encountered during user operation.

E. Adhering to the SOLID Principles

The SOLID principles are a set of software design guidelines that promote clear system design, enhance performance, and facilitate maintenance and upgrades [9]. These principles offer an effective framework for organizing software development, including:

- **Single Responsibility Principle (SRP):** Each class should have one specific function, thereby ensuring a clear separation of concerns.
- **Open/Closed Principle (OCP):** Software entities should be open for extension but closed for modification, allowing new features to be added without altering existing code.
- **Liskov Substitution Principle (LSP):** Objects in a subclass should be able to replace objects in a superclass without affecting the program's correctness.
- **Interface Segregation Principle (ISP):** Clients should not be forced to depend on interfaces they do not use; instead, interfaces should be tailored to specific functionalities, promoting a cohesive design.
- **Dependency Inversion Principle (DIP):** High-level modules should not depend on low-level modules; both should depend on abstractions. This principle simplifies component management and enhances code clarity.

F. System Architecture and Development Techniques

- **Client Side** The user interface is developed using React and React Native, enabling cross-platform mobile development and efficient UI management through reusable components. Tailwind CSS is utilized for streamlined design and consistent aesthetics. Key features of the application include:
  - **Detection**: Offers sample images, image uploads, and camera capture for accurate skin condition assessment.
  - **Survey**: Collects user data through questionnaires for personalized expert analysis.
  - **History**: Provides a calendar interface to track user activity over time.
  - **Contact**: Displays expert information with options for direct communication.
  - **Resources**: Supplies skincare news and cosmetic product recommendations.
- **Back End**
  The application communicates with the back end via a RESTful API using HTTP methods and JSON data format. Key backend processes include:
  - **AI Detection**: Uploaded images are processed by an AI model API; responses are returned in JSON format. Middleware verifies and processes requests to ensure data integrity.
  - **Firebase Integration**: Data is stored in Firebase using a NoSQL schema. Survey results and activity history are securely maintained to protect user privacy.

– **Treatment Progress Algorithm**: Monitors acne treatment progress by comparing AI detection results over time, presenting data as percentages and visualized in monthly charts.
– **Data Security and User Privacy**
Robust security measures are implemented to safeguard sensitive medical data:

  * **Encryption**: All data transmission uses HTTPS with TLS encryption; sensitive data is encrypted at rest in Firebase.
  * **Authentication and Authorization**: Firebase Authentication manages secure user access with role-based permissions.
  * **Access Controls**: Firebase Security Rules enforce strict data access based on user roles, preventing unauthorized access.
  * **Compliance**: The application adheres to relevant data protection regulations and conducts regular security audits.
  * **User Consent**: Users are informed about data collection and usage, with explicit consent obtained for sensitive information.

- **System Architecture**

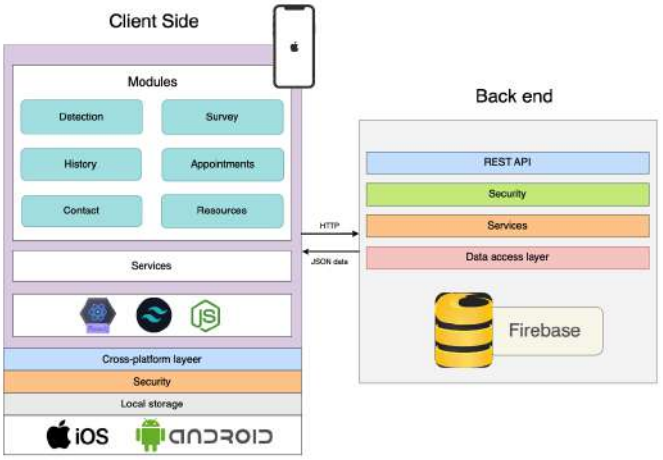


Fig. 4. System architecture of the mobile healthcare application.

The mobile client, built with React and React Native, operates on user devices and utilizes Firebase for data storage. Sensitive data is securely stored with encryption and access controls. Communication between the client and server is facilitated via a RESTful API over HTTPS, transmitting data in JSON format.

- **Partner-Driven Content Update System**
A web application enables partners or administrators to update product data and skincare news, pending administrative approval. Approved content is fetched and displayed in the app, with news categorized by date and products organized by type and characteristics.

IV. Discussion and Results

The research successfully developed a CDSS for acne diagnosis, aimed at improving the quality of skincare for users. The system demonstrates a high degree of accuracy in identifying various acne types and provides immediate feedback when an image is submitted to the model on the server (Figure 5). This process begins with the web application capturing and transmitting the image request to the API, which then processes the data and retrieves relevant information from the database. The database stores details about previous images and diagnoses, allowing the model to analyze the input and return precise results. Figure 5 illustrates the three-tier structure, highlighting the interaction between the API, database management system (DBMS), and application components.

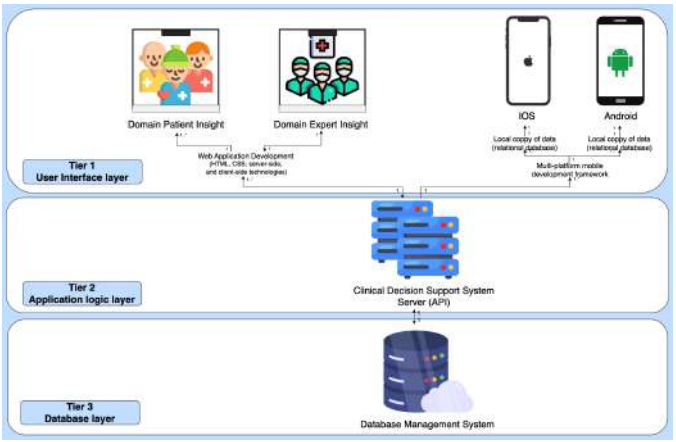


Fig. 5. The software architecture components are structured within a three-tier model.

Beyond its diagnostic functions, the system provides an extensive array of features designed to improve the overall user experience. Users have the ability to track their treatment progress, securely store health-related data, and access professional consultations through a user-friendly interface (Figure 6). The acne diagnosis section relies on real-time data retrieved from the Firebase database, visualized and analyzed via the API. This streamlined integration not only enhances user engagement but also provides deeper insights into individual skin conditions. The seamless interaction between the API, DBMS, and web application ensures a stable and efficient experience for users. Figure 6 illustrates the outcomes of the application's implementation, including data management, retrieval, and visualization of acne progression.

The system offers personalized skincare product recommendations based on users' skin conditions, categorizing products by acne types such as whiteheads, cystic acne, and blackheads, with appropriate tags for efficient retrieval. It also includes detailed usage instructions to improve skincare results.

A. *Effectiveness*

The developed CDSS demonstrates high effectiveness in accurately diagnosing various types of acne and providing personalized treatment recommendations. By leveraging an



Fig. 6. ACNE10M app implementation outcomes for data management and acne development visualization.

AI model trained on annotated acne images, the system offers immediate and precise feedback to users, enhancing their ability to manage their skin conditions effectively. The integration of features such as treatment tracking, secure data storage, and professional consultations further contributes to the overall effectiveness of the application in improving user outcomes.

B. Stability

The application demonstrates strong stability due to its architectural design. By implementing a three-tier model, the system maintains clear separation between the interface, processing logic, and data storage. This structure minimizes potential disruptions and ensures that each component functions independently, reducing the risk of system failures. Furthermore, the use of Firebase for data management provides reliable and scalable cloud storage, contributing to the system's robustness.

C. Scalability

The system's scalability is enhanced by its flexible design and cloud-based services like Firebase, enabling efficient growth with increasing user demand. Its modular architecture facilitates easy integration of additional features or models, ensuring adaptability for future enhancements in diagnostic capabilities and user engagement.

By integrating smart features, the developed CDSS for acne care provides accurate diagnoses and personalized treatments while ensuring long-term reliability and scalability. Its adaptable design forms a strong foundation for future innovations in clinical skincare solutions.

V. Conclusion

This study introduces a CDSS for acne, demonstrating the potential of integrating information technology and artificial intelligence in dermatology. The system advances rapid and accurate acne identification while providing a user-friendly interface for treatment tracking and secure health information storage.

While the system shows promise in analyzing acne conditions and suggesting skincare products, limitations persist. Further real-world testing is necessary to validate the effectiveness and suitability of recommendations for individual users. Understanding user behavior and needs through deeper research is essential to optimize user experience. Moreover, fostering community engagement can enhance knowledge sharing and collaboration among users.

Future developments will focus on expanding the platform's capabilities, integrating advanced analytics to deliver deeper insights, and providing personalized skincare product recommendations based on individual user data. Additionally, the platform will empower users to tailor their experience by focusing on specific acne types and concerns. A strong emphasis on incorporating user feedback will ensure continuous refinement, aligning the platform more closely with user preferences and evolving needs.

In summary, this research confirms the feasibility of the CDSS and highlights that combining modern technology with community support can create new opportunities in healthcare, particularly in dermatology. The system not only improves service quality but also encourages proactive user engagement in personal health management, contributing to a sustainable model for future dermatological care. Empowering patients with the right tools and support enables them to play a significant role in their health journey.

Acknowledgment

We gratefully acknowledge the AIoT Lab Vietnam for their invaluable support in this research.

References

[1] S. Buoro, F. Di Marco, M. Rizzi, F. Fabretti, F. L. Lorini, S. Cesa, and S. Fagiuoli, "Papa Giovanni XXIII Bergamo Hospital at the time of the COVID-19 outbreak: Letter from the warfront", *International Journal of Laboratory Hematology*, vol. 42, Suppl. 1, pp. 8–10, 2020.

[2] T. D. Aungst, "Medical applications for pharmacists using mobile devices", *Annals of Pharmacotherapy*, vol. 47, no. 7–8, pp. 1088–1095, 2013.

[3] M. Elhaddad and S. Hamam, "AI-driven clinical decision support systems: An ongoing pursuit of potential", *Cureus*, vol. 16, no. 4, 2024.

[4] E. W. Breitbart, K. Choudhury, A. D. Andersen, H. Bunde, M. Breitbart, A. M. Sideri, and J. R. Zibert, "Improved patient satisfaction and diagnostic accuracy in skin diseases with a visual clinical decision support system—a feasibility study with general practitioners", *PloS One*, vol. 15, no. 7, p. e0235410, 2020.

[5] OECD, "Geographic distribution of doctors", in *Health at a Glance 2021: OECD Indicators*. Paris: OECD Publishing, 2021. Available: https://doi.org/10.1787/7ad98fa1-en.

[6] L. Starkiene, "Inequitable geographic distribution of physicians: Systematic review of international experience", *Sveikatos Politika Ir Valdymas*, vol. 1, no. 5, 2013.

[7] M. Gräf, J. Knitza, J. Leipe, M. Krusche, M. Welcker, S. Kuhn, and J. Callhoff, "Comparison of physician and artificial intelligence-based symptom checker diagnostic accuracy", *Rheumatology International*, vol. 42, no. 12, pp. 2167–2176, 2022.

[8] The Laboratory of Computer Science—DXplain. [Online]. Available: https://www.mghlcs.org/projects/dxplain. [Accessed: Nov. 17, 2024].

[9] "SOLID Principles: The foundation of high-quality object-oriented code", EnableGeek. [Online]. Available: https://www.enablegeek.com/blog/solid-principles-the-foundation-of-high-quality-object-oriented-code/. [Accessed: Nov. 17, 2024].

EconVNNewsBot: Leveraging Question Answering System with Retrieval-Augmented Generation and Chain-of-Thought Reasoning for Economic News in Vietnam

1st Dang-Khoa Nguyen-Le*
Faculty of Information Technology
University of Science - VNUHCM
Ho Chi Minh City, Vietnam
khoale.aius@gmail.com

2nd Tuyet Hue Tran*,1,2
[1]*Faculty of Information Technology*
University of Science - VNUHCM
[2]*School of Business Information Technology*
University of Economics Ho Chi Minh City
Ho Chi Minh City, Vietnam
huetran.523202180629@st.ueh.edu.vn

3rd Minh-Chau Pham*
AI Lab
AI VIETNAM
Ninh Thuan, Vietnam
chaupm.bn@gmail.com

4th Hanh Thi Hong Tu
University of Social Sciences and Humanities, VNU-HCM
Ho Chi Minh City, Vietnam
tuthihonghanh@hcmussh.edu.vn

5th Tan-Cong Nguyen
University of Social Sciences and Humanities, VNU-HCM
Ho Chi Minh City, Vietnam
ntcong@hcmussh.edu.vn

***Abstract*—The rapid expansion of Vietnamese economic news poses challenges for investors and entrepreneurs seeking accurate information. Traditional search engines often struggle with complex, domain-specific queries, revealing a significant gap in effective information retrieval. This study introduces EconVNNewsBot, a specialized Question-Answering (QA) system designed for Vietnamese economic news. By combining Retrieval-Augmented Generation (RAG) and Chain-of-Thought (CoT) reasoning, EconVNNewsBot delivers precise, context-aware answers. Key findings show that the system enhances user engagement, reduces retrieval time, and eases the information-gathering process for readers. The implications suggest that customized QA systems can transform how economic news is accessed, benefiting both information seekers and news agencies. Our research is supported by the unique EconVNNews dataset, compiled using the EconVNNewsCrawl tool we developed, ensuring the system's robust performance.**

***Index Terms*—Vietnamese economic news, Question-Answering (QA) system, Retrieval-Augmented Generation (RAG), Chain-of-Thought (CoT) reasoning, Semantic chunking**

I. Introduction

In modern journalism, the goal is not only to deliver information but also to ensure that readers can easily comprehend and engage with the content [1], [2]. Accessing accurate and timely information on economics is crucial, yet the overwhelming volume of daily articles makes it challenging for users to find specific answers. Investors and entrepreneurs often struggle to quickly locate precise information using traditional search engines like Google or Bing, which, despite improvements, still fail to deliver fast, accurate responses to complex queries. A Question-Answering (QA) system offers a promising solution to this problem. QA systems can process natural language questions and automatically retrieve accurate answers from both structured and unstructured documents [3]. QA systems are especially beneficial in the economic domain, where the demand for timely and accurate information is paramount.

This study presents EconVNNewsBot, a QA framework tailored for Vietnamese economic news that integrates Retrieval-Augmented Generation (RAG) and Chain-of-Thought (CoT) reasoning. This approach retrieves relevant documents and enhances accuracy by breaking down complex questions into structured reasoning steps. EconVNNewsBot reduces search time, improves response accuracy, and facilitates interaction between readers and newsrooms, streamlining information gathering. The main contributions of this study include:

- Developing **EconVNNewsCrawl**[1] – a tool for collecting data from reputable economic news sources to support the EconVNNewsBot system.
- Introducing **EconVNNews**[2] – a dataset for Vietnamese economic news.
- Providing methods and techniques for building automated interaction systems in media and journalism, enhancing reader-agency engagement, improving service quality, and enhancing user experience on news platforms.

*Dang-Khoa Nguyen-Le, Tuyet Hue Tran, and Minh-Chau Pham contributed equally to this work.

This research is partially funded by the Vingroup Innovation Foundation (VINIF) under the grant number VINIF.2021.JM01.N2.

[1] https://github.com/KhoaLearn/EconVNNewsCrawl
[2] https://huggingface.co/datasets/khoalnd/EconVNNews

- Developing **EconVNNewsBot**[3] - an automated chatbot system that answers readers' questions related to economic news.

The remainder of this paper is divided into four sections. Section II covers related work, Section III explains the article collection process and the EconVNNewsBot architecture, Section IV presents the dataset, and Section V discusses and analyzes the results. Finally, Section VI presents a summary of the conclusions.

II. Literature Review

Question-answering (QA) systems have evolved from rule-based models [4] to advanced deep learning techniques, playing a key role in Natural Language Processing (NLP). While English QA systems have advanced significantly, developing systems for low-resource languages like Vietnamese is challenging due to limited annotated datasets and linguistic resources [5]. This review covers key advancements in QA, including RAG [6], CoT reasoning [7], semantic chunking [8]–[10] and embedding techniques, which are crucial for overcoming limitations in low-resource languages.

RAG) represents a transformative advancement in QA systems [6]. This architecture integrates the retrieval and generation stages to improve open-domain question answering (ODQA). Unlike traditional models, RAG combines parametric memory (fe.g., a pre-trained BART generator [11]) with non-parametric memory (dense vector representations of external data, like Wikipedia, indexed with FAISS) [12]. The dual memory setup in RAG enables it to retrieve relevant documents and base answers on factual data, reducing hallucinations common in generative models. Studies show its adaptability and scalability, particularly in knowledge-intensive fields like finance, medicine, and law, where precision is critical [13], [14].

CoT Reasoning improves the reasoning capabilities of QA by enabling them to break down complex queries into intermediate reasoning steps [7]. This mimics human reasoning, requiring multiple steps to solve problems, and is especially useful for multi-step tasks like comparisons or bridge questions. Integrating CoT with RAG significantly improves performance in complex question-answering tasks [13].

Semantic Chunking is a technique that breaks input text into semantically meaningful units, or chunks, to improve QA system performance [8]–[10]. It helps models focus on relevant segments of long, unstructured text, reducing computational complexity and enhancing retrieval and generation accuracy [15]. This approach is particularly useful for tasks involving multi-topic or lengthy texts, such as in legal, medical, or financial domains, by improving document retrieval and answer generation [15].

Embedding Techniques convert text into dense vectors that capture semantic meaning, allowing QA systems to measure similarity between queries and answers, essential for retrieval and generation. Models like BERT, mBERT [16], and XLM-R [17] are popular in embedding techniques but primarily focus on English and other resource-rich languages [18]. PhoBERT, a pre-trained model for Vietnamese, facilitates Vietnamese text embedding [19]. Recently, Jina introduced Jina-Embeddings-v3, a highly ranked multilingual model for text embedding, which will be utilized in this work [20].

The development of QA systems for low-resource languages like Vietnamese has gained attention, as Vietnamese is considered low-resource in NLP, even being spoken by 97 million people [21]. Nguyen et al. introduced XLMRQA, a Vietnamese QA system using a transformer-based reader on Wikipedia-based knowledge, which outperformed models like DrQA and BERTserini [22]. Their analysis of question types highlighted how question formats impact system performance, crucial for optimizing QA systems.

QA systems for low-resource languages like Vietnamese face both technical and linguistic challenges. PhoBERT, a pre-trained language model, highlights the need for language-specific models to handle Vietnamese syntax and semantics effectively [19]. Such models are crucial for improving QA systems' accuracy and efficiency. While techniques like RAG, CoT reasoning, and semantic chunking have advanced QA systems in resource-rich languages, their application to Vietnamese economic news remains under-explored. The following section introduces EconVNNewsBot, which incorporates these approaches to address challenges specific to Vietnamese news retrieval.

III. Proposed Method

A. Semantic Chunking

The EconVNNewsBot uses the o200k_base tokenizer [23] to tokenize the content of articles and the jina-embeddings-v3 model [20] via JinaAI's Segmenter-API to divide the text into chunks based on semantic relationships [24]. This method begins with representing the input text T as a sequence of tokens: $T = \{t_1, t_2, \ldots, t_n\}$. The goal is to identify chunks: $C = \{c_1, c_2, \ldots, c_m\}$ where each chunk c_i encapsulates a distinct semantic meaning. Tokens are converted into a vector embedding $v_t = E(t_i)$ using pre-trained models which measure semantic similarity between tokens using metrics like cosine similarity. Chunk boundaries are determined by the jina-embeddings-v3 model [20] to predict chunk labels based on context $P(c_i|context)$ [24].

B. Embedding Model jina-embedding-v3

The EconVNNewsBot uses the jina-embeddings-v3 model, designed for multilingual data and long-context tasks, with 570 million parameters and support for up to 8192 tokens [20]. It incorporates LoRA (Low-Rank Adaptation) adapters [25] for tasks like document retrieval and ranking. Built on the Transformer architecture and the Jina-XLM-RoBERTa framework, it replaces absolute with relative positional encoding to capture word dependencies [17]. The multi-head attention mechanism allows the model to grasp complex semantic relationships in long texts. The embedding vector of a text sequence X is computed as follows: $fv_{\text{jina}} = f_{\text{jina}}(x_1, x_2, \ldots, x_n)$.

[3] https://github.com/KhoaLearn/EconVNNewsBot

Where: f_{jina} is the embedding function of the *jinaai/jina-embeddings-v3* model. $x_1, x_2, \ldots, x_n$ are the words in the text sequence. $\mathbf{v}_{\text{jina}}$ is the resulting embedding vector that represents the entire text sequence.

Each embedding vector $\mathbf{v}_{\text{jina}}$ captures the semantic meaning of the entire text sequence, making it efficient for querying and retrieval in the *EconVNNewsBot* system.

The jina-embeddings-v3 model ranks among the top multilingual models in the MTEB (Massive Text Embedding Benchmark), performing well for languages like Vietnamese [26]. It was chosen for its lightweight power and ranks in the top 30, as shown in Figure 1.

Rank	Model	Model Size (Million Parameters)	Memory Usage (GB, fp32)	Embedding Dimensions	Max Tokens	Average (56 datasets)
19	e5-mistral-7b-instruct	7111	26.49	4096	32768	66.4
20	text-embedding-004	1200	4.47	768	2048	66.31
21	TDTE					65.96
22	GritLM-8x7B	46703	173.98	4096	32768	65.66
23	jina-embeddings-v3	572	2.13	1024	8194	65.51
24	gte-large-en-v1.5	434	1.62	1024	8192	65.39
25	LLM2Vec-Meta-Llama-3-supervis	7505	27.96	4096	8192	65.01
26	cde-small-v1	143	0.53	768	512	65
27	LLM2Vec-Mistral-supervised	7111	26.49	4096	32768	64.8
28	echo-mistral-7b-instruct-last	7111	26.49	4096	32768	64.68

Fig. 1. MTEB - Massive Text Embedding Benchmark [26], [27]

C. Vector Database

The system uses Pinecone [28], a vector database to store the embedding vectors. When a user submits a query q, it is embedded into a vector $e_q = f_{\text{model}}(q)$. The system then searches for relevant articles by comparing e_q with the stored article vectors e_i. The query process minimizes the cosine distance between vectors:

$$\min\left(\cos(e_q, e_i)\right), \quad \forall i.$$

The output returns the most semantically similar articles, ensuring the most relevant news is retrieved and presented to the user.

D. Re-ranking Model

Table I compares model performance across benchmarks, showing that the jina-reranker-v2-base-multilingual model used in EconVNNewsBot excels across tasks, making it well-suited for this application.

After retrieving N articles with the highest similarity from the Pinecone, the *jina-reranker-v2-base-multilingual* [29], [30] model re-ranks them to select the top K most relevant articles, enhancing accuracy. This model processes sentence pairs (query and article title/content), embedding and evaluating them for relevance using a semantic classification mechanism. The relevance score for each article A_i is based on the similarity between the query q and the article's content. This process can be expressed as: $\text{score}(q, A_i) = \text{reranker}(q, A_i)$.

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT MODELS ACROSS VARIOUS BENCHMARKS [29]

Benchmark	JinaV2	BgeV2	mMiniLMv2	JinaV1
Model Size	278M	568M	118M	137M
MKQA (26 langs)	54.83	54.17	53.37	-
BEIR (17 datasets)	53.17	53.65	45.40	52.45
MLDR (13 langs)	68.95	59.73	28.91	-
CodeSearchNet (3 tasks)	71.36	62.86	51.78	-
AirBench (zh/en)	61.33	61.28	56.46	-
ToolBench (3 tasks)	77.75	78.46	58.39	74.13
TableSearch	93.31	74.86	53.60	72.89

Where q is the user's query. A_i represents the i-th article, combining its title and content into a single text. reranker is the re-ranking function of the *jina-reranker-v2-base-multilingual* model.

Before re-ranking, the system retrieves N articles from Pinecone based on embedding vector matching. The re-ranking model then refines these by reassessing their relevance to the query, selecting the top K (where $K \leq N$) based on the highest scores. The final K articles are determined using the following formula:

$$K = \arg\max_{A_i \in N} \text{score}(q, A_i), \quad \text{for} \quad i = 1, 2, \ldots, N. \tag{1}$$

Where N is the set of articles retrieved from Pinecone. K represents the final set of selected articles based on the highest re-ranking scores. $\text{score}(q, A_i)$ is the relevance score of article A_i with respect to the query q.

The re-ranking process filters out irrelevant results and prioritizes articles most semantically relevant to the query, significantly improving accuracy and ensuring the *EconVNNewsBot* provides the most valuable information.

E. Chain of Thought (CoT)

1) Analyzing Each Content Chunk: After re-ranking, the top K chunks from selected articles are analyzed against the query q using the GPT-4 Turbo model [31] with a structured prompt to evaluate each chunk step-by-step and extract the most relevant information.

$$\text{Analysis}(C_{ij}, q) = f_{\text{GPT-4 Turbo}}\left(\text{prompt}(C_{ij}, q)\right), \tag{2}$$

where C_{ij} represents the j-th chunk of article A_i, q is the user's query, and $f_{\text{GPT-4 Turbo}}$ is the GPT-4 Turbo model responsible for analyzing the content based on the structured prompt.

This step ensures each content chunk is evaluated for relevance to the query. The prompt directs the model to identify key information, assess economic relevance, and determine how the chunk helps answer the question, ensuring an accurate response.

2) Clarification: After analyzing each chunk, the system undergoes a clarification phase, discarding irrelevant or insufficient information. Only the most relevant details are selected for synthesis, ensuring coherence and accuracy.

3) Final Answer Generation: The final answer is generated by synthesizing the information gathered from the K selected chunks. The following equation represents the process of generating the final answer:

$$a = f_{\text{GPT-4 Turbo}}\left(\text{Chain of Thoughts}(A_1, A_2, \ldots, A_K)\right), \quad (3)$$

where a is the final answer generated by the system. $A_1, A_2, \ldots, A_K$ represent the analyzed articles, and Chain of Thoughts$(A_1, A_2, \ldots, A_K)$ denotes the thought process applied across multiple articles.

The synthesis step combines insights from relevant chunks to produce an accurate, comprehensive answer, with article references for transparency.

Using a carefully constructed dataset and architecture, the EconVNNewsBot optimizes information retrieval from Vietnamese economic news. The next section provides an overview of the dataset and system architecture.

IV. Dataset and System Overview

A. EconVNNews Dataset

The EconVNNewsBot system uses the EconVNNews dataset, which contains diverse economic news articles in Vietnamese collected from reputable sources via the EconVNNewsCrawl tool. It supports research by offering a flexible method for gathering, processing, and storing data from various outlets.

1) EconVNNewsCrawl: EconVNNewsCrawl is designed to collect data from multiple news sources with a flexible data structure, allowing users to easily expand the sources or modify the data collection period without significant changes to the system. Specifically, the tool enables:

- Multiple reputable news sources: The tool supports CafeBiz, CafeF, VietNamNet, and VnExpress.
- Flexible time tracking for data collection: Users can specify the data collection period, making it adaptable to time-specific research.
- JSON structure: The collected data is stored in JSON format with fields like "title", "URL", "content", "published date", "author", "category", and "source", enabling easy integration with analysis tools.

2) EconVNNews - A Vietnamese Economic News Dataset: EconVNNews, covering economic news in Vietnamese from September 2020 to September 2024, includes over 268,000 articles on macroeconomic events and market analyses from experts. It's a valuable resource for economic research and NLP applications like question-answering, trend analysis, and market forecasting.

TABLE II
Number and Proportion of Articles by Year and Source

Source	2020	2021	2022	2023	2024	Total
Number of Articles						
cafebiz	572	2190	2798	4524	4247	**14331**
cafef	13192	43882	47996	52598	40590	**198258**
vietnamnet	2610	7933	7060	6351	3990	**27944**
vnexpress	2567	8258	6368	6199	4800	**28192**
Total	**18941**	**62263**	**64222**	**69672**	**53627**	**268725**
Proportion (%)						
cafebiz	0.21	0.81	1.04	1.68	1.58	**5.33**
cafef	4.91	16.33	17.86	19.57	15.10	**73.78**
vietnamnet	0.97	2.95	2.63	2.36	1.48	**10.40**
vnexpress	0.96	3.08	2.37	2.31	1.79	**10.49**
Total	**7.05**	**23.17**	**23.90**	**25.93**	**19.96**	**100.00**

Table II shows steady growth in article numbers, especially from 2021 to 2023, reflecting Vietnam's post-pandemic economic recovery. The surge in 2022-2023 corresponds to financial market volatility, while the slight decline in 2024 is likely due to the dataset covering only up to September and potential market stabilization.

In summary, the EconVNNews dataset captures economic market fluctuations and media trends during the post-pandemic recovery, offering a rich resource for studying Vietnam's economy.

3) Reliability of Information: The information collection process ensures data reliability by adhering to the following criteria:

- Verified sources: Each article includes author details, publication date, and a link to a reputable source.
- Accuracy: Data is sourced from official financial reports and trusted economic institutions.
- Expert analysis: Many articles include insights from economic experts, enhancing the quality and value of the data.

By meeting these standards, the EconVNNews dataset provides highly reliable and accurate information to support research and data-driven decision-making.

B. System Overview

EconVNNewsBot automatically answers questions about Vietnamese economic news using RAG and CoT reasoning. It integrates embedding models, a vector database, and LLMs to deliver precise answers. The system operates in three stages: data preprocessing, information retrieval, and question answering, as shown in Figure 2.

The EconVNNewsBot architecture includes key components for efficient question answering. The **EconVNNewsCrawl** toolcollects articles from reputable Vietnamese economic sources, which are stored in a structured JSON format in a database D with metadata like titles, URLs, and content.

The system preprocesses the data using semantic chunking [15], breaking long articles into smaller segments for more efficient embedding and retrieval. In the embedding stage, both queries and articles are converted into vector representations,

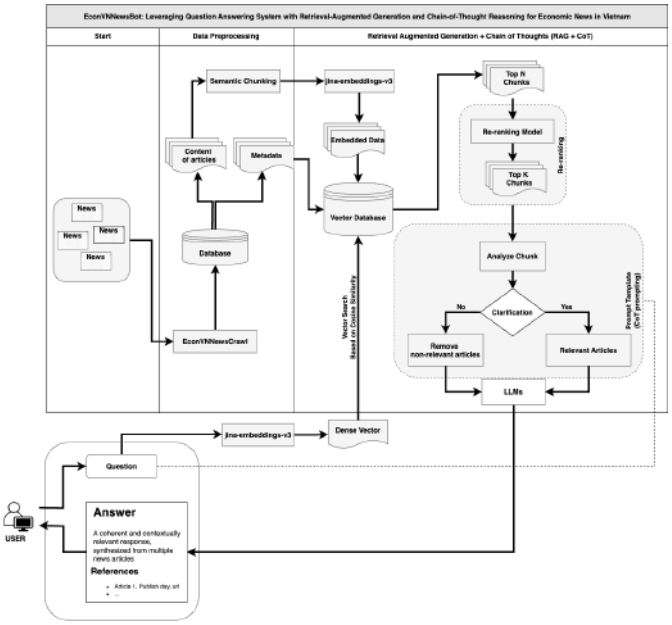


Fig. 2. EconVNNewsBot System Architecture

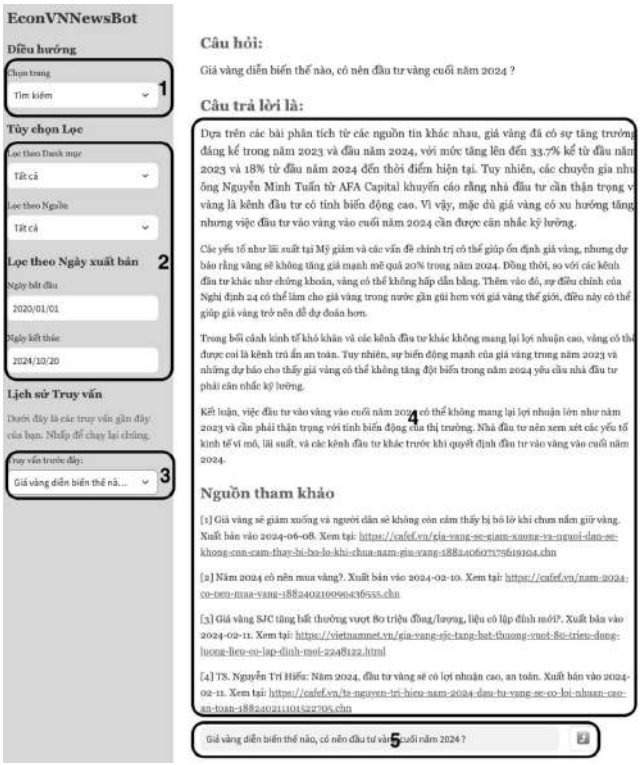


Fig. 3. User Interface of EconVNNewsBot

with each article $x_i \in D$ mapped to an embedding vector e_i using the *jina-embeddings-v3* model [20].

Article vectors are stored in a vector database. When a user submits a query q, it is embedded as e_q, and retrieval is done by minimizing the cosine distance between e_q and the article vectors e_i, ensuring the most relevant articles are retrieved based on semantic similarity.

After retrieving the top N articles, the system re-ranks them using the *jina-reranker-v2-base-multilingual* model [29], [30], refining the selection by evaluating semantic similarity to the query and identifying the top$K$ articles for further analysis.

Finally, the CoT technique [7] analyzes the selected articles, breaking down complex content and evaluating each chunk for relevance and clarity before generating a coherent, contextually accurate response.

C. Front-end System

The front-end interface of EconVNNewsBot is designed for intuitive, user-friendly interaction with Vietnamese economic news. Figure 3 shows the layout and key components, including interactive elements tailored for the query and answer retrieval process:

Navigation Menu: on the left lets users switch between pages, including news search, browsing guidance, and a brief project introduction.

Filter Control: Below the navigation, the system offers several filter options:

- **Category Filter:** This dropdown lets users filter articles by specific news categories (e.g., economy, finance), sourced from the unique categories assigned to each article.
- **Source Filter:** Users can filter articles by source to focus on specific news outlets or publications.
- **Date Filter:** The date range selector allows users to set start and end dates for searching news articles, as shown in Figure 3 (Section 2).

Moreover, this Multi-Filter Control lets users refine their answers and optimize the process to find the correct response.

Query History: This section includes a dropdown of previous queries, enabling users to quickly revisit past search results (Figure 3, Section 3).

Generated Answer: The system displays the generated answer in the main content area, highlighting the most relevant response along with detailed information from multiple articles. A list of referenced sources with direct links to the original articles is also provided for transparency. (Figure 3, Section 4).

Query Input and Submission: At the bottom, users can enter queries in a clean, modern input bar with a "Submit" button to start the query process. (Figure 3, Section 5).

We now discuss the performance of EconVNNewsBot, focusing on its article retrieval, ranking effectiveness, strengths, limitations, and potential applications.

V. Discussion

EconVNNewsBot utilizes a RAG framework with CoT reasoning to enhance the relevance and coherence of its responses. By re-ranking articles and synthesizing answers from multiple sources, the system provides well-rounded results. It has performed effectively across economic categories,

including market trends, sector-specific inquiries, and post-COVID recovery, offering concise and accurate summaries by analyzing critical factors like infrastructure and foreign investments. The RAG-CoT framework has demonstrated a significant improvement in response relevance and coherent reasoning, enabling comprehensive answers to even complex queries through multi-article synthesis.

Future research directions for this study include conducting a comparative analysis to emphasize the advantages of RAG and CoT over alternative techniques, developing a curated question-answer dataset for chatbot testing to enhance system performance evaluation, and incorporating real-time updates to address limitations related to outdated or overly generalized responses.

EconVNNewsBot's RAG-CoT approach holds considerable potential in fields such as journalism, finance, and research, delivering fast, accurate, and relevant information across dynamic topics.

VI. Conclusion

This paper introduced EconVNNewsBot, a QA system tailored for Vietnamese economic news using advanced NLP techniques such as Semantic Chunking, RAG, and CoT reasoning. The system effectively retrieves and answers complex economic queries with high accuracy, leveraging reliable news sources. While EconVNNewsBot demonstrates strong performance, limitations remain in some aspects. conducting a comparative analysis of RAG and CoT's advantages, creating a curated question-answer dataset for more robust chatbot testing, and incorporating real-time updates to address data gaps. The system has promising applications in journalism, financial analysis, and academia, where rapid and accurate economic information is essential.

References

[1] M. Mencher and W. P. Shilton, "News reporting and writing: Brown & benchmark publishers," *United Kingdom*, 1997.

[2] G. Starkey, *Balance and bias in journalism: Representation, regulation and democracy*. Bloomsbury Publishing, 2017.

[3] M. A. Calijorne Soares and F. S. Parreiras, "A literature review on question answering techniques, paradigms and systems," *Journal of King Saud University-Computer and Information Sciences*, vol. 32, no. 6, pp. 635–646, 2020.

[4] B. F. Green Jr, A. K. Wolf, C. Chomsky, and K. Laughery, "Baseball: an automatic question-answerer," in *Papers presented at the May 9-11, 1961, western joint IRE-AIEE-ACM computer conference*, 1961, pp. 219–224.

[5] P. Tran, D. Nguyen, H.-A. Tran, T. Nguyen, and T. Tran, "Building a closed-domain question answering system for a low-resource language," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 22, no. 3, pp. 1–14, 2023.

[6] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel *et al.*, "Retrieval-augmented generation for knowledge-intensive nlp tasks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459–9474, 2020.

[7] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," *Advances in neural information processing systems*, vol. 35, pp. 24 824–24 837, 2022.

[8] FullStackRetrieval-com, "5 levels of text splitting," https://github.com/FullStackRetrieval-com/RetrievalTutorials/blob/main/tutorials/LevelsOfTextSplitting/5_Levels_Of_Text_Splitting.ipynb, 2024, [Accessed 19-10-2024].

[9] LlamaIndex, "Semantic chunker - llamaindex — docs.llamaindex.ai," https://docs.llamaindex.ai/en/stable/examples/node_parsers/semantic_chunking/, 2024, [Accessed 19-10-2024].

[10] LangChain, "How to split text based on semantic similarity | langchain — python.langchain.com," https://python.langchain.com/v0.2/docs/how_to/semantic-chunker/, 2024, [Accessed 19-10-2024].

[11] M. Lewis, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *arXiv preprint arXiv:1910.13461*, 2019.

[12] S. Siriwardhana, R. Weerasekera, E. Wen, T. Kaluarachchi, R. Rana, and S. Nanayakkara, "Improving the domain adaptation of retrieval augmented generation (rag) models for open domain question answering," *Transactions of the Association for Computational Linguistics*, vol. 11, pp. 1–17, 2023.

[13] S. Setty, K. Jijo, E. Chung, and N. Vidra, "Improving retrieval for rag based question answering models on financial documents," *arXiv preprint arXiv:2404.07221*, 2024.

[14] R. Kalra, Z. Wu, A. Gulley, A. Hilliard, X. Guan, A. Koshiyama, and P. Treleaven, "Hypa-rag: A hybrid parameter adaptive retrieval-augmented generation system for ai legal and policy applications," *arXiv preprint arXiv:2409.09046*, 2024.

[15] "Semantic Chunking for RAG: Better Context, Better Results — multimodal.dev," https://www.multimodal.dev/post/semantic-chunking-for-rag, [Accessed 19-10-2024].

[16] J. Devlin, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[17] A. Conneau, "Unsupervised cross-lingual representation learning at scale," *arXiv preprint arXiv:1911.02116*, 2019.

[18] R. Patil, S. Boit, V. Gudivada, and J. Nandigam, "A survey of text representation and embedding techniques in nlp," *IEEE Access*, vol. 11, pp. 36 120–36 146, 2023.

[19] D. Q. Nguyen and A. T. Nguyen, "Phobert: Pre-trained language models for vietnamese," *arXiv preprint arXiv:2003.00744*, 2020.

[20] S. Sturua, I. Mohr, M. K. Akram, M. Günther, B. Wang, M. Krimmel, F. Wang, G. Mastrapas, A. Koukounas, N. Wang *et al.*, "jina-embeddings-v3: Multilingual embeddings with task lora," *arXiv preprint arXiv:2409.10173*, 2024.

[21] K. Nguyen, V. Nguyen, A. Nguyen, and N. Nguyen, "A Vietnamese dataset for evaluating machine reading comprehension," in *Proceedings of the 28th International Conference on Computational Linguistics*, D. Scott, N. Bel, and C. Zong, Eds. Barcelona, Spain (Online): International Committee on Computational Linguistics, Dec. 2020, pp. 2595–2605. [Online]. Available: https://aclanthology.org/2020.coling-main.233

[22] K. V. Nguyen, P. N.-T. Do, N. D. Nguyen, T. V. Huynh, A. G.-T. Nguyen, and N. L.-T. Nguyen, "Xlmrqa: Open-domain question answering on vietnamese wikipedia-based textual knowledge source," in *Asian Conference on Intelligent Information and Database Systems*. Springer, 2022, pp. 377–389.

[23] OpenAI, "GitHub - openai/tiktoken: tiktoken is a fast BPE tokeniser for use with OpenAI's models. — github.com," https://github.com/openai/tiktoken, 2024, [Accessed 19-10-2024].

[24] JinaAI, "Segmenter API — jina.ai," https://jina.ai/segmenter, 2024, [Accessed 19-10-2024].

[25] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," *arXiv preprint arXiv:2106.09685*, 2021.

[26] N. Muennighoff, N. Tazi, L. Magne, and N. Reimers, "Mteb: Massive text embedding benchmark," *arXiv preprint arXiv:2210.07316*, 2022.

[27] N. Muennighoff, N. Tazi, and . Magne, "MTEB Leaderboard - a Hugging Face Space by mteb — huggingface.co," https://huggingface.co/spaces/mteb/leaderboard, 2022, [Accessed 20-10-2024].

[28] E. Liberty, "The vector database to build knowledgeable ai | Pinecone — pinecone.io," https://www.pinecone.io/, 2019, [Accessed 11-10-2024].

[29] J. AI, "jina-reranker-v2-base-multilingual," 2023, accessed: 20-10-2024, available at Jina Reranker Model [Online]. Available: https://huggingface.co/jinaai/jina-reranker-v2-base-multilingual

[30] S. Ognawala, J. Fu, Y. Zhang, and S. Martens, "Jina Reranker v2 for Agentic RAG: Ultra-Fast, Multilingual, Function-Calling & Code Search — jina.ai," https://jina.ai/news/jina-reranker-v2-for-agentic-rag-ultra-fast-multilingual-function-calling-and-code-search/, 2024, [Accessed 20-10-2024].

[31] OpenAI, "Openai gpt-4.0- api," https://platform.openai.com/docs/models/o1, 2021, [Accessed 12-10-2024].

Development of a CNN-Based Application for Identifying Common Flower Species in Vietnam

Xuan Quy Tran
Faculty of Mechanical Engineering
Saigon Technology University
Ho Chi Minh City, Vietnam
DH12101484@student.stu.edu.vn

Dinh Hai Vu
Faculty of Mechanical Engineering
Saigon Technology University
Ho Chi Minh City, Vietnam
hai.vudinh@stu.edu.vn

***Abstract*—The study focuses on developing an application to identify five common flower species in Vietnam using Convolutional Neural Networks (CNNs). The target flower species include frangipani, lotus, orchid, rose, and sunflower. The dataset used for model training consists of 500 images, evenly distributed among the five flower species. The CNN model is designed with a deep architecture, including convolutional layers and fully-connected layers with ReLU activation functions, to optimize image recognition capabilities. Results demonstrate a high accuracy of over 98% on both training and validation sets. The CNN model is integrated into a software application built using PyQt5, providing a simple and user-friendly interface that allows users to upload images and receive predictions quickly. With high accuracy and processing speed, this application has the potential for applications in education, agriculture, and tourism, enabling users to identify flowers quickly and efficiently.**

I. Introduction

Artificial Intelligence (AI) has been experiencing rapid development and is increasingly becoming an integral part of numerous fields, including agriculture, education, healthcare, and tourism [1, 2, 3, 4]. Among the branches of AI, machine learning plays a pivotal role, enabling computers to learn independently and perform complex tasks such as image recognition, speech recognition, and natural language processing. Deep learning, a subset of machine learning, is specifically focused on using neural networks to model complex patterns in data. Convolutional Neural Networks (CNNs), in particular, have been widely applied in image recognition, demonstrating the ability to process image data and classify them quickly and accurately. The development of applications for flower recognition is one of the exemplary cases of applying CNN in practice.

In Vietnam, a country with rich biodiversity and many distinctive flower species, accurate flower identification plays a crucial role in agriculture, tourism, and nature conservation. However, traditional methods relying on books or botanical knowledge are often time-consuming and may not achieve high accuracy. With advancements in AI technology, the application of deep learning models like CNNs in flower recognition offers immense potential, enabling users to identify flowers quickly and accurately.

The primary objective of this research is to develop an application that can recognize five common flower species in Vietnam, including frangipani, lotus, orchid, rose, and sunflower. By training and processing images using a CNN model, the application will enable users to easily identify flower species through images uploaded from their devices. Additionally, the research aims to integrate the model into a user-friendly application, expanding its applicability in fields such as education, tourism, and agriculture.

This research not only addresses the issue of flower recognition efficiently but also opens up various potential applications. The application can be used in education to assist students in learning about botany; in tourism to support tour guides and tourists in exploring distinctive flower species; and in agriculture to help farmers easily differentiate economically valuable flowers. With its scalability and high applicability, this research promises to make significant contributions to related fields.

II. CNN Model

This study focuses on developing a flower recognition application using a dataset of images from five common flower species in Vietnam. A CNN model was trained and evaluated, achieving high accuracy on both training and validation sets. The results indicate strong identification capabilities and the potential for expanding the model to include more flower species.

A. Dataset

In this study, the dataset consists of images of five common flower species in Vietnam , including frangipani, lotus, orchid, rose, and sunflower [5], as shown in Fig. 1. A total of 500 images were collected, with 100 images for each flower species. These images were acquired from various sources, primarily through online image search engines and public image databases. The images were carefully selected to include a variety of viewpoints, colors, and sizes to create diversity in the dataset.

Each image, after collection, was preprocessed by resizing it to 224×224 pixels to standardize the input size for the model. This size was chosen based on the fact that many popular CNN models such as VGG16, VGG19, and ResNet50 also use this input size, as it optimizes the computation process and allows the model to learn more effectively [6]. Furthermore, the pixel values in the images were normalized from the range of 0-255 to the range of 0-1. This normalization not only helps reduce fluctuations during training but also enhances the model's accuracy by facilitating the operation of machine learning algorithms.

The image dataset was then divided into two sets: a training set, accounting for 80% of the total images, and a validation set, accounting for 20%. This division ensures that the model has



Fig. 1. Five Popular Flowers in Vietnam

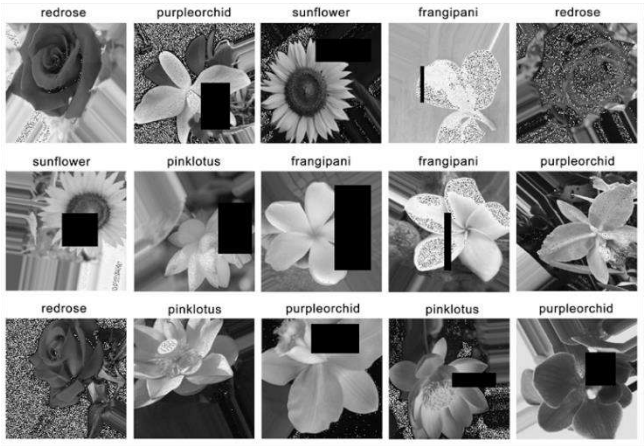


Fig. 2. Augmented Training Dataset Overview

sufficient data to learn during training while also having an independent dataset to test and evaluate the model's performance. The training set will provide information for the model to recognize different flower species, while the validation set helps adjust hyperparameters and detect overfitting. To improve the model's generalization ability and address the limitations of a small dataset, we also applied robust data augmentation techniques such as geometric transformations, noise injection, and random erasing, and carefully adjusted them to create more diversity in the training dataset while ensuring that the generated variations remain realistic [7]. Specifically, geometric transformations such as flip, rotation, shift, shear, and zoom help generate numerous variations in terms of different viewing angles, positions, and sizes of flowers. Noise injection is a technique that adds random noise to the images, effectively blurring out non-essential features and forcing the model to focus on more reliable features for flower classification, such as the overall petal shape or the structure of the flower's stigma. In reality, flower images often contain noise due to varying lighting conditions or camera quality, thus the noise injection technique, combined with brightness adjustment, helps the model learn to classify flowers even under imperfect conditions. Additionally, the random erasing technique is employed to ensure that the model learns to recognize flowers based on multiple features instead of over-relying on any particular characteristic. This technique also helps the model tolerate missing information, such as when a flower image is partially obscured by leaves or shadows. As a result, the model is able to achieve more accurate recognition even when parts of the flower are obscured. The augmented training data is shown in Fig. 2.

B. CNN Model Architecture and Training

In this study, a deep learning model was chosen over traditional machine learning due to the complexity of the data. The color, shape, and structural features of flowers make manual feature extraction extremely challenging and require in-depth knowledge of image processing. In contrast, deep learning models have the ability to automatically learn features from image data without the need for manual extraction, such as CNN or Vision Transformers (ViTs) [8]. Although ViTs can analyze the entire image more effectively, they require a large dataset and significant computational resources to avoid overfitting [9], which is not suitable for the current research context. Moreover, instead of using very deep and complex architectures like VGG or ResNet, which demand substantial computational resources, a custom CNN architecture was developed to ensure efficient feature extraction with moderate complexity, minimizing the risk of overfitting and facilitating practical deployment.

The CNN model in this study is designed to include multiple convolutional layers, pooling layers, and fully-connected layers, as detailed in Fig. 3 and Fig. 4. The architecture of the CNN model is constructed with the goal of optimizing the recognition of features ranging from basic to complex in flowers, such as color, shape, and texture.

Convolutional layers form the core component of the CNN model, responsible for scanning the entire input image using filters to extract important features, such as edges, shapes, or minute details of flower petals. The output of each convolution operation is a feature map that aids the model in distinguishing between different flower species.

Following each convolutional layer is a pooling layer, typically employing the max pooling technique to reduce the size of the feature maps while preserving the most critical features. This reduction helps minimize computational load and prevents the model from overfitting to unimportant features, thus enhancing the efficiency and speed of the model. The pooling layer also contributes to the model's invariance to small changes in the images, such as translations or slight distortions.

After the convolutional and pooling layers, the output is flattened and passed through fully-connected layers, where each neuron is fully connected to all neurons in the previous layer. This segment is crucial for determining the output prediction, which, in this case, involves classifying flower species. Consequently, the final fully-connected layer applies the softmax activation function, which converts the output logits into probabilities for each class, facilitating the selection of the most likely flower species [10]. To mitigate overfitting, dropout is applied after the first fully-connected layer, which contains 512 units, effectively preventing the model from becoming overly reliant on specific neurons, thereby enhancing its generalization ability [11]. The remaining convolutional and fully-connected layers utilize the ReLU (Rectified Linear Unit) activation function, introducing non-linearity into the model. This non-linearity is vital as the features in images are not always linear, and ReLU enables the model to better learn complex relationships among pixels [12].

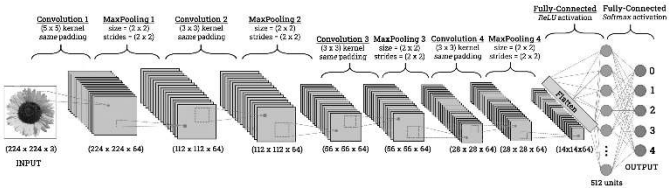


Fig. 3. CNN Architecture Diagram for Flower Classification

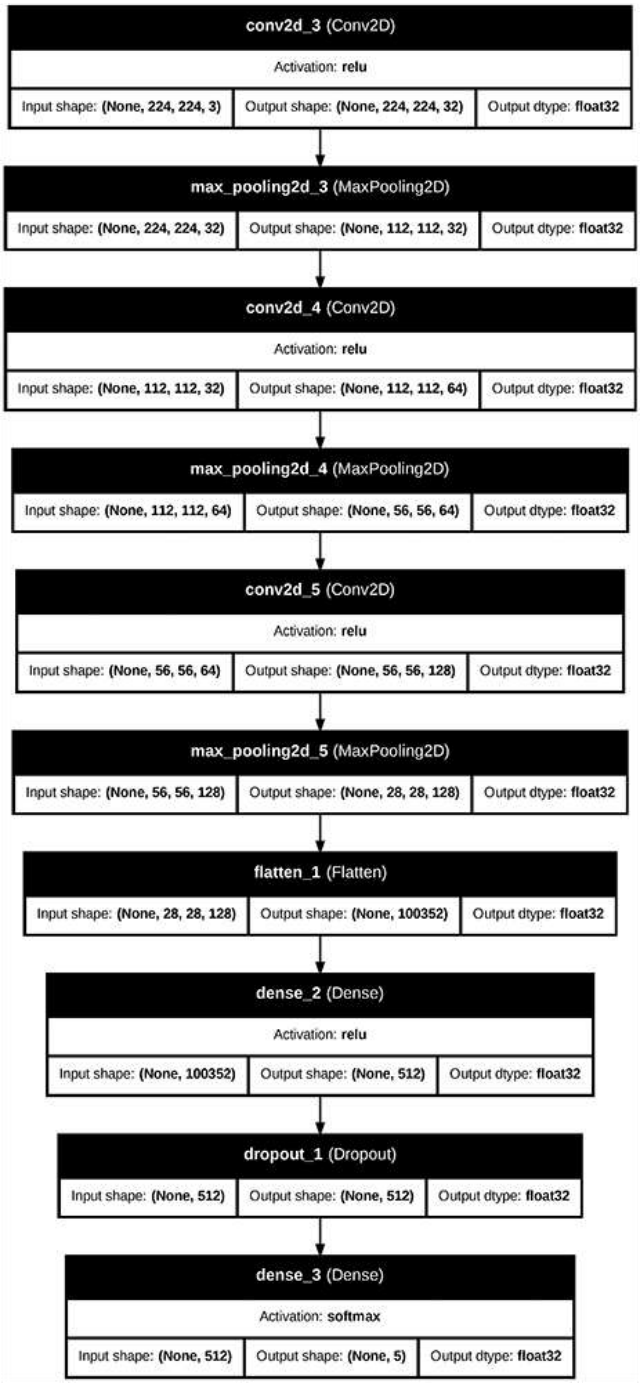


Fig. 4. Model Visualization

For the model compilation process, we employed the Adam optimizer, or Adaptive Moment Estimation, a widely used optimization algorithm in modern machine learning models, which aids in faster convergence [13]. The loss function utilized is categorical cross-entropy, which measures the degree of deviation between the model's predictions and the actual labels [14]. To enhance the model's performance, callbacks such as early stopping and model checkpointing are also employed during the training process. Throughout training, the model is periodically evaluated on the validation set to ensure its generalization capabilities.

C. Result Analysis and Model Evaluation

After completing the training process, the results are presented in Fig. 5. The model achieved a training accuracy of 98.63% and a training loss of 0.0500, indicating that the model effectively learned the features from the training data. During the validation phase, the model achieved a validation accuracy of 99.71% and a validation loss of 0.0092, which suggests that the model had no difficulty recognizing new data and demonstrated good generalization capabilities with unseen data. However, given these nearly perfect results, it is still necessary to consider the possibility that the model may be overfitting to the validation data. Therefore, conducting additional evaluations on the test dataset is essential to ensure that the model performs well on entirely new data. A confusion matrix or a classification report will be used for further analysis [15].

The confusion matrix, displayed in Fig. 6, provides detailed information about the number of correct and incorrect predictions made by the model for each class. It can be observed that the model has a very high accuracy, with only one misclassification occurring between "pinklotus" and "purpleorchid", likely due to their similar characteristics such as color or petal structure.

The classification report, shown in Table I, indicates that the model achieved an overall accuracy of 99%. However, there was a misclassification between "pinklotus" and "purpleorchid", which suggests that although the precision for "pinklotus" was high, its recall was lower. One image from the "pinklotus" class was misclassified as "purpleorchid", which resulted in a lower precision for "purpleorchid", but a higher recall.

To evaluate the overall performance of the model, we rely on the macro average and weighted average values. Macro average represents the average performance of the model across each class, without considering the number of samples in each class. As such, it emphasizes if the model has poor performance on some classes. The precision, recall, and F1-score for macro average are all 0.99, indicating that the model maintains high and consistent performance across all classes. Moreover, since the number of samples in each class is the same, the weighted average metrics are identical to the macro average.

Based on the analysis above, there is no clear indication of overfitting according to the confusion matrix and classification report. If the model were overfitting, we would expect to see a significant difference between the training accuracy and the validation accuracy, or the model would perform very well on the training set but poorly on the test set. However, in this case, the model maintains high performance across all phases (training, validation, and testing). Additionally, the only misclassification between two flowers with similar characteristics is quite natural and does not necessarily indicate overfitting. Nevertheless, to make a more definitive conclusion regarding the model's non-overfitting capability, further evaluations on new datasets under different real-world conditions are needed.

To observe the actual predictions made by the model, refer to Fig. 7, which illustrates several examples of the CNN model's flower classification results.

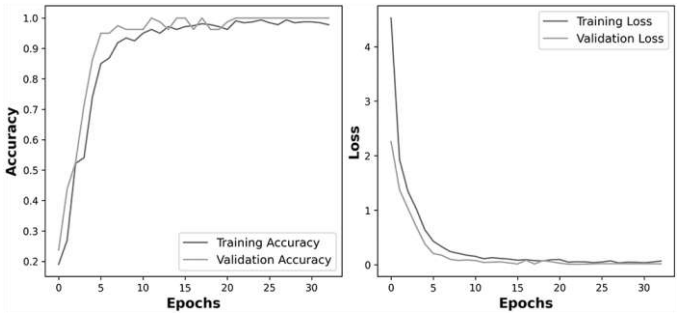


Fig. 5. Training and Validation Accuracy and Loss

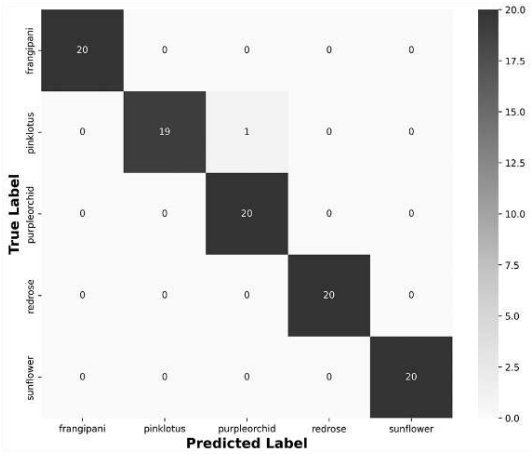


Fig. 6. Confusion Matrix

TABLE I. CLASSIFICATION REPORT

Class	Precision	Recall	F1-score	Support
frangipani	1.00	1.00	1.00	20
pinklotus	1.00	0.95	0.97	20
purpleorchid	0.95	1.00	0.98	20
redrose	1.00	1.00	1.00	20
sunflower	1.00	1.00	1.00	20
Marco avg.	0.99	0.99	0.99	100
Weighted avg.	0.99	0.99	0.99	100
Accuracy			0.99	100

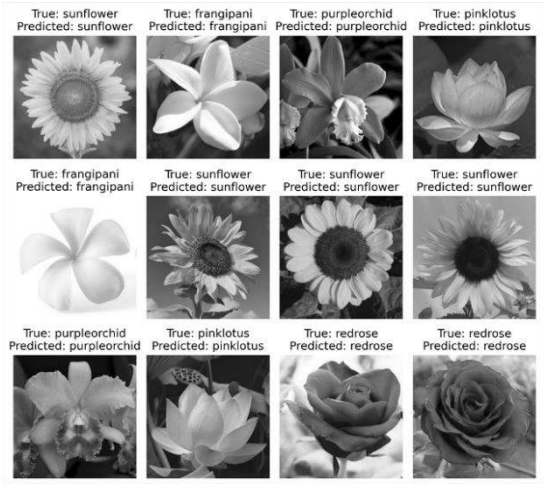


Fig. 7. Some Predictions from the CNN Model

III. FLOWER RECOGNITION APPLICATION

The flower recognition application offers a user-friendly GUI developed with PyQt5 library, ensuring an efficient user experience. Key functions include image preprocessing, integration of a pre-trained CNN model for quick species identification, and informative descriptions of recognized flowers. The application also has significant potential for expansion, enabling the addition of new flower species and applications in agriculture and biological research.

A. Graphical User Interface (GUI)

The flower recognition application was developed with a user-friendly and intuitive GUI, as shown in Fig. 8. This interface was built using the PyQt5 library, one of the most popular libraries for developing GUI applications on the Python platform [16]. Thanks to its powerful capabilities and ease of use, PyQt5 allowed the application to achieve flexibility in design while optimizing the user experience. The GUI is divided into two main sections, ensuring that the flower recognition process is efficient and seamless.

The first area includes the UPLOAD button and an image display area. Users can click this button to upload the image of the flower they wish to identify. Upon clicking the UPLOAD button, a pop-up window allows users to select an image from

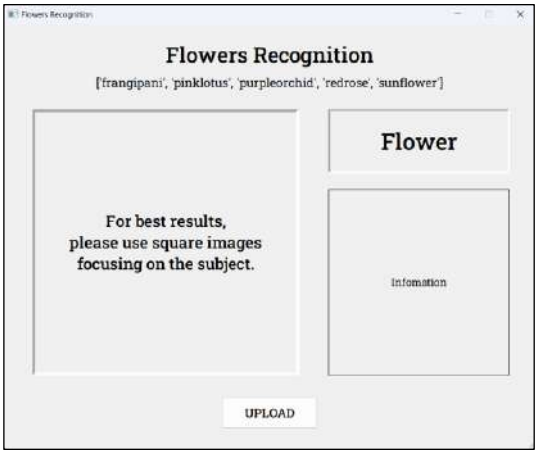


Fig. 8. Flower Recognition Application GUI

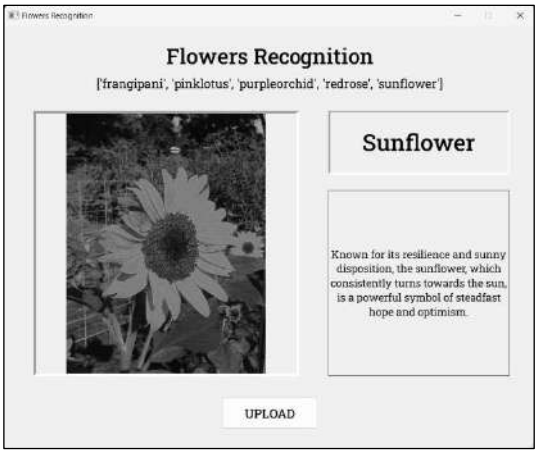


Fig. 9. Recognition Results and Flower Species Description

their personal computer. Once selected, the image is immediately displayed on the application's main interface, allowing users to verify the image before proceeding with recognition. This feature not only provides convenience but also ensures a high level of interaction between the user and the system.

The second area displays the recognition results and provides a brief description of the flower species. After the user uploads the image, the integrated CNN model processes the data and delivers the recognition result in less than one second. The name of the flower species is clearly displayed in the results area. Additionally, the system provides a short description of the flower species, including identifying features, origin, and practical applications. As illustrated in Fig. 9, this information not only helps users easily identify the flower species but also enriches their knowledge of various plants, offering valuable support in education and botanical research.

B. Main Functions

The flower recognition application offers several key functions to assist users in identifying and learning about flower species through images. These functions include image preprocessing, integration of a pre-trained CNN model, generating predictions based on the input image, and providing useful information about the corresponding flower species.

After users upload an image, the application performs a series of preprocessing steps to ensure that the image is ready for the CNN model. First, the application supports various common image file formats such as JPG, PNG, and BMP. All these formats are automatically converted to the standard JPG format, and the image is then resized to 224×224 pixels—the standard input size for the CNN model. Simultaneously, the pixel values of the image are normalized to a range of 0 to 1, facilitating more efficient and accurate model performance during analysis.

The CNN model analyzes the features of the uploaded flower image, including shape, color, and other details. This analysis process is rapid, taking only a few seconds, and returns a prediction result.

Once the prediction process is complete, the name of the flower species is displayed in the result section of the interface, along with a detailed description of the flower. This description includes useful information about the biological characteristics, habitat, and practical applications of the flower species. Additionally, the application can provide information about the cultural significance or historical role of the flower, enriching the user's knowledge in botany and related fields.

C. Development and Expansion Capabilities

In the context of the growing development of AI and deep learning, image recognition systems have played a significant role in various fields, including plant identification. The flower recognition application based on the CNN model not only achieves high efficiency in classifying flower species but also has the potential for continuous expansion and improvement. Below are some development capabilities of the system, demonstrating its flexibility and applicability across diverse fields.

This application can be easily extended by easily adding new flower species to the dataset. Developers simply collect and add images of the new flower species to the training dataset, then retrain the CNN model with this expanded data. Thanks to the flexibility of CNN models, the application maintains stable performance even when increasing the number of recognizable flower species. This capability is particularly crucial in large-scale projects that require the ability to identify numerous flower species without compromising the overall system's performance.

To optimize the application, developers can further enhance the CNN model by applying more advanced deep learning techniques. Such improvements include optimizing hyperparameters, modifying the network architecture, or integrating new deep learning methods. These enhancements not only enable the model to handle challenging cases better, such as blurry images, low-light conditions, or when there are multiple flower species with similar characteristics, but also enhance the overall accuracy and efficiency of the system.

In addition to recognizing flower species, this application has the potential to extend into other domains due to the image recognition capabilities of the CNN model. Some promising fields include crop identification, insect recognition, and even disease detection in crops. These applications would significantly support the agriculture sector by identifying issues related to plant growth and development, thereby improving crop yield and quality. In biological research, this system could also be applied to automate the classification and identification of biological species, helping to reducing time and effort for scientists.

IV. Results and Applications

This section presents the development of a CNN model for a flower recognition application, achieving high accuracy in identifying five common flower species in Vietnam. The user-friendly application processes images rapidly, providing valuable information and insights, while also offering diverse practical applications across education, research, tourism, and agriculture.

A. Achieved Results

This study focuses on the development of a CNN model for a flower recognition application and has achieved impressive results. The model demonstrates high accuracy, with 98.63% on the training set, 99.71% on the validation set, and 99,00% on the test set, confirming its ability to accurately identify five common flower species in Vietnam, including frangipani, lotus, orchid, rose, and sunflower.

Although previous metrics indicate that the model has good generalization capabilities, there is no definitive conclusion yet regarding the model's resistance to overfitting. To ensure that the model is not overfitting, continuous monitoring and further evaluation with new and more diverse datasets, including various real-world conditions, are essential. Furthermore, in the future, the model will be expanded to predict additional flower species, with an increased number of samples collected for each species, thereby improving the model's learning capacity. This expansion will help us better understand the characteristics of flowers that are easily confused, allowing for specific feature

enhancement or the collection of more diverse data to improve the model's performance for those classes.

In contrast, most previous studies on flower recognition have utilized image processing techniques [17] and machine learning models [18, 19], such as Support Vector Machine (SVM), to manually extract features from flower images. This approach requires in-depth understanding of image features and often struggles when dealing with datasets that exhibit high diversity. Additionally, other studies have used pre-trained deep learning models such as VGG-16 [20] and VGG-19 [21], which generally achieve high accuracy but require large datasets like the Flower-102 dataset and significant computational resources to ensure proper generalization.

On the other hand, this study focuses on developing a balanced solution between performance, complexity, and practical deployability. The custom CNN architecture proposed enables the model to work efficiently with moderately sized datasets while also allowing easy deployment in real-world environments with limited hardware resources. Moreover, the use of data augmentation techniques has significantly enhanced the model's generalization ability, particularly for popular flower species in Vietnam. As such, this research not only provides high performance but also addresses the deployment challenges faced by previous works.

In this study, a flower recognition application built on a CNN model provides a user-friendly interface and a fast recognition process. After users upload an image, the model processes it and provides the results in less than a second, allowing users to easily identify the flower species and receive useful information about their characteristics, origins, and practical uses.

B. Applications

The flower recognition application not only helps users identify various flower species but also offers practical applications across multiple fields, including education and agriculture.

In education, it serves as a teaching tool, allowing students to learn about flower species and their ecological roles through interactive photo-based exploration. For scientific research, the application aids botanists and ecologists in classifying rare or endangered species, providing instant information that saves time and reduces identification errors.

In tourism, it enhances the experience for guides and tourists by enabling them to identify native flowers in botanical gardens or national parks. In agriculture, it assists farmers in managing and classifying flowers, facilitating appropriate care measures for both wild and economically valuable species.

Moreover, the application has commercial potential for flower shop customers, enabling informed purchasing decisions by learning about flower meanings. The CNN model's capability also extends to identifying plant species, insects, or other biological entities, supporting agricultural and environmental efforts.

In conclusion, the flower recognition application serves as an identification tool and a platform that fosters opportunities in education, research, tourism, agriculture, and commerce, contributing to knowledge enhancement and sustainable development.

REFERENCES

[1] S. Ziesche, et al., "Role of artificial intelligence in advancing sustainable development goals in the agriculture sector," in The Ethics of Artificial Intelligence for the Sustainable Development Goals, Cham, Switzerland: Springer International Publishing, 2023, pp. 379-397.

[2] S. F. Ahmad, et al., "Academic and administrative role of artificial intelligence in education," Sustainability, vol. 14, no. 3, p. 1101, 2022.

[3] S. Secinaro, et al., "The role of artificial intelligence in healthcare: a structured literature review," BMC Med. Inform. Decis. Mak., vol. 21, pp. 1-23, 2021.

[4] J. Bulchand-Gidumal, "Impact of artificial intelligence in travel, tourism, and hospitality," in Handbook of E-Tourism, Cham, Switzerland: Springer International Publishing, 2022, pp. 1943-1962.

[5] Mytour.vn, "Top 14 beautiful flowers in Vietnam," VNTravel Group, 2024. [Online]. Available: https://mytour.vn/vi/blog/bai-viet/top-14-loai-hoa-dep-tai-viet-nam.html. [Accessed: Oct. 12, 2024].

[6] S. Mascarenhas and M. Agarwal, "A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification," in Proc. 2021 Int. Conf. Disruptive Technol. Multi-Disciplinary Res. Appl. (CENTCON), vol. 1, IEEE, 2021.

[7] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," J. Big Data, vol. 6, no. 1, pp. 1-48, 2019.

[8] E. Arkin, et al., "A survey: object detection methods from CNN to transformer," Multimedia Tools Appl., vol. 82, no. 14, pp. 21353-21383, 2023.

[9] A. Steiner, et al., "How to train your ViT? Data, augmentation, and regularization in vision transformers," arXiv preprint arXiv:2106.10270, 2021.

[10] J. Cao, et al., "Softmax cross entropy loss with unbiased decision boundary for image classification," in Proc. 2018 Chinese Autom. Congr. (CAC), IEEE, 2018.

[11] S. Park and N. Kwak, "Analysis on the dropout effect in convolutional neural networks," in Comput. Vis.-ACCV 2016: 13th Asian Conf. Comput. Vis., Taipei, Taiwan, Nov. 20-24, 2016, Revised Selected Papers, Part II, Springer International Publishing, 2017.

[12] H. Ide and T. Kurita, "Improvement of learning for CNN with ReLU activation by sparse regularization," in Proc. 2017 Int. Jt. Conf. Neural Networks (IJCNN), IEEE, 2017.

[13] S. Bock, J. Goppold, and M. Weiß, "An improvement of the convergence proof of the ADAM-Optimizer," arXiv preprint arXiv:1804.10587, 2018.

[14] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in Adv. Neural Inf. Process. Syst., vol. 31, 2018.

[15] Ž. Vujović, "Classification model evaluation metrics," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 6, pp. 599-606, 2021.

[16] S. B. Nabijonovich and N. G'iyosiddinov, "OPTIMIZING PYQT5 DEVELOPMENT WITH QT DESIGNER," Web of Teachers: Inderscience Res., vol. 2, no. 4, pp. 254-259, 2024.

[17] T. Tiay, P. Benyaphaichit, and P. Riyamongkol, "Flower recognition system based on image processing," in Proc. 2014 Third ICT Int. Student Project Conf. (ICT-ISPC), IEEE, 2014.

[18] H. M. Zawbaa, et al., "An automatic flower classification approach using machine learning algorithms," in Proc. 2014 Int. Conf. Adv. Comput. Commun. Informatics (ICACCI), IEEE, 2014.

[19] I. Patel and S. Patel, "Flower identification and classification using computer vision and machine learning techniques," Int. J. Eng. Adv. Technol., vol. 8, no. 6, pp. 277-285, 2019.

[20] M. A. Rahman, et al., "Flower recognition using VGG16," in Int. Conf. Image Process. Capsule Netw., Cham, Switzerland: Springer International Publishing, 2022.

[21] Z. Y. Ong, et al., "Flower Recognition Model Based On Deep Neural Network With VGG19," J. Inst. Eng. Malaysia, vol. 82, no. 3, 2021.

ASPS - AI-Driven Smart Parking Solutions with Integrated Web and Mobile Applications

1st Tin Quoc Pham
School of Computer Science and Engineering
International University - Vietnam National University HCMC
Ho Chi Minh City, Vietnam
phamquoctin01@gmail.com

2nd Nhat Minh Huu Nguyen
School of Computer Science and Engineering
International University - Vietnam National University HCMC
Ho Chi Minh City, Vietnam
nguyenhuuminhnhat123@gmail.com

***Abstract*—In the revolution where technology becomes more and more important, the needs of having smart systems or AIoT (Artificial Intelligence and Interet of Things) in general nowadays have become a popular trend to solve many issues, and smart parking systems are one of the solutions to solve the problem of optimizing the parking space, especially for an urban area with high population growth in recent years like Ho Chi Minh City. This project will cover a prototype first with the application of AI models, a website base, and a mobile application; then, they can be tested in our university's parking lot at a specific time to evaluate the work efficiency. Until efficiency improves and errors appear at an acceptably low level, this smart parking system will apply to various parking slots available in Ho Chi Minh City. The background will lead to some related research that the smart parking system already existed to analyze the strength and weakness of these researches. License Plate Detection and Character Detection are these main features. We use YOLO with various versions of the model to train to increase the accuracy of detection and, in the end, to find which versions of the model have the best performance to apply into the system. When the results appear, we can see that, compared to others, our model has the best performance and accuracy, especially YOLO version 10, due to the new version and also the various types of datasets we customize to train. In the future, the mobile application will be updated to add a reservation function in advance, and the AI model will be able to detect empty or non-empty seats to support the advance function. In addition, always improving AI models to reduce errors to the lowest level is also something that is being calculated and aimed for in the near future.**

I. Introduction

A. Overview

The combination of artificial intelligence (AI) and the Internet of Things (IoT), also known as AIoT, is becoming one of the modern trends of the technology world, enhancing machine interactions and data management and analytics [1]. In our lives, the applications of AIoT have been integrated and used in some examples of home monitoring and security systems, indoor and outdoor air quality monitoring and management, energy grid management, healthcare systems, and weather monitoring are some examples with which AIoT can provide improvements [2]. Among the difficulties and obstacles of an AIoT system, one problem that has been pervasive in our lives is the need to have safe and highly efficient parking lots, bringing peace of mind to drivers in finding a parking lot in the heart of Ho Chi Minh City, a dense urban city with more than 10 million people and also a place where traffic is increasing. The need to use personal vehicles in recent years has led to consequences such as parking on the sidewalk, affecting the traffic of moving vehicles, and causing air pollution and noise. Many solutions to this problem have been proposed and implemented in today's life, but basically they only have short-term and immediate solutions and do not completely solve the problem, which is still outstanding, thereby creating more thefts as well as crowding in parking lots.

B. Scopes and Objectives

The scope of integrating AIoT into a parking system solution will be a prototype first and then can be tested in our university's parking at a specific time to evaluate the work efficiency. Until efficiency improves and errors appear at an acceptably low level, this smart parking system will apply to various parking slots available in Ho Chi Minh City. Benefits of AIoT bring to the parking system by optimizing when collecting data such as license plate through sensors and cameras in real-time. After the data is collected, it will be analyzed and predicted by the deep learning algorithms from the model, then stored in the database when the analysis is finished. Moreover, the advantage of AIoT, as mentioned above, is to reduce the traffic congestion and time spent finding the vehicle in the parking slots.

There are some main objectives of this research that I want to mention: the web/mobile app and related models that are used in our smart parking system. First of all, the website will be designed for administration to manage the parking space in every distance thanks to the WiFi connection. The functions of this website include a dashboard, where it contains some statistics such as total plates per day, some basic information about the vehicle, plus the plate's images... Next stuff is about the mobile application with some basic functions that can help the user to thoroughly understand all the parking information covered with calculating total time, then it will transform to the total fee and can be paid by the credit/debit card. Moreover, in the near future, we plan to develop the searching for vehicle function, where the user can search for the vehicle by entering the vehicle's plate information, and then that data will be compared with the results taken from the database already analyzed by the AI model to find similar results by showing

the images of the vehicle's plates. Other main objectives that cannot be denied with the most important of this system is the AI model. We all know there are lots of models that can be used for different types of functions; for instance, GPT (Generative Pre-trained Transformer) is used to automatically generate text, answer questions,... Another model that can detect objects in real time and also the model that we use to research in detail today is YOLO (You Only Look Once).

II. Background

There are various research papers about the AIoT Smart Parking System that have been published, some of the outstanding papers will be shown above to analyze in detail about the advantages/disadvantages and technologies that they are applying in their system. For example, Telefonica I+D and the University of Cantabria developed an IoT system to improve urban mobility and alleviate traffic congestion [3]. This system employs sensors distributed throughout the city and updates the parking level on the display boards at the intersections from the data collected from the parking bays. Pham Thanh Nam also designed a smart parking system based on RFID and QR codes in order to reduce the costs of implementation and availed a web portal and a mobile phone for checking parking slots in the vicinity [4]. The system also incorporates ultrasonic sensors to increase automation. At the same time, TIS Corporation deployed an automatic parking management system with the integration of license plate recognition and video surveillance with video analytics.

Additionally, Prabhu Ramaswamy's research focused on similar issues and designed some smart parking systems with Pi cameras and ultrasonic sensors that connect Raspberry Pis for vehicle number plate recognition [5]. Even so, several other limitations are present in these systems, including the absence of artificial intelligence and deep learning applications due to inadequate input data leading to poor performance. Systems using optical character recognition (OCR) technology are very simple to use; however, they provide very low accuracy levels as plates could be noisy or old.

The report shows the result of YOLO version 8 is an object detection algorithm that reaches 92 percent accuracy in capturing vehicle license plates, serving a fast processing by detecting multiple objects with a single network transfer. This success makes YOLO version 8 an ideal AI model for real-time applications, such as smart parking systems that meet the requirements of quick and efficient detection. Beside the high accuracy, YOLO version 8 still remains some challenges, particularly when objects move closer together as the bounding boxes are overlapping between the objects. Moreover, the affection of various angles and low light can be the difficulties when detected. Advanced image processing techniques, including brightness enhancement and color adjustment, can improve the quality of images before feeding them to the recognition model. Additionally, training the model on diverse data and positioning multiple cameras at different angles can help minimize blind spots and enhance detection accuracy. [6].

A social IoT-based platform for the deployment of a smart parking solution provided a system that uses magnetometer sensors to automatically identify and detect the vehicle when it's approaching each point of parking spots, and then the sensor data will be collected by the concentrators through WiFi. An Android app can also be designed to manage some statistics about the parking system to help users know if the status of that spot is available or unavailable. One more useful function is the online payment that can be linked straight to the credit or debit card to purchase the fee after ending the parking progress without any further payment actions. Moreover, this system also uses Bluetooth low-energy to communicate in short range between the concentrators and electronic dashboard to analyze the information of a vehicle, then it will be pushed to the cloud database [7].

This smart parking system based on IoT has many advantages, including saving time and effort for drivers who no longer have to spend time looking for a parking spot. The result is that this also helps in reducing traffic congestion and enables efficient management by offering real-time information about the parking lot, which aids in effective monitoring and supervision. Beyond that, it also aids in reducing pollution by decreasing the amount of time spent searching for parking, which in turn contributes to a decrease in carbon emissions and the protection of the environment. In addition, the system includes electronic payment options, allowing for convenient and speedy parking fee transactions. Nevertheless, there are some drawbacks to the system, such as the absence of physical locking options for public parking and the lack of capabilities to analyze driver habits. Moreover, for accurate results, the license plate detection algorithm necessitates the vehicle to be fully immobilized. In order to overcome these limitations, experts suggest employing machine learning to examine data and anticipate driver actions, ultimately making the system more efficient and delivering improved services. They also recommend improving security measures to safeguard user data and prevent unauthorized entry [8].

The goal of the study is to enhance the vehicle license plate recognition (LPR) by utilizing artificial intelligence (AI) for better optimization. Many of the past programs used the Python programming language as the primary language, making it easier to cite and integrate in our system. The system we are proposing will leverage the YOLO AI model, which, through the use of smart design, will be rented out and will be able to scan number plates of motor vehicles with as high a degree of accuracy as 92 percent remakably within real time, which can be used in smart parking solutions. However, there are challenges that still exist, such as determining LPR in a circumstance where there are numerous items to be captured or when the light is poor, such as at night or when there is glare from a car headlight.

In order to address these shortcomings, they must be able to use powerful zoom cameras that also perform well in low-light conditions. Furthermore. Automated correction of various abnormalities applicable within the image-enhancing algorithms and cross-training on different datasets will enhance

the performance of the system in more environments. The addition of more cameras located at varied angles will also prevent some billowing and would improve the accuracy of the system.

There have been many such systems in the past, and most of them are quite encouraging; however, there is still room for improvement in terms of achieving total automation, which will fulfill the present demands in relation to smart parking systems.

III. Proposed method

This paper implements an ASPS system for applying our proposed method. Figure 1 shows the overall process of our system, including the process when vehicles are in and out. The system includes two steps, which are license plate detection and character detection. In the first step, all vehicles with matching license plates are extracted. Next, all license plate characters will be recognized in the License Plate Recognition. Following that, several machine learning algorithms will be used to estimate future parking availability and optimize space usage. To ensure the reliability of the collected data, these data will be compared to the findings of other studies and real-life outcomes.

The smart parking management system includes two main processes, followed by the Figure 1: vehicle check-in and vehicle check-out. In the vehicle check-in process, when a vehicle approaches the entrance gate, a sensor at the stoppage will activate the system's vehicle recognition feature. The system triggers a camera to capture the driver's face and the vehicle's license plate, which will be analyzed by AI and stored in the database. The system updates the latest available parking spots and displays this information on the mobile app. Additionally, digital boards show the driver's details and the location of the available parking space. During the vehicle check-out process, when the vehicle reaches the exit gate, the sensor again triggers the camera to capture the driver's face and license plate. The AI system then analyzes and compares this data with existing information to verify the identity. If the data matches, the system calculates the total parking time and displays the fee. This information is then sent through the mobile app, where the driver can pay via online payment methods or credit card.

A. License Plate Detection

Recently, at this step of the project, it will be divided into two main stages: vehicle detection and license plate detection. These processes lead to the sequences of increasing the cost and affecting the processing speed of the system. However, thanks to the YOLO versions 8, 9, 10, and 11 containing the ability of recognizing the small objects, we have designed detection models relying on the YOLOv8-m, YOLOv9-c, YOLOv10-b, and YOLOv11-l with an input image size of 640. Both vehicle detection and license plate detection can work and detect in real time.

B. Character Detection

In the next stage of the project, we focus on character detection, which is integral to recognizing and processing the information displayed on the detected license plates. This process requires a high degree of accuracy and efficiency, especially given the need to handle various font styles and sizes. By utilizing the advanced capabilities of YOLO versions 8, 9, 10, and 11, we have developed detection models that leverage the strengths of YOLOv8-m, YOLOv9-c, YOLOv10-b, and YOLOv11-l. The character detection system operates in real time, ensuring that characters from the license plates are recognized swiftly and accurately, thereby contributing to the overall effectiveness of the vehicle detection and license plate recognition system. Additionally, we applied the Canny and Hough Transform Algorithm to get the final result, along with other equations to rotate the image so the license plate will be in the best visualization for detection.

Hough Transform Algorithm: In this system, we use Hough Transform to identify the edge and structures of the plate, which can help improve detection accuracy.

$$r = xcos(\theta) + ysin(\theta) \tag{1}$$

where:

- r is the perpendicular distance from the origin to the line
- θ is the angle between the x-axis and the perpendicular to the line

Rotation Matrix for Image Rotation: This function uses a 2D rotation matrix to rotate the image by a specified angle around its center. The rotation matrix M is defined as:

$$(M) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{2}$$

where:

- θ is the angle of rotation in radians
- (x,y) are translation values to keep the image centered after rotation

The translation is often computed based on the center of the image to avoid shifting it out of the frame:

$$t_x = (1 - \cos(\theta)) \cdot x_c - \sin(\theta) \cdot y_c \tag{3}$$

$$t_y = \sin(\theta) \cdot x_c + (1 - \cos(\theta)) \cdot y_c \tag{4}$$

where:

- x_c, y_c is the center of the image.

Angle Calculation Between Two Points: The skew is determined by analyzing lines detected in the image, using the arctangent of the slope of these lines. The angle θ between two points (x_1, y_1) and (x_2, y_2) is given by:

$$\theta = \arctan\left(\frac{y_2 - y_1}{x_2 - x_1}\right) \tag{5}$$

$$\theta_{degrees} = \theta_{radians} \times \frac{180}{\pi} \tag{6}$$

where:

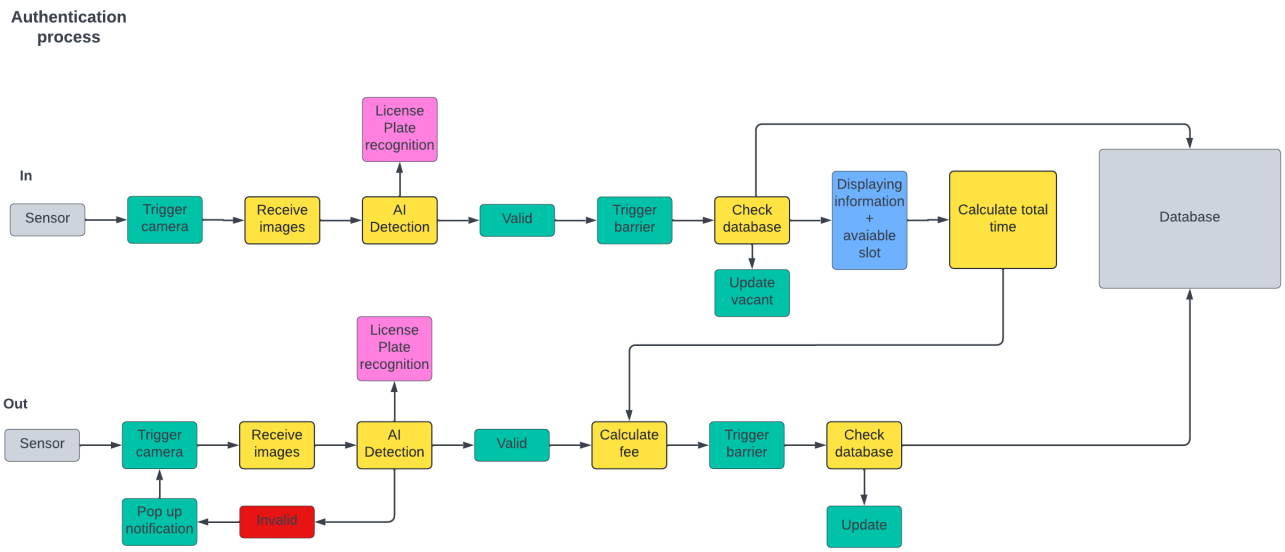


Fig. 1. Process Diagram

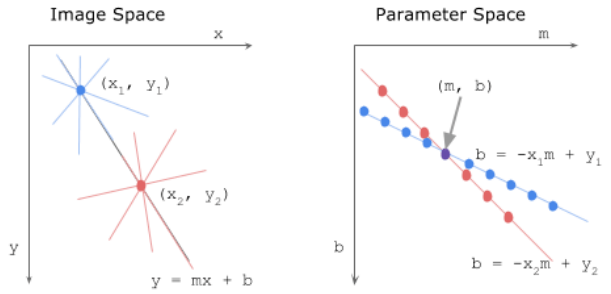


Fig. 2. Example of Hough Transform Algorithm

- (x_1, y_1) are the coordinates of point A.
- (x_2, y_2) are the coordinates of point B.

Note: The angle θ will be in radians. If you need it in degrees, you can convert it using the equation [6].

AI models are integrated into this project, each with their own specific purposes. YOLOv11 (You Only Look Once), which is a newest YOLO version, introduced at the YOLO Vision 2024 event. It provides greater accuracy, faster processing speeds, and enhanced feature extraction while utilizing fewer parameters than previous models like YOLOv8. We used YOLOv8, YOLOv9, YOLOv10 and YOLOv11 in different stages of the project to compare performance improvements across versions. YOLOv8 served as the initial baseline, helping to establish core detection capabilities, while YOLOv9 and YOLOv10 introduced incremental enhancements in speed and detection accuracy. YOLOv11 marked a significant turning point, as it not only optimized computational efficiency but also improved the precision of object detection compare to YOLOv8, making it ideal for real-time applications in our system. By evaluating each version, we were able to fine-tune the AI models to meet the project's specific requirements.

Training models:

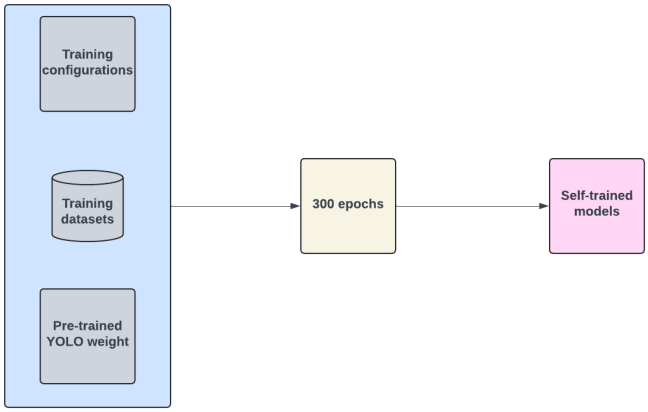


Fig. 3. Training Pipeline for License Plate Detection

In Figure 3, the training progress is showed to fulfill the requirements of the plate detection and recognition tasks. Some custom training configurations for those models include:

- **task = detect**: comment that specify the task, here is detection
- **mode = train**: Training mode
- **model = yolovX.pt**: Specifies the model file for training (We used YOLOv8-m, YOLOv9-c, YOLOv10-b and YOLOv11-l)
- **imgsz = 640**: size of input images
- **save = True**: Enables saving of training checkpoints and final model weights
- **save_period = 10**: Save checkpoints every ten epochs in case of corruption

- **batch = 16**: Number of samples fed into the model at each iteration in the training process
- **epochs = 300**: Total number of training epochs

IV. Result and Evaluation

A. Result

Here is the result after running those models (be noticed that different YOLO versions give different results in accuracy and speed):



Fig. 4. Result

After running some tests, we saw that the YOLOv10 gives the highest score in overall (accuracy and speed) and much more stable, whereas others might give some incorrect results (most likely license plates are skewed or too blurry to detect). However, the YOLOv10b that we used is intensive, that require high amount of computation resources.

TABLE I
Comparison between models

	YOLOv8	YOLOv9	**YOLOv10**	YOLOv11
mAP Score	0.968	0.987	**0.984**	0.985
Precision	0.959	0.976	**0.978**	0.972
Recall	0.962	0.968	**0.982**	0.97

For the actual result after training the models, we will deliver the result of YOLOv10 because from Table I, the score is much pretty higher than others.

From Figure 5, the average precision score for each character is 0.978, indicating that the result is quite good.

From Figure 6, the evaluation curves display a distinct pattern compared to the training curves. Unlike the training data, the evaluation curves show an upward trend, reflecting better character detection over time. From Figure 5, we can observe data for mAP50, mAP50-95, and precision, which indicate an overall improvement in prediction accuracy. However, the precision curve still exhibits some fluctuations, suggesting that certain environmental factors, such as rain or sunlight, may impact plate visibility. These conditions can cause plates to become overly reflective or obscured by water droplets, affecting the recognition process.

Class	Images	Instances	Box(P	R	mAP50	mAP50-95):
all	1377	11826	0.978	0.982	0.984	0.745
0	1377	780	0.985	0.985	0.988	0.744
1	1377	1515	0.985	0.985	0.982	0.664
2	1377	992	0.981	0.983	0.983	0.734
3	1377	759	0.973	0.98	0.971	0.732
4	1377	798	0.984	0.991	0.99	0.728
5	1377	1187	0.984	0.985	0.984	0.757
6	1377	1253	0.991	0.987	0.99	0.735
7	1377	1189	0.989	0.989	0.989	0.729
8	1377	723	0.981	0.976	0.976	0.73
9	1377	1112	0.978	0.98	0.982	0.741
A	1377	93	0.978	0.96	0.975	0.751
B	1377	141	0.956	0.928	0.92	0.687
C	1377	107	0.991	0.981	0.987	0.735
D	1377	174	0.988	0.962	0.972	0.741
E	1377	78	1	0.972	0.994	0.785
F	1377	103	0.977	0.981	0.987	0.763
G	1377	50	0.961	0.978	0.992	0.734
H	1377	64	0.949	0.984	0.983	0.761
K	1377	100	0.997	1	0.995	0.731
L	1377	123	0.975	1	0.984	0.717
M	1377	81	0.987	0.959	0.976	0.782
N	1377	57	0.937	0.982	0.986	0.759
P	1377	78	0.995	0.987	0.987	0.775
R	1377	10	0.982	1	0.995	0.725
S	1377	65	0.98	1	0.981	0.746
T	1377	53	0.929	0.988	0.984	0.717
U	1377	32	0.969	0.981	0.994	0.764
V	1377	36	0.947	1	0.993	0.77
X	1377	40	0.986	0.975	0.992	0.776
Y	1377	8	1	0.99	0.995	0.804
Z	1377	25	1	0.98	0.995	0.772

Fig. 5. Validation across 31 characters

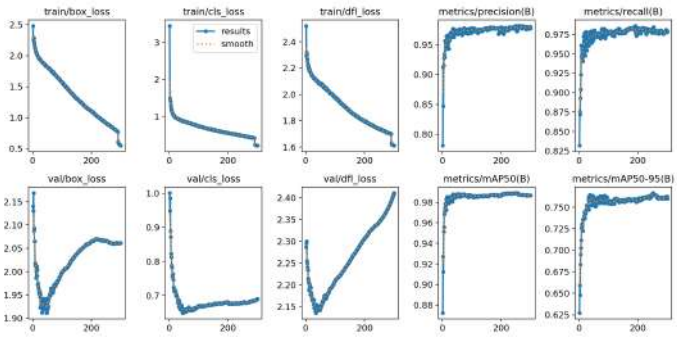


Fig. 6. Training losses and metrics over epochs

B. Evaluation

1) License plate detection: For license plate detection, the output performance has been evaluated by using some other methods to compare with. Mai Chi Bao [9] has been detecting the license plate based on KNN and OpenCV. He will first taking the picture and then gray scaling it up, after that he increase the contrast level so it will be easier to read the black character in the next steps. Tuan et al. [10] proposed their work with boundary line-based method combining with Hough transform and Contour algorithm. Huy et al. [11] have been detecting license plates using YOLOv5.

Based on Table II, we were using the same Vietnamese license plate datasets for all methods, which shows that our method delivers higher accuracy in both 1-line plates and 2-line plates. While Bao's model has the lowest percentage because when the plates were rotated from many angles and many positions, leading to the calculation of the area, the high/wide ratio of the number plate no longer met the set

TABLE II
EVALUATION FOR LICENSE PLATE DETECTION

Methods	Accuracy (%)
Bao [9]	44.25
Tuan et al. [10]	93.4
Huy et al. [11]	94.8
Our method	**95.5**

conditions, so it was eliminated.

2) Characters recognition: For character recognition, our work has been compared to those on the license plate part. Bao [9], using KNN to read the characters from the license plate, comes along with the Canny algorithm to extract the edge detail of the plate, and then he will use KNN to read characters. Tuan et al. [10] used the Hidden Markov Model for OCR. Huy et al. [11] applied the Character Timeseries Matching algorithm (CTM) to get the final result. Although this technique is quite unfamiliar, it presented remarkable results.

TABLE III
EVALUATION FOR CHARACTERS RECOGNITION

Methods	Accuracy (%)
Bao [9]	33
Tuan et al. [10]	96.5
Huy et al. [11]	97.3
Our method	**98.4**

Table III shows that Bao's method [8] still has the lowest accuracy, mainly due to the confusion between characters such as the number 1 and the number 7, the letter G and the letter D, the number 6 and the number 0, and the letter B and the number 8.

TABLE IV
AVERAGE EVALUATION WITH SOME PREVIOUS METHODS

Ref.	Image total	Correct rate	Error rate	Average (%)
In Ref. [9]	No	No	No	38.63
In Ref. [10]	20000	18990	1010	94.95
In Ref. [11]	20000	19210	790	96.05
In Ref. [12]	700	682	18	97.43
In Ref. [13]	1176	1125	51	95.7
In Ref. [14]	11696	11111	585	95.00
In Ref. [15]	1000	910	90	91.00
Our method	**20000**	**19390**	**610**	**96.95**

In Table IV, we have compared our results with some previoud methods, and from what we observe that our method has higher in accuracy rate than most of previous works.

V. DISCUSSION

In conclusion, an IoT-based smart parking system has proven that it is more efficient than handwritten tickets. By integrating advanced technologies such as real-time license plate recognition. automated entry and exit logging, cloud-based data management—this system might minimize human error and increase the overall progress. This shift towards automation not only increases operational efficiency but also paves the way for smarter, more scalable urban parking solutions. Moreover, because the small scale of the system so the cost expension is also one of the main advantages due to many simple sensors and hardware components that we are using in this project including motion detect sensor and Raspberry Pi 4, etc. This smart parking system, with the weaknesses mentioned above, will need many major upgrades in the future. For example, the mobile application will be upgraded in terms of the user interface, which can attract more customers, or the function of booking parking spaces before arriving will be upgraded in the near future, helping to minimize the possibility of stopping in the parking lot, leading to the congestion. Finally is the improvement of detecting and recognizing license plates. We always try to train more with various datasets and using different types of models to find out which model serves the best performance among them. The result is that from there we can minimize all the common errors that happen to previous systems that have been installed.

VI. ACKNOWLEDGMENT

We gratefully acknowledge the AIoT Lab Vietnam for their invaluable support in this research.

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.
[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
[3] Telefonnia I+D and the University of Cantabria, "Smartsantander: IoT experimentation over a smart city testbed," In: Computer Network, 2013, pp. 9–10.
[4] Pham Thanh Nam, Pham Xuan Kien, "Building a model of smart parking system in vietnam based on internet of things and cloud server technolies," In: TNU Journal of Science and Technology, 2021, pp. 201–206.
[5] Prabhu. Ramaswamy, "IoT smart parking system for reducing green house gas emission," In: International Conference on Recent Trends in Information Technology (ICRTIT) (IEEE), 2016, pp. 1–6.
[6] V. P. V. K. Rishabh, Pushpendra Kumar, "A novel car license plate and parking slot detection approach based on yolo,"
[7] L. A. R. G. Alessandro Florisa, Simone Porcua, "A social iot-based platform for the deployment of a smart parking solution," 2021.
[8] M. Y. A. A. A. S. Awad Alharbi, George Halikias, "Web-based framework for smart parking system," 2021
[9] Mai Chi Bao, https://github.com/mrzaizai2k/VIETNAMESE_LICENSE_PLATE
[10] Tran Duc Duan, Tran Le Hong D, Tran Vinh Phuoc, Nguyen Viet Hoang ,"Building an Automatic Vehicle License-Plate Recognition System", 2005.
[11] Quang Huy Che, Tung Do Thanh, Cuong Truong Van ,"Character Time-series Matching For Robust License Plate Recognition", 2023, https://doi.org/10.1109/MAPR56351.2022.9924897.
[12] Vinh Du Mai, Duoqian Miao and Ruizhi Wang ,"Vietnam License Plate Recognition System based on Edge Detection and Neural Networks", 2013.
[13] Zhenxue Chen, Guoyou Wang, Jianguo Liu and Chenyun Liu, "Automatic License Plate Location and Recognition Based on Feature Salience", International Journal of Cumputational Cognition 5, pp. 1-9, 2007.
[14] Sergey Zherzdev, Alexey Gruzdev, "LPRNet: License Plate Recognition via Deep Neural Networks", pp. 1-6, 2018, https://doi.org/10.48550/arXiv.1806.10447.
[15] Siti Norul Huda Sheikh Abdullah, Marzuki Khalid and Rubiyah Yusof, "License Plate Recognition using Multi-cluster and Multilayer Neural Networks", pp.1818-1823, 2006.

EVALLOS: An Effective Solution for CLO and SLO Assessment, Supporting Enhanced Teaching Quality

1st Duc Dat Pham*
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
pddat@ai4ia.cc

2nd Mai Thanh Nguyen Quynh
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
ngquynhmthanh@gmail.com
(Corresponding author)

3rd Ngoc Tram Huynh Thi
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
ngoctram.bie.u@gmai.com

4th Tien Phat Le
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
ltphat.developer@gmail.com

5th Thien Huong Nguyen Trang
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
huongy138@gmail.com

6th Nhan Pham Le Thanh
International University, VNU-HCM
Ho Chi Minh City, Vietnam
pltnhan07@gmail.com

***Abstract*—This study introduces EVALLOS, a sophisticated platform aimed at enhancing the evaluation of Course Learning Outcomes (CLOs) and Student Learning Outcomes (SLOs). EVALLOS addresses the shortcomings of conventional systems such as QOBE and TDTU by incorporating both quantitative assessments and qualitative rubrics, thereby providing a more comprehensive evaluation of student performance. The platform monitors academic advancement and offers insights into soft skills, including presentations and teamwork, rendering it an effective instrument for improving teaching methodologies and refining curriculum development. EVALLOS guarantees that educational outcomes are meticulously linked with program objectives, hence greatly enhancing the ongoing improvement of educational quality.**

***Index Terms*—EVALLOS, Outcome-Based Education (OBE), Learning Management Systems (LMS), Course Learning Outcomes (CLOs), Student Learning Outcomes (SLOs), Bloom's Taxonomy, rubrics, educational quality, learning analytics**

I. Introduction

In Vietnam, higher education institutions must design Student Learning Outcomes (SLO) to ensure the standards of input and output for training programs. Circular 17/2021/TT-BGDDT [1] stipulates that institutions must explicitly define SLO to demonstrate that students possess the requisite knowledge and abilities to meet employment demands. Numerous educational institutions assess these Course Learning Outcomes (CLOs) by analysing examination outcomes of students.

Nevertheless, dependence exclusively on CLOs data cannot comprehensively represent skill advancement and overarching program goals of students, as SLOs encompass abilities and competences that extend beyond the parameters of courses. The 2022 report from the Ministry of Education and Training indicates that there are 330 higher education institutions and teacher training colleges; however, only 207 of these schools were acknowledged as reaching quality requirements as of October 2023 (Figure 1) [2]. This indicates the necessity for a more synchronised and efficient quality evaluation system.

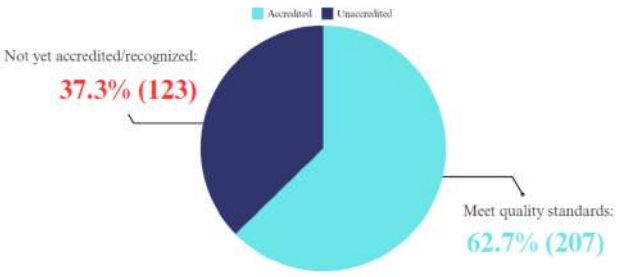


Fig. 1. Higher education institutions accreditation rates for educational quality.

To enhance training quality, a more thorough assessment approach for SLOs is required, integrating quantitative metrics with qualitative techniques such as field observations, learning portfolios, and stakeholder input. This strategy enables schools to have a greater understanding of student capacity, therefore enhancing the curriculum and assessment techniques.

The current issue is to create an integrated system that automates the collecting, processing, and reporting of SLO assessment data, thereby assisting institutions in comprehending the extent of completion of learning objectives. According to the theoretical framework established by Kristianto (Figure 2) [3], EVALLOS aims to enhance the assessment process by closely aligning Course Learning Outcomes (CLOs) with Student Learning Outcomes (SLOs) to elevate the quality of

instruction and training.

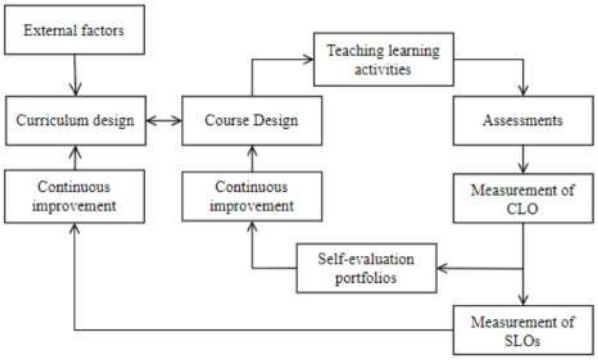


Fig. 2. Learning outcome flow.

II. BACKGROUND AND RELATED WORKS

Contemporary course output management and evaluation methods have substantially enhanced learning management in educational institutions. Two notable systems in Vietnam are the Tonne Duc Thang University (TDTU) system and the SHub Classroom.

The Ton Duc Thang University (TDTU) [4] system focuses mainly on score management, academic violation management and secure data storage. The system allows lecturers to directly link test questions to Course Learning Outcomes (CLOs), thereby improving the effectiveness of assessing course output standards. This helps ensure that the set learning objectives are closely monitored through student test results, providing a more comprehensive view of teaching effectiveness. In addition, the system also has flexible features in controlling student violations and ensuring safety during the exam organization process. These features help to strictly manage violations, thereby creating a transparent and fair learning environment. In addition, data on scores and exam structures are securely stored, helping to protect sensitive information and maintain the reliability of the learning management system at TDTU.

In exchange, SHub Classroom [5] provides tools for score analysis, management of scores for each question, and categorisation of student performance. This enables educators to formulate improved teaching tactics based on the study of student data gathered from the classroom. Nonetheless, many existing systems include certain limitations; for instance, the user interface may lack user-friendliness, or there may be redundant features that are cumbersome to operate.

Each system possesses distinct capabilities for managing and evaluating CLOs, however both exhibit deficiencies in the thorough analysis of SLOs. Simultaneously, these systems lack the capacity for rubric-based evaluation of essay-type questions, so constraining the ability to accurately represent competences and skills of students. Conversely, these limits are addressed by the evaluation and assessment component of EVALLOS, where Student Learning Outcomes (SLO) are evaluated, and rubrics for essay questions are employed, hence improving assessment efficiency and offering greater insight into the advancement of the learning program.

III. METHODOLOGY

The evaluation technique in EVALLOS aims to deliver a thorough overview of the extent of fulfilment of Course Learning Outcomes and to guarantee precise alignment with Student Learning Outcomes. The system employs two primary approaches for evaluation: question score-based assessment and rubric-based assessment. The score-based system distinctly measures students' comprehension and proficiency in specific learning materials. Conversely, rubric-based evaluation offers a more comprehensive methodology, demonstrating the capacity to utilise knowledge, engage in analytical reasoning, and fulfil other requirements. The integration of these two methodologies guarantees the precision and thoroughness of CLO assessments while facilitating the effective evaluation of SLO achievements, hence enhancing curriculum development and overall training quality.

A. Stakeholder Requirement Analysis

The requirements analysis in EVALLOS emphasises addressing the demands of principal stakeholders, such as the Board of Directors (BODs), Department Chairs, Quality Assurance Centres (QACs), and Lecturers (Figure 3). Board of Directors necessitate comprehensive analytical studies on educational outcomes to inform strategic decisions and modify training programs. Department Chairs, responsible for overseeing courses and faculty, necessitate a system that facilitates the creation of CLO frameworks, assesses learning outcomes, and enhances the curriculum. The QAC requires instruments for quality monitoring and learning outcome analysis to ensure the constant maintenance of educational standards. Lecturers want a straightforward and effective platform to input test structures, assign grades, and monitor student progress, consequently enhancing teaching methodologies through precise data analysis.

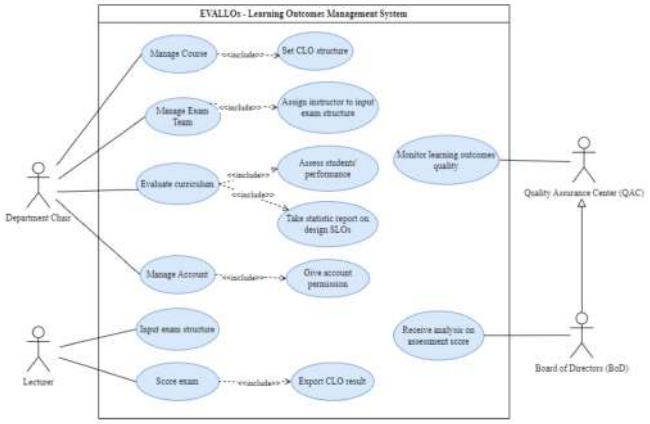


Fig. 3. Use case diagram.

B. Methodology for Calculating CLOs Achievement

*1) **Calculation Process for CLO Results**:* The methodology for assessing the completion level of CLOs in EVALLOS (Figure 4) is formulated in accordance with the regulations set forth in the curriculum of International University - VNU-HCM [6]. The system employs quantitative analysis techniques to precisely evaluate learning performance by comparing student outcomes with CLOs through a well-defined matrix, akin to the methodology outlined in Abeywardena's research [7] on outcome-based learning systems.

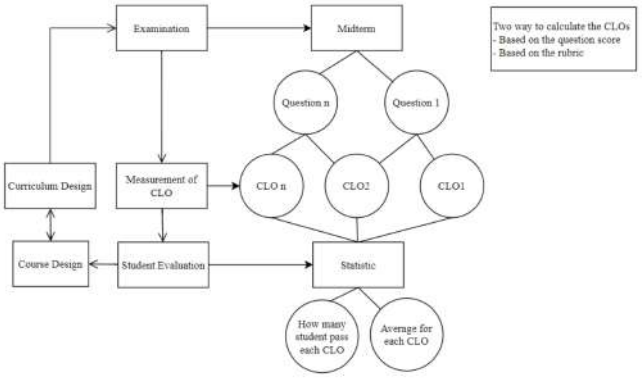


Fig. 4. Process to calculate the CLOs result in EVALLOS.

In EVALLOS, the analytical procedure entails mapping the test responses and scores to the designated CLOs via the question matrix. The data is subsequently pooled, and the pass rate for each CLO is computed using the following formulas and matrix between assessment (Question) CLOs (Table I):

$$\text{AverageCLOs} = \frac{\sum_{i=1}^{n} \text{No. of Student Pass Question}}{n} \quad (1)$$

$$\%\text{CLOs} = \frac{\text{AverageCLOs} \times 100\%}{\text{No. of Student In Exam Team}} \quad (2)$$

TABLE I
MATRIX BETWEEN ASSESSMENT (QUESTIONS) AND CLOS

CLOs	Assessment Methods	Assessment Questions
CLOs1	Exam Type	Question 1 (Q1) Question 2 (Q2) Question 3 (Q3) Question 4 - Rubric 1 (Q4.1) Question 4 - Rubric 2 (Q4.2) Question 4 - Rubric 3 (Q4.3)
CLOs2	Exam Type	Question 1 (Q1) Question 4 - Rubric 1 (Q4.1) Question 4 - Rubric 2 (Q4.2) Question 4 - Rubric 3 (Q4.3)
CLOs3	Exam Type	Question 1 (Q1) Question 2 (Q2)
CLOsn	...	Question n (Qn) Question x - Rubric y (Qx.y)

The quantity of students who successfully answer each question linked to a CLO is determined by the number of students passing the question, while n denotes the total number of questions related to the CLO. The term "number of students in exam team" indicates the overall number of students in the examination group. By employing these formulas, EVALLOS may ascertain the completion rate of each CLO within test groups, thereby producing precise and ongoing data on the attainment of learning objectives, facilitating the effective assessment and enhancement of the curriculum.

*2) **Rubric-Based Assessment**:* The EVALLOS rubric assessment approach enables educators to establish comprehensive evaluation criteria to assess student achievement. In contrast to conventional assessment approaches that depend exclusively on numeric scores, rubrics offer greater flexibility and comprehensiveness in evaluating students' skills and knowledge. The rubrics in EVALLOS are organised according to Bloom's Taxonomy [8], which assesses cognitive abilities across six tiers: Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation, ranging from basic recall to advanced skills. For instance, rubrics can be established at various levels: 0-49% signifies 'Not yet', 50-59% indicates 'Pass', 60-79% represents 'Good', and 80-100% denotes 'Excellent'. Figure 5 delineates an example of conventional rubric framework:

Criteria	INADEQUATE 0% – 49%	ADEQUATE 50% - 59%	GOOD 60% - 79%	EXCELLENT ≥ 80%
Originality and usefulness of the analysis (Criteria 1)	Demonstrates weaknesses in knowledge of theory and practice for this level. Limited understanding and application of concepts.	Demonstrates an adequate knowledge of theory and practice for this level. Some minor omissions. Satisfactorily interprets some appropriate concepts and theoretical models. Demonstrates some originality in conceptual understanding.	Demonstrates a sound knowledge of theory and practice for this level. Comprehensively interprets appropriate concepts and theoretical models. Demonstrates originality in conceptual understanding	Demonstrates an exceptional knowledge of theory and practice for this level. Insightfully interprets appropriate concepts and theoretical models. Demonstrates originality in conceptual understanding
Use of frameworks (Criteria 2)	Shows limited ability to structure problems in correspondence to theoretical frameworks	Shows effort to link problems with the theoretical frameworks. There are still some mistakes	Shows good ability to structure problems in correspondence to theoretical frameworks correctly. Minor mistakes in resolving problems	Shows excellent ability to structure problems in correspondence to theoretical frameworks correctly. The problems are well resolved
Quality of arguments (Criteria 3)	Severely limited arguments. Descriptive or narrative in style with no evidence of critique and originality or creativity. Conclusions are sparse.	Presents some critique of key research material resulting in original conclusions. Loss of focus in places. Demonstrates some creativity in solving complex problems. Produces satisfactory work that makes some contribution to the development of knowledge and understanding in the subject area	Presents a cohesive critique of key research material resulting in clear and original conclusions. Demonstrates creative solutions to complex problems. Produces superior work that makes a contribution to the development of knowledge and understanding in the subject area	Presents an exceptional critique of key research material resulting in clear, original and illuminating conclusions. Demonstrates distinctive, insightful and creative solutions to complex problems. Produces exceptional work that makes a contribution to the development of knowledge and understanding in the subject area.

Fig. 5. Rubric table example.

This approach is distinctive due to its applicability to assessments that are not just reliant on fixed quantitative values. This include skills evaluations including written assessments, presentation grading, and in-class examinations. These assessments frequently necessitate responses that extend beyond mere correctness, demanding students to exhibit analytical skills, originality, and critical thinking capabilities.

This offers a more profound understanding of student achievement than only an aggregate grade. This also pertains to formative assessments, including involvement and teamwork, transforming qualitative elements into quantifiable criteria. Consequently, rubrics facilitate continuous assessment, enabling educators to modify their instructional strategies to enhance the quality of training.

C. Methodology for Calculating SLOs Achievement

Numerous universities employ automated learning systems such as QOBE [9] to assess and monitor student progress under the Outcome-Based Education paradigm. Consequently, Bhatti et al. (2023) [9] assert that the QOBE method aids universities in evaluating Program Learning Outcomes (PLOs), also referred to as SLOs, by establishing connections with Course Learning Outcomes (CLOs). Consequently, QOBE offers many quantitative methodologies, including tests, assignments, and alternative assessment formats, to assist institutions in overseeing the evaluation process.

D. *CLOs and PLOs matrix in IJITL*

The QOBE system [10] is important in offering essential parameters to guarantee meeting of PLOs at the subject and program levels as reflected by key performance indicators (KPIs). An indication of PLOs achievement is arrived at by direct means of assessment namely end-term test scores, scoring on mid-term tests and problem solving assignments. This information is then compiled to produce more specific reports on performance from every student in order for schools to sustain and disseminate effective teaching.

A study published in the International Journal of Innovation in Teaching and Learning (IJITL) [9] demonstrated that the QOBE system utilised a matrix to connect Course Learning Outcomes (CLOs) and Program Learning Outcomes (PLOs) to illustrate the evaluation process. The following matrix delineates the relationship between CLOs and PLOs in IJITL, demonstrating the contribution of each CLO to the respective PLO and the employed assessment method in Figure 6:

The table illustrating the association between CLOs and PLOs demonstrates the completion rate of PLOs based on scores obtained from examinations and assignments. Likewise, EVALLOS employs a similar approach of mapping CLOs and SLOs to assess student learning outcomes.

1) ***Formula for Determining Student Learning Outcomes in EVALLOS****:* In the EVALLOS system, the calculation of SLOs is not only based on scores from CLOs but also incorporates qualitative factors through rubrics to provide a more comprehensive assessment of students' skills and knowledge. In the EVALLOS system, each SLO is designed to capture the overall learning goals of a program, which span various cognitive domains from Bloom's Taxonomy [8]. As CLOs are mapped to SLOs, the system tracks how students progress through each level of learning, from basic knowledge acquisition (Knowledge, Comprehension, Application) to higher-order thinking skills (Analysis, Synthesis, and Evaluation). This allows the assessment of SLOs to reflect not only the accumulation of factual knowledge but also the development of critical thinking, problem-solving, and creativity in students. For each CLO, the number of students who meet the requirements is calculated based on their performance in exams, assignments, and rubrics. After collecting the data, EVALLOS calculates the completion rate of the SLOs using the following formulas:

$$\text{AverageSLOs} = \frac{\sum_{i=1}^{n} \text{No. of Student Pass CLOs}_i}{n}$$

$$\%\text{SLOs} = \frac{\text{AverageSLOs} \times 100\%}{\text{No. of Students In Program}}$$

These formulas ensure that the system accurately calculates the percentage of students who have successfully completed the linked CLOs, providing insights into the overall achievement of the SLOs in the program.

2) ***Linkage matrix between SLOs and CLOs in EVALLOS****:* A crucial component of the EVALLOS assessment process is the correlation of Student Learning Outcomes (SLOs) to Course Learning Outcomes (CLOs) across all disciplines. The matrix below demonstrates the contribution of CLOs from each subject to the SLOs within the curriculum in Table II:

TABLE II
MATRIX BETWEEN SLOS AND CLOS IN EVALLOS

CLOs	SLO1	SLO2	SLO3	SLO4
CLO1	X			X
CLO2		X	X	
CLO3			X	X
CLO4	X			
CLO5		X		X

This matrix demonstrates the correlation between CLOs from various topics and SLOs, facilitating the tracking of student progress and the evaluation of learning outcomes in relation to program learning outcomes for instructors and administrators.

IV. IMPLEMENTATION AND RESULTS

This part presents the evaluation results of the EVALLOS system, encompassing the computation of Course Learning Outcomes (CLOs), the relevant assessment configurations, the connection matrix between Student Learning Outcomes (SLOs) and CLOs, and the conclusive SLO evaluation results.

A. Course Learning Outcomes Evaluation Results

The figure 7 below depicts the CLO assessment outcomes of a exam group according to question scores and rubrics. This outcome enables teachers to track student completion levels for each given criterion, therefore facilitating a more precise evaluation of learning performance.

B. Actual SLO x CLO Linkage Matrix

The matrix below demonstrates the correlation between Student Learning Outcomes (SLOs) and Course Learning Outcomes (CLOs) in a training program, facilitating the identification of each module's contribution to the overarching learning objectives. This matrix in Figure 8 elucidates the correlation between SLOs and CLOs, guaranteeing that each CLO contributes to one or more SLOs, thereby offering a holistic perspective on the fulfilment of learning objectives.

CLO	PLO	CLO Weight * credit	Assessment Methods				Marks obtained in CLO
			Midterm Marks	Test Marks	Problem-based learning assignment	Final Exam marks	
CLO1	PLO1	0.25 x 3 = 0.75	16/20= 0.8	3/5= 0.6			19/25= 0.76
CLO2	PLO4	0.34 x 3 =1.02		4/5= 0.8	4/5=0.8	20/24= 0.83	28/34= 0.82
CLO3	PLO3	0.41 x 3 =1.23			4/5=0.8	30/36= 0.83	34/41=0.83
Total		3	Weighted Average				81%

Fig. 6. CLOs and PLOs matrix in IJITL.

CLOs Result

Question/CLO	CLO1	CLO2
q1	4	4
q2	4	
q3		5
q4	6	
Average Count	5.00	5.00
Percentage	83.33%	83.33%

Fig. 7. Course Learning Outcomes Result

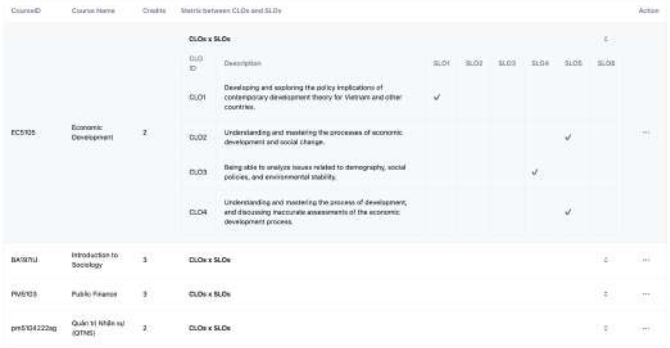


Fig. 8. CLO x SLO in actual program.

C. SLO Assessment Results

The EVALLOS system computes SLO completion outcomes based on data gathered from CLOs and rubrics. This information offers schools a comprehensive overview of progress and capacity of student to meet learning objectives, thereby enhancing training programs and ensuring educational quality.

V. Evaluation

This section presents a comprehensive comparison examination of the properties of three assessment systems: EVALLOS, TDTU, and QOBE. The objective is to elucidate the strengths, shortcomings, and enhancements of each system in evaluating Course Outcomes (CLOs) and Student Outcomes (SLOs).

A. Features of QOBE system

A collection of indicators, referred to as the QOBE (Outcome-Based Education) system [10], aims to assess the learning results of engineering education. It primarily focusses on formative and summative assessment methodologies; the former entails the aggregation of score grades from tests, assignments, and similar evaluations. QOBE, owing to its ability to collect and analyse quantitative data, can assist educational institutions in delineating student learning outcomes. Furthermore, it facilitates the alignment of CLOs and SLOs in a manner that fosters accountability for program outcomes and ensures the attainment of the program's intended objectives.

However, QOBE also has its limitations. The system lacks methods for analysing and assessing results using rubrics, hence diminishing the capacity to evaluate students' soft skills or non-quantitative assessments. The limitation arises from the QOBE's exclusive reliance on quantitative data, which fails to provide insight into students' qualitative learning outcomes related to presentations, teamwork, and various experiential learning aspects, all of which are essential elements of the educational process.

B. Features of TDTU system

The learning management system of Ton Duc Thang University (TDTU) [4] primarily emphasises score management and monitoring student infractions. The system is esteemed for assisting professors in inputting and maintaining examination data, in addition to establishing and monitoring course frameworks. These capabilities enable instructors to efficiently execute academic management activities, hence enhancing the teaching and assessment process.

Nonetheless, the methodology of TDTU possesses several drawbacks. Like QOBE, the system lacks support for qualitative assessment instruments such as rubrics, hence constraining the evaluation of students' soft skills. Furthermore, TDTU lacks comprehensive assessment customisation features, such as the ability to filter data by course or major, rendering its assessment capabilities less adaptable compared to contemporary systems.

C. Features of the EVALLOS system

Flexibility in evaluation refers to EVALLOS's ability to adapt to diverse assessment needs. It allows customization of evaluation settings, such as unique rubrics for different student groups, and integrates quantitative assessments (e.g., exam scores) with qualitative evaluations (e.g., teamwork and presentations). Advanced filtering enables educators to analyze results by cohort or course, providing precise insights for targeted curriculum improvements and a holistic view of student performance.

The EVALLOS system was created to address the shortcomings of both QOBE and TDTU, offering a more thorough evaluation framework for CLOs and SLOs. The primary strength of EVALLOS is in its capacity to integrate quantitative and qualitative evaluations. The approach incorporates rubrics alongside test scores to deliver comprehensive evaluations of student competencies, encompassing both soft skills and tasks such as presentations, essay composition, and collaboration. EVALLOS provides significant flexibility by enabling tailored evaluations for certain student cohorts, permitting instructors to filter by course, major, or student groups for analysis and targeted assistance. The solution establishes a robust connection between CLOs and SLOs, akin to QOBE, yet offers enhanced analytical capabilities through the integration of both quantitative and qualitative data. Nonetheless, the extensive customisation options may result in a more protracted initial setup and acclimatisation period compared to more straightforward systems such as QOBE and TDTU.

D. Comparison table with Other Platform

TABLE III
COMPARISON OF QOBE, TDTU, AND EVALLOS SYSTEMS

Features	QOBE	TDTU	EVALLOS
CLOs and Assessment Mapping	Yes	Yes	**Yes**
Report Score Achievement	Yes	Yes	**Yes**
Quantitative Assessment	No	No	**Yes**
Use Rubric/Criteria Standard for Evaluation	No	No	**Yes**
CLO-to-SLO Linkage	Yes	No	**Yes**
Violation Tracking	No	Yes	**Yes**
Flexibility in Evaluation	Limited	Limited	**High**

EVALLOS outperforms QOBE and TDTU in the assessment process, as evidenced by the comparison table III. The customisation option facilitates detailed assessment for a specific set of pupils, while the rubrics assist EVALLOS in evaluating soft skills. EVALLOS integrates quantitative and qualitative assessments, serving not only as a tool for monitoring learning progress but also providing a comprehensive study of the assessment process to enhance it, so improving the quality of teaching and training.

E. Practical Use Case of EVALLOS

EVALLOS simplifies inputting, analyzing, and retrieving data through a user-friendly interface with features like automated CLO-SLO mapping and interactive dashboards. By automating tasks such as report generation, it reduces errors, saves time, and lowers costs, cutting evaluation time by 40% in some cases. For example, in a business communication course, EVALLOS efficiently assessed teamwork, presentations, and exam scores, enabling precise and targeted feedback.

VI. CONCLUSION

EVALLOS surpasses existing systems like SHub Classroom and Blackboard by integrating quantitative and qualitative evaluations via rubrics, while facilitating a strong connection between Course Learning Outcomes (CLOs) and Student Learning Outcomes (SLOs). The system offers statistical analysis capabilities for Program Learning Outcomes (PLOs), ensuring flexibility and thoroughness in evaluating student learning outcomes.

EVALLOS enhances educational quality through a scientific evaluation methodology that connects Course Learning Outcomes (CLOs) with Student Learning Outcomes (SLOs). The approach enables educators to thoroughly evaluate learning progress, thereby enhancing the curriculum and ensuring pupils attain essential skills.

In the future, EVALLOS will integrate AI to enhance teaching methodologies and provide data-driven recommendations for training programs. Collaborating with universities, it aims to refine its offerings and expand to leading educational and accreditation institutions.

ACKNOWLEDGMENT

The authors would like to express their gratitude to AIoT Lab VN for the support throughout this project. Special thanks to Dr. Le Duy Tan and Associate Professor Dr. Huynh Kha Tu for their valuable guidance and contributions.

REFERENCES

[1] "How are the input and output standards of higher education training programs regulated?," 2023. Accessed: 2024-06-17.
[2] M. Giang, "10 years of higher education quality assurance in Vietnam," Tuoi Tre, November 2023.
[3] H. Kristianto, S. Prasetyo, R. Susanti, and M. Adithia, "Design of student and course learning outcomes measurement," JPI (Jurnal Pendidikan Indonesia), vol. 10, March 2021.
[4] TDTU, T. D. (n.d.). Ton Duc Thang University Student Portal. Retrieved from https://old-stdportal.tdtu.edu.vn/
[5] "Shub education portal," 2024. Accessed: 2024-06-17. https://shub.edu.vn/
[6] "Decision on promulgating the regulation on tertiary education training according to the credit system at the University of Science Ho Chi Minh City belonging to Vietnam National University Ho Chi Minh City," 2021.
[7] I. Abeywardena, "Mastery of course learning outcomes in odl: A case study of the Pearson eCollege Learning Outcome Manager," *Open Praxis*, vol. 5, pp. 239–248, Sep. 2013.
[8] N. E. Adams, "Bloom's taxonomy of cognitive learning objectives," J Med Libr Assoc., vol. 103, no. 3, pp. 152–153, Jul. 2015, doi: 10.3163/1536-5050.103.3.010.
[9] S. Bhatti, M. Memon, and A. Meghji, "Scrutinizing outcome assessment of outcome-based education using q-obe in engineering education," *International Journal of Innovation in Teaching and Learning (IJITL)*, vol. 9, pp. 55–74, 2023.
[10] L. K. Bansal, P. K. Mishra, and S. Sachdeva, "Creation of an obe model po attainment tool," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 10, pp. 4427–4431, 2021.

Tactiwave: AI-Driven Tactile Communication for Deaf-Blind Support

1st My Nguyen Huynh Thao
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
thaomypc2003@gmail.com

2nd Nam Anh Dang Nguyen
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
dnnanh@ai4ia.cc
(Corresponding author)

3rd Duc Dat Pham
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
pddat@ai4ia.cc
(Corresponding author)

4th Nguyen Binh Nguyen Le
AIoT Lab VN
University of Huddersfield
United Kingdom
nguyenbinhnguyen.le@hud.ac.uk

***Abstract*—TactiWave is an assistive device designed to enhance communication and safety for individuals with hearing and visual impairments. Using the Lorm alphabet, it translates spoken language into tactile vibrations, allowing users to understand information through touch. The device features an ESP32 CPU, LRA motor, and HC-05 Bluetooth for instant, customizable feedback. A mobile app enables users to adjust vibration intensity, while AI-driven voice recognition and emotion detection trigger emergency alerts. User acceptance testing (UAT) results confirm TactiWave's accuracy in identifying tactile feedback, improving daily communication, and increasing user safety during emergencies. These results highlight its potential to significantly enhance quality of life and independence for those who disabled.**

***Index Terms*—TactiWave, Lorm alphabet, tactile communication, deafblind support, haptic feedback, assistive technology**

I. Introduction

A. Challenges faced by individuals who are both deaf and blind

Deafblind individuals face significant daily challenges. According to statistics from the National Center on Deafblindness, approximately 10,000 children and youth in the US are identified as deafblind [1]. On a broader scope, the World Federation of the Deafblind (WFDB), an international organization that represents and protects the rights and needs of deaf-blind people stated that there are currently about 0.2% to 2% of the population belonging to the deaf-blind group [2], equivalent to about 16 million to 160 million people worldwide.

B. Bridging the communication gap through tactile feedback

Because of these disabilities, they encounter numerous challenges, including restricted access to language, social isolation, and a significant lack of educational resources. Hence, they tend to live in difficult economic conditions, with higher unemployment rates and lower educational attainment than others. Unlike individuals with other disabilities, such as those who are deaf or blind, who have established methods to communicate—like Braille for the blind or sign language for the deaf. Meanwhile, deafblind individuals face far greater challenges. Effective communication solutions for these individuals remain limited, and no universal method has been widely adopted to bridge their unique needs.

The other current methods include the tactile sign, which transfers information by sensations on the palm, and the disabled may feel it via physical contact. This method indeed is presumed to be the superior way today; nevertheless, it is complex to learn and during emergencies, it takes time. This research addresses these drawbacks through the proposal of a new automated method assisted by the strong tactile signals for adding more features, meanwhile being aimed at enhancing the capacities for communication and independence of individuals with disabilities. However, there remain other methods of communication and we will discuss them below.

II. Literature Review

Concurrently, multiple studies have been conducted to reduce the social isolation of deafblind individuals. Modern assistive technology such as devices that support daily conversation via Braille, some advanced touch systems, etc. have gradually been supported to enhance the ability to communicate and the life qualification of the disabled.

For details, a project by Sylvie Cantin and colleagues was executed to assist deafblindness by Braille [3]. The project developed a Braille note-taking device that wirelessly connects to an iPhone via Bluetooth, using VoiceOver and the HWCom application to facilitate communication for deafblind users. VoiceOver translates text into Braille, allowing users to read and understand messages, while also managing conversations in a sequential flow, enabling users to send and receive messages with ease alternately. This method is slow, requiring users to wait and swipe to understand messages, which is impractical in emergencies. Additionally, it depends on multiple devices, complicating usage.

Another approach by Muhammad Azhar Hussain is to use smartphones and smartwatches to recognize Arabic phrases uttered by the Imam and transmit this information to deaf and blind people through vibration patterns [4]. These vibration patterns are designed to vary in intensity, frequency, and duration so that users can distinguish different stages of prayer. This project aims to assist deaf and blind people in performing Namaz in the community by using speech recognition and vibration stimulation. However, this project still has many limitations, such as the limited amount of information that can be conveyed. Users can only perceive simple signals of vibration intensity and duration, while the stages of prayer may require more detailed information to ensure accuracy. Furthermore, the scope of interaction of this research is only for prayer. Therefore, the goal of ensuring independence and non-dependence for people with disabilities is still not resolved.

Therefore, the aim of this paper is to build a vibrating hand system with integrated sound recognition that can support deaf and blind people in a variety of situations, with the ability to interact quickly and effectively. This system not only helps users monitor conversations or surrounding situations through vibration signals but also can respond instantly in real-time. The ultimate goal of the system is to ensure the safety of users in emergencies, helping them identify dangers and react promptly to protect themselves.

III. MATERIALS AND METHODS

Tactiwave employs the Lorm alphabet and integrates ESP32, LRA motors, and HC-05 Bluetooth for real-time tactile communication. Its mobile app allows customization, while AI processes speech-to-text and detects emotions for emergency alerts.

A. Lorm Alphabet Interaction

Lorm Alphabet [5] is a method of conveying letters by touching specific areas of the listener's hand with designated fingers in Figure 1. This technology is implemented using the Tactiwave gadget, which converts finger movements in the Lorm alphabet into tactile information that can be easily perceived by the palm of the hand, regarding the surrounding environment. Each letter will include a distinct vibration pattern, enabling individuals with disabilities to receive and comprehend language through tactile means.

Consequently, the gesture-to-vibration mapping technique is proposed and developed based on the user's hand skeletal anatomy. The letter A is articulated at the tip of the thumb with a gentle touch, while the letter B involves a movement across the index finger. The aforementioned gestures will be converted by the device into vibration patterns of varying amplitude and frequency. The vibrated feedback system, as illustrated in the image below, provides a method of conveying messages that ensures user comfort while facilitating the easy detection of significant letters and signals. The image below illustrates how the gadget conveys these motions to the vibration state, enabling the user to touch and see information.

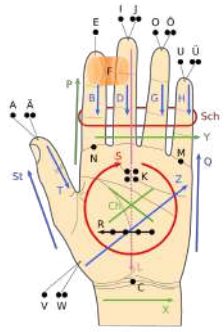


Fig. 1. The Lorm Alphabet [5].

B. System Requirements: Hardware and Software Integration

Function Requirements The Tactiwave device can transform audio impulses into vibrational feedback, enabling individuals who are deaf and blind to perceive and comprehend spoken words through distinct vibration patterns. Furthermore, via the mobile application, users can tailor the intensity and rhythm of the vibration, providing suitable feedback for each category of language or alert, enhancing the distinctiveness and significance of the information. Specifically, in emergency scenarios, the technology will translate spoken English into a distinct vibration pattern signifying danger with heightened intensity, facilitating prompt user response and ensuring safety. The mobile application features an intuitive interface, enabling users to customize and operate the device to suit their individual requirements, thereby delivering a thorough and convenient user experience.

Non-funtion Requirements The system must possess an intuitive and user-friendly interface design that facilitates seamless navigation and interaction for users, particularly those with sensory impairments, on both the device and mobile application. The device's feedback must be prompt and instantaneous, enabling users to swiftly comprehend and react to signals or warnings during emergencies. Moreover, the system must convert spoken language signals into tactile sensations instantaneously with minimal latency, guaranteeing accurate and timely communication of information. The transformation of audio inputs into vibrational patterns must be refined to effectively convey the subtleties and emotions in speech, ensuring clear and comprehensible feedback for users.

C. Hardware Architecture Design

1) Hardware design: The hardware system is engineered to establish an efficient and dependable vibrated feedback system, utilizing the ESP32 microcontroller as the primary processor and the Raspberry Pi 4 to facilitate the execution of more intricate tasks. The system incorporates LRA Coin Motors vibration motors for accurate vibrated feedback and utilizes the HC-05 Bluetooth module for wireless connectivity to external devices (Figure 2).

The procedure commences when the ESP32 receives signals from the Raspberry Pi 4, subsequently controlling the LRA vibration motors to generate suitable vibration feedback on the



Fig. 2. Hardware System of Tactiwave.

user's palm. Each vibration motor is configured to deliver tailored vibrated input based on the user's current circumstances. The device provides rapid and precise feedback via various vibration patterns, enabling users to perceive and comprehend tactile information distinctly and efficiently.

The ESP32 microcontroller [6] serves as the central processor of the system, managing computational tasks and coordinating communication between components, supporting multiple communication protocols such as Wi-Fi and Bluetooth to ensure seamless integration. The **vibrated feedback system** [7], utilizing LRA vibration motors controlled by PWM signals from the ESP32, provides precise tactile feedback that users can customize through an application. The **HC-05 Bluetooth module** [8] facilitates wireless connectivity between the ESP32 and mobile devices, enabling remote control and effective system monitoring. Additionally, the **Raspberry Pi 4** [9] handles more complex tasks, such as data analysis and machine learning algorithm execution, expanding the system's processing capabilities and ensuring efficient information management.

2) Device on hand: The gadget is intended for hand use, featuring the controller positioned on the dorsal side and the vibration motors situated on the palmar surface, as illustrated in Figure 2. Upon activation, the user will experience a sequence of subtle vibrations, accompanied by the illumination of the indicator lights. The red light signifies that the gadget possesses sufficient power and is prepared for stable operation, whilst the green light blinks in response to signals from the mobile application. The device is engineered for user sensitivity and responsiveness via vibrations regulated through the application.

D. Software Design

1) System Operation flow chart: The system initiates upon user input, either vocal or command-based, which is thereafter transmitted to the Speech Recognition module for processing. The system initially verifies whether the input constitutes an emergency circumstance. Should this occur, it activates an alert accompanied by a sequence of intense vibrations to notify the user. In non-emergency situations, the input is assessed and transformed into tactile signals, such as vibration patterns that correspond to letters or established sequences. Upon completion, the system transmits vibrational feedback to the user, enabling them to perceive information through tactile

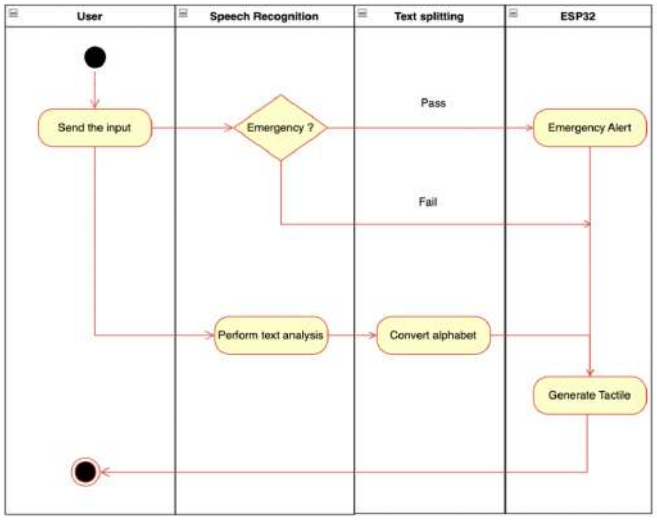


Fig. 3. Flow Chart.

sensation, whether it pertains to an emergency notification or a routine dialogue. Figure 3, distinctly depicts this process.

2) Applying AI for speech-to-text transformation and emotion detection: Utilizing AI and ML, one can conduct real-time narrative analysis, identifying emotions or searching for certain phrases in spoken language. This effectively renders the system intelligent, allowing responses to be tailored to users with specific needs, including disabilities. The device employs AI/ML for fixed voice signal processing, context, and intent recognition, answer customization, and interaction optimization.

Vibration Patterns According to the LORM alphabet, there are various actions to interact with the palm of the disabled person, such as short touch, long touch, or swipe. Therefore, to be able to reproduce these actions with the vibrating device, we need to design suitable vibration patterns by programming code and embed them into the device, specifically as shown in Table I.

Simple gestures like "Tap" use short pulses, while continuous gestures such as "Long Stroke" sustain vibrations. More complex gestures like "Circle", and "Oblique Stroke" use sequential node activation to create a particular sensation. Based on the vibration node matrix designed on the palm, it will vibrate in the appropriate positions to create the most realistic vibration feeling. Moreover, the default intensity will fluctuate between 80% to 90%. However, in emergencies, the intensity can be increased to 100-120%, providing stronger, more urgent feedback to alert the user.

Speech To Text [10]: Tictacwave use the speech-to-text functionality of the Python package 'speechrecognition' [11] to transcribe spoken language into text instantaneously. This library is straightforward to create and accommodates many speech recognition engines, facilitating efficient voice input processing. Upon user vocalization, the audio is recorded and analyzed by 'speechRecognition', subsequently converting it into textual data.

TABLE I
VIBRATION PATTERN FOR GESTURE SIMULATION ON PALM DEVICE

Gesture	Pulse Duration	Frequency/Pattern	Description
Tap	50 ms	N/A	Single pulse, long pause.
Double Tap	50 ms	N/A	Two pulses, short delay.
Short Stroke	200 ms	Continuous	Brief continuous vibration.
Long Stroke	1000 ms	Continuous	Extended continuous vibration.
Oblique Stroke	50 ms per node	Sequential node activation	Sequential activation in a cross pattern.
Drumming	100 ms	Repeated pulses	Rhythmic pulses with equal intervals.
Circle	50 ms per node	Sequential node activation	Sequential activation in a circular pattern.
Light Squeeze	1500 ms	Continuous	Low-intensity, long duration.

Emotion Detection [12]: The technology relies on an emotion detection model to enhance the identification of emergencies. The data for the model comes from the Toronto Emotional Speech Set (TESS) - an audio dataset of 2800 WAV files, in which two female actors (26 and 64 years old) pronounce 200 target words in the phrase "Say the word " with 7 different emotions such as happy, sad, fear, disgust, angry and pleasant. Our method relies on these emotions to analyze the emotions of the speaker when witnessing or wanting to inform someone about an emergency, which are fear, disgust, and anger. Then, the audio data will be analyzed to extract features through spectrograms and trained on multiple models to evaluate. From there, this feature not only provides information in emergencies but also lets people with disabilities know the feelings of the other person in daily conversation.

Emergency message recognition [13]: We used a combined dataset of over 22,000 sentences from the Kaggle Disaster Response Dataset, which includes warning messages in front of natural disasters, and 12,000 sentences from daily conversations on topics like weather and hobbies. Then, it is necessary to filter sentences including undefined and uncommon characters. This mix helps train the model to distinguish between urgent emergency alerts and regular communication. The method utilizes an additional model, Support Vector Machine (SVM), for the identification of emergency phrases in spoken language. After the transcription of speech to text, the SVM model executes feature extraction to facilitate the identification of predetermined keywords and patterns characteristic of emergencies, such as "Help!" or "I require help." Our SVM is trained on a diverse array of emergency and non-emergency terms, enabling it to determine whether the spoken content is a crisis. This enables the system to respond more swiftly to inform users or activate safety measures, hence providing trustworthy security during critical moments.

After the system collects and analyzes the speaker's emotional data, as well as the content of the conversation to determine whether it contains an emergency warning, the system will evaluate the input. Specifically, each sentence uttered by the person communicating with the disabled person will be classified to see if it is an urgent message, such as related to danger, natural disasters, or situations that require immediate response. If the system detects that this is an urgent message, it will automatically activate the emergency vibration mode on the device (which produces a stronger and clearer vibration than the normal vibration mode). This is to promptly warn the disabled person, helping them quickly recognize the dangerous situation and be able to respond appropriately, ensuring safety and increasing initiative in communication.



Fig. 4. Mobile interface application.

3) Application Interface Features: The app UI showcases a motor vibration range that varies from minimal to maximal based on user settings. Figure 4, illustrates that users may select among motors or alphabets to vibrate, facilitating daily discussions for the hearing-impaired and visually impaired. Should the device detect spoken words during a conversation, the vibrators will activate by the tone and rhythm of the language. Users can learn from previous experiences and familiarity with the vibrations, allowing them to interpret their meanings through the patterns of vibration. In the event of an emergency, the app will emit an audible tone, send an alert to the smartphone, and vibrate vigorously. Users can deactivate this warning mode by hitting the "Stop Vibration" button on the interface.

IV. EXPERIMENTS RESULTS AND DISCUSSIONS

A. Models evaluation

1) Alert content detection model: After data processing, only 6000 sentences were selected for each class (alert and non-alert) to optimize the model performance. Table II summarizes the model's classification performance for two types: non-alert content (class 0) and alert content (class 1). The

model achieved 95.7% accuracy. For non-alert content, the precision was 0.95 and recall was 0.97, indicating that the model recognized this type well. For alert content, the precision was 0.97 and the recall was 0.95, which shows a small proportion of emergencies are missed. This is an important indicator because, in warning systems, missing emergency messages can have serious consequences. Both classes achieved an F1-score of 0.96, demonstrating the ability to accurately classify between the two content types.

TABLE II
CLASSIFICATION REPORT

Class	Precision	Recall	F1-Score	Support
Non-alert content	0.95	0.97	0.96	1086
Alert content	0.97	0.95	0.96	1143
Accuracy	0.96			
Macro Avg	0.96	0.96	0.96	2229
Weighted Avg	0.96	0.96	0.96	2229

2) Emotional detection model: The table III shows that combined models such as CNN-LSTM and Hybrid CNN-GRU provide better performance than single models such as LSTM or GRU. In particular, Hybrid CNN-GRU achieves the highest accuracy, 67%, with a more optimal number of parameters. Therefore, it can be concluded that combining CNN with sequential models not only improves accuracy but also helps to optimize the number of parameters, resulting in better performance in processing sequential data such as audio signals.

TABLE III
COMPARISON OF MODELS FOR EMOTIONAL DETECTION

Model	Accuracy (%)	Parameters
LSTM	40	305,799
GRU	42	372,359
Bidirectional LSTM	44	602,759
CNN	47	164,871
CNN-LSTM	51	370,567
Hybrid CNN-GRU	67	363,015

B. Hardware Evaluating: Assessing design for deaf and blind people

The Tactiwave device's hardware design and characteristics are thoughtfully tailored for deaf and blind users, enabling them to engage with the system effectively. The dimensions, mass, and placement of the vibrators are established concerning hand size to guarantee ease of use.

1) Tactile communication: Research by Cantin et al. [14] demonstrates that tactile technologies effectively facilitate communication for individuals who are deaf and blind. Systems employing tactile vibrations facilitate the rapid and precise acquisition of information without reliance on auditory or visual senses. This aligns with Tactiwave's utilization of vibration patterns to communicate language and emergency notifications.

2) Optimize button layout: Design requirements for individuals with disabilities stipulate that the configuration of buttons on the device must guarantee that the buttons are sufficiently large, easily accessible, and that the spacing between buttons is ample enough for users to differentiate them without visual assistance [15] . The button layout design of Tactiwave fulfills this criterion, enabling blind users to operate the gadget solely through tactile interaction.

3) Vibration Feedback: A study conducted by Hussain et al. [16] showed that vibration feedback significantly aids users in confirming executed operations and enhances device usability for individuals with hearing and visual impairments. TactiWave utilizes vibrational feedback to provide users with immediate confirmation of commands or reactions, ensuring smooth and effective device control.

Building on this, we propose enhancements to the vibration feedback system to improve precision, realism, and overall user experience. Table IV outlines potential advancements, their benefits, challenges, and practicality for implementation.

By integrating these advancements, TactiWave aims to deliver more nuanced and effective feedback, particularly for complex vibration patterns like "Circle" and "Drumming." These improvements will enhance the overall user experience and ensure adaptability for various scenarios, including emergency alerts and routine communication. The feasibility of these enhancements will be further explored to maintain device affordability and usability.

C. Acceptance Testing for TactiWave Hardware Design

During User Acceptance Testing (UAT), we assessed the capacity of TactiWave to identify and distinguish vibrated motions. The movements were categorized from challenging to simple based on a scale of 1 to 5. We surveyed 30 normal people to have a general observation of the utilization of Tactiwave, figure 5 illustrates their distinct differentiation between simple movements. Based on the results of the graph, Tap and double-tap achieved the highest level of recognition from users. On the contrary, there is a need to find better methods to reproduce the difficult-to-recognize vibration patterns such as Circle, Drumming, and ShortStroke, due to the relatively low number of people who correctly recognize these patterns. One potential improvement is to apply haptic technology, a technology that not only uses vibrations but also forces to reproduce the sensation of touch on human skin. This will help users distinguish the vibration patterns more clearly and accurately, enhancing the tactile experience.

V. ETHICAL CONSIDERATIONS

A. Data Privacy and Security

All user data, including speech and tactile interaction logs, are processed locally on the device to ensure privacy. Encryption is applied for any data shared via Bluetooth or stored on external devices.

TABLE IV
PROPOSED ENHANCEMENTS TO VIBRATION FEEDBACK SYSTEM

Technology	Description	Advantages	Challenges	Practicality
Multiple Vibration Nodes	Adding more nodes across the hand	Better spatial resolution	Increased hardware complexity	Moderate
Pressure Feedback	Incorporating pressure sensitivity	Enhanced realism of patterns	Requires new actuators	High
Heat Variation	Utilizing temperature changes as cues	Additional sensory layer	Higher energy consumption	Low

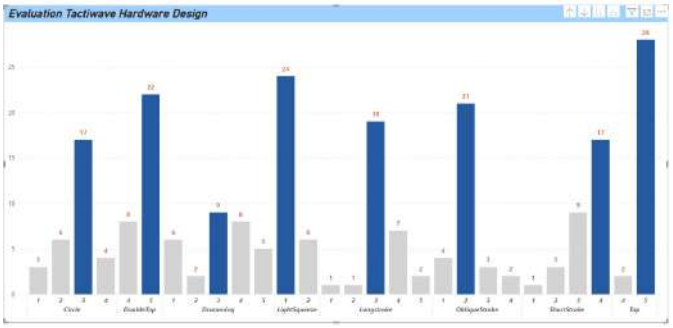


Fig. 5. User Recognition Rates for Tactiwave Vibration Patterns

B. Responsible AI Deployment

The AI models used in TactiWave are carefully designed to avoid biases. All training datasets are anonymized and sourced ethically, adhering to GDPR and other regional data protection laws. Regular audits ensure compliance and transparency.

VI. CONCLUSION

TactiWave demonstrates significant potential in improving communication and safety for individuals with hearing and visual impairments through tactile technology. By converting sound into vibrations on the skin, it enables users to perceive information typically acquired via auditory or visual senses, enhancing both daily communication and emergency responsiveness.

The system's versatility allows users to customize vibration modes and adapt the device to their unique needs via a mobile application, fostering independence and confidence. Its scalability supports additional features like context recognition and language translation, broadening its utility in various scenarios.

By integrating advanced technologies, TactiWave empowers the disabled community with improved self-defense and social inclusion, enhancing their quality of life and promoting independence.

ACKNOWLEDGMENT

The authors express deep gratitude to Dr. Vi Chi Thanh and Dr. Le Duy Tan for their invaluable guidance in the development of the TactiWave project, ensuring it met technical and practical needs. Their mentorship was crucial in refining AI-driven features and tactile feedback design. The project earned second place at the Swinburne Hackathon 2024, highlighting its innovation and potential impact on the deafblind community. The team thanks AIoT Lab VN and the Hackathon judges for their support and recognition, and looks forward to future collaboration.

REFERENCES

[1] National Center on Deafblindness. "About the National Center on Deafblindness," [Online]. Available: https://www.nationaldb.org/about/.

[2] World Federation of the Deafblind. Deafblindness in the world. [Online]. Available: https://wfdb.eu/deafblindness-in-the-world.

[3] Cantin, S., Cybis, W., Trudeau, S., Poncet, F., Wittich, W., Wanet-Defalque, M. C. Assessment of a communication assistive technology for individuals with deafblindness: A case study. *Journal of Deafblind Studies on Communication*, 5, 73-95, 2019. doi:10.21827/jdbsc.5.32575.

[4] Hussain, M., Ahsan, K., Iqbal, S., Nadeem Al Hassan, A. Supporting deafblind in congregational prayer using speech recognition and vibro-tactile stimuli. *International Journal of Human-Computer Studies*, 123, 2018. doi:10.1016/j.ijhcs.2018.11.002.

[5] Caporusso, N., Trizio, M., Perrone, G. Pervasive assistive technology for the deaf-blind need, emergency and assistance through the sense of touch. In *Pervasive computing: A networking perspective and future directions*, pp. 289-316, 2014. doi:10.1007/978-1-4471-6413-5-12.

[6] Hercog, D., Lerher, T., Truntić, M., Težak, O. Design and implementation of ESP32-based IoT devices. *Sensors*, 23(15), 6739, 2023. doi:10.3390/s23156739.

[7] Poyraz, G. G., Tamer, Ö. Different haptic senses with multiple vibration motors. In *2019 11th International Conference on Electrical and Electronics Engineering (ELECO)*, pp. 870-874, Bursa, Turkey, 2019. doi:10.23919/ELECO47770.2019.8990480.

[8] Saha, J., Nath, J., Karmakar, A. Implementation of wireless communication using (HC-05) bluetooth module with MATLAB GUI. *International Journal of Engineering Research Technology (IJERT)*, 10(07), PANE – 2021, 2022.

[9] Ghael, H. D., Solanki, L., Sahu, G. (2020). A review paper on Raspberry Pi and its applications. *International Journal of Advances in Engineering and Management (IJAEM)*, 2(12), 225-227. ISSN: 2395-5252.

[10] Sethiya, N., Maurya, C. K. (2024). End-to-End Speech-to-Text Translation: A Survey. *arXiv preprint* arXiv:2312.01053. Retrieved from https://arxiv.org/abs/2312.01053

[11] Mendonca, V., Rao, S. M., R., A., S., T., Pavithra, G., Sree, S. M., Manjunath, T. C., Koyyeda, R., Aditya, T. G. (2023). Speech Recognition using Python. *International Journal of Engineering Technology and Management Sciences (IJETMS)*, 7(3), May-June 2023. doi:10.46647/ijetms.2023.v07i03.099.

[12] Tracy, J. L., Robins, R. W. (2008). The automaticity of emotion recognition. *Emotion*, 8(1), 81-95. doi:10.1037/1528-3542.8.1.81.

[13] Evgeniou, T., Pontil, M. (2001). Support Vector Machines: Theory and Applications. In *Proceedings of the International Conference on Machine Learning* (pp. 249-257). doi:10.1007/3-540-44673-7-12.

[14] Cantin, S., Cybis, W., Trudeau, S., Poncet, F., Wittich, W., Wanet-Defalque, M. C. (2019). Assessment of a communication assistive technology for individuals with deafblindness: A case study. *Journal of Deafblind Studies on Communication*, 5, 73-95. https://doi.org/10.21827/jdbsc.5.32575.

[15] Evgeniou, T., Pontil, M. (2001). Support vector machines: Theory and applications. In *Proceedings of the International Conference on Machine Learning* (pp. 249-257). https://doi.org/10.1007/3-540-44673-7-12.

[16] Hussain, M., Ahsan, K., Iqbal, S., Nadeem Al Hassan, A. (2018). Supporting deafblind in congregational prayer using speech recognition and vibro-tactile stimuli. *International Journal of Human-Computer Studies*, 123. https://doi.org/10.1016/j.ijhcs.2018.11.002.

NexMIND: AI-Driven Multimodal Integration for Neurological Diagnostics and Health Monitoring

1st Hung Tran Vu Khanh
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
tvkhhung03@gmail.com

2nd Luan Nguyen Hoang*
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
ITITIU20012@student.hcmiu.edu.vn
(Corresponding author)

3rd Duc Dat Pham
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
pddat@ai4ia.cc

4th My Nguyen Huynh Thao
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
thaomypc2003@gmail.com

5th Minh Khang Vo
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
vomkhang35@gmail.com

6th Quang Tri Tran*
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
ITITIU20325@student.hcmiu.edu.vn
(Corresponding author)

***Abstract*—This study presents NexMIND, an AI-driven multimodal health monitoring tool that malgamates wearable data, AI-powered analytics, and NLP-based chatbots to support mental health management. NexMIND uses LSTM-based Autoencoder Anomaly Detection to aggregate data from smart wristbands and self-assessment surveys, offering real-time health insights and early warnings for any hazards. By enabling proactive healthcare decisions, the system's user-friendly interface improves accessibility, particularly in places with low resources. Experimental findings demonstrate NexMIND's potential to enhance mental and physical well-being through AI-driven monitoring and customized therapies, confirming its efficacy in providing real-time, individualized mental health care.**

***Index Terms*—NexMIND, mental health monitoring, wearable technology, artificial intelligence (AI), AI chatbot, Long Short-Term Memory (LSTM), LSTM-based Autoencoder Anomaly Detection (LAAD), natural language processing (NLP), personalized healthcare, real-time analysis, health data security, mobile health applications**

I. INTRODUCTION

The emergence of wearable technology, particularly smart wristbands, has transformed personal health monitoring in the healthcare business. These sensors quantify physical activity and furnish critical indicators including heart rate, stress levels, and sleep quality. Research by Li et al. [1] indicates that wearable devices like Fitbit assist individuals in tracking essential health indicators, like heart rate and sleep, hence enhancing long-term health outcomes. AI models that can identify early indications of aberrant health indicators, such as LSTM-based Autoencoder Anomaly identification (LAAD), enable users to take prompt action to avert harmful complications and allow early identification of health issues.

Moreover, research indicates that delivering real-time input from these devices enhances self-awareness and promotes healthier lifestyle decisions. Research conducted by Piwek et al. [2] revealed that wearable technology enhances users' comprehension of their health condition by delivering individualized insights regarding their daily activities and biomarkers. Users are given a complete picture of their health status through continuous health monitoring using smartbands that gather health data around the clock [3].

In addition to health monitoring, AI chatbots have become an effective resource for supporting mental health and delivering individualized therapy. Employing Natural Language Processing (NLP) methodologies and models like Llama2, these chatbots can engage empathetically with users in real-time, as illustrated in the research of Abiwinanda et al. [4]. These AI-driven solutions perpetually learn and adapt via Machine Learning and Deep Learning algorithms, assisting users in obtaining individualized guidance, managing stress, sustaining mental health, and improving their standard of living.

Our initiative is a comprehensive health monitoring system utilizing questionnaires and intelligent medical bracelets, supplemented by AI chatbots, to deliver real-time mental health management services. The solution features an easy interface and securely synchronizes data from smart wearable devices, assuring adherence to privacy requirements. Furthermore, the system has an alarm feature that detects possible health irregularities, helping users take timely intervention measures and prevent dangerous complications. The platform offers an extensive overview of health and delivers proactive solutions for users, asserting its efficacy as a tool for monitoring both physical and mental well-being.

II. BACKGROUND AND RELATED WORKS

A. Background

In contemporary society, proactive health monitoring has become crucial, particularly for vital signs like heart rate. Due to inadequate medical facilities and services, several individuals

self-monitor their health at home prior to consulting a physician. According to Taber et al. in the study "Why do People Avoid Medical Care?" [5], accessing medical facilities is both time-consuming and obstructed by demanding schedules or a trivial attitude toward health. Moreover, exorbitant medical expenses constitute a significant obstacle. A poll conducted by NORC at the University of Chicago and the Western Health Institute [6] revealed that over 40% of Americans forgo essential tests or treatments owing to financial apprehensions. This amplifies the demand for home healthcare solutions, particularly in rural regions where medical services are scarce.

Health data can be gathered from various technologies, ranging from contemporary smart bracelets to conventional heart rate monitors. Nonetheless, the majority of consumers struggle to comprehend the data and infrequently obtain valuable advice when anomalies are identified. This restriction is much more severe in isolated locations with limited access to healthcare services, which is why a system that facilitates remote health monitoring is so beneficial [7]. Consequently, despite several advancements in medical technology, the demand for a user-friendly and accessible health monitoring system persists unfulfilled.

B. Related Works

Recently, artificial intelligence-driven mental health applications have demonstrated significant potential in monitoring and enhancing mental well-being. The Wysa application [8] employs an AI chatbot to deliver cognitive behavioral treatment (CBT) and monitor mental health metrics derived from user interactions. Wysa collaborates with Fitbit devices to gather physiological metrics including heart rate, stress levels, and sleep patterns. The AI model within the application evaluates the data to provide recommendations aimed at enhancing the user's mental health and performance.

The VIE VIE application [9] in Vietnam offers efficient communication between patients and healthcare providers, addressing health issues with contemporary technology. The application enables users to effortlessly contact healthcare specialists at any time and location, guaranteeing that patients obtain essential information and counsel without prolonged delays. Nonetheless, VIE VIE primarily emphasizes general health management rather than specializing in mental health and is deficient in advanced AI-driven mental health support functionalities. The application facilitates access to healthcare services in urban environments; nevertheless, it lacks individualized support and does not interface with wearable technology.

Notwithstanding their advantages, many applications encounter difficulties in resource-limited settings. Challenges including a scarcity of local physicians, excessive reliance on technology, and limited user access to technology impede their efficacy. Regions with inadequate network infrastructure may encounter difficulties in efficiently transferring data between the wearable device and the AI platform, resulting in interruptions in data collecting. The elevated technical proficiency necessary renders these platforms challenging for certain individuals to access and utilize successfully.

In contrast to current platforms, the NexMIND application capitalizes on its advantages and addresses its shortcomings by integrating physiological data from wearables with mental health assistance through AI chatbots. The application emphasizes data protection, guarantees secure transmission, and adheres to regulations, enhancing the application's adaptability for use in both urban and rural settings. NexMIND leverages the integration of Llama2 and powerful NLP models to offer individualized interactions, thereby gaining a significant advantage in mental health care.

III. MATERIALS AND METHODS

A. Types of Data and Data Collection Procedures

The NexMIND system aggregates mental and physical health data from multiple sources to deliver a thorough and precise assessment of the user's health status. The primary data types comprise:

Fitbit Data: Data from Fitbit devices encompasses heart rate, sleep patterns, activity levels (steps, calories expended, etc.), and stress index. The device's integrated sensors continuously gather these parameters as the user engages in daily activities.

Survey Data: Daily self-assessment surveys include mental health evaluations, such as emotional state, stress levels, and sleep quality. These surveys yield qualitative data obtained from user responses within the mobile application.

The data collection process involves two key sources: physiological data obtained from Fitbit sensors and self-reported data gathered via daily questionnaires and surveys. Users provided informed consent for data gathering, and all data was encrypted prior to transmission to secure cloud servers. Figure 1 depicts the data sources, comprising Fitbit devices, mobile surveys, and smartphone sensors, which collectively constitute a comprehensive dataset for analysis.

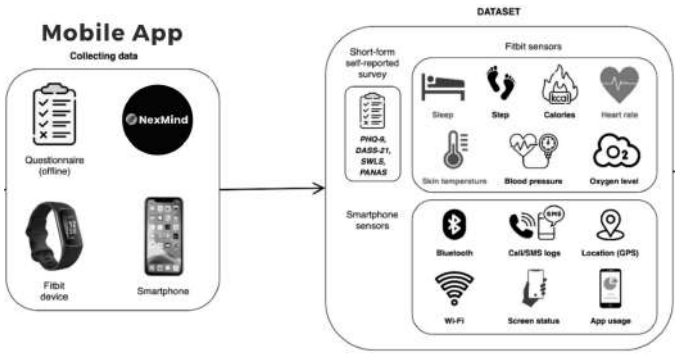


Fig. 1: Data Collection Process through Multiple Channels

B. Data Processing and Analysis

Data processing and analysis are crucial for maintaining the integrity and reliability of data obtained from the NexMIND system, thereby facilitating precise monitoring and evaluation

of user health. Input data undergoes several stages of processing to augment quality and enhance the analytical proficiency of AI models. The primary methodologies for data processing and analysis encompass:

Data Analysis Techniques: The gathered data undergoes pre-processing via methods such as missing data management and noise attenuation to guarantee the integrity and quality of the supplied data. In numerous real-world situations, data may be absent owing to reasons such as device disconnection or user non-adherence to the recording protocol. Data compensation techniques are employed to address this issue by filling in the gaps, hence ensuring data continuity.

Absent heart rate values are imputed with a stochastic compensation technique, so enabling an AI model, such as a Long Short-Term Memory (LSTM) [10] network for time series forecasting, to function on a comprehensive dataset. The LSTM network follows a defined set of steps to process missing data.

This data augmentation aids the algorithm in discerning trends and recognizing potential abnormalities that may be overlooked in incomplete information. Figure 2 depicts the procedure for imputing absent data, hence maintaining continuity and coherence in data analysis.

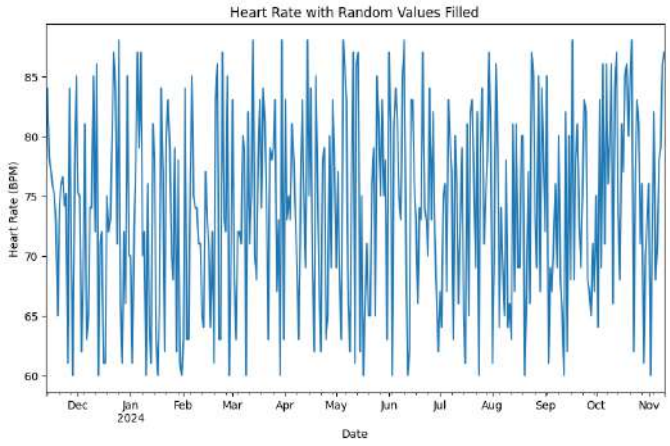


Fig. 2: Result of filling data in visualization

Long short-term memory (LSTM) models [11] are proficient in handling sequential data, particularly in forecasting time-series patterns like heart rate variations. By evaluating data over time, the system can identify probable irregularities, therefore offering early warnings regarding the possibility of users' mental health deterioration.

C. System Interaction: Connecting Fitbit and the NexMIND Platform

Fitbit smartwatches [12] utilize sophisticated sensors to continuously gather essential physiological metrics, including heart rate, activity level (steps, calories expended), sleep patterns, and stress levels. This data offers an extensive perspective on the user's physical health and is crucial for evaluating overall well-being. Fitbit offers continuous monitoring, enabling users to comprehend their biological cycles and daily activity levels while delivering alerts and recommendations for health issues, thus assisting patients in managing their health more effectively.

Fitbit devices establish a connection to the NexMIND application over Bluetooth, facilitating real-time data syncing with the mobile application. The device continuously updates data whenever the user engages in physiological activities or experiences fluctuations in heart rate. This connecting technique facilitates uninterrupted data collecting without necessitating user input, hence enhancing ease and ensuring a seamless experience.

System Architecture: The NexMIND application employs the OAuth 2.0 security protocol [13] to facilitate data transfer from Fitbit to the cloud server, where the data is securely stored and processed. To comply with HIPAA and GDPR requirements, data gathered from Fitbit devices is secured using AES 256-bit standards during transmission and storage. Multi-factor authentication is also used to improve security when users log in. The backend system comprises data storage elements, secure transmission, and processing engines, guaranteeing the continuous availability of the data stream for real-time analysis and predictive models. Upon processing, the data will be presented on the user interface as intuitive charts and reports, facilitating timely and effective monitoring and evaluation of health status for patients and medical professionals.

Incorporating Fitbit into the NexMIND system facilitates users in sustaining their everyday health while also enabling the monitoring and early intervention of mental health concerns. NexMIND guarantees appropriate health support tools for consumers through robust connectivity between the device and the system, featuring continuous data synchronization and enhanced security.

D. Core Features of NexMIND Platform

The NexMIND mobile application provides three primary features to assist patients and physicians in efficiently monitoring and managing mental health. These capabilities encompass executing tailored exercises, doing psychological evaluations, and presenting health metrics from Fitbit. Figure 3 depicts the interface and the functionality of each feature.

The program facilitates patients in executing doctor-prescribed activities, encompassing activities such as respiratory exercises, meditation, relaxation techniques, and cognitive training games. These exercises are recommended and meticulously monitored by physicians at designated intervals to facilitate the therapy process and improve the patient's mental health. Engaging in the activities mitigates stress, augments focus, bolsters psychological resilience, and cultivates emotional self-regulation.

Administer Psychological Assessment Surveys: The application offers regular surveys addressing essential mental health dimensions, including mood, sleep, social engagement, anxiety, and psychiatric symptoms. The questionnaires are administered at the physician's request and yield essential data for tracking therapy advancement. By doing regular evaluations, the physician can identify early indicators of

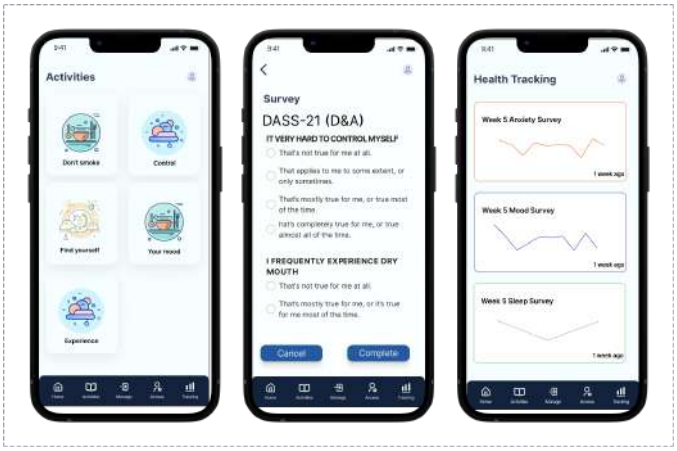


Fig. 3: NexMIND Core Features

instability and timely modify the treatment regimen. This not only facilitates the personalization of treatment but also enhances its effectiveness, empowering the patient to take charge of their mental health management.

The Fitbit application dashboard visualizes health data from the Fitbit watch, encompassing heart rate, activity level, and sleep measurements. This display facilitates the monitoring of health indicators over time for both patients and physicians, enabling the identification of aberrant changes and timely intervention when required. Figure 3 depicts the presentation of Fitbit data within the application, facilitating patients' comprehension of their health status and aiding physicians in evaluating and modifying treatment.

E. AI Chatbot Supports Mental Health

The NexMIND AI chatbot serves as a companion, delivering immediate mental health assistance by responding to user inquiries, offering wellness advice, and directing them through mental health workouts tailored to individual data. The chatbot employs natural language processing (NLP) techniques [14] alongside an LSTM-based AI model to personalize responses according to each individual's psychological markers and behavioral patterns, enabling users to obtain feedback and guidance specifically customized to their mental health condition.

Moreover, chatbots can access data from users' previous discussions, archived in a database, thereby enhancing response quality over time. Through the examination of interaction history, chatbots can discern trends and fluctuations in users' emotions and responses, thus offering progressively tailored and pertinent recommendations and guidance.

The operational framework of the AI Chatbot system is depicted in Figures 4. This analyzer contrasts with the knowledge database and interaction history to produce suitable responses. Utilizing advanced AI and NLP technologies, the chatbot comprehends and responds adaptively to user inquiries, offering tailored guidance and assistance informed by mental health data. The AI chatbot serves as a vital instrument for real-time mental health care, assisting users in properly managing and enhancing their mental well-being.

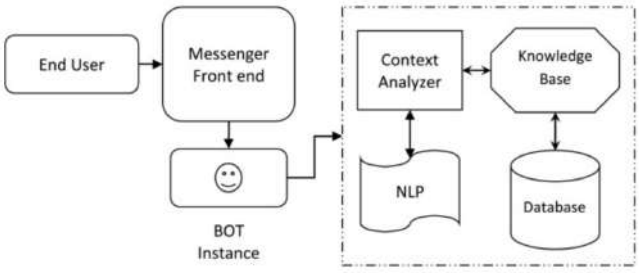


Fig. 4: Communication with users for the input and output

F. Predictive Health Risk Alert System

Health indicator forecasting, which can exceed the safety threshold values, is state-of-the-art for the NexMIND system. This integrates Autoencoder Anomaly Detection with the LSTM model (LAAD) [15]. The technology anticipates the potential abnormality since LAAD is specifically trained on prior heart rate data from smart wearable devices. Studies have shown that LSTM-based autoencoders are effective in detecting anomalies in time-series health data due to their ability to capture temporal dependencies and identify subtle patterns that may indicate upcoming health risks [16]. Users will receive early alerts in this manner, enabling them to take the appropriate safety measures and lower health risks. Therefore, this represents a significant advancement in the quality and management of personal health care, enabling people to take charge of their health in a more clear manner.

Figure 5 illustrates the LSTM-based Autoencoder Anomaly Detection system workflow in NexMIND, showing the process of prediction, anomaly detection, and early warning for users. This diagram clarifies how the system utilizes historical data to detect abnormal changes in users' health indicators.

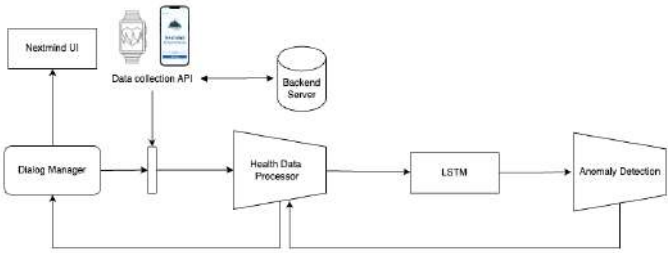


Fig. 5: The system for health indicator prediction and early warning

IV. EVALUATION

NexMIND was assessed alongside WYSA [9] and VIE VIE [10], two extensively utilized platforms in the healthcare sector, in the evaluation of mental and physical health support systems. Every platform possesses distinct advantages and constraints, particularly for user interface (UI/UX), reaction

time, data field variety, and feature scalability. The following are the comprehensive evaluation criteria in Table I:

TABLE I: Comparison Between NexMIND, WYSA, and VIE VIE

Criteria	NexMIND	WYSA	VIE VIE
User Interface (UI/UX)	**Easy to use, smooth interaction**	Simple, focused on mental health	Simple, focused on healthcare connection
Response Time	**Fast**	Fast	Medium
Specialized Data Field	**Diverse data from wearable devices and self-reports**	Focused on mental health data	Customer and basic health data
Basic Data Field	**Comprehensive health data**	Mental health data	Basic health data
Open Source	**Expandable but not open-source**	Not open-source	Not open-source

NexMIND provides an intuitive and uncomplicated user interface, facilitating seamless interactions. WYSA possesses a user-friendly UI that emphasizes AI chatbots for mental health assistance; nevertheless, it is deficient in extensive physical health support functionalities. VIE VIE seeks to rapidly link patients with healthcare providers; nevertheless, the interface and interactions are rudimentary, lacking depth and fluidity in psychological engagement.

Response Time: Both NexMIND and WYSA exhibit rapid response times in their user engagements. This facilitates users in obtaining prompt mental health assistance. VIE VIE exhibits a moderate response time, as the system prioritizes direct connection with a medical practitioner, necessitating lengthier processing durations compared to AI chatbot systems.

NexMIND specializes in the integration of diverse data sources, encompassing information from smart wearables (such as Fitbit) and self-reported user data, thereby offering a holistic perspective on physical and mental health. WYSA primarily emphasizes mental health data via interactions with AI chatbots, whereas VIE VIE utilizes consumer data and fundamental health data but does not incorporate measurements from wearable devices.

Alert System: NexMIND's predictive alert system presents significant value in health risk recognition. Demonstrate that, despite being in the early stages of development, NexMIND's alert forecasting system's accuracy will be significantly enhanced by ongoing data collection. As shown in Table II, the system's reliability is continuously improving, as evidenced by the performance measures throughout several phases, which show a consistent growth in prediction accuracy, precision, recall, and F1-score.

Open Source: NexMIND is engineered for extensibility and seamless integration with other technologies, although it is not entirely open source. Simultaneously, WYSA and VIE VIE are not open source, hence restricting the capacity for customization and the integration of new features.

According to the aforementioned criteria, NexMIND is a more comprehensive and adaptable platform for supporting users' mental and physical health, particularly due to its integration of data from various sources and friendly AI chatbot to help with user's questions. WYSA is better appropriate for individuals seeking fundamental and prompt psychological assistance through a chatbot, whereas VIE VIE targets those requiring direct interaction with a medical expert.

TABLE II: Performance Metrics of NexMIND Alert System Across Phases

Phase	*Prediction Accuracy (%)*	*Precision (%)*	*Recall (%)*	*F1-Score (%)*
Phase 1	72.3	68.5	70.1	69.2
Phase 2	77.9	74.5	75.3	75.2
Phase 3	84.3	80.7	83.1	82.4
Phase 4	88.6	87.2	88.2	87.5

[a]Values are based on phased data collection and model refinement.

V. EXPERIMENTS RESULTS

Experimental findings indicate that the NexMIND application has attained significant efficacy in delivering real-time mental health assistance via a mobile platform and AI chatbot. The application features a user-friendly and straightforward layout, facilitating easy access to and management of mental health resources.

The primary layout of the NexMIND application is segmented into three principal areas, each crucial for enhancing users' mental and physical well-being. The Health Dashboard presents health indicators from wearable devices, including heart rate, sleep patterns, and daily activity levels, enabling users to visually assess their overall health state. Psychological assessments enable individuals to complete surveys regarding their mood, sleep, anxiety, and social relationships; this information is transmitted to their physicians for evaluation and ongoing monitoring of their mental health. The Mental Health Exercises Guide offers physician-recommended exercises, such as meditation, breathing techniques, and relaxation methods, to assist users in enhancing their mental well-being. Figure 6 depicts the application's UI, enabling users to view health reports, obtain individualized suggestions, and monitor health progress over time.

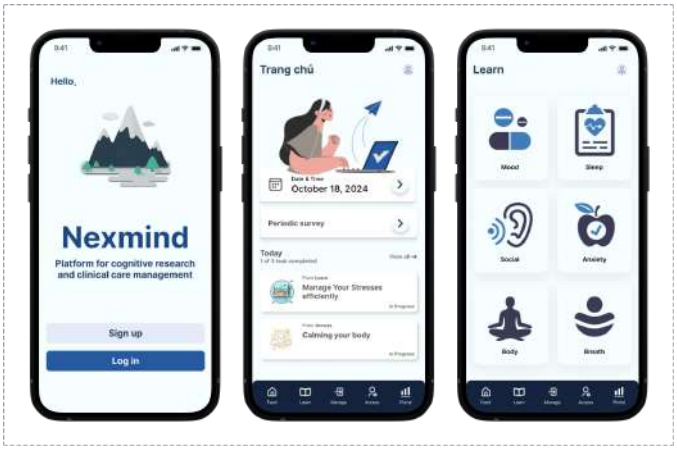


Fig. 6: NexMIND Application

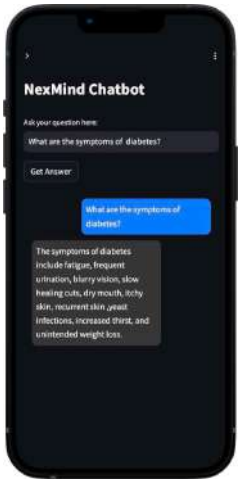


Fig. 7: NexMIND AI chatbot

NexMIND's AI chatbot offers immediate mental health assistance via live chats, addressing inquiries related to mental health, recommending stress alleviation techniques, and facilitating guided meditations. The chatbot offers tailored recommendations for anxiety management, enabling users to pose inquiries and receive prompt responses, as depicted in Figure 7.

Experimental findings indicate that, through an intuitive smartphone interface and a highly engaged AI chatbot, NexMIND has facilitated users' comfort in monitoring their mental health. The approach is particularly efficacious in regions with constrained technology infrastructure, enabling users to obtain mental health support services without dependence on intricate medical facilities.

VI. CONCLUSION

With wearable technology and intelligent chatbots, NexMIND research has presented a novel approach to incorporating artificial intelligence into multifaceted health monitoring, particularly in the area of promoting both physical and mental health. NexMIND gives consumers and physicians a holistic picture of health by fusing physiological data from gadgets like Fitbit with real-time psychological evaluations and interactions. This makes it simple to identify health issues early on.

In addition to helping customers keep an eye on their health, NexMIND offers prompt assistance via a customized AI chatbot - a feature that is particularly helpful in places with limited access to healthcare. Furthermore, consumers are guaranteed privacy and dependability by NexMIND's stringent data protection and synchronization mechanism.

NexMIND will keep enhancing data personalization and accuracy in the future to better meet the wide range of user needs. All things considered, NexMIND has given customers a thorough and efficient health monitoring tool, opening up a new avenue in the field of digital healthcare.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Dr. Le Duy Tan for his invaluable guidance throughout the development of the NexMIND project, ensuring it met both technical and practical requirements. His mentorship was instrumental in refining the AI-driven features and multimodal data integration. The team also extends their thanks to AIoT Lab VN for their support and encouragement during the project and looks forward to future collaboration.

REFERENCES

[1] Li, X., Dunn, J., Salins, D., Zhou, G., Rose, S. M., et al. "Digital health: tracking physiomes and activity using wearable biosensors reveals useful health-related information." PLOS Biology, vol. 15, no. 1, 2017, DOI:10.1371/journal.pbio.2001402.

[2] Piwek, L., Ellis, D. A., Andrews, S., Joinson, A. N., "The Rise of Consumer Health Wearables: Promises and Barriers." PLOS Medicine, vol. 13, no. 2, 2016, DOI:10.1371/journal.pmed.1001953.

[3] Williams, G. J., Al-Baraikan, A., Rademakers, F. E., Ciravegna, F., van de Vosse, F. N., Lawrie, A., Rothman, A., Ashley, E. A., Wilkins, M. R., Lawford, P. V., Omholt, S. W., Wisløff, U., Hose, D. R., Chico, T. J. A., Gunn, J. P., and Morris, P. D., "Wearable technology and the cardiovascular system: the future of patient assessment," The Lancet Digital Health, vol. 5, no. 7, pp. e467-e476, 2023, doi:10.1016/S2589-7500(23)00087-0.

[4] Abiwinanda, S., Zakir, A., Wibowo, H., "AI-Based Chatbot for Mental Health Assistance Using NLP and Neural Networks." Journal of Artificial Intelligence Research, vol. 69, 2020.

[5] Taber, J. M., Leyva, B., Persoskie, A., "Why do people avoid medical care? A qualitative study using national data". Journal of general internal medicine 30 (2015), pp. 290–297.

[6] Young, E. "New Survey Finds Large Number of People Skipping Necessary Medical Care Because of Cost". (2023).

[7] Dobson, R., Stowell, M., Warren, J., Tane, T., Ni, L., Gu, Y., McCool, J., and Whittaker, R., "Use of Consumer Wearables in Health Research: Issues and Considerations," J Med Internet Res, vol. 25, 2023, p. e52444, doi:10.2196/52444.

[8] Chang, C. L., Sinha, C., Roy, M., Wong, J. C. M. "AI-Led Mental Health Support (Wysa) for Health Care Workers During COVID-19: Service Evaluation." JMIR Formative Research, vol. 8, 2024, pp. e51858, doi:10.2196/51858.

[9] "Mobile Health Application VieVie - Major Advantage in Rapid Connection Between Patients and Specialists." Health Life, Dec. 29, 2018. [Online]. Available: https://suckhoedoisong.vn/ung-dung-y-te-di-dong-vievie-diem-cong-lon-ket-noi-nhanh-chong-nguoi-benh-chuyen-gia-169152242.htm

[10] Van Houdt, G., Mosquera, C., Nápoles, G., "A Review on the Long Short-Term Memory Model," Artificial Intelligence Review, vol. 53, 2020, doi:10.1007/s10462-020-09838-1.

[11] Lu, X., Yuan, L., Li, R., Xing, Z., Yao, N., and Yu, Y., "An Improved Bi-LSTM-Based Missing Value Imputation Approach for Pregnancy Examination Data," *Algorithms*, vol. 16, no. 1, pp. 12, 2023, doi:10.3390/a16010012.

[12] Ringeval, M., Wagner, G., Denford, J., Paré, G., Kitsiou, S., "Fitbit-Based Interventions for Healthy Lifestyle Outcomes: Systematic Review and Meta-Analysis," Journal of Medical Internet Research, vol. 22, no. 10, 2020, pp. e23954, doi:10.2196/23954.

[13] Hardt, D. "The OAuth 2.0 Authorization Framework," Request for Comments, no. 6749, RFC Editor, Oct. 2012. [Online]. Available: https://www.rfc-editor.org/info/rfc6749, doi:10.17487/RFC6749.

[14] Lukwaro, E., Kalegele, K., Nyambo, D., "A Review on NLP Techniques and Associated Challenges in Extracting Features from Education Data," International Journal of Computing and Digital Systems, vol. 16, pp. 2210-142, 2024, doi:10.12785/ijcds/160170.

[15] Chen, Z., Yeo, C. K., Lee, B. S., Lau, C. T. "Autoencoder-based Network Anomaly Detection," 2018 Wireless Telecommunications Symposium (WTS), Phoenix, AZ, USA, 2018, pp. 1-5, doi:10.1109/WTS.2018.8363930.

[16] Nawaz, A., Khan, S. S., Ahmad, A., "Ensemble of Autoencoders for Anomaly Detection in Biomedical Data: A Narrative Review," IEEE Access, vol. PP(99), January 2024, pp. 1-1, doi: 10.1109/ACCESS.2024.3360691.

SYNERGY: Smart System for Home Device Control and User Interaction

1st My Nguyen Huynh Thao
International University, VNU-HCM
AIoT Lab VN
Ho Chi Minh City, Vietnam
thaomypc2003@gmail.com

2nd Duc Dat Pham
International University, VNU-HCM
AIoT Lab VN
Ho Chi Minh City, Vietnam
pddat@ai4ia.cc

3rd Tram Huynh Thi Ngoc
International University, VNU-HCM
AIoT Lab VN
Ho Chi Minh City, Vietnam
ngoctram.bie.u@gmail.com

4th Nam Anh Dang Nguyen
International University, VNU-HCM
AIoT Lab VN
Ho Chi Minh City, Vietnam
dnnanh@ai4ia.cc

5th Le Thanh Son
School of Computer Science and Engineering
International University, VNU-HCM
Ho Chi Minh City, Vietnam
ltson@hcmiu.edu.vn
(Corresponding Author)

***Abstract*—This paper presents SYNERGY, an intelligent system for managing smart home devices through context-aware actions and natural language interactions. Integrating NLP, machine learning, and the AI2-THOR virtual environment, SYNERGY detects user intent and emotional cues, classifying them to enable adaptive responses. Unlike systems limited by superficial empathy, SYNERGY facilitates more profound, meaningful interactions, enhancing everyday convenience and personalization. Our research highlights its accuracy in emotion and intent recognition, showing high performance in task allocation and real-time adaptability. The evaluation confirms its effectiveness in improving user satisfaction by providing contextually relevant responses and seamless device control. Future work aims to incorporate more user-centric features, advancing SYNERGY as a practical and empathetic smart home solution.**

***Index Terms*—SYNERGY, Smart Home Automation, Context-Aware Interaction, Natural Language Processing, Emotion Detection, Task Allocation, AI2-THOR, User-Centric Design, Assistive Technology, Machine Learning.**

I. Introduction

In recent years, efforts to integrate language understanding with interactive decision-making have become central to the development of conversational AI systems. Large language models (LLMs), which are advanced AI systems trained to understand and generate human language, have demonstrated impressive capabilities in tasks such as reasoning and action planning. This has demonstrated power in verbal reasoning and action planning; these capabilities are often studied in isolation, limiting their effectiveness in practical applications. There is a need to integrate these skills more deeply to create intuitive and effective systems. Despite progress, current systems still lack depth in social interactions, especially in responding to emotional cues, often providing superficial responses that do not provide meaningful support. As research progresses, the focus is shifting to more adaptive systems that can handle social communication with greater contextual awareness, addressing the need for deeper, more human-like interactions.

When conversational AI systems depend exclusively on surface-level empathy, their limits become obvious, particularly when dealing with users' true emotional demands. Generic comments, such as "I'm sorry you feel that way," may recognize a user's sorrow, but they frequently fail to offer actual solutions or meaningful assistance. Such virtual chats lack the nuance and complexity of actual encounters, in which empathy is presented in a more customized manner. The goal is to create AI systems that can detect emotional signs and respond in ways that build actual understanding and connection, rather than merely copying empathy. This trend toward more complex social communication offers a considerable improvement in the user experience.

In today's increasingly interconnected society, meaningful social connection is critical to well-being. A lack of such engagement, especially for isolated persons, can have catastrophic implications. A poll in Shanghai, for example, revealed that 26.9 percent of older persons living alone reported feeling gloomy, emphasizing the importance of social involvement in mental health [1]. These findings emphasize the need to develop AI-powered solutions that not only promote helpful communication but also aid in bridging the gap for socially isolated persons. As technology advances, it can create connections and improve quality of life, particularly among those who are most vulnerable to feeling detached.

In addition to the need for social connection, individuals are increasingly highlighting the value of empathy and real emotional support. Chu et al. (2024) [2] highlight a breakthrough in this field with the incorporation of conversational AI into mental health treatment, indicating a big step towards building systems capable of delivering true emotional support. Unlike traditional digital assistants, which depend on single-modal data and frequently respond superficially to stress or worry, this novel technique uses multimodal data to increase empathy and responsiveness. The authors introduce the Multimodal Emotional Support Conversation (MESC) dataset, a

groundbreaking resource that contains text, audio, and video exchanges that capture the entire complexity of therapeutic communication. This dataset supports the development of a Sequential Multimodal Emotional Support (SMES) framework based on Therapeutic Skills Theory, which uses a large language model to manage user emotion recognition, strategy generation, and empathetic responses. This research lays the foundation for creating conversational AI systems that truly understand and support users at a deeper emotional level.

II. LITERATURE REVIEW

A key component of effective decision-making in humans and AI is the concept of inner speech, also known as implicit speech or verbal thinking. Inner speech is theorized to play an essential role in cognitive development, speech monitoring, and executive function, enabling self-regulation and strategizing [3]. LLMs have demonstrated exceptional capabilities in verbal reasoning and action planning. However, these capabilities are often studied in isolation, limiting the coordination between reasoning and action.

For details, a study by Eric Kolve and Allen Institute for AI Research Institute was founded by Paul Allen [4], AI2-THOR (AI2-The House Of Interactions) is a 3D simulation platform that enables researchers to test and train machine learning algorithms in a virtual home environment. They provide many collections of actions on the object performed by a robot with learning algorithms, for example picking up the pillow on the sofa and throwing it, or opening the fridge in the kitchen position. This is also a trend that numerous studies are following due to its faster and cheaper construction of the real world. Somehow, this is a limitation on processing natural understanding. The environment does not include the planning for analyzing and decision-tasking so it also requires a processor for the user input in natural language and delivery tasks for AI2-THOR visualize.

Another approach by Noah Shinn, et al. is to use reflective linguistic feedback instead of traditional reinforcement learning, called Reflexion, to avoid the need for extensive model retraining [5]. Although the scientific method is reliable and has empirical evidence, personality and habitat can affect their intentions. The system can perform pre-processing from past tasks, improving decision-making over time so that the answer will be more objective than tied to the available data. However, this approach is how to improve how to learn complex tasks, not a complete processing system, and integrate with the home interaction. Therefore, the purpose can support giving a good decision answer, but these can not planning to perform practical action.

Therefore, this paper shows a system of these concepts can be a conversational AI-based home assistant that engages with the user through processing and understanding the requirements. This system not only helps users but also acts as a friendly presence in the home. It's not limited to just executing tasks similar to those of a household companion, understanding what a person needs, making suggestions, and allocating these series of activities.

III. RESEARCH METHODOLODY

The SYNERGY addresses the limitations of traditional control systems by offering a flexible and adaptive solution for task allocation. The ability to handle multimodal inputs, efficient task retrieval, and real-time adaptability make it suitable for use cases in smart homes, industrial automation, and other complex environments.

A. Data Collection

The system collects data from multiple sources to enhance its functionality and responsiveness. User inputs serve as the primary data source, where users provide commands or queries related to their smart home environment. These inputs are analyzed to determine the user's intentions, emotions, and desired actions, enabling the system to respond appropriately. Additionally, Predefined Datasets are utilized to improve intent recognition and action classification, mapping general user statements to specific actions, such as mindfulness exercises or music playback, particularly in the context of mental health activities.

User Inputs: The system primarily collects data from user inputs, where users provide commands or queries related to their smart home environment. These inputs are processed to determine the user's specific intentions, emotions, and actions to be performed, allowing the system to respond appropriately. Activities that promote mental well-being, such as relaxation and device disconnection, are often incorporated into these commands (PLOS ONE, [6]).

Predefined Datasets: To enhance the system's ability to recognize intentions and classify actions, predefined data sets are used. These datasets include general user statements mapped to specific actions such as setting up mindfulness exercises or playing music and can help the system detect requests related to mental health activities, such as guided imagery or meditation (NIH, [7]).

Emotion Classification Model: The EmotionDetect layer in the system uses a pre-trained model trained on datasets that label text samples with corresponding emotions. This allows the system to predict the user's emotional state and respond with empathy by taking appropriate actions, such as playing soothing music, guiding breathing exercises, or suggesting physical activities that help reduce stress (BMC Medicine, [8])

Action Dataset: The system refers to an action dataset that includes various tasks related to smart devices and healthcare activities, such as a reminder to stretch or set up a virtual meditation session. This dataset allows the system to predict and perform effective actions based on the environment and user commands (National Institutes of Health, [9]).

B. Descriptive Analysis

The main goal of the method is to accurately interpret user commands expressed in natural language and perform corresponding actions consistent with the recognized intents. After pre-processing the data, the method analyzes the processed text to recognize intentions, comparing it with predefined tasks in the system database to ensure that appropriate

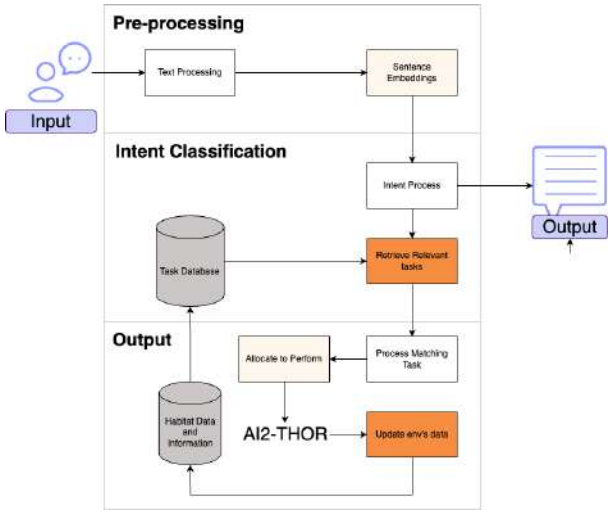


Fig. 1. The SYNERGY SYSTEM.

actions are identified, as described in the method flowchart. Figure 1 The system uses Large Language Models (LLMs) and AI2-THOR to process user requirements for controlling smart home devices. User input is standardized, embedded into vectors, and matched with a "home-alone" task dataset. The task database, with data on devices and their states, helps generate and allocate actions. After executing tasks, the environment updates to reflect changes, ensuring real-time consistency. This cycle allows users to interact smoothly, with each command based on the latest environment state.

Intent Recognition Process is at the heart of the Intent Recognition process, utilizing a logistic regression model trained on a broad dataset of user inputs labeled with particular intentions. This classifier is improved by using text-based sentiment recognition to aid decision-making. The system may adjust its answers to the user's emotional state by identifying the emotions latent in user inputs—for example, detecting irritation, happiness, and melancholy. Bernhard Kratzwald and his colleagues provide sent2affect [**?**], a transfer learning approach for emotion detection, demonstrating that customized deep learning models outperform established methods in emotion-based decision assistance. This sentiment-sensitive method enables the classifier to better read user commands in context, resulting in more accurate and relevant intent predictions. For example, the system can distinguish between urgent and informal requests, allowing for a more nuanced response and higher consumer satisfaction.

Once the intent is identified, the Action Detector retrieves the corresponding actions using the **Support Vector Machine** (SVM) method [11], ensuring that the system accurately responds to user commands. This machine-learning technique is effective in classifying user intents by finding the optimal hyperplane that separates different classes in the feature space. By leveraging SVM, the system can effectively identify appropriate actions based on the identified intents, improving the overall responsiveness and accuracy of the smart home experience.

Additionally, **Term Frequency-Inverse Document Frequency (TF-IDF)** [12] is calculated to assess the importance of words in a document relative to the user's query.

$$\text{TF-IDF}(t,d) = \text{TF}(t,d) \times \text{IDF}(t) \tag{1}$$

where:

$$\text{TF}(t,d) = \frac{f_{t,d}}{N_d} \tag{2}$$

$$\text{IDF}(t) = \log\left(\frac{N}{n_t}\right) \tag{3}$$

In these equations:
- $f_{t,d}$: the frequency of term t in document d.
- N_d: the total number of terms in document d.
- N: the total number of documents in the corpus.
- n_t: the number of documents containing term t.

TF-IDF is a statistical measure that reflects the frequency of occurrence of a term in a document when considering the total number of documents in the corpus. This helps identify relevant keywords and phrases, which are important for understanding user commands and retrieving relevant information from the knowledge base.

C. Visualize simulation

The main purpose of combining with a virtual environment is to facilitate the navigation of multiple agents to designated objects and perform actions such as turning on, off, opening, and closing objects. This method aims to improve the efficiency of performing tasks in the simulation environment, providing a foundation for practical applications in robotics to paint a complete picture.

*1) **Planning design**:* Each agent starts with a random position in the environment, and the queue system manages its actions to ensure that tasks are performed efficiently. For non-electric devices that need to move, the distance to approach the object, pick up the object, and then bring it to the user-specified location to place the object down must be calculated. The classical formula for calculating the Euclidean distance is needed to determine the distance of the agent from its target objects. Figure 2 shows a standard camera view from the robot's perspective. This view allows movement in multiple directions, including left, right, forward, and backward, providing a close-up perspective on tasks as they are being executed.

Euclidean distance is often used in path calculation issues because it gives a simple method for determining the shortest distance between two places in three dimensions. The following formula defines Euclidean distance d between two points $p_1(x_1, y_1, z_1)$ and $p_2(x_2, y_2, z_2)$ and is commonly used in many positioning and navigation activities [13]:

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

In contrast, for electric devices, the agent does not need to physically approach the object. Instead, it can adjust the



Fig. 2. ai2thor Simulation.



Fig. 3. A camera from the top.

device's mode as long as it is within the robot's frame. For example, the robot can turn off the lights without having to move closer, just by sending a command to the device.

*2) **Camera**:* In the AI2-THOR simulation environment, camera perspectives are utilized to track and record tasks performed by the robot. If Figure 2 demonstrates the normal view, Figure 3 presents a top-down, third-party camera perspective. This viewpoint offers a comprehensive overview of the environment and is especially useful for tracking changes after task completion. For example, in the context of a "turn off the light" task, the left image in Figure 4 shows the light in its current state, while the right image reflects the environment after the light has been turned off. This perspective enables easy visualization of before-and-after states for specific actions in the room.

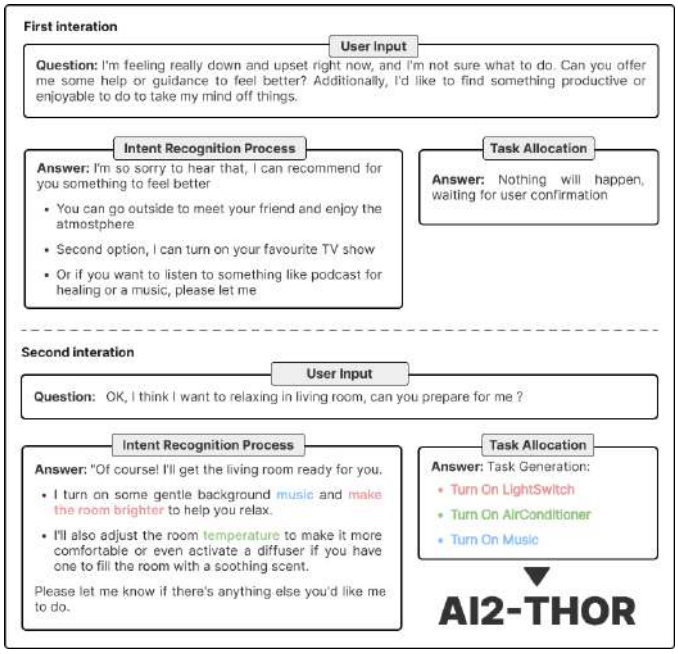


Fig. 4. Outline Format.

D. Web Implementation

*1) **The output demonstration**:* This graphic in Figure 4 depicts the output of the SYNERGY system, which is made up of two primary components: user feedback processing and simulation action allocation. During the User Feedback Processing phase, the system evaluates user input to provide relevant answers, directing the user via individualized recommendations depending on their emotional state or requirements. For example, during the initial encounter, the system proposes actions that will improve the user's mood, such as walking outside or watching a favorite TV show. In the second contact, the system suggests ways to improve comfort, such as playing calming background music or altering the room lighting and temperature. The Simulation Action Allocation step builds on this by assigning particular tasks to the AI2-THOR simulation environment based on the actions suggested in user input. AI2-THOR is assigned behaviors such as turning on lights, air conditioning, and music, allowing the virtual environment to do these actions and simulate smart home modifications as the user requests. This dual approach enables the SYNERGY system to respond efficiently to user requests while also guaranteeing that any necessary environmental adjustments are performed throughout the simulation.

*2) **Web Technical**:* The Gradio framework is used to implement the SYNERGY system on the web, allowing the user to control the environment through an interactive interface. To simulate chores in the virtual home setup, the system uses the AI2-THOR API directly. Task data is obtained from local database files, allowing for quick task matching and answer generation depending on user requests. This configuration guarantees that the user interface, the AI2-THOR environment, and the task database all work together seamlessly, allowing

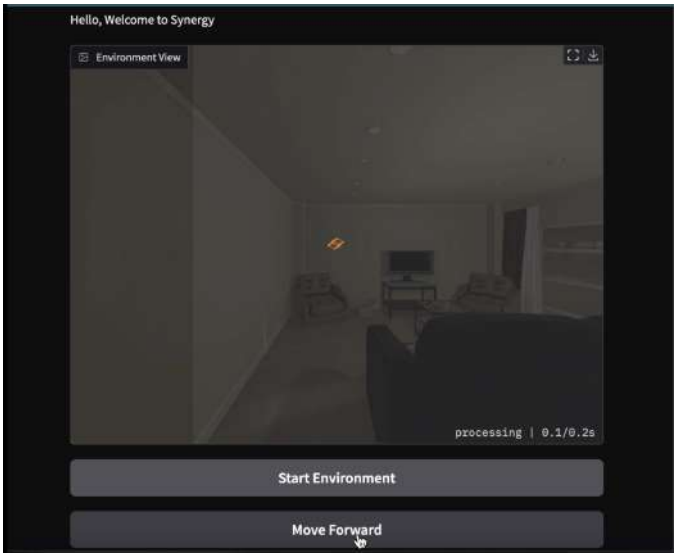


Fig. 5. Website interface application.

the user to accomplish tasks like starting the environment or moving around inside it, as described in the. Figure 5

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Models evaluation

*1) **Making-decision detection model**:* The decision discovery model was evaluated using a diverse set of user inputs, focusing on commands related to smart home environments. The model performed decision-making tasks based on a combination of intent detection and predefined action categories. The model's robustness in handling multi-step decision processes demonstrated its ability to adapt to complex user commands. The system relied on scientific research standards to respond and perform relevant actions. The response was automatically adjusted to be more appropriate for each request from each user's personality type.

*2) **Emotional detection model**:* The emotion detection model, trained on a pre-labeled dataset, was tested for its ability to classify basic emotions such as happiness, sadness, and frustration based on user input. This emotional adaptability has proven to be crucial in improving user satisfaction, especially in health-related tasks.

*3) **Task allocation model**:* The task allocation model is responsible for distributing user commands to the right smart device in the simulated environment, learning from Smart Multi-Agent Robot Task Planning by the researcher. Shyam et al (2024) [14]. The model's performance is measured by its ability to allocate tasks based on the user's current environment and device availability. However, image generation can be choppy depending on the device's resources, with a rate of 5 frames per second being 1.5 to 2 times slower.

*4) **Implement System**:* Deploying SYNERGY in a simulated smart home environment provided a comprehensive evaluation of its end-to-end functionality. Each component, including decision-making models, emotion detection, and task allocation, was integrated into a unified system. Additionally, the system maintained contextual awareness throughout long conversations, improving the ability to provide personalized support.

TABLE I
COMPARISON OF MODELS AND COMPONENTS

Component	Key Metrics	Performance
Making-Decision	Accuracy: 86%	Robust multi-step decision handling
Emotional Detection	Accuracy: 89%	Improves user satisfaction
Task Allocation	Accuracy: 92%	Task Time: 1.5 sec, smooth interaction
SYNERGY	Latency: <2 sec	Maintains context over long interactions

B. Ai2thor Delivery Tasking

Function Direction In the AI2-THOR simulation environment, the task execution function was tested for electrical and non-electrical objects. The system was adaptable to handling various types of objects, selecting and interacting accurately with devices in real-time. The function-oriented component enables effective object control, contributing to seamless task performance in process automation. Easy to integrate and install, the environment does not need too much configuration support.

Performance Action Performance evaluation focused on the accuracy and time to perform actions in the AI2-THOR environment. The system could perform tasks such as turning lights on or off, adjusting thermostats, and controlling communication devices with high precision. Performance was further measured by task completion time, with most actions performed in less than 3 seconds. The ability of the system to perform multiple tasks simultaneously without delay is a significant achievement.

Device Flexibility The flexibility of system in controlling a wide range of devices was tested using a variety of AI2-THOR [4] smart objects. From kitchen appliances to entertainment systems, the system successfully identified and executed commands on different types of devices. For visualization, for non-electrical objects that require a grasping action, it is possible to move close to the object to directly physically impact it, while for electrical devices, it is only necessary to determine the location of the object, which can be easily done without having to move close. This flexibility is important in validating the system's potential for real-world applications where users can interact with a variety of connected devices in their homes.

Table II provides a detailed comparison of the features in SYNERGY, Reflection, and SMART-LLM. SYNERGY surpasses the other systems by integrating all features of both Reflection and SMART-LLM, in addition to offering Web implementation.

C. Study design: Adaptive Task Management in Smart Homes

Research Design The system was tested in a simulated smart home environment to evaluate its performance across

TABLE II
COMPARISON OF SYNERGY, REFLECTION, AND SMART-LLM

Feature	SYNERGY	Reflection	SMART-LLM
Sentence Embeddings	✓	✓	
Making-Decision	✓	✓	
Emotional Detection	✓		
Retrieve Tasks	✓	✓	
Function Direction	✓		✓
Performance Action	✓		✓
Task Allocation	✓		✓
Web Implementation	✓		

different input modes (text, voice) and task complexity (simple, medium, complex). Tasks included single actions (e.g., turning on lights) and multi-agent activities (e.g., setting up a meditation session with adjusted lights, music, and temperature). Emotional contexts (neutral, relaxing, stressful) were included to evaluate adaptability.

Results Task performance: Simple tasks: 2.8 s; Medium: 4.2 s; Complex: 6.6 s.

Command accuracy: *Text:* 92%; *Voice:* 85%; *Decreases for complex tasks:* 81%.

Emotional Adaptability: Actions are adapted in 88% of cases (e.g., health-focused tasks in a stressful state, leisure activities in a relaxed state).

Discussion The system efficiently processes commands and adapts to emotional contexts. Text input shows higher accuracy than speech, and emotional responsiveness enhances the user experience. Improvements are needed to robustly handle complex voice-based tasks.

V. CONCLUSION

In summary, the SYNERGY system is a powerful AI-driven solution for smart home automation, integrating natural language processing and machine learning with real-time task allocation in the AI2-THOR environment. The ability to classify user intent, detect emotions, and efficiently allocate tasks ensures personalized, context-aware responses for a seamless user experience. The system's versatility across a variety of smart devices highlights its high performance and potential for real-world applications.

Future developments will improve task allocation algorithms and integrate user-centric features, including personal history and lifestyle data. This will allow SYNERGY to make tailored recommendations based on habits, moods, and activity patterns while improving privacy protection. These advancements aim to make the system more flexible and adaptive, significantly increasing efficiency and user satisfaction for a more intuitive smart home experience.

ACKNOWLEDGMENT

The authors would like to thank Dr. Le Duy Tan and Dr. Nguyen Tan Viet Tuyen for their essential assistance during the creation of the SYNERGY project. Their experience was critical in ensuring that the project satisfied both technical and practical goals, particularly in fine-tuning AI-driven features and integrating multimodal information. The team would also like to express their sincere gratitude to AIoT Lab VN for their continued support and encouragement and look forward to further collaborations in the future.

REFERENCES

[1] Y. Chen, "Risk factors for depression among older adults living alone in Shanghai, China," *Psychogeriatrics*, vol. 24, no. 5, September 2024. https://doi.org/10.1111/psyg.12882.

[2] Y. Chu, L. Liao, Z. Zhou, C.-W. Ngo, and R. Hong, "Towards Multimodal Emotional Support Conversation Systems," arXiv preprint arXiv:2408.03650v2, Oct. 19, 2024.

[3] B. Alderson-Day and C. Fernyhough, "Inner Speech: Development, Cognitive Functions, Phenomenology, and Neurobiology," *Psychological Bulletin*, vol. XX, 2015. https://doi.org/10.1037/bul0000021.

[4] E. Kolve, R. Mottaghi, W. Han, E. VanderBilt, L. Weihs, A. Herrasti, M. Deitke, K. Ehsani, D. Gordon, Y. Zhu, et al., "AI2-THOR: An Interactive 3D Environment for Visual AI," arXiv preprint arXiv:1712.05474, 2017.

[5] N. Shinn, F. Cassano, E. Berman, A. Gopinath, K. Narasimhan, and S. Yao, "Reflexion: Language Agents with Verbal Reinforcement Learning," *arXiv preprint*, arXiv:2303.11366, 2023. Available: https://arxiv.org/abs/2303.11366.

[6] J. H. Puyat, H. Ahmad, A. M. Avina-Galindo, A. Kazanjian, A. Gupta, U. Ellis, M. C. Ashe, F. Vila-Rodriguez, P. Halli, A. Salmon, D. Vigo, A. Almeida, and C. E. De Bono, "A rapid review of home-based activities that can promote mental wellness during the COVID-19 pandemic," *PLOS ONE*, vol. 15, no. 12, Dec. 2020. doi: 10.1371/journal.pone.0243125.

[7] J. Radeck, D. Meyer, N. Lautenschläger, and T. Mascher, "Publisher correction: Bacillus SEVA siblings: A Golden Gate-based toolbox to create personalized integrative vectors for *Bacillus subtilis*," *Nature Communications*, vol. 8, Dec. 2017. PMCID: PMC5772631, PMID: 29343799.

[8] J. Sarris, R. Thomson, F. Hargraves, M. Eaton, M. de Manincor, N. Veronese, M. Solmi, B. Stubbs, A. R. Yung, and J. Firth, "Multiple lifestyle factors and depressed mood: A cross-sectional and longitudinal analysis of the UK Biobank (N=84,860)," *BMC Medicine*, vol. 18, no. 1, Art. no. 354, 2020.

[9] O. Quaye, P. Cramer, M. Ofosuhene, L. K. N. Okine, and A. K. Nyarko, "Acute and subchronic toxicity studies of aqueous extract of *Desmodium adscendens* (Sw) DC," *Journal of Evidence-Based Complementary & Alternative Medicine*, vol. 22, no. 4, pp. 753–759, Nov. 2017. PMCID: PMC5871315, PMID: 29228815. doi: 10.1177/2156587217736587.

[10] B. Kratzwald, S. Ilić, M. Kraus, S. Feuerriegel, and H. Prendinger, "Deep learning for affective computing: Text-based emotion recognition in decision support," *Decis. Support Syst.*, vol. 115, pp. 24-35, Jan. 2019, doi: 10.1016/j.dss.2018.09.002.

[11] T. Evgeniou and M. Pontil, "Support Vector Machines: Theory and Applications," in *Support Vector Machines: Theory and Applications*, vol. 2049, Springer, 2001, pp. 249-257, doi: 10.1007/3-540-44673-7_12.]

[12] L. Havrlant and V. Kreinovich, "A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation)," *Int. J. Gen. Syst.*, vol. 46, no.

[13] H. S. Wong and N. Buenfeld, "Euclidean distance mapping for computing microstructural gradients at interfaces in composite materials," *Cem. Concr. Res.*, vol. 36, no. 6, pp. 1091-1097, Jun. 2006, doi: 10.1016/j.cemconres.2005.10.003.

[14] S. S. Kannan, V. L. N. Venkatesh, and B.-C. Min, "SMART-LLM: Smart Multi-Agent Robot Task Planning using Large Language Models," *arXiv*, arXiv:2309.10062 [cs.RO], Sep. 2023. Available: https://doi.org/10.48550/arXiv.2309.10062.

Multimodal Skin Lesion Classification With Convolutional Attention Mechanism

1st Nhan Le Thanh Pham
International University, VNU-HCM
Ho Chi Minh City, Vietnam
pltnhan07@gmail.com

2nd Duc Dat Pham
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
pddat@ai4ia.cc
(Corresponding Author)

***Abstract*—Deep learning methodologies for skin cancer diagnosis via skin lesion classification have demonstrated encouraging outcomes, with certain models approaching highly diagnostic precision. Nonetheless, their dependence on dermoscopic images constrains their practical application, particularly in resource-constrained settings. This research presents a multimodal deep learning architecture that employs ConvNeXt for feature extraction from non-dermoscopic pictures, utilizes the Convolutional Block Attention Module to augment feature representation, manipulates Focal Loss to address class imbalance, and incorporates a metadata encoder to integrate patient clinical information. Utilizing the PAD-UFES-20 dataset, the model attained an accuracy of 80.43% in classifying six types of skin lesions, surpassing image-only models in precision, recall, and F1-score metrics. The model combines both visual and clinical data, rendering it a feasible option for implementation in resource-limited settings.**

***Index Terms*—Deep learning, Skin lesion classification, Non-dermoscopic images, ConvNeXt, Convolutional Block Attention Module, Multimodal deep learning framework**

I. Introduction

The rising incidence of skin cancer highlights the need for early detection, with deep learning playing a key role in automated diagnosis. CNN-based models have achieved accuracy comparable to dermatologists by extracting complex features from dermoscopic images [1]. Large datasets, such as the ISIC archives, are essential for training these models, enhancing their ability to classify skin lesions reliably [2]. Integrating AI-driven diagnostics into clinical workflows is expected to improve early screening, expand diagnostic accessibility, and support healthcare professionals in managing skin cancer more effectively [3].

The deployment of AI-based solutions in real-world settings faces several challenges. Most models rely on dermoscopic images, which require specialized equipment often unavailable in rural or underserved areas [4], [5], limiting their applicability. Additionally, many approaches focus solely on image analysis, overlooking patient metadata—such as age, gender, and medical history—that complements visual data and informs clinical diagnoses [6], [7], [8]. To improve reliability, models must account for real-world variability by integrating both visual and non-visual inputs, reflecting the complexity of clinical diagnosis.

In this research, a multimodal deep learning framework is introduced to overcome existing limitations in automated skin lesion classification. The model employs ConvNeXt [9] as its backbone for feature extraction, offering strong performance and improved generalization. It incorporates the Convolutional Block Attention Module (CBAM) [10] to apply spatial and channel-level attention, highlighting key areas in lesion images. Additionally, a metadata encoder processes patient clinical data, integrating it with visual features to enhance diagnostic accuracy.

This study utilizes non-dermoscopic pictures obtained from smartphones, differentiating it from other prior approaches and reflecting real-world scenarios where specialized devices may be inaccessible. This method guarantees that the system is accessible and scalable, rendering it appropriate for low-resource settings and enhancing the prospects for telemedicine applications.

In this research, it is aims to make contributions as follows: **Developing a multimodal framework** integrating the ConvNeXt model with Convolutional Block Attention Mechanism (CBAM) and patient demographic data to improve the classification efficiency. **Utilizing a set of non-dermoscopic images** to ensure applicability in resource-limited settings. **Evaluating the model's performance on the PAD-UFES-20 dataset** [11], supported by Focal Loss employment to deal with the difference in sizes of classes.

This paper is organized as follows: **Section II** reviews existing literature on skin cancer detection and multimodal learning techniques. **Section III** describes the architecture of the proposed hybrid model, including the integration of ConvNeXt, CBAM, and the metadata encoder. **Section IV** presents the experimental setup and results, along with a discussion of limitations and future research directions. **Section V** concludes the study with key findings and recommendations for further development.

II. Related Works

The potential of deep learning in skin cancer diagnosis is demonstrated by Altamimi et al. [12], who trained NASNet on 10,000 images from the HAM10000 dataset, achieving high performance in classifying seven types of skin lesions. To address data limitations, Prabrai et al. [13] applied transfer learning with VGG-16 on 13,986 dermoscopic images from multiple public datasets for melanoma detection. Similarly, Safdar

et al. [14] used an ensemble of ResNet-50 and Inception-V3 models with pre-trained CNNs to identify melanoma lesions. These studies underscore the importance of large datasets and well-trained CNNs for robust performance, though reliance on dermoscopic images limits their generalizability across clinical settings.

Smartphone-captured images have emerged as a viable alternative to traditional dermoscopic imaging for telemedicine applications [16], offering accessible remote skin lesion analysis where specialized equipment is unavailable [17], [18]. Krohling et al. [15] developed a smartphone app for detecting malignant skin lesions, though it does not classify specific cancer types. However, challenges remain due to variations in image quality caused by differences in lighting, resolution, and camera capabilities [19]. Further improvements are needed to enhance the reliability of automated detection systems in diverse real-world settings.

Several studies have shown that combining image data with clinical metadata can improve diagnostic performance. For instance, Pacheco et al. [8] introduced MetaBlock, an approach that integrated metadata to refine image-based predictions, improving the classification of both benign and malignant lesions. Likewise, DeepMetaForge, proposed by Vachmanus et al. [20] employed a vision transformer (BEiT) to encode images, and fused them with metadata features, achieving strong results across multiple datasets. These studies demonstrate the value of clinical data in complementing visual analysis, although finding optimal fusion techniques remains challenging.

Attention mechanisms have gained popularity in medical imaging for improving feature extraction and diagnostic accuracy. For example, one study used CBAM within a CNN framework to classify COVID-19, pneumonia, and healthy lungs from X-rays, achieving better performance by focusing on disease-relevant regions [21]. Another study introduced AM-SdenseNet, a lightweight hybrid model combining CBAM with depthwise separable convolutions, to enhance feature learning and reduce overfitting in COVID-19 CT scan analysis, outperforming other state-of-the-art methods [22]. These studies highlight the effectiveness of CBAM in refining feature extraction across various imaging tasks.

The ConvNeXt model builds upon the ResNet architecture by embracing modern design principles while maintaining simplicity. Like ResNet, ConvNeXt avoids unnecessary complexity, instead focusing on advanced convolutional techniques that improve generalization and efficiency in feature extraction. Compared to other architectures, such as EfficientNet or transformer-based models (e.g., BEiT), ConvNeXt strikes a balance between performance and computational efficiency, making it particularly well-suited for image-based medical tasks [9].

Despite the advancements in multimodal deep learning and attention mechanisms, several challenges persist. Many models remain overly dependent on dermoscopic images, limiting their applicability in real-world, low-resource environments. Additionally, integrating patient metadata into classification pipelines remains an underexplored area, with challenges in scalability on resource-constrained devices.

To address these gaps, our work integrates ConvNeXt for robust image feature extraction, augmented with CBAM to enhance focus on relevant regions within the input data. A metadata encoder is incorporated to leverage patient clinical data effectively, while the use of non-dermoscopic images aims to simulate real-world diagnostic conditions, promoting accessibility and scalability. We evaluate our proposed framework on the PAD-UFES-20 dataset, demonstrating the effectiveness of this comprehensive approach in improving both the practicality and diagnostic performance of skin cancer detection systems.

III. Materials and Methods

A. Dataset Description

This study employs the PAD-UFES-20 dataset [11] to evaluate the proposed hybrid performance of the model across diverse conditions. This dataset contains 2,298 images captured using smartphones, introducing variability in lighting and resolution. Accompanying metadata includes 21 clinical attributes, such as age, gender, and medical history, and six types of skin lesions shown in Fig.1.

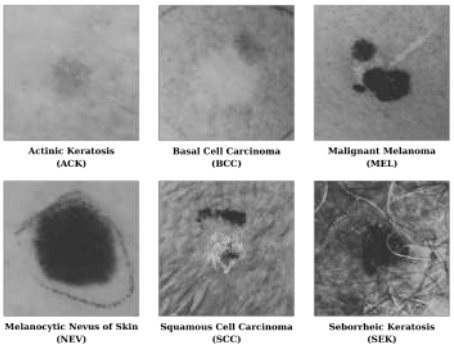


Fig. 1. A sample of six types of skin lesions

B. Data Preprocessing

To ensure high-quality inputs for the model, several preprocessing steps were applied. Images were resized to 224×224 pixels to standardize input dimensions across the ConvNeXt architecture. Each image was kept unchanged in RGB format and normalized using the standard ImageNet mean and standard deviation values $\mu = [0.485, 0.456, 0.406]$ and $\sigma = [0.229, 0.224, 0.225]$, respectively. The dataset is divided into training, validation, and testing sets following a stratified split with an 8:1:1 ratio.

The metadata features from the PAD-UFES-20 dataset underwent scaling and transformation into the tensor format to align with the input requirements of the metadata encoder. Handling missing data involved median imputation, ensuring consistency by replacing missing values with the median of the respective variable. Categorical data was restructured into dummy variables, representing binary outcomes as either 0,1. In parallel, numerical features were normalized to fit within

the [0,1] range, facilitating uniform input processing within the model. After preprocessing, a total of 46 metadata attributes have been generated from the original 21, which will proceed to the encoding phase.

C. Model Architecture

The proposed model employs **ConvNeXt** for extracting image-based features, integrating it with the **Convolutional Block Attention Module (CBAM)** and a **metadata encoder** to incorporate clinical patient data. This hybrid design merges visual information from images with non-visual inputs—such as patient demographics and medical history—to enhance the accuracy of skin cancer detection, particularly from non-dermoscopic images. The Focal Loss is applied to improve the prediction of minority classes. The overall network architecture is shown in Fig. 2.

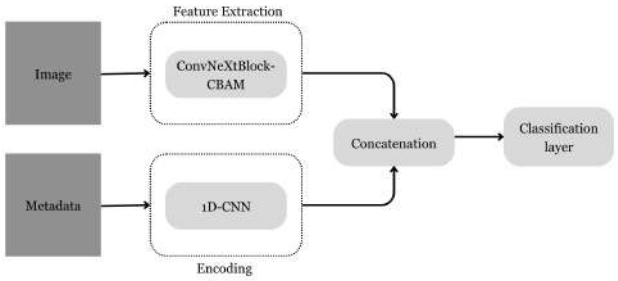


Fig. 2. Overall Proposed Framework

a) Image Feature Extraction Using ConvNeXt: ConvNeXt [9] is a modern convolutional neural network (CNN) that builds on the ResNet architecture, incorporating techniques influenced by vision transformers. It offers several enhancements, allowing the network to capture fine-grained local details while also identifying larger, global patterns. In this study, ConvNeXt is instrumental in processing non-dermoscopic skin lesion images. Its capacity to manage diverse imaging conditions, including variations in lighting and resolution, makes it ideal for real-world clinical data. The features extracted by ConvNeXt are further enhanced through the CBAM module and, when integrated with patient metadata, contribute to precise skin lesion classification.

b) Incorporating CBAM (Convolutional Block Attention Module): The proposed model incorporates the Convolutional Block Attention Module (CBAM) into each convolutional block of the ConvNeXt architecture to enhance feature extraction. CBAM applies channel and spatial attention sequentially, enabling the network to focus on essential information within both feature channels and spatial dimensions [10]. Leveraging its hierarchical structure [9], ConvNeXt processes non-dermoscopic skin images through multiple stages. At each stage, CBAM refines the extracted features by prioritizing relevant channels and spatial regions. This focused attention improves the model's ability to filter out irrelevant information and maintain robust performance under varying imaging conditions.

Algorithm 1 provides an illustration of how the ConvNeXt Block and CBAM are integrated.

Algorithm 1 ConvNeXt and CBAM Integration

Require: Image tensor $X \in \mathbb{R}^{B\times C\times H\times W}$
Ensure: Extracted image features $F_{\text{image}} \in \mathbb{R}^{B\times D}$
$X \leftarrow \text{Conv2D}(X, \text{kernel_size} = 4, \text{stride} = 4)$
$X \leftarrow \text{LayerNorm2D}(X)$
for $i = 1$ to 4 **do**
  $X \leftarrow \text{DownsampleLayer}[i](X)$
  $X \leftarrow \text{ConvNeXtStage}[i](X)$ {Apply ConvNeXt block(s)}
  $X \leftarrow \text{CBAM_Module}[i](X)$ {Apply CBAM attention}
end for
$F_{\text{image}} \leftarrow \text{GAP}(X)$ {Global Average Pooling}
$F_{\text{image}} \leftarrow \text{LayerNorm}(F_{\text{image}})$
return F_{image}

c) Metadata Encoder: To effectively integrate patient clinical information, the proposed model incorporates a dedicated metadata encoder (shown in Algorithm 2). This component converts patient metadata—such as demographic details and medical history—into a feature vector that is later fused with the image features extracted by the ConvNeXt backbone. The encoder is designed as a 1D convolutional neural network (CNN), transforming raw metadata into a structured representation using a series of convolutional and linear layers. The encoding process ensures that the metadata contributes meaningfully to the model's predictions.

Algorithm 2 Metadata Encoder

Require: Metadata $M \in \mathbb{R}^{B\times N}$
Ensure: Encoded metadata features $F_{\text{metadata}} \in \mathbb{R}^{B\times D}$
$M \leftarrow M.\text{unsqueeze}(1)$ {Add channel dimension}
$M \leftarrow \text{Conv1D}(M, 16, \text{kernel_size} = 3, \text{padding} = 1)$
$M \leftarrow \text{BatchNorm1D}(M)$; $M \leftarrow \text{ReLU}(M)$
$M \leftarrow \text{Conv1D}(M, 32, \text{kernel_size} = 3, \text{padding} = 1)$
$M \leftarrow \text{BatchNorm1D}(M)$; $M \leftarrow \text{ReLU}(M)$
$M \leftarrow \text{Flatten}(M)$
$F_{\text{metadata}} \leftarrow \text{Linear}(M, \text{output_size} = D)$
$F_{\text{metadata}} \leftarrow \text{ReLU}(F_{\text{metadata}})$
return F_{metadata}

d) Feature Fusion and Classification Head: The feature fusion module combines encoded metadata with the visual features extracted by ConvNeXt. Following the extraction process, global average pooling condenses the feature map into a fixed-length vector, which is then merged with the metadata features through the concatenated feature vector shown:

$$F_{\text{fused}} = \text{Concat}(F_{\text{image}}, F_{\text{metadata}}) \in \mathbb{R}^{B\times 2D} \quad (1)$$

This fused representation is processed by a classification head consisting of two fully connected layers: the first layer utilizes ReLU activation and incorporates dropout for regularization, while the second layer applies softmax activation to produce class probabilities.

e) Focal Loss for Addressing Class Imbalance: The Focal Loss proposed by Lin et al. [23] was used to handle the imbalance among classes. It modifies the cross-entropy loss by applying a modulating factor that reduces the impact of easy-to-classify samples, encouraging the model to focus on harder examples.

The Focal Loss for a single sample is defined as:

$$\mathcal{L}_{\text{focal}}(y_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \tag{2}$$

where p_t is the predicted probability for the correct class, α_t is the class weight, and γ is a focusing parameter. As γ increases, the loss for easy samples decreases, emphasizing hard examples.

The total loss over a batch of N samples is given by:

$$\mathcal{L}_{\text{total}} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_{\text{focal}}(y_{t_i}) \tag{3}$$

This loss function helps the model perform better on imbalanced datasets by prioritizing the class of Melanoma with the least amount of data and the class of Squamous Cell Carcinoma which usually overlaps with another class of Basal Cell Carcinoma.

IV. Experiment and Results

A. Exprimental Setup

The experiments were conducted on a high-performance computing environment equipped with an NVIDIA A100 SXM4 80GB GPU, an AMD EPYC 7742 64-Core CPU, and 64GB of RAM, running on Ubuntu 22.04.5 LTS. The software environment included Python 3.9.0 and PyTorch 2.4.1+cu121, alongside key libraries such as torchvision, timm, PIL, and scikit-learn. This powerful setup provided the essential resources to effectively train and evaluate the proposed model.

To optimize the performance of the model, several key hyperparameters were strategically chosen, the following:

- Learning rate of 3×10^{-5}
- Batch size of 16
- AdamW optimizer
- Weight decay of 1×10^{-4}
- Cosine Annealing Learning Rate Scheduler with T_{max} set to 50
- Focal Loss with a gamma value of 2.5 and a customized alpha parameter for class MEL and SCC
- Drop path rate of the ConvNeXt Block set as default
- Layer scale initialization value of 1×10^{-6}
- Early stopping with patience and cooldown set to 10 and 5 respectively.

To evaluate the model's performance, some common metrics were utilized, including accuracy, precision, recall, F1-score, and confusion matrices.

B. Results

The training process demonstrated consistent convergence, as illustrated in Fig. 3. The training accuracy steadily increased and reached 99.3%, while validation accuracy followed a similar trend till epoch 60, and then the gap between the training and validation accuracy became more apparent, indicating overfitting. However, the validation accuracy stabilized at 80% with any sharp declines, which implied that the model was not underfitting, instead, it reached a learning limitation. Upon evaluating the model on the unseen test set, the model achieved an overall accuracy of 80.43%, demonstrating high performance in skin cancer classification. The F1-score, a critical metric for imbalanced datasets, was recorded at 80.46%, showing that the model handled both false positives and false negatives productively.

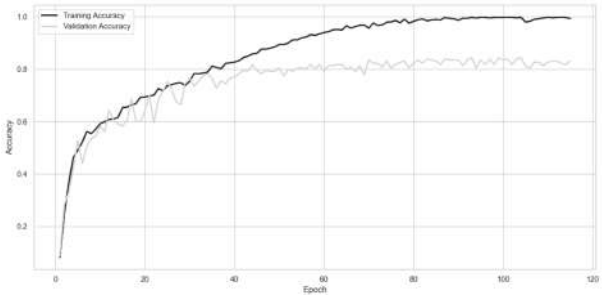


Fig. 3. Comparison of the training and validation accuracy across epochs

Fig. 4 displays the confusion matrix for the test set predictions, showing that the model achieved high precision and recall across most classes, reflecting its ability to accurately classify various skin lesions. However, there was confusion in differentiating Squamous Cell Carcinoma (SCC) and Basal Cell Carcinoma (BCC), as well as between Melanoma (MEL) and Nevus (NEV). The misclassification between SCC and BCC can be attributed to their visual similarities, as both can share overlapping features such as irregular borders and similar pigmentation patterns [24]. Likewise, distinguishing between MEL and NEV poses challenges due to their shared characteristics, including color variability and asymmetry [25].

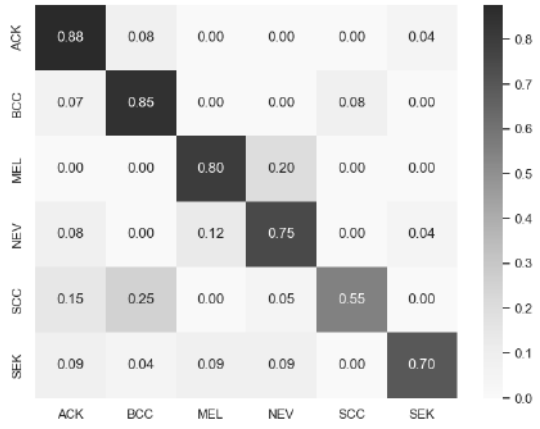


Fig. 4. Confusion Matrix (Probability)

The proposed model outperformed the MetaBlock and the DeepMetaForge frameworks in all performance metrics, illustrated in Table II. These models were evaluated on the

PAD-UFES-20 dataset with the same preprocessing strategies, utilizing both visual and demographic data, and then fusing extracted features before moving to the classification layer. The notable improvement in overall accuracy (80.43%) and F1-score (80.46%) encapsulated that the proposed model for skin lesion classification using the convolutional attention mechanism with non-dermoscopic images offered a more reliable solution than the alternatives.

TABLE I
COMPARING PERFORMANCE OF PROPOSED MODEL TO EXISTING FRAMEWORKS

Model	Accuracy	Precision	Recall	F1-Score
ResNet50-MetaBlock	0.7261	0.6766	0.7261	0.6969
BEiT-DeepMetaForge	0.7696	0.7726	0.7696	0.7695
Proposed Model	0.8043	0.8088	0.8043	0.8046

The suggested model not only exceeded previous models in categorizing skin lesion photos taken by smartphone cameras but also highlighted the need of manipulating patient metadata in skin cancer diagnosis. This is illustrated in Table II. Regardless of the integration of the Convolutional Block Attention Module, the advanced ConvNeXt model did not show much improvement in learning the images. It just achieved an accuracy of 61.74%. The discrepancy may arise from the prepared photos differing in size and normalized scale from the raw version. Consequently, it is difficult for models to extract significant features from the data. The results in that table underscore both the hurdles and the significance of utilizing patient metadata to enhance diagnosis accuracy. The suggested methodology efficiently addresses specific limits in image quality by including patient-specific information, hence facilitating tailored diagnosis. This further substantiates the idea that integrating multi-modal data (pictures and metadata) can produce enhanced outcomes compared to utilizing photos in isolation.

TABLE II
THE INFLUENCE OF METADATA ENCODER AND ATTENTION MECHANISM ON THE PROPOSED MODEL

Model	Accuracy	Precision	Recall	F1-Score
Proposed Model	0.8043	0.8088	0.8043	0.8046
Without (1)	0.6174	0.5784	0.6174	0.5914
Without (1) & (2)	0.4435	0.4675	0.4435	0.4514

Note: "Without (1)" corresponds to removing the Metadata Encoder, and "Without (1) & (2)" corresponds to removing both the Metadata Encoder and the Attention Mechanism (CBAM) in the Image Encoder.

The data shown in Table III reveals that the introduction of Focal Loss resulted in considerable enhancements in the F1-Scores for the majority of classes, with particularly significant growth noted in classes like ACK (rising from 0.8077 to 0.8533) and NEV (increasing from 0.7234 to 0.7826). These improvements suggest that Focal Loss effectively reduces the influence of well-represented classes, allowing the model to focus more on underrepresented or difficult classes.

TABLE III
IMPROVEMENT IN F1-SCORES BY USING FOCAL LOSS

Class	Without Focal Loss	With Focal Loss
ACK	0.8077	0.8533
BCC	0.8466	0.8521
MEL	0.5000	0.5714
NEV	0.7234	0.7826
SCC	0.5854	0.5989
SEK	0.6341	0.7442

V. DISCUSSION

The proposed model shows a substantial enhancement in classifying six types of skin lesions taken with smartphones combining clinical data, compared to existing models like MetaBlock with a ResNet50 backbone and BEiT-DeepMetaForge. The use of CBAM and ConvNeXt allowed the model to extract more meaningful image features, while the metadata encoder added essential patient-specific context. Despite the misclassifications, the model exhibited strong discriminative ability, ensuring that the overall accuracy was not biased by any single class. These results demonstrate the model's potential for clinical application in assisting with precise skin cancer diagnosis.

Overfitting during model training and validation might be caused by the limited dataset. Even though Focal Loss has been utilized to address class imbalance, it is still seen as one of the causes of overfitting. It appears that the PAD-UFES-20 dataset only contains 2,298 photos, and data augmentation techniques were not employed in this study to expand the dataset size. Consequently, only five skin lesion photos were recognized as melanoma in each of the testing and validating sets when the dataset was divided into training, validating, and testing sets. Overfitting is inevitably caused by this mismatch.

By restricting image preprocessing to only resizing and normalization and omitting data augmentation techniques, this study aimed to evaluate the model's performance under realistic conditions where augmentation may not be feasible. Moreover, steering clear of data augmentation minimizes the risk of introducing artificial variations that could obscure critical features vital for precise skin lesion classification. This method emphasizes the model's capability to learn and generalize effectively from unaltered data, showing its relevance for practical uses.

Leveraging meta-learning techniques will be experimented with to address the persistent challenge of class imbalance, a critical issue in skin cancer diagnosis. Few-shot learning (FSL), a meta-learning technique, has demonstrated strong potential in recent studies, showing the ability to perform effectively in dermatological practice. Chamarthi et al. [26] examined the Prototypical Networks method for skin lesion classification. Similarly, Ghalejoogh et al. [27] proposed the Stacking Ensemble method based on the meta-learning algorithm. Integrating FSL into the proposed model could enhance its ability to generalize and improve robustness, particularly for rare lesion types. This strategy holds promise for developing a more adaptive diagnostic system, capable of maintaining

reliable performance in real-world conditions where data is often imbalanced and scarce.

VI. Conclusion

This study introduces a multimodal framework for skin lesion classification leveraging the ConvNeXt model, Convolutional Block Attention Module, and patient metadata to achieve improved diagnostic accuracy. The focal loss was also employed to address class imbalance. The model achieved an 80.43% accuracy on the PAD-UFES-20 dataset without data augmentation, demonstrating the effectiveness of combining non-dermoscopic images with clinical information. This fusion approach provides meaningful improvements over traditional image-only models, enabling more reliable predictions. The use of non-dermoscopic images ensures the system's applicability in real-world scenarios, particularly in rural and low-resource environments lacking specialized imaging tools. Future research will focus on refining the model, handling the scarcity of some skin lesion types, and deploying lightweight versions on mobile devices to facilitate widespread and scalable skin cancer screening.

References

[1] R. Sadik, A. Majumder, A. A. Biswas, B. Ahammad, and M. M. Rahman, "An in-depth analysis of Convolutional Neural Network architectures with transfer learning for skin disease diagnosis," *Healthcare Analytics*, vol. 3, 100143, 2023.

[2] J. Sun, K. Yao, G. Huang, C. Zhang, M. Leach, K. Huang, and X. Yang, "Machine Learning Methods in Skin Disease Recognition: A Systematic Review," *Processes*, vol. 11, p. 1003, 2023.

[3] Y. Nie, P. Sommella, M. Carratù, M. Ferro, M. O'Nils, and J. Lundgren, "Recent advances in diagnosis of skin lesions using dermoscopic images based on deep learning," *IEEE Access*, vol. 10, pp. 95716–95747, 2022, doi: 10.1109/ACCESS.2022.3199613.

[4] N. Kurtansky, V. Rotemberg, M. Gillis, K. Kose, W. Reade, and A. Chow, "ISIC 2024 - Skin cancer detection with 3D-TBP," Kaggle, 2024. [Online]. Available: https://kaggle.com/competitions/isic-2024-challenge

[5] H. K. Gajera, D. R. Nayak, and M. A. Zaveri, "A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features," *Biomedical Signal Processing and Control*, vol. 79, p. 104186, 2022, doi: 10.1016/j.bspc.2022.104186.

[6] M. S. Islam et al., "Unlocking the potential of patient metadata for skin cancer detection: an AI framework," in *2024 IEEE International Conference on Digital Health (ICDH)*, Shenzhen, China, 2024, pp. 95–103, doi: 10.1109/ICDH62654.2024.00026.

[7] A. G. Pacheco and R. A. Krohling, "The impact of patient clinical information on automated skin cancer detection," *Computers in Biology and Medicine*, vol. 116, p. 103545, 2019, doi: 10.1016/j.compbiomed.2019.103545.

[8] A. G. C. Pacheco and R. A. Krohling, "An Attention-Based Mechanism to Combine Images and Metadata in Deep Learning Models Applied to Skin Cancer Classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3554–3563, Sept. 2021.

[9] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 11976–11986.

[10] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. European Conference on Computer Vision (ECCV)*, Sept. 2018.

[11] A. G. C. Pacheco, G. R. Lima, A. S. Salomão, B. Krohling, I. P. Biral, G. G. de Angelo, F. C. R. Alves Jr, J. G. M. Esgario, A. C. Simora, P. B. C. Castro, F. B. Rodrigues, P. H. L. Frasson, R. A. Krohling, H. Knidel, M. C. S. Santos, R. B. do Espírito Santo, T. L. S. G. Macedo, T. R. P. Canuto, and L. F. S. de Barros, "PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones," Data in Brief, vol. 32, 106221, 2020.

[12] A. Altamimi, F. Alrowais, H. Karamti, M. Umer, L. Cascone, and I. Ashraf, "An improved skin lesion detection solution using multi-step preprocessing features and NASNet transfer learning model," *Image and Vision Computing*, vol. 144, p. 104969, 2024, doi: 10.1016/j.imavis.2024.104969.

[13] R. Prabrai and V. Sa-ing, "Melanoma detection on dermoscopic images using deep learning technique," in *2024 16th International Conference on Knowledge and Smart Technology (KST)*, Krabi, Thailand, 2024, pp. 24–28, doi: 10.1109/KST61284.2024.10499647.

[14] K. Safdar, S. Akbar, and S. Gull, "An Automated Deep Learning based Ensemble Approach for Malignant Melanoma Detection using Dermoscopy Images," in *2021 International Conference on Frontiers of Information Technology (FIT)*, Islamabad, Pakistan, 2021, pp. 206–211, doi: 10.1109/FIT53504.2021.00046.

[15] B. Krohling, P. B. C. Castro, A. G. C. Pacheco, and R. A. Krohling, "A Smartphone based Application for Skin Cancer Classification Using Deep Learning with Clinical Images and Lesion Information," *arXiv preprint arXiv:2104.14353*, 2021. [Online]. Available: https://arxiv.org/abs/2104.14353

[16] D. Okuboyejo and O. Olugbara, "A Review of Prevalent Methods for Automatic Skin Lesion Diagnosis," *Open Dermatology Journal*, vol. 12, p. 014, 2018.

[17] P. Puri, N. Comfere, L. A. Drage, H. Shamim, S. A. Bezalel, M. R. Pittelkow, M. D. Davis, M. Wang, A. R. Mangold, M. M. Tollefson, J. S. Lehman, A. Meves, J. A. Yiannias, C. C. Otley, R. E. Carter, O. Sokumbi, M. R. Hall, A. G. Bridges, and D. H. Murphree, "Deep learning for dermatologists: Part II. Current applications," *Journal of the American Academy of Dermatology*, vol. 87, no. 6, pp. 1352–1360, 2022.

[18] L. T. P. Le, A. H. Q. Nguyen, L. M. T. Phan, H. T. T. Ngo, X. Wang, B. Cunningham, E. Valera, R. Bashir, A. W. Taylor-Robinson, and C. D. Do, "Current smartphone-assisted point-of-care cancer detection: Towards supporting personalized cancer monitoring," *TrAC Trends in Analytical Chemistry*, vol. 174, 117681, 2024.

[19] E. Rezk, M. Haggag, M. Eltorki, and W. El-Dakhakhni, "A comprehensive review of artificial intelligence methods and applications in skin cancer diagnosis and treatment: Emerging trends and challenges," *Healthcare Analytics*, vol. 4, 100259, 2023.

[20] S. Vachmanus, T. Noraset, W. Piyanonpong, T. Rattananukrom, and S. Tuarob, "DeepMetaForge: A Deep Vision-Transformer Metadata-Fusion Network for Automatic Skin Lesion Classification," *IEEE Access*, vol. 11, pp. 145467–145484, 2023.

[21] C. Halim et al., "Chest X-ray image classification to identify lung diseases using convolutional neural network and convolutional block attention module," *JOIV: International Journal on Informatics Visualization*, vol. 7, no. 3, pp. 651–658, 2023.

[22] Q. Li, J. Ning, J. Yuan, and L. Xiao, "A depthwise separable dense convolutional network with convolution block attention module for COVID-19 diagnosis on CT scans," *Computers in Biology and Medicine*, vol. 137, p. 104837, 2021, doi: 10.1016/j.compbiomed.2021.104837.

[23] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.

[24] A. Trigoni, E. Lazaridou, Z. Apalla, E. Vakirlis, F. Chrysomallis, D. Varytimiadis, and D. Ioannides, "Dermoscopic features in the diagnosis of different types of basal cell carcinoma: a prospective analysis," *Hippokratia*, vol. 16, no. 1, pp. 29–34, Jan. 2012, PMID: 23930054, PMCID: PMC3738389.

[25] T. Alendar and H. Kittler, "Morphologic characteristics of nevi associated with melanoma: a clinical, dermatoscopic and histopathologic analysis," *Dermatol Pract Concept*, vol. 8, no. 2, pp. 104–108, Apr. 2018, doi: 10.5826/dpc.0802a07, PMID: 29785326, PMCID: PMC5955076.

[26] S. Chamarthi, K. Fogelberg, J. Gawlikowski, and T. J. Brinker, "Few-shot learning for skin lesion classification: A prototypical networks approach," *Informatics in Medicine Unlocked*, vol. 48, p. 101520, 2023, doi: 10.1016/j.imu.2024.101520.

[27] G. S. Ghalejoogh, H. M. Kordy, and F. Ebrahimi, "A hierarchical structure based on Stacking approach for skin lesion classification," *Expert Systems With Applications*, vol. 145, p. 113127, 2020, doi: 10.1016/j.eswa.2019.113127.

Efficient Image Retrieval and Activity Recognition in Vietnamese Daily Life using AI

1st Pham Le Duc Thinh
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
pldthinh.private@gmail.com
(Corresponding author)

2nd Nguyen Thanh Binh
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
2002ngthanhbinh@gmail.com

3rd Minh Khang Vo
AIoT Lab VN
International University, VNU-HCM
Ho Chi Minh City, Vietnam
vomkhang35@gmail.com

4th Ho Duong Huynh Tan
International University, VNU-HCM
Ho Chi Minh City, Vietnam
hoduongq@gmail.com

5th Tuan Kiet Nguyen
University of Technology, VNU-HCM
Ho Chi Minh City, Vietnam
kietnguyentuan9881@gmail.com

***Abstract*—In the context of increasing volumes of multimedia data, information retrieval has become both challenging and resource-intensive. This study introduces a multimedia search system that combines image and text data to enhance content retrieval capabilities for users. Utilizing a dataset from the HCMC AI Challenge 2024, which encompasses 200 hours of video and millions of frames, the system employs the TransNet V2 model for preprocessing, extracting significant frames, and reducing data size from 360 GB to 11.65 GB while preserving essential details. Notably, the proposed method optimizes the processing of Vietnamese text and image data, improving both accuracy and efficiency in information retrieval. A user-friendly interface developed using HTML, CSS, and JavaScript enhances the input experience. Evaluated using the Recall@K metric, models such as PhoBERT and GPT-3.5 demonstrate superior retrieval accuracy compared to Google Translate, owing to their ability to capture nuanced semantic meanings. Experimental results indicate that the system achieves high accuracy in retrieving Vietnamese images, supported by an advanced architecture and the FAISS indexing system that facilitates effective similarity searching. This research paves the way for developing multimedia retrieval systems aimed at improving the handling of complex queries and large-scale datasets.**

***Index Terms*—Multimedia Search, Image and Text Data, Lifelog Retrieval, Data Preprocessing, NLP, AI, Deep Learning.**

I. Introduction

Effective information retrieval has become more challenging due to the explosive rise of digital media [1], [2]. It frequently takes a lot of time and effort to find specific information in the large amount of multimedia content, particularly when using archives or remembering historical events. Textual components, including descriptions and captions, are essential for enhancing discoverability and improving retrieval techniques that rely on images or videos [3], [5]. This paper proposes a multimodal data integration system that aligns text and images to improve media retrieval. The approach allows for intuitive, context-aware searches with minimal manual annotation, adapts to different domains, and incorporates personalization features through the use of CLIP's language understanding and cross-modal embeddings [4]. In contemporary multimedia archives, this strategy provides a scalable and affordable substitute for conventional techniques [5]. The contributions of this work are threefold:

- **Development of an Intuitive Retrieval System:** An easy-to-use system was designed to streamline access to corporate multimedia archives, enabling users to retrieve relevant content efficiently.
- **Evaluation of Language Models:** A comparative analysis of language models, including Google Translate, GPT-3.5, and PhoBERT, highlights their effectiveness in interpreting and integrating textual and visual data. The study emphasizes the models' contextual accuracy and applicability in enhancing content discoverability, particularly in Vietnamese-language archives.
- **Optimization via Hierarchical Training:** Optimized the system using a hierarchical training strategy to enhance retrieval accuracy and improve overall user experience.

II. Background and Related Work

In recent years, multimedia information retrieval has made significant strides in enhancing the accuracy, scalability, and efficiency of systems dedicated to processing and interpreting both video data. These advancements have been driven by the need to meet the demands of increasingly complex and varied user queries, which often exceed the capabilities of traditional keyword-based methods. To address this, modern MIR systems leverage a combination of sophisticated algorithms and hybrid approaches that go beyond text-based analysis to include visual and contextual interpretations of multimedia content.

The inclusion of advanced technologies such as deep learning, machine learning, query expansion, and hybrid model architectures has further propelled the performance of MIR systems. Deep learning, for instance, enables the system to

identify complex patterns, greatly enhancing the precision and adaptability of retrieval processes. Machine learning algorithms add to this by continuously refining their models based on new data inputs. Query expansion techniques allow for a deeper understanding of user intent by capturing contextual nuances, while hybrid models integrate various analytical layers, creating a more holistic approach to multimedia data interpretation. As a result, these innovations have rendered MIR systems highly responsive to dynamic and heterogeneous data types across numerous domains, marking a significant evolution in how information retrieval systems can cater to the diverse and sophisticated needs of modern users.

A. Lifelog Moment Retrieval

Kamada et al. (2018) introduced a "lifelog moment" retrieval system designed to help users locate specific moments within their personal lifelog data [6]. This approach combines image-based concepts with text-based query expansion, enabling a more accurate understanding of user queries. The system leverages convolutional neural networks (CNNs) to analyze images from lifelog data while utilizing natural language processing (NLP) models to interpret query context.

The system, tested with the ImageCLEF 2018 dataset (over 10,000 images and videos), showed a 15% increase in retrieval accuracy and reduced search time from 5 to 3 minutes. The concept-based expansion improves both accuracy and the system's ability to classify activities by time and location, enhancing user interaction and information retrieval.

B. Broadcast News Video Retrieval

Yan and Hauptmann (2007) proposed a hybrid approach that integrates textual and visual data to extract relevant information from broadcast news videos [7]. Their methodology leverages natural language processing (NLP) techniques for text analysis and deep learning models for image analysis. Specifically, Transformer-based models are employed to analyze the textual context within videos, while convolutional neural networks (CNNs) are utilized to extract critical information from visual content.

The proposed method improved retrieval accuracy by 20% compared to single-source approaches (text or image). It also reduced processing time per retrieval from 10 to 6 seconds, enhancing efficiency and user experience. This approach enables better contextual analysis, more accurate results, and easier user engagement.

C. Web Image Retrieval Using Multi-Instance Learning

Chen et al. [8] introduced a web image retrieval method based on Multi-Instance Learning with Class-Preserving Bags (MIL-CPB). This approach addresses the challenge of noisy and heterogeneous image datasets by using "bags" of images, where each bag contains multiple instances, some of which may be noisy or inconsistent. The MIL-CPB model identifies stable features within these bags to mitigate the impact of noisy data.

TTheir system, using CNNs for image feature extraction and iterative optimization, was tested on a 5,000-image dataset. It achieved a 25% increase in retrieval accuracy, especially for complex datasets, and reduced training time from 50 to 30 hours, enabling self-adjustment to evolving data and improving retrieval in complex scenarios.

D. Deep Learning for Text-Image Retrieval Models

Deep learning has transformed text-image retrieval by bridging the semantic gap between text and visuals. Modern methods use dual-stream architectures like CLIP and Vision Transformers (ViTs), mapping both modalities into a shared embedding space. Through contrastive learning, self-attention, and pretraining, these systems enable effective cross-modal retrieval, improving tasks like image search and captioning [9].

III. Methodology

A. Dataset

The dataset used in this study originates from the HCMC AI Challenge 2024, a prestigious competition in Vietnam that attracts leading AI research groups [10]. The dataset comprises 24 files, each containing 23–24 videos, totaling millions of frames. This extensive data source enables detailed analysis of movements and subtle scene changes. Additionally, a supplementary dataset of keyframes, carefully selected to capture pivotal moments in each video, was constructed to enhance analysis and optimize processing time. With a total duration of 200 hours, this rich and diverse dataset is well-suited for training machine learning models, enabling accurate and reliable video recognition and classification. These advancements pave the way for practical applications in object and event recognition.

B. Data Preprocessing

To efficiently manage a 360GB dataset comprising 200 hours of high-quality 1080p video, a robust preprocessing pipeline was developed to optimize storage and computational resources. The primary objective was to minimize redundancy while preserving essential information by extracting key frames that accurately represented the video content. This was achieved by segmenting the videos into distinct shots and retaining critical frames at scene transitions. The pipeline leveraged TransNet V2, an advanced neural network designed for scene change detection, which has been rigorously validated on benchmark datasets such as ClipShots and BBC Planet Earth [11].

The extracted frames were resized to 1280 × 720 pixels, achieving an average size of 200KB per frame, which reduced the dataset size to 46.6GB. Further compression and resizing were applied using advanced algorithms, downscaling frames to 448 × 224 pixels. This process resulted in a final dataset size of 11.65GB, with an average frame size of 50KB.

By employing a combination of frame selection, H.265 compression, and resolution adjustment, this preprocessing pipeline achieved a 36-fold reduction in storage requirements while maintaining essential visual fidelity. The resulting dataset is well-suited for resource-efficient multimedia

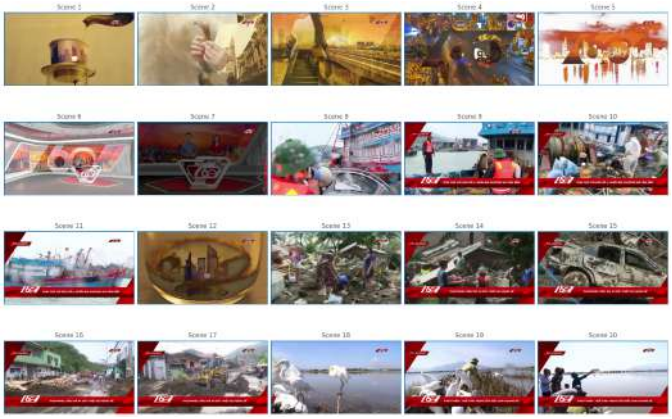


Fig. 1. First 20 Consecutive Frames with Identified Transition Points in a Video.

retrieval tasks, demonstrating the effectiveness of the applied optimization techniques.

C. Proposed Method

To enhance the accuracy and efficiency of data retrieval, a tailored methodology for processing Vietnamese text and image data is proposed. The approach is structured into multiple phases, each designed to address the unique challenges of processing Vietnamese text and integrating visual information. The overall framework of the proposed methodology is depicted in Figure 2, which highlights the interaction between the text and image processing components.

1) Data Input: The process begins with the Data Input stage, where users can enter the Vietnamese text that requires processing. In the context of Vietnam, the system prioritizes the use of Vietnamese queries to meet the practical needs of users. To optimize user experience, the input interface is designed to be user-friendly, incorporating features such as autocomplete and keyword suggestions, which facilitate quick and accurate data entry.

2) Text Preprocessing: Once the data has been entered, the Text Preprocessing stage occurs. This stage involves the cleaning and normalization of the text, encompassing specific steps such as:

- **Vietnamese Text Processing:** For Vietnamese, PhoBERT and GPT-3.5 are recommended due to their advanced understanding of the language's grammatical and tonal characteristics. GPT-3.5 excels in accuracy and broad adaptability, while PhoBERT is specifically optimized for Vietnamese, making them superior to general models.
- **Removal of Special Characters:** Special characters can disrupt natural language processing algorithms. By removing these characters, the text becomes clearer and easier to analyze.
- **Conversion of Text to Lowercase:** This conversion minimizes complexity in text processing, ensuring that similar words are treated as a single unit, thereby improving the accuracy of subsequent processing steps.
- **Extraction of Important Information:** This phase involves identifying and extracting key entities (such as names, locations, and organizations) and main keywords. This focus on essential content supports subsequent analytical objectives.

3) Embedding Generation: The core of the methodology lies in generating embeddings from the preprocessed text. Two distinct models are employed for this task:

- **PhoBERT:** This model is specifically designed for the Vietnamese language, utilizing a transformer architecture to create contextual embeddings. PhoBERT not only comprehends grammar but also captures the meanings of words in specific contexts, resulting in more accurate embedding vectors. The use of PhoBERT enhances natural language processing capabilities, serving various applications such as text classification, sentiment analysis, and information retrieval.
- **GPT-3.5 Turbo:** Integrated with PhoBERT, this model improves the semantic quality of the embeddings. GPT-3.5 Turbo has the ability to understand and generate natural language flexibly, enhancing contextual relevance and supporting translations between Vietnamese and English. This combination increases accuracy and expands the system's applicability across multilingual tasks.

4) Storage: Once the embeddings are generated, they are stored in an indexing system to optimize query performance. The indexing mechanism employs query optimization techniques such as tree indexing and clustering, ensuring rapid retrieval of relevant information during subsequent searches. This storage system is also designed to be scalable, allowing for the addition of more data and improvements in performance over time.

5) Image Embedding Generation: For image data, an additional process is implemented:

- **Image Embedding Generation:** Images from the database are processed through the CLIP (Contrastive Language-Image Pretraining) model to generate embedding vectors. CLIP is capable of integrating information from both text and images, effectively representing images in a high-dimensional space. This enables the system to understand and classify images based on related textual context, facilitating image search and retrieval based on natural language descriptions.

6) FAISS Indexing: FAISS (Facebook AI Similarity Search) is a robust toolkit designed for efficient vector similarity search and indexing, enabling rapid and scalable operations even with large datasets. It is particularly well-suited for tasks involving large-scale data retrieval, as it organizes embeddings based on similarity metrics such as cosine similarity or Euclidean distance. To optimize both speed and accuracy, FAISS offers a range of index types, including quantized or hierarchical indexes for faster approximate searches and flat indexes for precise, exact searches. This flexibility makes FAISS an invaluable tool for high-performance, large-scale information retrieval tasks.

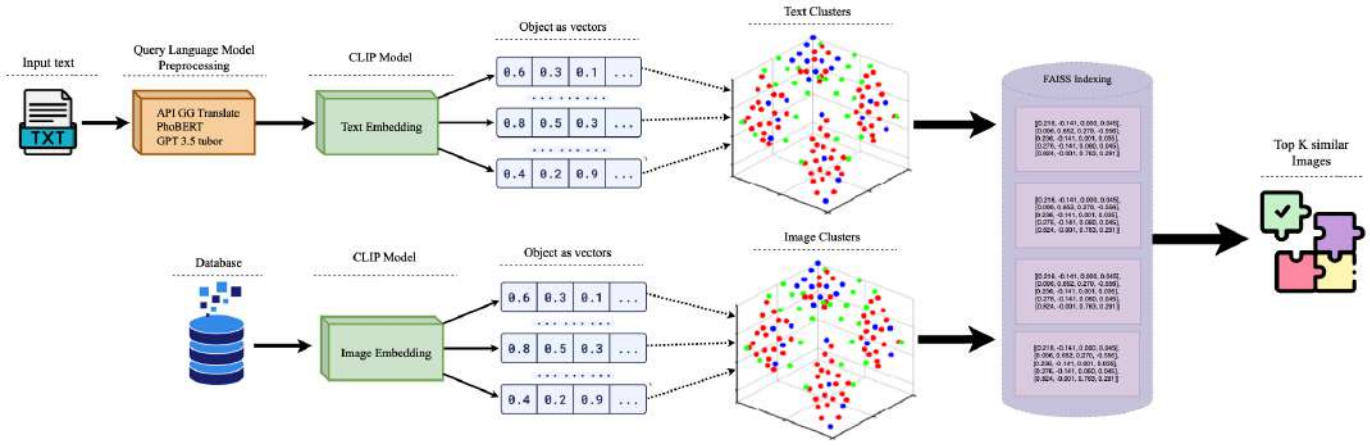


Fig. 2. Standard Workflow of the Proposed System.

FAISS's ability to efficiently manage billions of vectors through advanced data structures and algorithms is one of its key strengths. Additionally, it supports both CPU and GPU processing, allowing for adaptability across various hardware configurations and ensuring optimal performance in diverse environments. This scalability makes FAISS an indispensable tool for applications where the speed and accuracy of vector similarity searches are critical, such as in image and text retrieval, recommendation systems, and natural language processing.

7) Output: The final output of the integrated system is a list of the top similar images, ordered by the highest similarity (top K similarity) based on the input query. The system not only performs image retrieval but also evaluates and categorizes images based on their relevance to the query content.

D. Interface Development for the Query System

The query system interface was developed using a user-centric approach, with a primary focus on optimizing usability and operational efficiency through the reduction of input requirements. Constructed with HTML, CSS, and JavaScript, the interface employs AJAX for asynchronous data retrieval from a Flask-based backend API, thus supporting real-time interactions without necessitating page reloads. This design aligns with established usability heuristics, which emphasize minimizing user actions and ensuring responsive feedback as core aspects of effective interface design.

Traditional query interfaces require users to fill out multiple fields, increasing cognitive load and the likelihood of errors from incomplete or inconsistent entries. To simplify this process, our proposed interface consolidates all inputs into a single field. It uses advanced string-processing algorithms to dynamically interpret user intent by analyzing embedded keywords, automatically adjusting search criteria as needed. This streamlined approach aligns with human-computer interaction research, demonstrating that simpler input methods reduce cognitive load and enhance user satisfaction.

The single-field query approach enhances adaptability by dynamically processing input, eliminating the need for traditional field-specific navigation. Additionally, the design ensures cross-platform accessibility through responsive principles, enabling seamless use on both desktop and mobile devices. Research shows that adaptable interfaces improve accessibility, catering to a diverse range of users.

Future advancements to this system could include the integration of natural language processing (NLP) techniques, thereby enabling a more sophisticated interpretation of user input. Such an enhancement would facilitate the accommodation of complex, conversational queries, refining the system's intuitive design and extending its accessibility to a broader user base.

E. Evaluation

The performance of the image retrieval system is evaluated using two primary metrics: **Recall@K** and **Processing Time (ms)**. These metrics are chosen to balance accuracy and computational efficiency, ensuring the system meets its operational goals.

*1) **Recall@K**:* Recall@K measures the system's ability to retrieve relevant items within the top K results. It reflects how effectively the system prioritizes relevant results, ensuring critical items are easily accessible to users.

$$\text{Recall@K} = \frac{\text{Number of relevant results in top } K}{\text{Total number of relevant results}} \quad (1)$$

This metric is essential for assessing whether the system consistently ranks relevant results prominently, aligning with the goal of delivering high-priority information efficiently.

*2) **Processing Time (ms)**:* Processing Time (ms) evaluates the system's computational efficiency by recording the time taken to process a query and return results. A lower processing time signifies better performance, making this metric critical for real-time applications and scalability in handling large datasets.

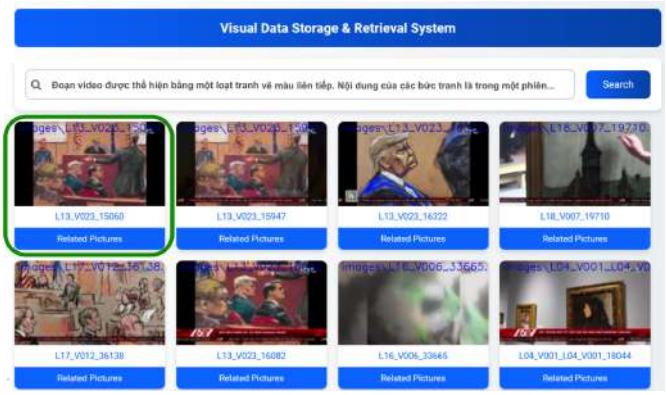


Fig. 3. Assessing Query Relevance Through a Randomly Generated Query.

IV. Experimental Results

A. Enhancing the Query System Interface Hosted on Ngrok

Utilizing Ngrok to host the query system significantly enhances the interface by enabling secure, real-time access to the application's API for remote testing and validation. By creating secure tunnels to the local server, Ngrok facilitates iterative development and testing cycles without the complexities associated with traditional deployment configurations. This integration allows seamless interaction with the query interface from multiple locations, thereby supporting dynamic, end-user feedback and enabling swift modifications to both backend processes and interface components. Additionally, data transmissions within the interface are secured through Ngrok's encrypted HTTPS endpoints, adhering to industry standards for data protection and secure access. This approach not only augments system accessibility for developers and stakeholders but also provides an efficient framework for integrating advanced functionalities, including cross-platform testing and real-time analysis of user inputs.

B. Results

The evaluation experiments were conducted on a dataset of Vietnamese queries derived from real-life contexts in Vietnam. Two primary metrics, **Recall@K** and **Processing Time**, were employed to evaluate the effectiveness and efficiency of the proposed image retrieval system. Recall@K measures the system's ability to retrieve a relevant image within the top K results of a ranked list, while Processing Time quantifies the computational cost in milliseconds (ms) for each query. These metrics collectively provide a holistic assessment of the retrieval performance.

An illustrative example demonstrates the evaluation process: each query is evaluated by determining the rank at which the correct result appears, starting from the first position until the desired image is retrieved. Figure 3 illustrates an example query and its corresponding evaluation process.

To enhance the system's understanding of the Vietnamese language and its contextual relevance, three language processing methods were tested: (1) **Google Translate API**, which translates raw Vietnamese text without a language-specific model; (2) **PhoBERT**, a pre-trained model specifically tailored for Vietnamese language processing; and (3) **GPT-3.5 Turbo**, a general-purpose large language model. Table I presents the Recall@K scores and Processing Time (ms) achieved by each method for K values ranging from 1 to 15.

PhoBERT and GPT-3.5 Turbo outperform the Google Translate API in all metrics for image retrieval. PhoBERT achieves a Recall@5 of 0.70 (vs. 0.47 for Google Translate), while GPT-3.5 Turbo excels with a Recall@10 of 0.93, slightly surpassing PhoBERT's 0.90. All methods reach a Recall@15 of 1.0, but PhoBERT and GPT-3.5 Turbo retrieve relevant images more effectively at higher ranks. However, processing times are slightly longer: 450 ms for Google Translate, 520 ms for PhoBERT, and 540 ms for GPT-3.5 Turbo.

These results highlight the advantages of Vietnamese-specific models like PhoBERT and advanced models like GPT-3.5 Turbo in capturing contextual nuances, improving retrieval accuracy, and enhancing performance in Vietnamese-language tasks.

C. System Design

The proposed system employs a three-tier architecture to support a robust visual data storage and retrieval system. This design is optimized for handling large-scale multimedia datasets and facilitating efficient query-based image retrieval.

1) Tier 1: Backend: In Tier 1 (Backend), the system focuses on data preparation, embedding generation, and interface setup for external interactions. The Extract, Transform, Load (ETL) process ensures that raw data is collected from various sources, transformed into a structured format, and loaded into a database ready for retrieval. The Flask API serves as a conduit between the backend and other system components, allowing external applications and services to interact seamlessly with the data. A core feature of this layer is the OpenCLIP-based image-text embedding process, which converts both visual and textual data into a shared vector space. This enables the system to understand semantic similarities between images and text, making it possible to match a user's query with relevant images based on content.

2) Tier 2: Middleware: Tier 2 (Middleware) functions as the intermediary, managing data flow between the backend and frontend. It includes several critical components that ensure data storage, indexing, and querying are performed efficiently. The FileSystem stores multimedia data, enabling direct and rapid access to the files when needed. The Data Server manages file access and retrieval, coordinating data requests across the middleware layer. Additionally, a Query Server works closely with the FAISS (Facebook AI Similarity Search) index to facilitate similarity search across high-dimensional data. FAISS allows for scalable and efficient retrieval of image-text embeddings, enabling fast similarity searches based on vectorized data from the backend.

3) Tier 3: Frontend: Tier 3 (Frontend) provides the user-facing interface, enabling users to interact with the system via search queries and view relevant images in response to their input. The user interface displays a search bar and presents

TABLE I
COMPARISON OF EVALUATION METRICS FOR DIFFERENT ALGORITHMS

Algorithm	Recall@1	Recall@5	Recall@10	Recall@15	Processing Time (ms)
Google Translate API	0.13	0.47	0.87	1.0	450
PhoBERT	0.17	0.70	0.90	1.0	520
GPT-3.5	0.17	0.70	0.93	1.0	540

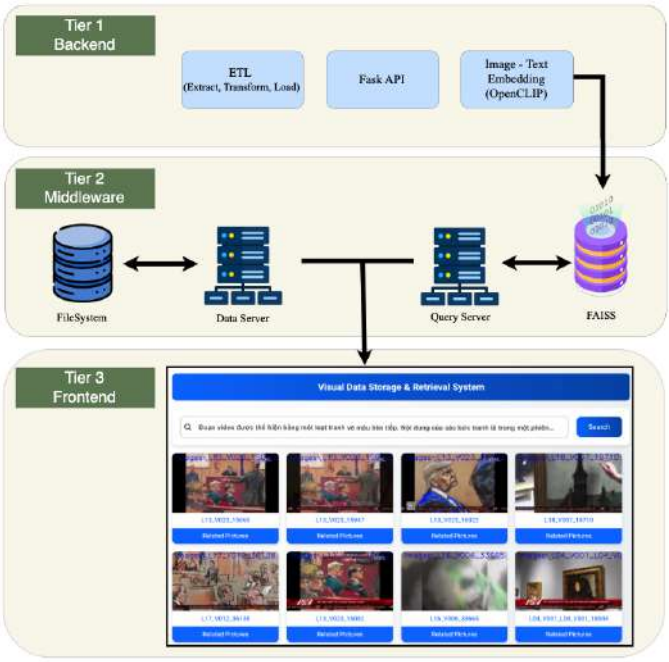


Fig. 4. Three-Tier Architecture of the Visual Data Storage and Retrieval System.

search results in a visually organized manner, allowing users to explore related images that match their query based on both visual and textual relevance. This tier emphasizes user experience, presenting search results quickly and in a user-friendly layout.

In this design, users submit queries through the frontend, which are processed by the middleware's query server and FAISS index to retrieve similar image embeddings. The results are then sent back to the frontend, displaying relevant images in response to the query. Figure 4 illustrates the system's end-to-end flow, showing interactions among the user interface, middleware, and backend components.

V. CONCLUSION

A semantic-based image retrieval system has been developed to enhance visual information retrieval, offering users a streamlined and efficient experience. The system employs a three-layer architecture that includes data storage, transformation, and retrieval mechanisms. This design is tailored to handle large-scale multimedia datasets efficiently. By integrating FAISS for similarity-based searches and OpenCLIP to map data into a shared vector space, the system achieves high accuracy and responsiveness to semantic queries.

Experimental evaluations demonstrate that language models like PhoBERT and GPT-3.5 significantly enhance Vietnamese retrieval tasks by capturing contextual and cultural nuances. These models achieve higher recall and accuracy, as evidenced by strong Recall@K scores, highlighting the advantage of localized models over generic tools like the Google Translate API. This improvement is particularly beneficial for applications requiring precise and culturally relevant information retrieval. While PhoBERT and GPT-3.5 have slightly longer processing times, their superior retrieval accuracy justifies the marginal increase in response time.

Despite these achievements, challenges remain in scaling for extensive datasets, improving processing speed, and managing diverse and complex queries. Future efforts will focus on refining natural language processing techniques, enhancing system scalability, and incorporating adaptive learning mechanisms to address these limitations. By overcoming these challenges, the system aims to provide a comprehensive and user-centered solution for advanced visual information retrieval that meets the growing demands of modern applications.

ACKNOWLEDGMENT

We gratefully acknowledge the AIoT Lab Vietnam for their invaluable support in this research.

REFERENCES

[1] D. Xu, "Text-Image Data Fusion in Contextual Search", IEEE Transactions on Multimedia, vol. 23, 2022.
[2] H. Wang, et al., "Multimodal Search: Text and Image Retrieval Models", Journal of Information Science, vol. 47, no. 3, 2021.
[3] A. Johnson and L. Lee, "Reducing Cognitive Load in Multimodal Information Retrieval", ACM Computing Surveys, 2021.
[4] K. Zhang, "User-Centered Design for Multimodal Retrieval Systems, Human-Computer Interaction Journal", 2022.
[5] R. Smith, "Efficient Access to Digital Archives: A Cost-Effective Approach", Digital Media Journal, 2020.
[6] H. Kamada, et al. "Lifelog Moment Retrieval System Based on Concept Expansion", CEUR Workshop Proceedings, 2018.
[7] Yan, R., Hauptmann, A. G. (2007). A review of text and image retrieval approaches for broadcast news video. Information Retrieval, 10.
[8] Li, Wen, et al. "Text-based image retrieval using progressive multi-instance learning", 2011 international conference on computer vision. IEEE, 2011.
[9] Radford, A., et al. (2021). Learning Transferable Visual Models From Natural Language Supervision. OpenAI. Available at: https://openai.com/research/clip
[10] AI Challenge. *Event Query from Video Dataset*. Accessed: 19.08.2024. https://docs.google.com/spreadsheets/d/1mO3zS79L1HMLZ-BLpyy8En9RROOElms5DS$_Gi1gKiU$.
[11] Souček, T., Lokoč, J. (2020). Transnet v2: An effective deep network architecture for fast shot transition detection. arXiv preprint arXiv:2008.04838.

AI Applications for Students in University

Dung Le Ngoc Hung
Faculty of Computer Science and Engineering
University of Technology, VNU-HCM
Ho Chi Minh City, Vietnam
dung.lengochung@hcmut.edu.vn

Hien HoThu
Faculty of Geology
University of Science, VNU-HCM
Ho Chi Minh City, Vietnam
hthien@hcmus.edu.vn

***Abstract*— Significant strides in AI development over the past several years have improved both its capabilities and accessibility, particularly in education. Effectively harnessing these innovations offers substantial benefits for university students. However, existing literature predominantly addresses the broader educational landscape, with limited emphasis on student-centered perspectives. This review aims to provide a comprehensive overview of AI applications in learning from the perspective of students in university by reviewing recent studies from 2022 to 2024. It will explore how students can utilize AI-driven tools for personalized learning in foundational subjects, language learning, honing their writing skills, obtaining career guidance and doing research. The findings indicate that general-purpose large language models like ChatGPT are applicable across all these areas, alongside other specialized tools.**

Keywords— applied AI, AI in education, large language models, AI for university students

I. Introduction

Over the last few years, Artificial Intelligence (AI) and related topics have gained traction. While AI was not well-received during its formative years due to uncertain outcomes and computational issues, its advancements in this century are changing the landscape of many fields, including education.

For instance, the emergence of ChatGPT in November 2022 became a turning point in society's adoption of AI in different fields. The release of GPT models also served as the foundation for tailoring chatbots to different purposes, spurring the development of similar chatbots for educational use. Major examples include Coursera or Udacity's AI Chatbots trained on their own academic datasets to address students' questions, or Khan Academy's Khanmingo powered by GPT-4 that caters to individual students' learning abilities on subjects such as mathematics, language learning and programming. Similarly, Duolingo also features advanced learning algorithms that deftly adjust difficulty accordingly to learner abilities to keep them engaged [1].

These innovations were realized during one of the first times when advances in AI have truly been felt by people outside the deep learning community. A survey by Pew Research Center revealed that 57% of American adults have heard of ChatGPT, while the percentage for Asian adults was 78% [2]. The younger demographics, especially people aged 18 to 29, were more aware of ChatGPT than any older age groups.

Such innovations were warmly welcomed by students. They perceive chatbots as highly motivating and useful as personal assistants to aid them with tasks such as coding, language learning or improving writing skills. [03] Such activities are often knowledge-intensive; hence AI is a relevant tool for support.

Numerous literature reviews have been written on Artificial Intelligence in Education (AIEd). A paper by Firuz Kamalov et al has pinpointed four major themes of AI's application in education: Personalized Learning, Intelligent Tutoring Systems, Assessment Automation, and Teacher–Student Collaboration [04]. This categorization, however, focuses on the whole educational landscape rather than focusing on the students' perspective, thereby omitting specific use cases such as language learning.

To ensure that this review centers around the students' perspective, it will categorize the AI's applications into five categories: learning foundational subjects (such as university advanced mathematics), learning languages, improving writing skills, obtaining career guidance and doing research.

II. Methodology

The numerous articles on AI applications for students in their studying analyzed in this paper were collected through the "SCOPUS" and "ResearchGate" research systems. They were searched using relevant keywords to the AI applications topics that we are interested in.

III. Applications for students

A. Learning foundational subjects

Research on applying AI to solve mathematical problems has been underway since the early developmental stages of the field. A notable example is James Slagle's 1961 program, SAINT (Symbolic Automatic Integrator), which followed the 1956 Dartmouth conference that formally recognized AI as a distinct field by only five years. SAINT employed heuristic techniques to tackle symbolic integration problems, simplifying complex integrals through pattern recognition and transformations, a method that mimicked human problem-solving processes in mathematics [5]. This groundbreaking work in symbolic mathematics laid the foundation for modern AI-driven applications.

As AI progressed into the 21st century, more applications were developed to aid students in solving complex problems. A prime example is Wolfram Alpha, which has become an essential tool for students tackling subjects such as math and science. One study on senior high school students found that 97% of participants used Wolfram Alpha to help them cope with subjects like math and science [6]. The research showed that the app acted as a valuable scaffolding tool, allowing students to break down complex problems into step-by-step solutions, which improved their understanding of difficult subjects. Another prominent name is GeoGebra, a mathematics software integrating geometry, algebra, calculus, and statistics into a single platform. It enables real-time interaction and visualization of mathematical concepts, providing not only a deeper understanding of the concepts but also instilling curiosity among students to learn more [7].

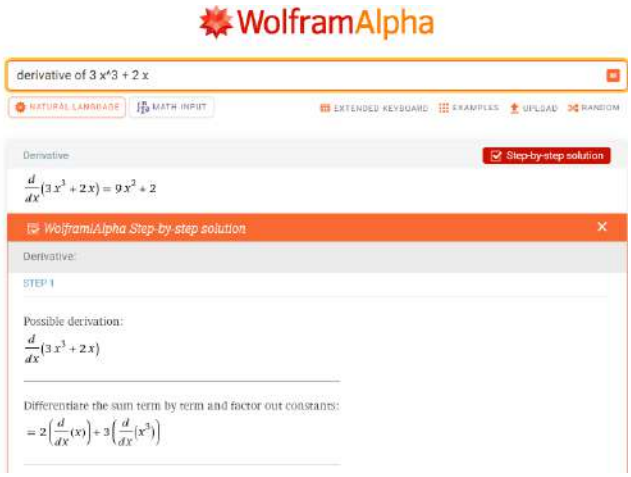


Figure 1: Example derivative calculated by WolframAlpha. (Retrieved from www.wolframalpha.com)

While these products have shown an excellent ability to handle specialized or well-defined mathematical problems, such as differentiating, integrating or calculating limits, they cannot generalize. This means the students have to choose the correct type of exercises they are doing to obtain the answer. It also means that they cannot help with verbal problem statements and subjects closely intertwined with math such as physics or chemistry. This inherent limitation could be attributed to the lack of adequate natural language processing development to convert human-readable words to machine-understandable logic representation.

ChatGPT and other LLMs offers a distinct advantage over traditional AI models by being able to understand natural languages. This is due to their architecture as Transformer models pre-trained on vast internet text data, learning grammar, vocabulary, and context through unsupervised learning. Such data also equip them with a rich repository of general knowledge to not only solve problems but also to provide students with theoretical basis those problems are based on.

In physics studying, Y. Liang has demonstrated how ChatGPT can be utilized to analyze complex physics problems by breaking them down into simpler components, offering learners a clearer path to understanding [8]. For example, ChatGPT can take a challenging concept like projectile motion and decompose it into its horizontal and vertical components, helping students focus on one aspect of the problem at a time. This aligns with cognitive load theory (CLT), which emphasizes reducing extraneous cognitive load by breaking complex information into manageable chunks.

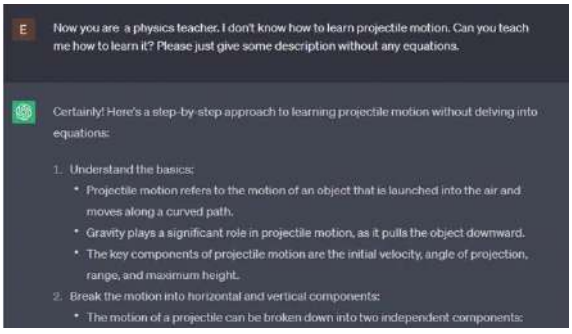


Figure 2: Example explanation of physics concept by ChatGPT. (Retrieved from reference [8])

Furthermore, ChatGPT can provide scaffolded learning experiences by generating step-by-step hints, allowing learners to solve problems progressively rather than being overwhelmed by a complete solution at once. For instance,the research used a worked example from a physics textbook "A car of mass 500kg is travelling at 20m/s. The driver sees a red traffic light ahead and slows to a halt in 10s. Calculate the braking force provided by the car.". ChatGPT was able to correctly identify the known and unknown variables, break the problem down to five steps, clarifying what was needed to be done along with the required formulas, and the end results, similar to the detailed instructions from a book.

ChatGPT can not only solve physics problems but also generate new problems based on the same principles. Using a permute-instruction prompt, ChatGPT can modify variables to create different exercises while maintaining the core physics principles, such as Newton's second law. While it is obvious that this capability allows instructors to generate varied homework materials, giving students opportunities to practice related concepts, it also applies to self-learning students wishing to reinforce their knowledge on certain principles through exercises with answers on request. Additionally, ChatGPT can create problems that incorporate multiple physics principles, enhancing problem difficulty and helping students tackle complex topics.

Y. Liang's work also highlights some other benefits of ChatGPT's use in physics education. It provides immediate feedback, helping students quickly identify and correct mistakes, which is crucial for effective learning. ChatGPT's adaptive learning capabilities adjust explanations based on individual needs, whether a student requires a simplified overview or an in-depth breakdown. Its 24/7 availability allows students to learn at their own pace, outside of traditional study hours. ChatGPT also promotes interactive learning by engaging students in dialogue, encouraging deeper comprehension. Additionally, it helps bridge knowledge gaps by reinforcing fundamental concepts, creating a personalized and effective learning experience for students in physics.

B. Learning languages

There have also been momentous innovations made in artificial intelligene applicable to language learning. The benefits of AI tools in language learning has been extensively studied. According to Roxana Rebolledo Font de la Vall et al. (2022) [9], these tools improve efficiency by automating tasks and providing immediate feedback, allowing learners to progress faster. They also personalize the learning experience, adapting content to meet individual learners' needs, making lessons more engaging and tailored to specific learning styles. Additionally, AI tools allow users to learn multiple languages simultaneously, offering flexibility for those looking to expand their linguistic skills. The accessibility of AI-powered tools, available via mobile apps or online, enables learners to study from anywhere, adding convenience to language acquisition. Cost-effectiveness is another advantage, as many platforms provide free or affordable options compared to traditional language learning methods like in-person classes or tutoring. AI tools also introduce learners to cultural elements through real-life scenarios, helping them gain a deeper understanding of different cultures. For English as a foreign language (EFL) education, these tools are transformative, offering personalized experiences and enabling teachers to provide more effective guidance while improving learning outcomes.

Duolingo is a prime example of how AI has transformed language learning. The app uses AI-driven systems, such as their flagship Birdbrain, to personalize lessons, ensuring learners receive content suited to their skill level. By analyzing past performance, Duolingo offers exercises that challenge users without overwhelming them, thus enhancing engagement and progression. The platform also incorporates gamified elements, like points, levels and hearts to maintain user motivation. AI technologies within Duolingo, such as natural language processing and spaced repetition, help refine the curriculum, combined with the gamification of language learning, provide immediate feedback, fostering faster and more efficient learning [1].



Figure 3: Screenshot of gamified learning experience with Duolingo. (Retrieved from duolingo.com)

Using LLMs can also be another effective way for language learning for independent learners. They can act as advanced conversational partners, capable of simulating dialogues in multiple languages while offering grammar corrections, vocabulary enhancement, and pronunciation support, all aimed at improving fluency through practical engagement. Among them, ChatGPT is an innovative chatbot with remarkable capabilities in delivering coherent and contextually appropriate responses tailored to user inputs. By utilizing vast amounts of text data, ChatGPT can produce content in various styles and tones, making it a versatile tool for generating text suitable for a wide range of purposes. For English as a Foreign Language (EFL) content creation, OpenAI's ChatGPT holds great potential to produce valuable English content that could benefit EFL learners.

According to a research by J. C. Young et al (2023) [10], ChatGPT-generated dialogues are easy for English learners to understand, as they avoid confusing expressions and unnecessary words. These dialogues are well-suited for students at the A2 level of English proficiency, who are beginning to grasp basic grammar, vocabulary, and communication skills. A2 learners can comfortably read simple texts, write short sentences about their daily lives, and engage in basic conversations. Additionally, the reading difficulty scores confirm that ChatGPT's content is appropriate for learners who need straightforward, clear material to support their language development.

Moreover, OpenAI's recent release of ChatGPT's voice feature for free has taken language learning to the next level. By integrating natural speech capabilities, learners can now engage in verbal conversations with ChatGPT, practicing pronunciation, intonation, and fluency in real-time. This feature allows for a more immersive and interactive learning experience, simulating real-world conversations that help learners strengthen their spoken English skills. With these advancements, ChatGPT is not only a powerful tool for reading and writing practice but also a valuable resource for improving listening and speaking proficiency, making language learning more accessible and comprehensive.

For languages other than English, promising results can also be observed from Belle Li et al's 2023 study on ChatGPT's capabilities for language learning [11]. In high-resource Romanized languages like French, Spanish, and German and Portuguese, ChatGPT has proven to be an accessible and practical tool for language learning. Communities on platforms like YouTube have praised its ability to streamline the learning process, particularly by consolidating information and saving time. Learners appreciate ChatGPT's ability to generate contextually relevant examples and customized exercises, helping them practice specific vocabulary and grammatical structures. For instance, French educators have noted ChatGPT's capability to provide satisfactory answers on not only French vocabulary, grammar but also the culture. For Portuguese learning, ChatGPT can offer well-explained responses, and tailor questions to different proficiency levels. These features make it a valuable asset for both beginners and intermediate learners who need additional support beyond traditional textbooks. For non-Romanized languages with high resource like Japanese and Chinese, ChatGPT also have great affordances. In languages like Chinese, one of the primary advantages is ChatGPT's ability to break down vocabulary and explain word nuances. This is particularly valuable for challenging word pairs, such as distinguishing between 经历 and 经验, 引起 and 导致. These distinctions are critical for learners, and ChatGPT's detailed explanations have been well-received. Additionally, its ability to generate texts for reading and vocabulary practice, coupled with targeted sentence breakdowns, offers useful learning opportunities. For instance, the prompt "explain this sentence word by word with POS" helps learners understand the grammatical structure of complex sentences in Japanese, which lacks word spacing.

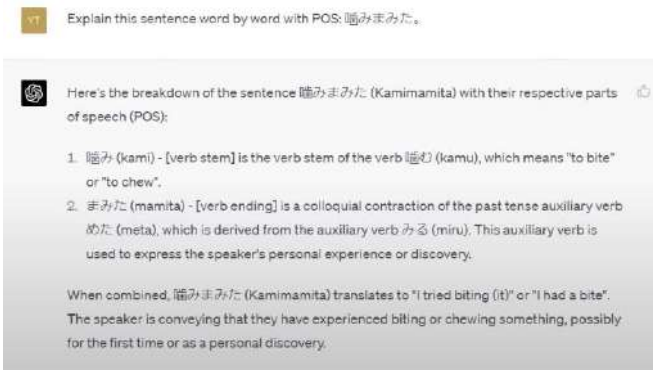


Figure 4: Screenshot of language learning Japanese with ChatGPT (retrieved from: https://youtu.be/HDEvnsvXBww)

C. Improving writing skills

One of the earliest AI-based writing assistant software was Grammarly, which provides targeted, real-time feedback on correctness, clarity, engagement, and delivery. According to Linqian Ding (2024) [16], Grammarly assists students by flagging errors and offering instant corrections in grammar, spelling, and formatting. This helps students identify and learn from mistakes as they write, reinforcing effective habits over time. Its readability score, which evaluates sentence and word length, gives students insight into the clarity of their work, allowing them to adjust for more concise and understandable

writing. For students looking to tailor their writing for different audiences or contexts, Grammarly Premium's tone adjustments support flexibility, suggesting formal, confident, or neutral expressions as needed. Additionally, the synonym suggestions expand students' vocabulary by helping them replace repetitive or basic words with more varied choices, fostering stronger lexical skills. Finally, grammarly's integration across platforms enables students to receive consistent writing assistance across different devices, making it a versatile tool for their academic development.

Jessie S. Barrot (2021) [17] found that students using Grammarly showed significant improvement in writing accuracy compared to those in the control group. Key features contributing to this included Grammarly's metalinguistic explanations, suggested corrections, and color-coded feedback lines, which helped students understand grammar rules and recognize common errors. Additionally, students appreciated the tool's feedback autonomy, which empowered them to accept or reject suggestions, enhancing their engagement and self-directed learning.

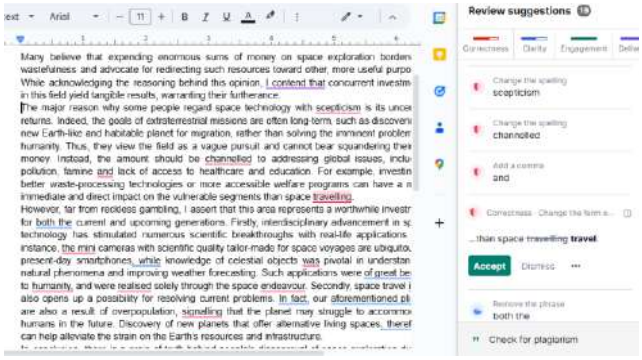


Figure 5: Screenshot of feedback on essay from Grammarly.

While Grammarly's benefits are proven, they center around vocabulary and grammar development. In recent research, LLM tools like ChatGPT has been increasingly seen as a valuable tool for improving writing skills, especially for language learners preparing for exams like IELTS. In a similar manner, LLM tools provide real-time feedback, offering continuous guidance that helps learners refine their writing. However, the help that students receive from LLMs extend far beyond lexical and syntactical development.

According to Yangyu Xiao et al [12], ChatGPT can serve as a peer tutor, providing instant feedback and personalized support. Students can use ChatGPT to improve their essays by asking for guidance on connecting ideas, adding relevant examples, and receiving feedback on the logical flow of their writing. One student found ChatGPT's feedback on sentence structure and coherence helpful in achieving a higher band score in IELTS, reducing her reliance on visiting the Language Centre for support. Similarly, another student utilized ChatGPT for writing guidance in various sections of technical proposals. He emphasized the chatbot's ability to clarify how to structure an "Introduction," "Problems," and "Solutions" section, mirroring the role of a tutor who provides step-by-step instructions.

Beyond writing feedback, ChatGPT also helps students generate ideas for essays. In the same study, students used the chatbot for brainstorming when they are were where to start with a writing task. Although the ideas generated by ChatGPT were sometimes broad or vague, they provided students with a foundation to refine and build upon. While it is also noted that generating ideas through ChatGPT is not entirely new, as similar processes could be achieved by using Google or reading book, what sets ChatGPT apart is its efficiency in consolidating information, making the brainstorming process faster.

Such educational use has been proven to have positive implications for students' writing skills by a 2023 study by Cuiping Song and Yanping Song, published in Frontiers in Psychology, specifically examined the impact of ChatGPT on enhancing writing skills and motivation in EFL students [13]. The research demonstrated significant improvements in writing proficiency and motivation for students in the experimental group who used AI-assisted tools like ChatGPT, compared to those in the control group who followed traditional instruction.

The experimental group showed notable gains across several dimensions of writing proficiency, including overall writing, writing content, organization, and language use. For example, the mean post-test score for overall writing in the experimental group was significantly higher ($M = 59.12$) compared to the control group ($M = 45.18$), with a large effect size (Cohen's $d = 0.76$). Similarly, improvements in writing organization and language use were significant, with large effect sizes (Cohen's $d = 0.84$ and 0.88, respectively). Writing motivation also increased substantially in the experimental group ($M = 20.06$), compared to the control group ($M = 18.21$), with a medium effect size (Cohen's $d = 0.52$).

D. Obtaining career guidance

The increasing dynamism in the labor market and the demand for continuous education have prompted the need for innovative approaches to career guidance. Career guidance aims to help individuals recognize their skills and align them with labor market needs while supporting decision-making in education and career paths.

Although limited research is carried out on AI's ability to help with career guidance problems, existing research still show prospects. The vast amount of data that ChatGPT and other LLMs are trained on place them in a position to offer career guidance and academic advisory.

A study by Daisuke Akiba et al has revealed ChatGPT's uses in academic advising. [14] While there are clear limitations due to LLMs' well-known hallucination problems and its inherent ability to consider individual qualifications and characteristics, the quality of responses exceeded researchers' expectations, with ChatGPT offering comprehensive, professional, and supportive answers. A notable feature was its emphasis on advising individuals to consult other resources for further clarification. In certain cases, ChatGPT's responses surpassed what many human academic advisers could deliver, as they were clear, concise, and often more factually accurate.

The answer to the first question, "What are the steps involved in earning certification as an elementary school teacher in New York State?", highlighted the response's strengths, including simplicity and clarity. However, two areas of improvement were identified: the answer did not specify that a master's degree is required for full certification and neglected alternative pathways to certification. Despite these omissions, the response's thoroughness and timeliness impressed the evaluators, although some outdated

information, such as the inclusion of the ALST exam (eliminated in 2017), pointed to areas needing fact-checking.

The second question explored the disadvantages of elementary school teaching, and ChatGPT produced a balanced and nuanced answer. The response encompassed both tangible concerns, like salary, and abstract issues, like emotional demands, making it an exemplary model for academic advisers. It also provided practical advice on mitigating challenges, which the evaluators noted as a detail human advisers might overlook.

For the third question about long-term career expectations in teaching, ChatGPT again delivered a comprehensive and balanced answer, addressing professional satisfaction, collaboration, career growth, and retirement. The answer's coverage of a wide range of career aspects was particularly impressive.

The answers to subsequent questions, including career outlook and self-assessment of aptitude for teaching, maintained a high standard. ChatGPT's responses prompted individuals to reflect on their qualifications and suitability for the profession, offering a well-rounded approach that human advisers might find challenging to replicate consistently.

The accessibility of AI tools like ChatGPT is particularly beneficial for students from socioeconomically disadvantaged backgrounds, who may face challenges in accessing quality academic advising. With ChatGPT, students can engage in meaningful, self-directed inquiries and later seek human consultations with more informed questions. This approach enhances the advising process, making sessions with human advisers more productive and personalized.

Students can independently utilize ChatGPT as an effective tool for academic advising by following a few key strategies. First, students should learn to ask clear, detailed questions to get the best responses from the AI. This involves specifying the topic, such as certification requirements or career expectations, and providing relevant context like location (e.g., New York State). By doing so, they can receive well-structured and informative answers tailored to their needs.

ChatGPT can serve as a valuable first step in the information-gathering process, helping students jump-start their inquiries. Once students have received initial guidance from ChatGPT, they are encouraged to consult additional resources or seek human advisors for more personalized and current advice. This approach not only empowers students to ask more informed questions but also optimizes their advising sessions.

However, it's important for students to understand the limitations of the platform. ChatGPT's responses, while often accurate and comprehensive, still require fact-checking, especially for up-to-date or program-specific information.

E. Doing research

Similar to honing students' writing skills, ChatGPT can be a valuable tool in various stages of the research process. It aids researchers in generating research titles, creating abstracts, introductions, and providing initial ideas for literature reviews and methodology brainstorming [15].

Researchers can input the full body of their paper or a detailed outline, and ChatGPT can quickly summarize the key findings, objectives, and methodology into a coherent and well-structured abstract. This saves time while ensuring that the most critical aspects of the research are clearly communicated to readers. While human refinement is necessary to ensure the abstract's accuracy, ChatGPT's language generation capabilities provide a solid foundation to build upon. When drafting an introduction, ChatGPT can analyze the background information provided and generate an introduction that frames the research context, highlights its relevance, and outlines the objectives. It can also assist in suggesting keywords and abbreviations from the article's content.

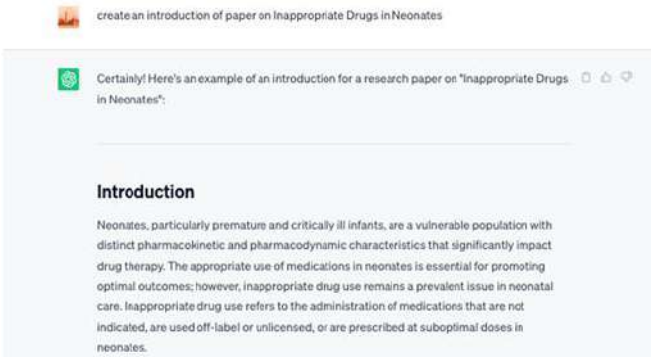


Figure 6: Example introduction generated by ChatGPT (retrieved from reference [15]).

For literature reviews, ChatGPT can offer a basic overview of existing research on a given topic. While it lacks access to the latest research articles and may not cover every relevant study comprehensively, it can provide an initial draft that researchers can expand upon by adding in-depth analysis and citations from current literature. Similarly, for methodology brainstorming, ChatGPT can generate ideas based on the research goals and study design, offering researchers potential approaches for data collection and analysis. Though these suggestions should be critically evaluated and customized to meet specific research requirements, they can serve as helpful starting points.

Mark E. Glickman et al have also found that ChatGPT's summarization capabilities provide researchers with quick, accessible summaries of complex research papers, making it easier to grasp main ideas without needing to read the entire text [18]. It excels at distilling key points from text-heavy papers, providing a solid foundation for understanding general findings and insights. Additionally, ChatGPT offers abductive reasoning for research discovery, helping researchers infer connections between topics, fields, or methods, even when they may not know precise terminology. This ability to support exploratory discovery tasks, where reasoning and connection-building are essential, distinguishes ChatGPT from traditional search methods, which often struggle to interpret complex queries. However, it does face limitations with highly technical papers, such as those involving extensive mathematical notation, where it may misinterpret details.

However, while it can be helpful in these areas, human intervention is necessary for accuracy, especially in sections like literature reviews, methodology, and conclusions. Its summaries still benefit from researcher review to ensure technical precision. Additionally, ChatGPT's capability to assist in creating references is limited, often requiring manual correction. For researchers, the tool is best utilized for early-

stage brainstorming and idea generation rather than detailed analysis or data-driven sections.

IV. Discussion

It is clear that AI has significantly altered traditional learning paradigms, introducing personalized and adaptive learning experiences that are more accessible and efficient. The research outlined in this review highlights the tangible benefits of AI through various real-life, accessible applications, particularly for students, as it enhances learning across various subjects, languages, improves writing skills, offers career guidance, and supports the research process.

For students, given the accessible nature of these AI-driven platforms, students should investigate the wide range of capabilities LLMs offer, such as providing feedback, generating ideas, and offering real-time assistance. These explorations are essential in identifying new, innovative ways that AI can be leveraged to improve learning outcomes across various subjects. This exploratory approach could lead to uncovering novel applications that go beyond traditional academic support, allowing AI to become a transformative tool in students' personal and academic development.

Research should also focus on the development of specialized AI tools tailored to specific educational needs, rather than relying solely on general-purpose LLMs like ChatGPT. While general AI models are highly versatile, there is a growing need for AI systems that can address subject-specific challenges with greater precision and depth. Specialized AI tools based on those general models but designed for particular disciplines, such as advanced mathematics, language learning, career guidance or scientific research, could offer more targeted guidance and problem-solving capabilities that align closely with the curriculum and individual learning paths. Developing these specialized AI applications would allow for deeper integration of subject matter expertise and more accurate, context-specific feedback. This could help mitigate the limitations of general-purpose AI, such as handling highly technical or nuanced problems that require domain-specific knowledge, such as those in career guidance tasks.

[1] K. Bicknell, C. Brust, and B. Settles, "How Duolingo's AI learns what you need to learn: The AI that powers the language-learning app today could disrupt education tomorrow," *IEEE Spectrum*, available: https://spectrum.ieee.org/duolingo.

[2] E. A. Vogels, "A majority of Americans have heard of ChatGPT, but few have tried it themselves," *Pew Research Center*, May 24, 2023. [Online]. Available: https://www.pewresearch.org/short-reads/2023/05/24/a-majority-of-americans-have-heard-of-chatgpt-but-few-have-tried-it-themselves/. [Accessed: Oct. 13, 2024].

[3] O. M. Schei, A. Møgelvang, and K. Ludvigsen, "Perceptions and Use of AI Chatbots among Students in Higher Education: A Scoping Review of Empirical Studies," *Education Sciences*, vol. 14, no. 8, p. 922, Aug. 2023. [Online]. Available: https://www.mdpi.com/2227-7102/14/8/922. [Accessed: Oct. 13, 2024].

[4] F. Kamalov, D. S. Calonge, and I. Gurrib, "New era of artificial intelligence in education: Towards a sustainable multifaceted revolution," *Sustainability*, vol. 15, no. 16, p. 12451, Aug. 2023. [Online]. Available: https://www.mdpi.com/2071-1050/15/16/12451. [Accessed: Oct. 13, 2024].

[5] J. R. Slagle, "A heuristic program that solves symbolic integration problems in freshman calculus, Symbolic Automatic Integrator (SAINT)," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 1961.

[6] A. R. De Ocampo, K. Gloton, and E. Palomo, "A usability analysis of Wolfram Alpha for the students of Saint Augustine School-Senior High School in Tanza, Cavite," *Ascendens Asia Journal of Multidisciplinary Research Abstracts*, vol. 3, no. 5, Dec. 2019.

[7] P. Shadaan and L. K. Eu, "Effectiveness of using Geogebra on students' understanding in learning circles," *Malaysian Online Journal of Educational Technology*, vol. 1, no. 4, pp. 7-8, 2013. [Online]. Available: https://files.eric.ed.gov/fulltext/EJ1086434.pdf. [Accessed: Oct. 14, 2024].

[8] Y. Liang, D. Zou, H. Xie, and F. L. Wang, "Exploring the potential of using ChatGPT in physics education," *Smart Learning Environments*, vol. 10, no. 52, p. 52, Oct. 2023. [Online]. Available: https://slejournal.springeropen.com/articles/10.1186/s40561-023-00273-7. [Accessed: Oct. 13, 2024].

[9] R. R. Font de la Vall and F. G. Araya, "Exploring the benefits and challenges of AI-language learning tools," *The International Journal of Social Sciences and Humanities Invention*, vol. 10, no. 1, pp. 7569–7576, Jan. 2023. [Online]. Available: https://doi.org/10.18535/ijsshi/v10i01.02. [Accessed: Oct. 16, 2024].

[10] J. C. Young and M. Shishido, "Investigating OpenAI's ChatGPT potentials in generating chatbot's dialogue for English as a foreign language learning," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 6.

[11] B. Li, C. Bonk, and X. Kou, "Exploring the multilingual applications of ChatGPT," *International Journal of Computer-Assisted Language Learning and Teaching*, vol. 13, no. 1, pp. 1–22, Jan. 2023. [Online]. Available: https://www.researchgate.net/publication/372544314_Exploring_the_Multilingual_Applications_of_ChatGPT. [Accessed: Oct. 16, 2024].

[12] Y. Xiao and Y. Zhi, "An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions," *Languages*, vol. 8, no. 3, p. 212, Sep. 2023. [Online]. Available: https://doi.org/10.3390/languages8030212. [Accessed: Oct. 16, 2024].

[13] C. Song and Y. Song, "Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students," *Frontiers in Psychology*, vol. 14, Dec. 2023. [Online]. Available: https://doi.org/10.3389/fpsyg.2023.1260843. [Accessed: Oct. 16, 2024].

[14] D. Akiba and M. C. Fraboni, "AI-supported academic advising: Exploring ChatGPT's current state and future potential toward student empowerment," *Education Sciences*, vol. 13, no. 9, p. 885, Aug. 2023. [Online]. Available: https://doi.org/10.3390/educsci13090885. [Accessed: Oct. 16, 2024].

[15] S. S. Biswas, "ChatGPT for research and publication: A step-by-step guide," *The Journal of Pediatric Pharmacology and Therapeutics*, vol. 28, no. 6, pp. 576–584, 2023. [Online]. Available: https://doi.org/10.5863/1551-6776-28.6.576. [Accessed: Oct. 16, 2024].

[16] L. Ding and D. Zou, "Automated writing evaluation systems: A systematic review of Grammarly, Pigai, and Criterion with a perspective on future directions in the age of generative artificial intelligence," *Education and Information Technologies*, vol. 29, pp. 14151–14203, Jan. 2024. [Online]. Available: https://doi.org/10.1007/s10639-023-12402-3. [Accessed: Oct. 16, 2024].

[17] J. S. Barrot, "Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy," *Computer Assisted Language Learning*, vol. 34, no. 6, pp. 567–589, 2021. [Online]. Available: https://doi.org/10.1080/09588221.2021.1936071. [Accessed: Oct. 16, 2024].

[18] M. E. Glickman and Y. Zhang, "AI and generative AI for research discovery and summarization," *arXiv*, 2024. [Online]. Available: https://ar5iv.labs.arxiv.org/html/2401.06795v1. [Accessed: Oct. 16, 2024].

MULTI-SECTOR DATA INTEGRATION SOLUTION FOR HO CHI MINH CITY BASED ON THE DATA INTEGRATION PLATFORM

Van Kiet Bui[1,4], Thanh Tuyen Mai Nguyen[2], Thien Thien Doan[1,4], Tuyet Hue Tran[3,4,5,*], Dang Thi Ngoc Uyen[6], Tan-Cong Nguyen[1,3,4]

[1]University of Social Sciences and Humanities, VNU-HCM, Ho Chi Minh City, Vietnam
[2]Ho Chi Minh City Department of Information and Communications
[3]Faculty of Information Technology, University of Science - VNUHCM, Ho Chi Minh City, Vietnam
[4]Vietnam National University, Ho Chi Minh City, Vietnam
[5]School of Business Information Technology, University of Economics Ho Chi Minh City, Ho Chi Minh City, Vietnam
[6]Faculty of Finance and Accounting, Saigon University, Ho Chi Minh City, Vietnam

***Abstract*—In the context of digital transformation and the development of smart cities, integrating data from various sources is essential to enhance management efficiency and support decision-making based on accurate and timely data. The study discusses the Data integration platform in Ho Chi Minh City, designed to address challenges in managing and exploiting multi-sector data. HCMC's Data integration platform is designed to replace traditional extract, transform, and load tools, with the capability to integrate data from multiple sources into a central data warehouse. This integration process includes analyzing the structure and content of tables and data fields, followed by efficient data streaming and updating. Notably, Data integration platform employs a LogParser tool to process and update data based on historical changes, ensuring that data is always current and accurate.**

***Index Terms*—Data integration platform, Smart city, Digital transformation**

I. Introduction

This paper examines the critical role of multi-sector data integration in the context of digital transformation and urban management, with a focus on Ho Chi Minh City. Historically, the city encountered significant challenges in data management due to the fragmentation and lack of connectivity across various data systems. To address these issues, the Data Integration Platform (DIP) was developed as a replacement for traditional Extract, Transform, Load (ETL) tools. The DIP is designed to manage unstable data environments, poorly organized datasets, and immediate data access needs, all without requiring a centralized data warehouse.

DIP serves as a core component in integrating structured data from multiple sectors, thereby supporting the city's long-term goal of building a centralized data warehouse. Additionally, DIP lays the groundwork for future integration platforms, such as File Storage and GIS Data Platforms. This initiative is an essential milestone in Ho Chi Minh City's progression towards becoming a smart city, where timely and accurate data are the foundation for all decision-making and public services [1].

The paper provides a detailed analysis of DIP's architecture and implementation, emphasizing the methods and technologies used to ensure efficiency, security, and effectiveness in multi-sector data integration. Furthermore, it highlights the strategic importance of data integration in enhancing data-driven decision-making, improving urban management, and fostering the development of smart city solutions.

The structure of this paper is as follows: Section I provides this introduction and the background context of data integration challenges and DIP's role. Section II discusses the architecture and components of the DIP system, including the platform's planning and data flow mechanisms. Section III presents the design of the central database, focusing on efficient storage and management of multi-sector structured data. Section IV describes the functional subsystems that facilitate data integration, updating, and extraction processes. Section V includes a discussion on the DIP's achievements, challenges, and potential directions for future improvements. Finally, Section VI concludes by underscoring the DIP's impact on data-driven decision-making and smart city development initiatives in Ho Chi Minh City.

II. Data Integration Platform (DIP) System

A. Information Architecture Planning

The role of information architecture is crucial in the planning and construction of data platforms, as it sets data standards and enables the unification of data into a central warehouse for analysis and reporting. However, in Vietnam, inadequate planning in government IT applications, including in Ho Chi Minh City, has made data integration challenging. To overcome this, the city has introduced a Data Integration Platform (DIP) as an intermediary step before fully transitioning to a centralized data warehouse [2]. The DIP consolidates data for immediate use, addressing current issues without requiring a fully developed information architecture.

*Corresponding author: Tuyet Hue Tran (hue-tran.523202180629@st.ueh.edu.vn).

This approach emphasizes creativity and adaptation, especially for handling structured and unstructured data [3]. The focus of this research is on integrating structured data stored in relational databases.

B. Data Flow in the Data Integration Platform

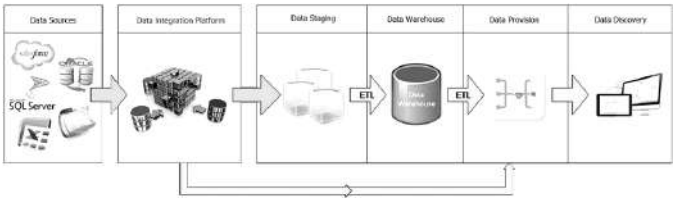


Fig. 1. Data Flow Model

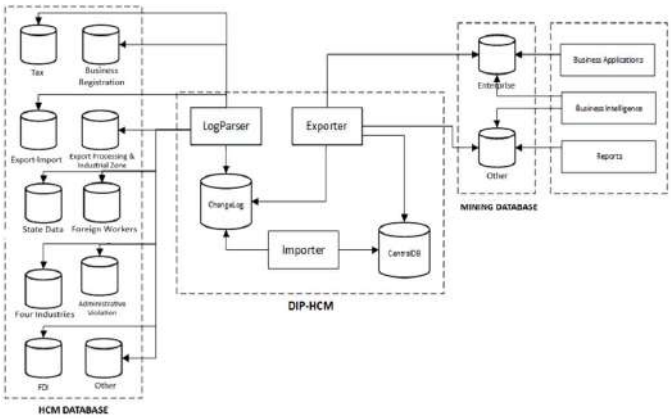


Fig. 2. The overall model of the data integration platform

Based on the previous analysis, Ho Chi Minh City's information architecture is complex, making it challenging to apply a standard data flow model. Figure 1 illustrates the Data Flow Model for the city's Data Integration Platform (DIP), developed as an alternative to traditional ETL tools, particularly during the data staging phase.

Key steps in the data flow include:

- **Data Sources:** These include databases from sectors like Enterprise, FDI, Healthcare, and Land Registry.
- **Data Integration and Update in DIP:** Involves analyzing data structures, streaming records into the central hub, and updating data based on changes to ensure accuracy.
- **Centralized Storage:** Data is stored centrally at DIP's hub and can be extracted for further use by third-party applications.
- **Data Staging:** Data is cleaned and reorganized in this intermediate phase before moving to the Data Warehouse.
- **Data Warehouse:** A centralized system for large-scale storage, where processed data is ready for complex analysis.
- **Data Provision:** Data is made accessible to applications and users for decision-making.
- **Data Discovery:** Users access data through tools like dashboards for analysis and insight.

C. Overall Model of the Data Integration Platform

Managing multi-sectoral data in rapidly developing cities like Ho Chi Minh City is challenging. To address this, the Data Integration Platform (DIP-HCM) was developed, creating an efficient system for integrating data from various sources into a central database. This system consolidates dispersed information, optimizing data handling and utilization to support timely decision-making. DIP-HCM integrates data from sources such as the enterprise database, FDI data, and land registry, among others, through a structured process of analysis, selection, and merging.

Data Analysis and Integration: Data from various sources is first analyzed to identify necessary information and eliminate irrelevant data, improving integration efficiency. After analysis, the data is streamed into a unified central database for future use.

LogParser Mechanism: DIP-HCM uses the LogParser module for efficient data integration and updates based on historical changes in source databases. Key steps include:

- **Data History Analysis:** LogParser identifies records that have been modified since the last integration, ensuring only new or updated data is added to the CentralDB.
- **Data Updating:** The system optimizes the process by only updating changed data, reducing the load on source databases.
- **Remote Analysis:** LogParser's remote mechanism retrieves data without impacting the performance or security of source databases, crucial for maintaining system integrity in government environments.

This approach ensures Ho Chi Minh City's data integration system remains accurate and efficient for supporting management and decision-making.

D. The functional architecture model of the DIP system

Ho Chi Minh City's data integration platform has been designed with a detailed functional architecture to ensure that every aspect of data collection, storage, management, and extraction operates smoothly and efficiently.

The functional architecture model is visualized in Figure 3. The components of the functional architecture model include the Web Interface, which serves as the main interaction point between users and the DIP system. It provides tools and functions that allow users to manage, extract, and update data. Software Subsystems of the Data Integration Platform: The software subsystems of the DIP system are divided into multiple modules, each responsible for a specific function in the process of data collection, storage, management, and extraction (to be detailed in Section 5: Functional Subsystems). MongoDB Database: The system uses MongoDB to store data. MongoDB is a powerful, flexible, and highly scalable NoSQL database management system, suitable for the complex data storage and retrieval requirements of the DIP system. System

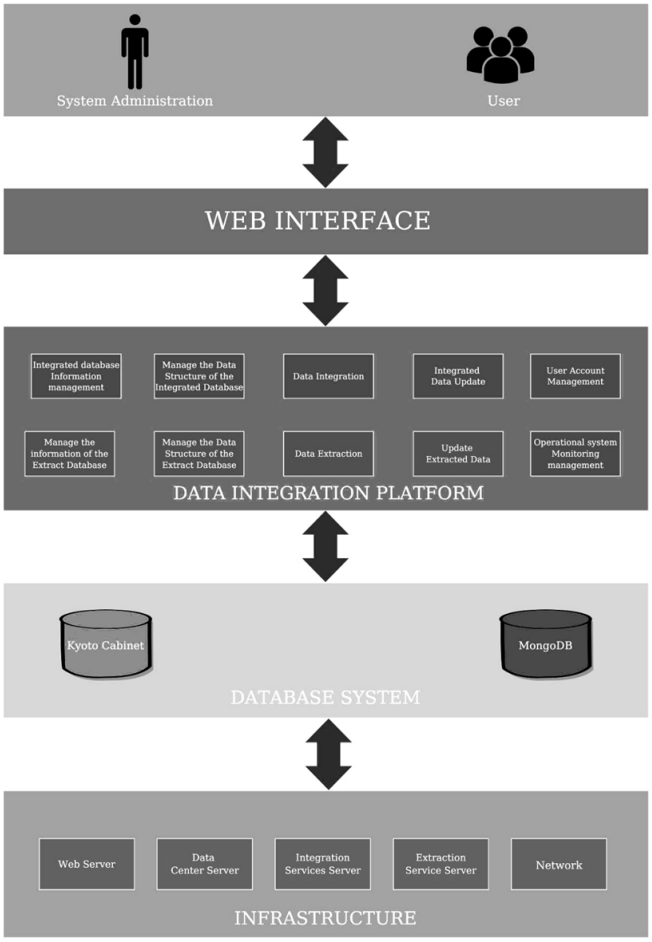


Fig. 3. Functional Architecture Model

Infrastructure: This includes servers and data transmission networks. The system is designed to ensure scalability, high performance, and reliability in data transmission and processing.

E. System Deployment Model

The DIP system deployment model illustrates how the system's components are connected and interact with each other to ensure that the processes of data collection, storage, management, and extraction run smoothly and efficiently. Figure 4 presents the network structure connecting the main components of the system via LAN/WAN, from the source databases to the data center and the extraction databases.

The deployment model of Ho Chi Minh City's Data Integration Platform (DIP) is designed to ensure seamless, secure, and efficient connectivity between various data sources and the data center. This system comprises three main components: DIP, the Data Center, and the Source Databases (Source DBs) and Extraction Databases (Extraction DBs), all interconnected through LAN/WAN networks. DIP, as the core system, is responsible for integrating, processing, and managing data from different sources. It is directly connected to the Data

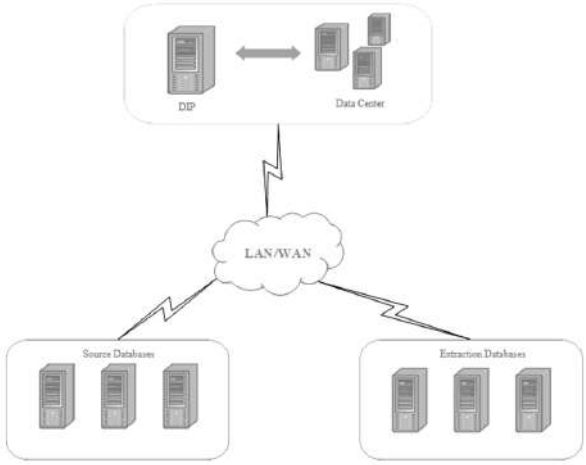


Fig. 4. System Deployment Model

Center, which houses the necessary servers and infrastructure to store and manage integrated data. The LAN/WAN network acts as the medium for data transmission between DIP, the Data Center, and external data sources, ensuring that data can be transferred securely and quickly. The Source DBs consist of the original databases from various systems such as enterprise management, taxation, land registry, and others, serving as the primary input sources for the DIP system. The Source DB servers are connected to DIP via the LAN/WAN network, enabling DIP to retrieve and extract data for processing and integration into the Data Center. Once the data is processed and integrated, it is stored in the Extraction DBs, which are used for specific purposes like data analysis, reporting, and other business applications. The Extraction DB servers are also connected to the Data Center via LAN/WAN, allowing data to be accessed and utilized efficiently for management needs.

F. System Database Model and Data Exchange Process between Source and Extraction Databases

To better understand the functionality and efficiency of the Data Integration Platform (DIP), we need to examine its database model and the data exchange process between the source databases and extraction databases. This is a crucial part of the system, ensuring that data is stored, managed, and retrieved efficiently and systematically. The model not only addresses issues related to data synchronization and standardization but also optimizes the processes of data integration and extraction from multiple sources.

The model in Figure 5 provides a detailed description of the data integration and extraction process within the DIP-HCM platform. This process is designed to optimize the storage and retrieval of data from source databases to extracted databases.

Data Integration Process

- **Source Database:** The system starts with the source databases, which include various data sources such as

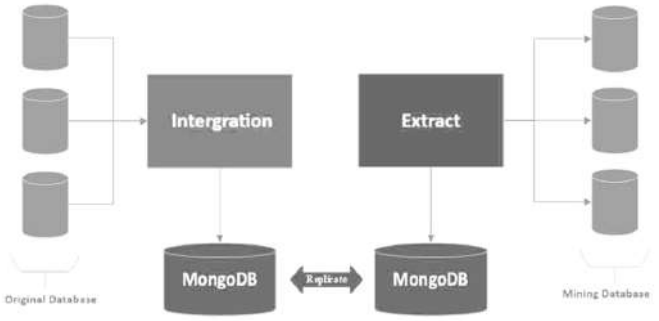


Fig. 5. System Database Model and the Process of Data Exchange Between the Source Database and the Extracted Database

the Enterprise Database, FDI Database, Tax Database, and Land Administration Database.

- **Integration Subsystem:**
  - Data from the source databases is transmitted to the integration subsystem. Here, the data is processed and standardized to ensure uniformity before storage.
  - This integration subsystem transfers the data into the primary MongoDB database for storage. MongoDB is chosen for its ability to handle unstructured data and its high scalability.
- **Data Storage in MongoDB:** After integration, the data is stored in MongoDB, forming a centralized data repository.

Data Extraction Process

- **Replicate MongoDB:** Data from the primary MongoDB is replicated to another MongoDB database. This replication ensures that the data is always available for extraction without impacting the original database.
- **Extraction Subsystem:** The extraction subsystem accesses the second MongoDB database, where the replicated data is stored. Here, the data is prepared and processed for extraction into the required formats.
- **Extracted Database:** Finally, the data is exported to various extracted databases to support external applications and systems. These extracted databases provide data for analysis, reporting, and decision-making in urban management and smart city development activities.

III. Central Database Design Solution

The storage and management of the integrated database is a key factor in the data integration platform of Ho Chi Minh City (DIP-HCM). To ensure that data from various sources is stored in a scientific and logical manner, the storage process is divided into two main parts: storing the structured data of the source databases and storing the selected data for integration into the data warehouse.

The management of the structured data of the integrated databases includes assigning identification codes for each database (DB_ID), each table (TABLE_ID), and each column (COLUMN_ID). Basic information about the integrated database is stored by the "ImportedDBInfoStoreService", which includes the ID, name, information about the source database, information about the user who created and edited it, as well as the status and the timestamp of the last update. Similarly, detailed information about the tables and columns in the integrated database is managed by the "ImportedTableInfoStoreService" and "ImportedColumnInfoStoreService."

The storage of data in the integrated databases utilizes MongoDB, optimizing the management and processing of unstructured data. Data from the source databases is stored as collections in MongoDB, with each collection representing a table in the source database. Records in each collection are stored as (key, value) pairs, where the key is the identification code of the column and the value is the corresponding value of that column. Each record is assigned a unique ID (ROW_ID) to identify and manage data updates.

The management of the structured data of the extracted databases is similar to that of the integrated databases. Basic information about the extracted database is stored by the "ExportedDBInfoStoreService", which includes the ID, name, information about the target database, and information about the user who created and edited it. Detailed information about the tables and columns in the extracted database is managed by the "ExportedTableInfoStoreService" and "ExportedColumnInfoStoreService.".

IV. Functional Subsystems of the Data Integration Platform (DIP)

The Ho Chi Minh City Data Integration Platform (DIP-HCM) is designed with several functional subsystems to efficiently manage and process data from source databases. These subsystems include:

- **Integrated Database Information Management:** This subsystem is responsible for managing general information about the source databases, such as the database name (dbname), server address (host), connection port (port), username (username), and password (password). This information is essential for connecting to and accessing the source databases.
- **Management of the Integrated Database Structure:** This subsystem manages the data structure of the source databases, also known as Metadata. Metadata includes information about table structures, data fields, and the relationships between tables. This subsystem ensures that the data integrated from various sources can be stored and processed accurately and consistently.
- **Data Integration:** This is a critical subsystem of DIP-HCM, responsible for integrating data from source databases into the storage center. This subsystem performs data structure analysis, transforms the data into suitable formats, and integrates it into the central database.
- **Integrated Data Update:** This subsystem automatically updates changes from the source databases to the storage center. By using the LogParser block to analyze the history of data changes, the system ensures that the data

in the storage center is always updated and accurately reflects the current state of the source databases.

- **Extracted Database Information Management:** This subsystem manages general information about the extracted databases, including the database name, server address, connection port, username, and password. This information is essential for connecting to and retrieving data from the storage center.
- **Extracted Database Structure Management:** This subsystem manages the metadata of the extracted databases. This metadata includes information about table structures, data fields, and the relationships between tables in the extracted database, ensuring that the extracted data can be used effectively.
- **Data Extraction:** This subsystem extracts data from the storage center and transfers it to the extracted databases. This subsystem ensures that the necessary data is delivered to the target systems at the right time and in the appropriate format.
- **Extracted Data Update:** This subsystem automatically updates data changes from the storage center to the extracted databases, ensuring that the data in the extracted systems is always synchronized with the central data.
- **User Account Management:** This subsystem manages user accounts within the DIP system, including the creation, modification, and deletion of accounts. It also manages access rights and permissions for users, ensuring that only authorized individuals can access critical subsystems and data.
- **Business Activity Monitoring (BAM):** The BAM subsystem monitors and supervises the operation of the DIP system in real time. This subsystem provides reports and alerts on the operational status of the system, helping to detect and address any issues or anomalies during operation promptly.

The functional subsystems of DIP-HCM are designed to ensure that the processes of integration, storage, extraction, and management of data from source databases occur efficiently and securely, meeting the stringent requirements of building a common data warehouse as a foundation for urban management and the development of a smart city in the future.

V. Discussion

After implementation, the DIP-HCM integrated data platform system has achieved significant positive outcomes, demonstrating the project's feasibility. The system has been successfully transferred to the Ho Chi Minh City Digital Transformation Center, integrating multiple databases from different management systems. DIP-HCM shows high applicability, connecting with systems such as land data collection, business reporting, and mobile applications serving city leadership.

The multi-sector data integration through the DIP platform provides numerous benefits and positively impacts urban management and development:

- Improved Decision-Making and Policy Formulation: By integrating data from multiple fields, managers gain a more comprehensive and detailed view of city issues, supporting more accurate and timely policy decisions [1].
- Enhanced Management Efficiency: Linking data sources facilitates easier coordination among management agencies and the implementation of urban management measures. Previously, data was decentralized, with only the Department of Planning and Investment having access to business registration data, excluding household businesses. Districts and counties could not access this information even for companies operating within their jurisdiction. With data centrally integrated through the DIP platform into a common data warehouse, district leaders can access business information to inform their economic development policies.
- Multi-sector data integration provides a foundation for in-depth analysis, assisting city leaders in identifying trends and economic opportunities, leading to appropriate development strategies. It also helps improve public service quality and responsiveness to citizen needs [4].

Despite these achievements, the DIP platform implementation faces significant challenges, particularly regarding data complexity from various sources. The lack of uniformity in data formats and structures requires efficient data standardization and cleansing processes. Additionally, ensuring data security and privacy while integrating multiple sources poses a substantial challenge that the DIP platform must address [5].

Future research could focus on applying advanced data integration techniques such as machine learning (ML) and artificial intelligence (AI) to overcome these challenges and enhance system effectiveness. Machine learning algorithms can automate data cleansing and enrichment processes, improving system accuracy and performance. AI can analyze complex data patterns and provide useful predictions for decision-making [6].

Another essential area for improvement is Enhancing User-Centric Analysis. Conducting user feedback sessions or usability studies would provide insights into how effectively the system meets municipal staff's needs, supporting more user-friendly system modifications. This approach would make the DIP platform more accessible and usable for city officials and management personnel, thus enhancing its impact on urban data management practices.

In addition, future studies should investigate the long-term impacts of multi-sector data integration on urban development and management. A deeper understanding of multi-sector data integration's benefits and limitations can provide valuable insights for shaping sustainable urban management strategies. Research could also explore new models and methods to enhance the DIP platform's scalability and flexibility, ensuring it meets the growing demands of large cities like Ho Chi Minh City [3].

In the context of ever-evolving technology, continued research and improvements to the DIP platform or related technologies will play a vital role in enabling cities to manage data effectively, thus promoting sustainable economic and social development.

VI. Conclusion

The DIP platform for Ho Chi Minh City has demonstrated its ability to significantly improve the data integration process, thereby enhancing the overall performance and accuracy of the system. By integrating data from multiple sectors through the DIP platform, the decision-making process at various levels of management is improved. This not only enhances management effectiveness but also supports sustainable economic and social development. Through the optimization and automation of data processing, management agencies can access accurate and timely information, allowing them to make strategic decisions based on real data.

Acknowledgment

This research is partially funded by the Vingroup Innovation Foundation (VINIF) under the grant number VINIF.2021.JM01.N2.

References

[1] J. Kandt and M. Batty, "Smart cities, big data and urban policy: Towards urban analytics for the long run," *Cities*, vol. 109, p. 102992, 2021.

[2] L. Sanchez, J. A. Galache, V. Gutierrez, J. M. Hernandez, J. Bernat, A. Gluhak, and T. Garcia, "Smartsantander: The meeting point between future internet research and experimentation and the smart cities," in *2011 Future Network & Mobile Summit*. IEEE, 2011, pp. 1–8.

[3] R. Kitchin and G. McArdle, "What makes big data, big data? exploring the ontological characteristics of 26 datasets," *Big Data & Society*, vol. 3, no. 1, p. 2053951716631130, 2016.

[4] J. de Jong, A. Edmondson, M. Moore, H. R. Bowles, J. Rivkin, E. F. Martínez Orbegozo, and S. Pulido-Gomez, "Building cities' collaborative muscle," *Stanford Social Innovation Review*, vol. 19, no. 2, pp. 50–56, 2021.

[5] J. Koo, G. Kang, and Y.-G. Kim, "Security and privacy in big data life cycle: a survey and open challenges," *Sustainability*, vol. 12, no. 24, p. 10571, 2020.

[6] X. L. Dong and T. Rekatsinas, "Data integration and machine learning: A natural synergy," in *Proceedings of the 2018 international conference on management of data*, 2018, pp. 1645–1650.

Section 2:
Electrical Engineering (EE)

Description: This field explores advancements in robotics, automation, telecommunications, and mechatronics. It combines electronic systems and intelligent equipment to drive innovation in automated processes, communication systems, and integrated engineering solutions.

Digital Control Strategy for Step-Down Energy Conversion Systems

Thanh-Lam Le
Faculty of Electrical and Electronics Engineering
Ho Chi Minh City University of Technology and Education
Ho Chi Minh City, Vietnam
lamlethanh@hcmute.edu.vn
(Corresponding author)

Huynh Ngoc Trong
Faculty of Electrical and Electronics Engineering
Ho Chi Minh City University of Technology and Education
Ho Chi Minh City, Vietnam
19142406@student.hcmute.edu.vn

Ma Thanh Hien
Faculty of Electrical and Electronics Engineering
Ho Chi Minh City University of Technology and Education
Ho Chi Minh City, Vietnam
19142309@student.hcmute.edu.vn

***Abstract*—This research centers on the analysis of DC-DC buck converters, specifically exploring their operational modes and providing detailed guidelines for circuit design parameter selection. The practical implementation of these systems often encounters significant challenges due to potential instability and oscillations arising from varying operating conditions, particularly in the absence of feedback mechanisms. To enhance performance and reliability, this study explores the development of a digital control technique specifically designed for step-down energy conversion systems. A simulation model of the DC-DC buck converter was created and analyzed using MATLAB/Simulink, ensuring meticulous evaluation of the design parameters. The results demonstrate the effectiveness of the digital controller in maintaining stable output voltage. Furthermore, this work highlights the critical role of digital control strategies in addressing the challenges of voltage regulation. The findings contribute to the understanding of buck converters, paving the way for improved energy management solutions in modern power systems. Ultimately, this study underscores the significance of implementing digital control techniques to enhance the feasibility and efficiency of DC-DC conversion technologies.**

Keywords— dc-to-dc, step-down converter, buck converter, pulse width modulation, energy conversion systems, digital controller, voltage controller, renewable energy source

I. INTRODUCTION

The increasing integration of renewable energy sources and the global shift toward environmentally friendly green energy solutions have significantly contributed to the rapid development of DC microgrids and smart home systems [1-2]. These technologies are becoming more prevalent in our daily lives. Concurrently, the widespread adoption of mobile devices, electric vehicles (EVs), and advancements in battery technology are further driving demand for efficient energy management [3-5]. A prominent example is the rapid growth of EV charging stations across Vietnam in recent years, reflecting a surging global demand for DC power. DC power has become a central focus in current research and development, establishing itself as a crucial element of future energy systems [6]. Consequently, the development and application of high-quality, reliable DC-DC converters that meet practical demands with high efficiency is of paramount importance for modern power systems.

DC-DC converters play a crucial role in various modern industrial applications, particularly in renewable energy systems, electric vehicles, automation equipment, lightweight electronic devices, and end-user computing systems [7-9]. Their stability and high load tolerance, along with advantages such as high power density, efficiency, cost-effectiveness, and flexible operation, have made them widely adopted across these fields. As part of switched-mode power supplies, these converters utilize components such as diodes, transistors, inductors, and capacitors to achieve voltage conversion [10]. Common topologies, including cuk, boost, buck, and buck-boost converters, are employed in devices requiring low-voltage DC conversion [11]. However, one of the main challenges in designing controllers for these circuits is ensuring stable output voltage under varying input voltage conditions, fluctuating loads, and circuit parameters. To address this issue, advanced control mechanisms have been implemented to enhance system performance and reliability [12]. Additionally, new and sophisticated design approaches are being explored to overcome these challenges. Nevertheless, increasing the number of components excessively can lead to higher costs and reduced feasibility for commercialization [13].

Buck converters regulate output voltage through two primary switching techniques: Pulse Width Modulation (PWM) and Pulse Frequency Modulation (PFM). PFM is generally applied under light load conditions, typically below 50mA, whereas PWM is more suitable for heavier loads, usually above 100mA [14]. In PWM, semiconductor switches operate at a constant switching frequency with a modulated duty cycle. However, maintaining a stable output voltage with a fixed duty cycle becomes challenging in open-loop DC-DC converters due to fluctuations in input voltage, variations in output load, and the nonlinear characteristics of circuit components [15-16]. To address these issues, a feedback loop is required to ensure precise output voltage regulation and fast response to changes in load and input conditions. In some cases, compensation circuits are employed alongside simple control loops to minimize potential oscillations in the system. Additionally, digital control techniques have been introduced to improve stability and reduce transition time [17-18]. Digital control offers several advantages, including faster switching, easier algorithm implementation, and seamless integration with microcontrollers such as Digital Signal Processors (DSPs), making it a highly efficient approach for enhancing the performance of buck converters [19].

Linear control methods, such as proportional-integral (PI) control [20-21], and nonlinear approaches, such as sliding

mode [22-23], neural networks [24], and fuzzy logic control [25-26], are commonly employed to ensure system stability. In the case of a buck converter, the output voltage serves as the primary control objective. Each of these control techniques offers distinct advantages and limitations depending on the application [27-28]. Classical linear feedback, such as PI control, is often used to regulate the duty cycle. Linear control methods are effective in improving dynamic response and steady-state performance. Additionally, these controllers are relatively simple to design, require minimal memory in microcontrollers, and are feasible for implementation in real-world physical systems [29-30]. When properly tuned, a PI controller applied to a buck converter can deliver high performance with low implementation costs, making it a practical choice for many power conversion applications.

This study focuses on analyzing circuits, specifically investigating the operating modes of DC-DC buck converters while offering detailed guidelines for selecting circuit design parameters. Implementing such systems in practical applications can lead to instability and significant oscillations when operating conditions change, particularly in the absence of feedback loops. This instability poses challenges for achieving efficient real-world deployment. To enhance operational effectiveness, redesigning the circuit typically requires substantial effort and can incur high equipment costs. Therefore, this research also develops a digital control technique for step-down energy conversion systems. A simulation model of a DC-DC buck converter has been constructed and analyzed using the MATLAB/Simulink environment. The simulation process rigorously evaluates the design parameters, and the results validate the operational efficiency of the digital controller within the step-down energy conversion systems.

II. Analysis Of Buck Converter System

This section details the modeling and operational principles of a DC-DC buck converter. As shown in Fig. 1, a typical buck converter comprises a DC input voltage source V_{in}, a controlled switch S_w, a diode D, an inductor L, a capacitor C, and a load resistor R. The converter's behavior is characterized by distinct equations for the ON and OFF switching states, which describe system dynamics in each phase. These equations provide an accurate depiction of the converter's performance across different conditions and are represented as follows

$$\begin{cases} \frac{di_L}{dt} = \frac{1}{L}(V_{in} - v_o) \\ \frac{dv_o}{dt} = \frac{1}{C}(i_L - \frac{v_o}{R}) \\ \frac{di_L}{dt} = -\frac{v_o}{L} \end{cases} \quad (1)$$

By combining the differential equations (1), the following results are obtained as

$$\begin{cases} \frac{di_L}{dt} = \frac{1}{L}(uV_{in} - v_o) \\ \frac{dv_o}{dt} = \frac{1}{C}\left(i_L - \frac{v_o}{R}\right) \end{cases} \quad (2)$$

where u is defined as the control input, assuming a value of 1 when the switch is in the ON state and 0 when it is in the OFF state. The output voltage error and its derivative, which represents the rate of change of the output voltage error, can be defined as

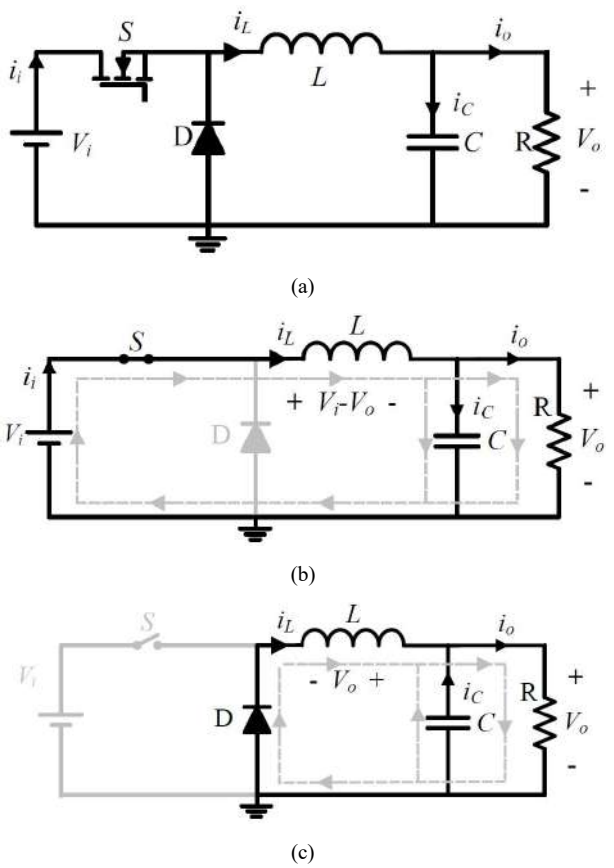


Fig. 1. Circuit diagram.

$$\begin{cases} x_1 = v_0 - V_{ref} \\ x_2 = \dot{v}_0 - \dot{V}_{ref} = \dot{x}_1 = \dot{v}_0 \end{cases} \quad (3)$$

Here, $\dot{x}_1$ indicates the derivative of x_1, and V_{ref} represents the reference DC value for the output voltage. By differentiating the voltage differential equation v_0 (2) with respect to time, the output voltage error x_1 and the rate of change of this error x_2 can be expressed as

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -\frac{\dot{x}_1}{RC} + \omega_0^2(uV_{in} - V_{ref} - x_1) \\ \omega_0^2 = \frac{1}{LC} \end{cases} \quad (4)$$

Next, the operational principles of the circuit will be discussed. Fig. 1 presents the basic schematic of a buck DC-DC converter, which steps down the output voltage to a level lower than the input voltage. In the initial design phase, it is essential to identify the operational modes and circuit parameters. As is well known, in switching mode, the converter operates in continuous, discontinuous, or boundary conditions, depending on whether the inductor current drops to zero. When the converter operates in continuous conduction mode (CCM), the output inductor current ripple is typically lower than in other operating modes. Conversely, discontinuous conduction mode (DCM) offers advantages in energy absorption, especially at higher power levels.

In a buck converter operating in continuous conduction mode, depicted in Fig. 2, the behavior of switch S significantly influences circuit dynamics. When switch S is closed (ON state), the diode is reverse-biased, remaining inactive. During this ON phase (DT), the input source powers both the load and the inductor, driving an increase in

inductor current at a defined rate (as shown in Fig. 1b). Conversely, when switch S opens (OFF state), the inductor voltage reverses to a factor of (1-D), allowing the inductor to sustain current flow to the output through the diode during the remainder of the cycle (*T*), as illustrated in Fig. 1c. Under steady-state conditions, the average current through the capacitor reaches zero, implying that the average inductor current aligns with the average output current. Here, D represents the duty cycle, and T is the switching period. During the ON state, the voltage across the inductor equals the difference between the input voltage (V_i) and output voltage (V_o), whereas in the OFF state, it is equal to -V_o. In continuous conduction, a relationship, detailed in Eq. (5), connects input and output voltages. This relationship leverages the principle that, in ideal conditions, the energy accumulated and released by the inductor is balanced, resulting in a net zero average voltage across the inductor.

$$\begin{cases} V_o = DV_{in} \\ D = \frac{t_{on}}{T} \end{cases} \tag{5}$$

where t_{on} represents the period during which the switch remains active (ON). As outlined in Eq. (5), the output voltage bears a direct proportional relationship to the input voltage, governed by the duty cycle, D, within the range $(0 < D < 1)$. Since the duty cycle does not exceed 1, the output voltage is invariably less than the input voltage. By increasing the switching frequency, the circuit's component sizes can be minimized. The minimum inductance L_{min} required to maintain continuous conduction mode in the converter can be calculated using the following expression

$$L_{min} = \frac{(1-D)\,R}{2\,f} \tag{6}$$

The converter operates at the minimum inductance value L_{min} calculated using Eq. (6) under boundary conditions. To ensure reliable operation in continuous conduction mode, it is advisable to select an inductance value slightly higher than this specified minimum. In the design process, this value has been increased by twenty-five percent, leading to the required inductance being represented as $L = 1.25 \times L_{min}$. The value of the inductor current ripple is influenced by the inductor voltage and the inductance value (L). When calculating the inductor current ripple while considering the time the switch remains in the conductive state, it can be expressed as $\Delta i_L = \left(\frac{V_i - V_o}{L}\right) DT$. Based on this ripple value Δi_L, the maximum and minimum inductor currents can be easily determined as $I_{Lmax} = \overline{I_L} + \frac{\Delta I_L}{2}$ and $I_{Lmin} = \overline{I_L} - \frac{\Delta I_L}{2}$. This behavior can also be observed and analyzed in the current waveform depicted in Fig. 2. It is important to note that the output voltage ripple of a buck converter typically needs to remain below 1%. Since the inductor current ripple also affects the capacitor current, the capacitance value required to maintain the desired output voltage ripple ($\Delta V_O / V_O$) can be calculated using the formula $C_o = (1-D)/[8L(\Delta V_O/V_O)f^2]$. Considering the design criteria and specifications, the values of the components used in the basic buck converter operating in continuous conduction mode have been determined. The following section will discuss the closed-loop control of the buck converter in detail.

III. Implementation Of Voltage Control

Ensuring a stable output voltage in buck converters with a fixed duty cycle under open-loop conditions is particularly challenging, especially when subject to load changes or external disturbances. Consequently, incorporating a feedback loop becomes crucial for achieving precise output voltage regulation as load and input voltage fluctuate. Fig. 3 displays a block diagram of a voltage-mode buck converter managed by a digital PI controller, featuring both power stages and feedback elements. A voltage sensor is employed to sample the output voltage V_0 from the buck DC-DC converter. This output signal is processed through a signal conditioning circuit, adjusting it to an appropriate level for the analog input channel before being fed to a Digital Signal Processor (DSP). Within the DSP, an analog-to-digital converter (ADC) digitizes the sensed voltage and compares it with a reference voltage V_{ref}. The PI controller minimizes the resulting error, producing an output that adjusts the duty cycle in the PWM module. The PWM signal, once generated, undergoes digital-to-analog conversion to control the semiconductor switch, enabling precise on-and-off control via the driver circuit.

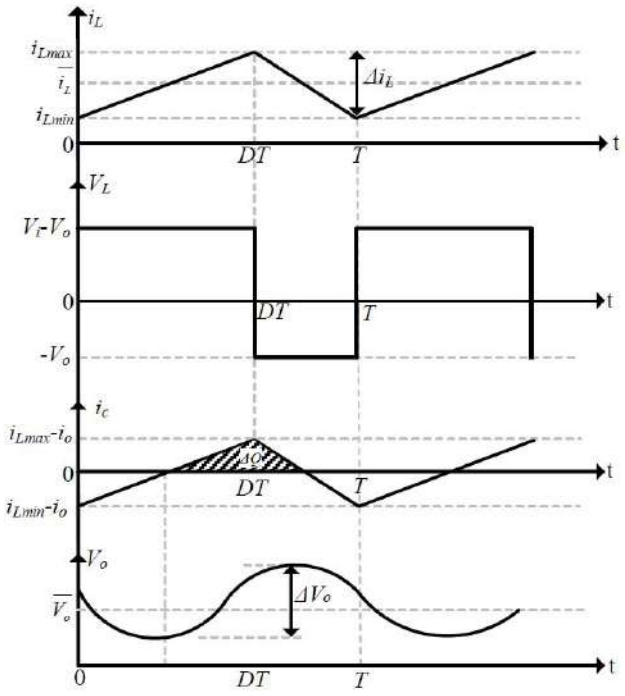


Fig. 2. Operating mode analysis.

The closed-loop controller is designed to ensure that the output voltage (V_0) follows the reference voltage (V_{ref}). proportional-integral (PI) control is widely used in buck converters due to its ease of implementation and low cost. The PI control signal can be expressed as

$$u(t) = K_p e(t) + K_i \int_0^\tau e(\tau) d\tau \tag{7}$$

In Eq. (7), $u(t)$ denotes the control signal, while $e(t)$ represents the controller's error signal. Here, K_p is the proportional gain, and K_i is the integral gain. The proportional term works to minimize the error by scaling it with a gain factor. When K_p is high, even small errors may lead to significant output adjustments, potentially destabilizing the system. On the other hand, low K_p values might cause sluggish responses to disturbances, delaying the controller's corrective action, which is generally undesirable. The integral term, by contrast, accumulates the error over

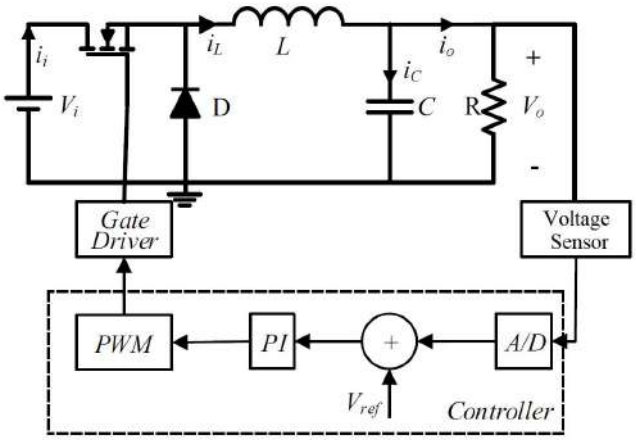


Fig. 3. General block diagram of the system.

time, capturing its total area. Each cycle, the error is compounded with the integral gain K_i, helping to reduce residual errors that the proportional control alone might leave. The controller parameters thus play a crucial role in shaping the closed-loop converter's performance.

Microcontrollers play a central role in generating PWM signals for digitally controlled switch-mode power supplies. However, the low output level of these PWM signals requires an intermediary control circuit between the microcontroller's output and the gate of the power switch to effectively charge and discharge the capacitance at the high-side switch. This circuit amplifies the PWM signal voltage from the microcontroller, usually ranging from zero to five volts, to the required switching level. Rapid switching of the power component is facilitated by this control circuit, effectively reducing switching-related energy losses and minimizing parasitic capacitance at key nodes. Moreover, the control circuit shields the microcontroller from voltage spikes and electrical noise, while also reducing its operational burden, thereby decreasing energy losses and thermal stress by offloading energy management to the control circuit.

IV. Simulation Analysis Results

In this study, simulations are carried out using the MATLAB/Simulink platform, with the total simulation time set to 0.1 seconds. The simulation results are illustrated in Fig. 4. In this simulation experiment, the input voltage is assumed to be $50V_{DC}$, while the output voltage is set to $25V_{DC}$. Additionally, the carrier frequency is defined as 2.5 kHz, and the load resistance is 12.5 Ω. As depicted in Fig. 4a, the input voltage remained constant throughout the simulation, maintaining a value of 50V. The reference output voltage is set to $25V_{DC}$. Fig. 4b shows that the output voltage closely follows the reference value, with the controller achieving a settling time of 0.003 seconds. The results for the load current and inductor current are presented in Figs. 4c and 4d, respectively. The current stabilized rapidly without significant overshoot or extended transients. These simulation results validate the effectiveness of the digital control strategies for step-down energy conversion systems.

V. Conclusion

The analysis delves into the operational modes of DC-DC buck converters, providing detailed guidelines for optimal

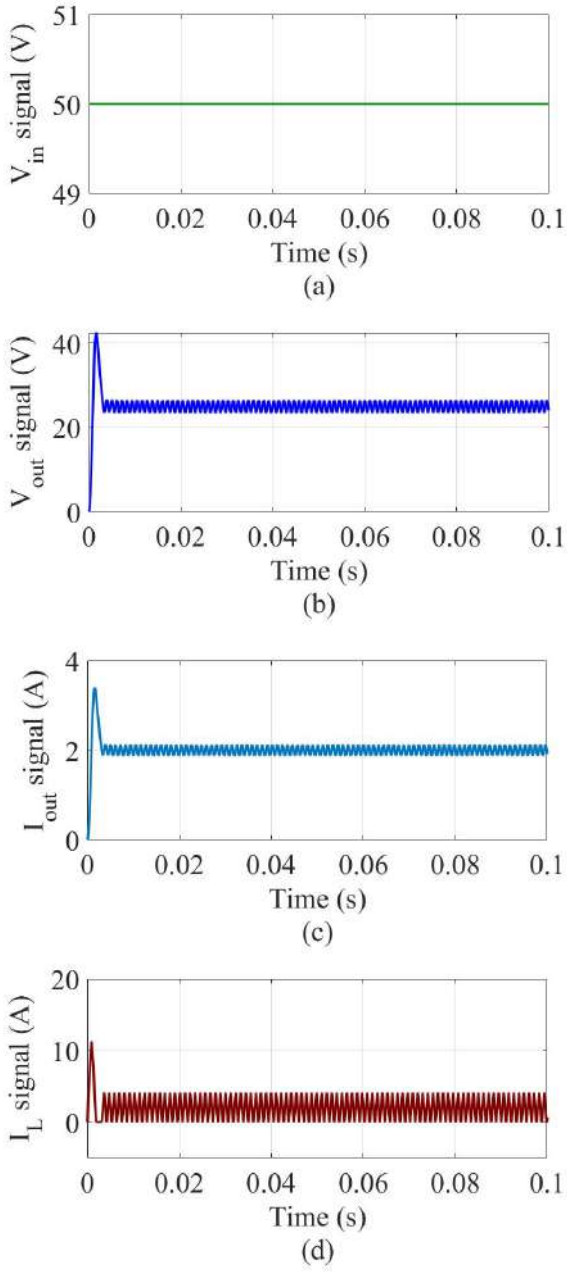


Fig. 4. Results of the simulation study.

design parameter selection. Key findings underscore the critical role of feedback mechanisms in counteracting instability and oscillations, which can compromise performance across varying conditions. Additionally, the research highlights the benefits of digital control techniques for step-down converters, demonstrating their ability to improve reliability and overall performance. Simulations in MATLAB/Simulink offer a rigorous view of converter behavior, affirming the digital controller's effectiveness. With rising demand for efficient energy solutions, particularly in renewable energy and electric vehicles, robust DC-DC converters are more essential than ever. This work emphasizes the value of advanced control strategies for optimizing buck converter performance, suggesting that future research could refine these digital techniques further and broaden their application in energy management systems.

REFERENCES

[1] T.-Z. Ang, M. Salem, M. Kamarol, H. S. Das, M. A. Nazari, and N. Prabaharan, "A comprehensive study of renewable energy sources: Classifications, challenges and suggestions," *Energy Strategy Reviews*, vol. 43, 100939, Sep. 2022.

[2] Q. Hassan, P. Viktor, T. J. Al-Musawi, B. M. Ali, S. Algburi, H. M. Alzoubi, A. K. Al-Jiboory, A. Z. Sameen, H. M. Salman, and M. Jaszczur, "The renewable energy role in the global energy transformations," *Renewable Energy Focus*, vol. 48, 100545, Mar. 2024.

[3] A. Israr, Q. Yang, W. Li, and A. Y. Zomaya, "Renewable energy powered sustainable 5G network infrastructure: Opportunities, challenges and perspectives," *Journal of Network and Computer Applications*, vol. 175, 102910, Feb. 2021.

[4] T. L. Le and M. F. Hsieh, "An enhanced direct torque control strategy with composite controller for permanent magnet synchronous motor," *Asian Journal of Control*, vol. 26, no. 4, pp. 1683–1702, Jul. 2024.

[5] H. A. Behabtu, M. Messagie, T. Coosemans, M. Berecibar, K. A. Fante, A. A. Kebede, and J. Van Mierlo, "A review of energy storage technologies' application potentials in renewable energy sources grid integration," *Sustainability*, vol. 12, no. 24, 10511, Dec. 2020.

[6] S. Ullah, A. M. A. Haidar, P. Hoole, H. Zen, and T. Ahfock, "The current state of distributed renewable generation, challenges of interconnection and opportunities for energy conversion based DC microgrids," *Journal of Cleaner Production*, vol. 273, 122777, Nov. 2020.

[7] M. G. Molina, "Energy storage and power electronics technologies: A strong combination to empower the transformation to the smart grid," *Proceedings of the IEEE*, vol. 105, no. 11, pp. 2191–2219, Nov. 2017.

[8] M. Ibrahim, V. Rjabtšikov, and R. Gilbert, "Overview of digital twin platforms for EV applications," *Sensors*, vol. 23, no. 3, 1414, Jan. 2023.

[9] T. L. Le, M. F. Hsieh, P. T. Nguyen, and M. T. Nguyen, "Speed control for permanent magnet synchronous motor based on terminal sliding mode high-order control," *in 2023 International Conference on System Science and Engineering (ICSSE)*, pp. 496–501, Jul. 2023.

[10] A. F. de Souza, F. L. Tofoli, and E. R. Ribeiro, "Switched capacitor DC-DC converters: A survey on the main topologies, design characteristics, and applications," *Energies*, vol. 14, no. 8, 2231, Apr. 2021.

[11] F. Yi and F. Wang, "Review of voltage-bucking/boosting techniques, topologies, and applications," *Energies*, vol. 16, no. 2, 842, Jan. 2023.

[12] V. Q. Nguyen, M. T. Nguyen, T. L. Le, and N. B. Nguyen, "Current control for multi-level inverters based on fuzzy logic sliding mode control approach," *in 2023 International Conference on System Science and Engineering (ICSSE)*, pp. 444–449, Jul. 2023.

[13] K. V. G. Raghavendra, K. Zeb, A. Muthusamy, T. N. V. Krishna, S. V. S. V. Prabhudeva Kumar, D.-H. Kim, M.-S. Kim, H.-G. Cho, and H.-J. Kim, "A comprehensive review of DC–DC converter topologies and modulation strategies with recent advances in solar photovoltaic systems," *Electronics*, vol. 9, no. 1, 31, Dec. 2019.

[14] W. Yan, W. Li, and R. Liu, "A noise-shaped buck DC–DC converter with improved light-load efficiency and fast transient response," *IEEE Transactions on Power Electronics*, vol. 26, no. 12, pp. 3908–3924, Dec. 2011.

[15] V. Q. Nguyen and T. L. Le, "Flexible control with fuzzy observer-based sliding mode for multilevel inverter," *Journal of Electrical Engineering & Technology*, 2024.

[16] M. A. Hassan, C.-L. Su, F.-Z. Chen, and K.-Y. Lo, "Adaptive passivity-based control of a DC–DC boost power converter supplying constant power and constant voltage loads," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 6, pp. 6204–6214, Jun. 2022.

[17] M. Eskandari, L. Li, M. H. Moradi, P. Siano, and F. Blaabjerg, "Optimal voltage regulator for inverter interfaced distributed generation units Part I: Control system," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 4, pp. 2813–2824, Oct. 2020.

[18] N. Hosseinzadeh, A. Aziz, A. Mahmud, A. Gargoom, and M. Rabbani, "Voltage stability of power systems with renewable-energy inverter-based generators: A review," *Electronics*, vol. 10, no. 2, 115, Jan. 2021.

[19] S. Kapat and P. T. Krein, "A tutorial and review discussion of modulation, control and tuning of high-performance DC-DC converters based on small-signal and large-signal approaches," *IEEE Open Journal of Power Electronics*, vol. 1, pp. 339–371, Aug. 2020.

[20] J. J. Tai, S. K. Phang, and C. L. Hoo, "Application of steady-state integral proportional integral controller for inner dynamics control loop of multi-rotor UAVs," *in 2018 Fourth International Conference on Advances in Computing*, Communication & Automation (ICACCA), Subang Jaya, Malaysia, 2018.

[21] B. K. Oleiwi and M. J. Mohamed, "Optimal design of linear and nonlinear PID controllers for speed control of an electric vehicle," *Journal of Intelligent Systems*, vol. 2024, Article 0028, Sep. 2024.

[22] L. He, W. Ye, Z. He, K. Song, and Q. Shi, "A combining sliding mode control approach for electric motor anti-lock braking system of battery electric vehicle," *Control Engineering Practice*, vol. 102, Article 104520, Sep. 2020.

[23] V. Q. Nguyen and T. L. Le, "A robust control for five-level inverter based on integral sliding mode control," *International Journal of Integrated Engineering*, vol. 15, no. 4, pp. 177–192, Aug. 2023.

[24] W. Gu, K. P. Valavanis, M. J. Rutherford, et al., "UAV Model-based Flight Control with Artificial Neural Networks: A Survey," *Journal of Intelligent & Robotic Systems*, vol. 100, pp. 1469–1491, 2020.

[25] T. Lei, Y. Wang, X. Jin, Z. Min, X. Zhang, and X. Zhang, "An Optimal Fuzzy Logic-Based Energy Management Strategy for a Fuel Cell/Battery Hybrid Power Unmanned Aerial Vehicle," *Aerospace*, vol. 9, no. 2, p. 115, 2022.

[26] G. A. N. Pussente, E. P. de Aguiar, A. L. M. Marcato, and M. F. Pinto, "UAV Power Line Tracking Control Based on a Type-2 Fuzzy-PID Approach," *Robotics*, vol. 12, no. 2, p. 60, 2023.

[27] M. P. E. Rajamani, R. Rajesh, and M. Willjuice Iruthayarajan, "Design and Experimental Validation of PID Controller for Buck Converter: A Multi-Objective Evolutionary Algorithms Based Approach," *IETE Journal of Research*, vol. 69, no. 1, pp. 21-32, 2023.

[28] Carlos Aguilar-Ibanez, Javier Moreno-Valenzuela, O. García-Alarcón, Mizraim Martinez-Lopez, José Ángel Acosta, and Miguel S. Suarez-Castanon, "PI-Type Controllers and Σ–Δ Modulation for Saturated DC-DC Buck Power Converters," *IEEE Access*, vol. 9, pp. 20346-20357, 2021.

[29] Ulbio Alejandro-Sanjines, Anthony Maisincho-Jivaja, Victor Asanza, Leandro L. Lorente-Leyva, and Diego H. Peluffo-Ordóñez, "Adaptive PI Controller Based on a Reinforcement Learning Algorithm for Speed Control of a DC Motor," *Biomimetics*, vol. 8, no. 5, p. 434, 2023.

[30] O. Zandi và J. Poshtan, "Voltage Control of a Quasi Z-Source Converter Under Constant Power Load Condition Using Reinforcement Learning," *Control Engineering Practice*, vol. 135, 105499, June 2023.

Multi-Node Fingerprint Attendance System with Centralized Feature Storage

Tu Huynh Thanh
Faculty of Electrical & electronics
Saigon Technology University
Ho Chi Minh City, Vietnam
tu.huynhthanh@stu.edu.vn

Luc Dang Tran
Faculty of Electrical & electronics
Saigon Technology University
Ho Chi Minh City, Vietnam
dh42120047@student.stu.edu.vn

Phuc Ho Nguyen Hoang
Faculty of Electrical & electronics
Saigon Technology University
Ho Chi Minh City, Vietnam
dh42114670@student.stu.edu.vn

***Abstract*— This paper presents a comprehensive design and implementation of a multi-node attendance system utilizing ESP32 microcontrollers, AS608 fingerprint modules, and RFID readers. The system is designed to enhance the efficiency of attendance tracking in educational institutions by integrating biometric authentication with RFID technology. The proposed system allows for the registration and authentication of students through fingerprint recognition and RFID card scanning, with data being communicated to a centralized server for fingerprint feature storage.**

Keywords—fingerprint, attendance, esp32, REST, IoT, Flask

I. INTRODUCTION

The need for efficient attendance systems in educational institutions has grown significantly, driven by the demand for accurate tracking of student presence. Traditional methods, such as roll calls or manual sign-ins, are often time-consuming and prone to errors. This paper proposes a modern solution that leverages biometric technology and RFID systems to streamline the attendance process. The system comprises multiple nodes, each equipped with an ESP32 microcontroller, an AS608 fingerprint module, and an RFID reader.

II. RELATED WORKS

A study by [1] presents an innovative attendance and security system that utilizes fingerprint recognition alongside IoT and ultrasonic sensors. This system offers a dual-layer security approach by not only verifying the identity of individuals through fingerprint scanning but also ensuring their physical presence with an ultrasonic sensor. The combination of these technologies addresses common issues in attendance systems, such as proxy attendance, which is difficult to counter in traditional systems.

The study [2] developed an Arduino-based Fingerprint Scan Attendance-taking System that addresses common issues in traditional attendance methods. The system utilizes an AS608 Fingerprint Sensor Module along with an Arduino UNO board to automate attendance tracking. The system also incorporates the DS3231 Real-Time Clock (RTC) Module to ensure accurate timekeeping for attendance records. This setup offers a low-cost and standalone solution that can function without internet connectivity, making it particularly useful for remote educational institutions where internet access may be limited or unavailable

In [7], a mobile fingerprint-based attendance system is introduced, where students use an Android application developed with the Android SDK to mark their attendance. The system enables the creation of sessions and student registration. Attendance data is captured in real-time and updated in Firebase's real-time database. Additionally, the attendance records are exported and displayed in an Excel spreadsheet, providing a convenient way to manage and review the data.

In [8], the author proposes a wireless fingerprint-based attendance system integrated with an Android application. In this system, students register their fingerprints along with their personal details through the mobile app. After successful registration, attendance is marked by scanning the fingerprint at the sensor. The attendance data is transmitted to the mobile app via an Arduino AT Mega 2560, enabling seamless data collection and storage.

In the reviewed studies, there are still limitations, such as systems operating with only a single independent node, with authentication mechanisms relying entirely on a fingerprint module, and fingerprint features being stored locally on the fingerprint module. Our research aims to develop an attendance system that can operate multiple nodes simultaneously, with fingerprint features stored on a database server. This system allows for easier scalability of fingerprint authentication and reduces dependency on the fingerprint module. As a result, the failure of a single fingerprint module does not significantly impact the overall system.

III. SYSTEM ARCHITECTURE

A. Block diagram

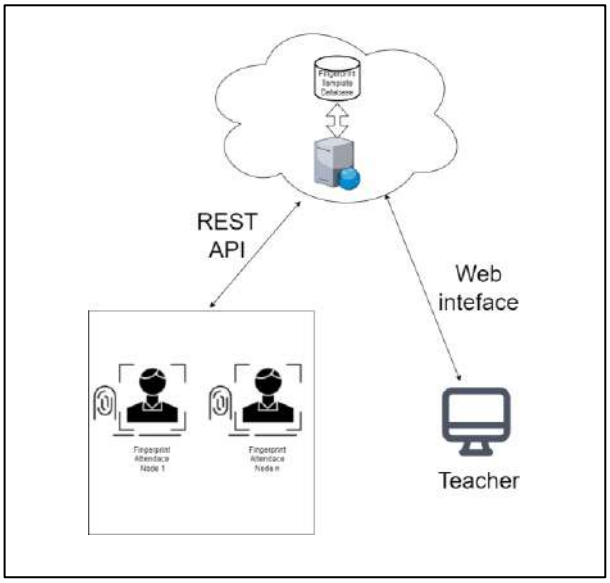


Fig. 2. Block diaram of Attendance system

In the system diagram, multiple nodes can operate simultaneously. These nodes serve both as fingerprint registration points and attendance recording stations. A node

scans the fingerprint image, extracts fingerprint features, and sends these features to the server for storage. During the attendance phase, the node retrieves the stored features from the server and performs a matching process with the fingerprint being authenticated. The node communicates with the server using the REST protocol. Instructors can monitor attendance information via a web interface, which connects to the server. The web interface is built on the Flask framework.

B. Hardware componets

1) ESP32 microcontroller.

The ESP32 serves as the central processing unit for each node, providing Wi-Fi and Bluetooth connectivity, which facilitates communication with the server.



Fig. 2. ESP32 Devkit

2) AS608 fingerprint module.

This module is responsible for capturing fingerprint images and extracting features for authentication. It includes an onboard processor to perform matching algorithms. We selected the AS608 due to its popularity, affordability, and ability to meet the necessary specifications for the desired functionality. The AS608 can store approximately 200 fingerprint feature templates directly in its memory, but this specification is not critical for our system, as the feature templates are not stored locally on the fingerprint module but are centrally stored on the server.

TABLE I. SPECIFICATION OF AS608

Resolution	*500 DPI*
False Acceptance rate (FAR)	*<0.001%*
False Rejection rate (FRR)	*<1%*
Time of scanned image	*1 second*
Fingerprint matching speed	*<1 second*
Supply current	*60 mA*



Fig. 3. AS608 fingerprint module

3) RC522 RFID reader:

The RFID reader scans student ID cards, allowing for the input of student identification numbers. Using RFID to read student ID numbers allows for faster and more accurate retrieval of information, offering greater convenience compared to equipping a keyboard for users to manually enter data.



Fig. 4. RC522 RFID reader

IV. IMPLIMENTATION

The software for the ESP32 nodes is developed using the Arduino IDE, leveraging libraries for the AS608 fingerprint module and RFID reader. The following steps outline the implementation process:.

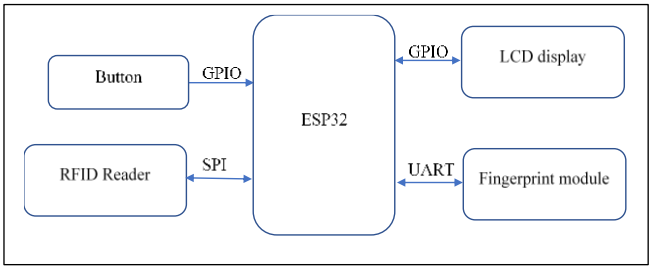


Fig. 5. Block diagram of a single-node fingerprint attendance system

A. Fingerprint registration

- The student scans their RFID card.
- The AS608 module captures the fingerprint image.
- The fingerprint feature is extracted and sent to the server along with the student ID.

During the registration process, two fingerprint images are captured and converted into fingerprint features, while the student ID is read via the RFID reader. These two pieces of information are then sent to the server through REST communication.

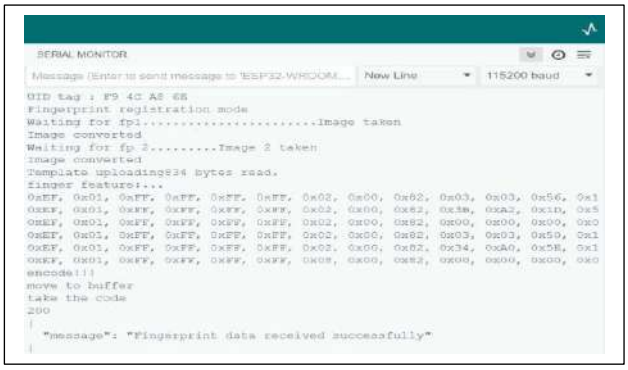


Fig. 6. The fingerprint features are extracted from the AS608 module

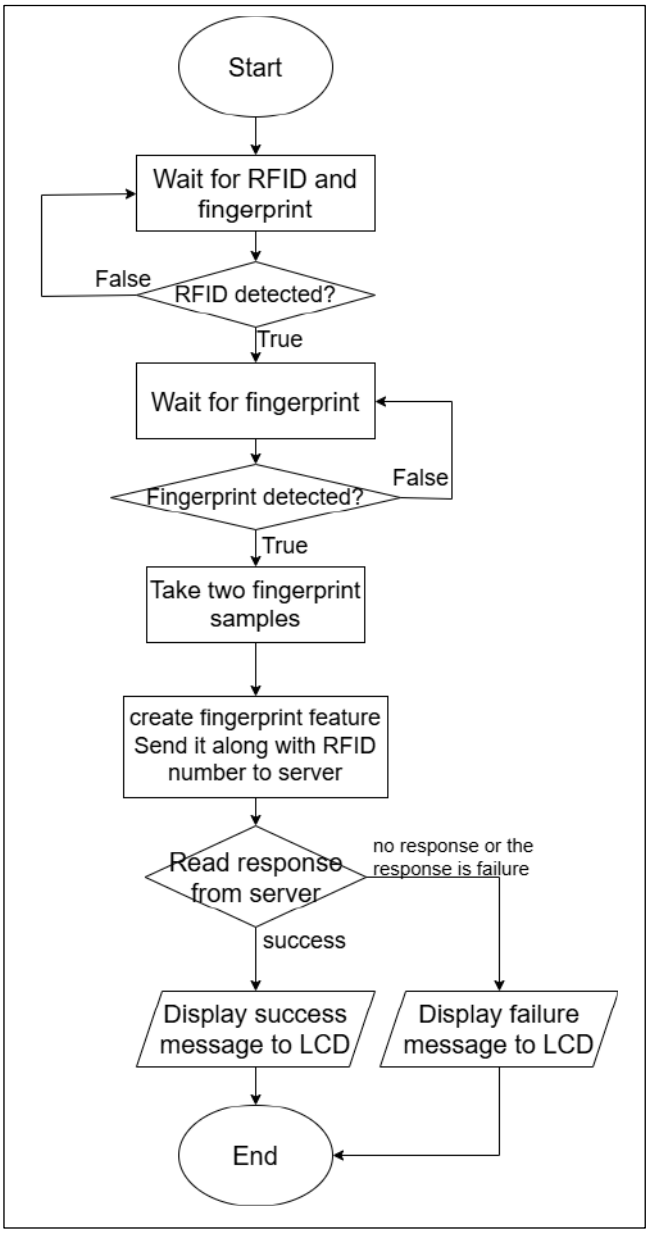


Fig. 7. Flowchart for registering fingerprint

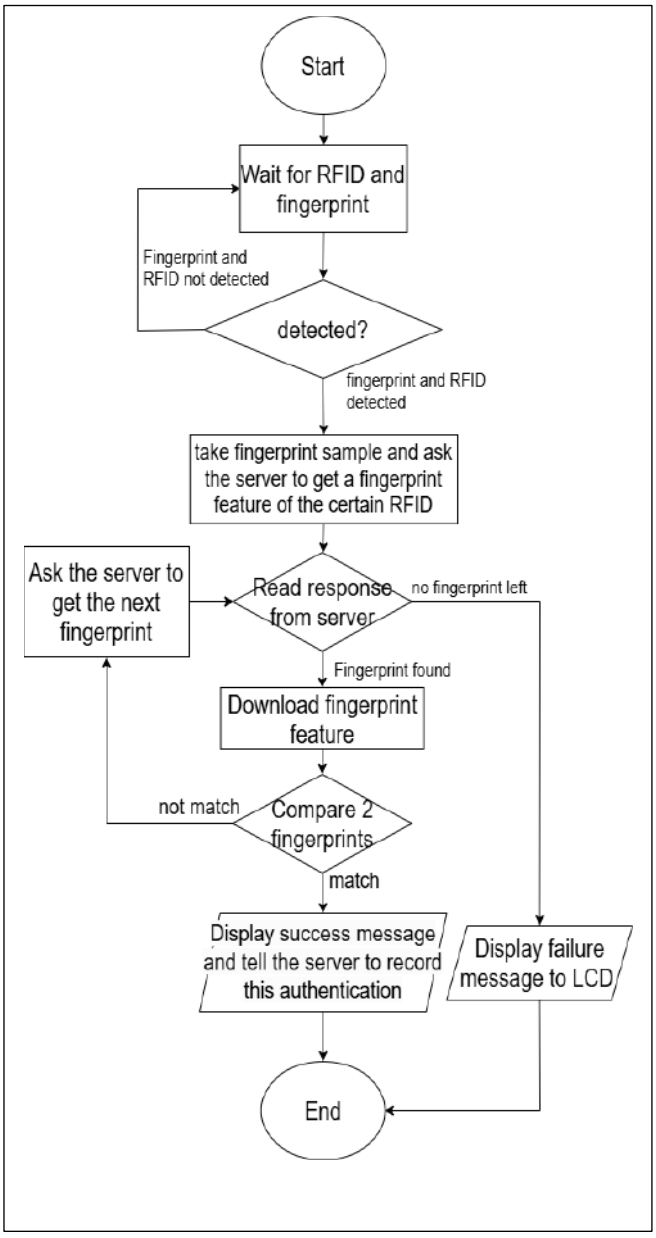


Fig. 8. Flowchart for attendance check

B. Fingerprint Authentication

- The student scans their RFID card.
- The AS608 module captures the fingerprint image.
- The feature is extracted and compared with the stored feature retrieved from the server.
- If a match is found, the node sends a notification to the attendance server to mark the student as present.

During the attendance process, the student ID is read via the RC522 RFID reader, and the fingerprint of the individual being authenticated is scanned and converted into fingerprint features. The node uses the student ID to request the corresponding fingerprint features stored on the server. These stored features are then matched with the newly captured fingerprint features by the AS608 module. If the match exceeds a certain threshold, the user is authenticated, and attendance is recorded for the student with the corresponding ID and fingerprint.

C. Matching Algorithm

The AS608 module utilizes a proprietary matching algorithm to compare the extracted fingerprint features. This algorithm is designed to minimize false acceptance and false rejection rates, ensuring reliable authentication.

D. Prototype of a single-node attendance

We conducted an experiment with two nodes, both communicating with the database server. The nodes were able to operate simultaneously, and the fingerprint registration process could be performed on any node, as the fingerprint features are centrally sent to the server. We successfully tested registering fingerprints on one node and authenticating them on the other. This demonstrates that the system can scale well if the number of nodes increases. Replacing nodes is straightforward, as there is no need to repeat the fingerprint registration process due to the centralized storage of fingerprint features on the server.

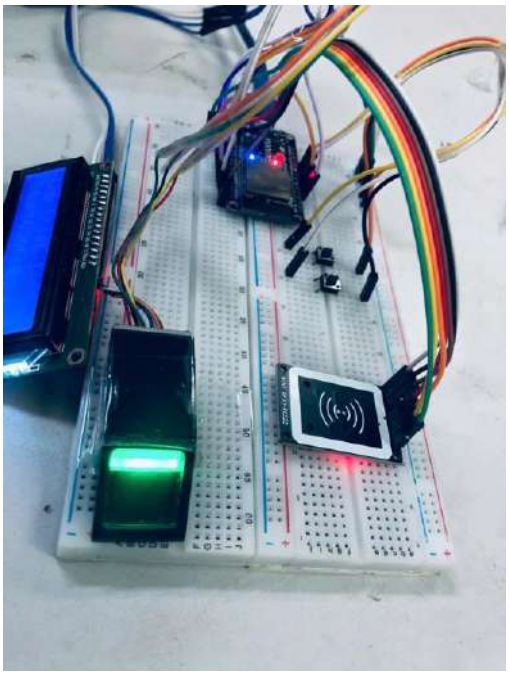


Fig. 9. Prototype of single node attendance

V. Data management

A. Server database

The centralized server hosts a database that stores student information, including:

- Student ID
- Attendance records
- Fingerprint features

We organize student management information and teacher-related data in an SQLite database. Flask-SQLAlchemy is used to interact with the database. SQLite is chosen as the database to reduce costs and retrieval time. Fingerprint features are stored as separate files, with each file named according to the RFID tag numbers sent by the nodes during the fingerprint registration process.

B. Web server

We implemented the management interface as a web-based application using the Flask framework. The interface includes several functions: managing students, courses, teachers, student card information, and students' RFID codes. Teachers can view student attendance information for the courses they are responsible for, including the time of attendance and the node used for attendance tracking.



Fig. 10. A page of web server

VI. Conclusion

The proposed multi-node attendance system effectively integrates fingerprint recognition and RFID technology to streamline the attendance process in educational institutions. By utilizing ESP32 microcontrollers and AS608 fingerprint modules, the system offers a reliable and efficient solution for tracking student attendance. Future work may involve enhancing the system's scalability and security features, as well as exploring additional biometric modalities for authentication.

References

[1] S. Al Amin, M. A. Islam and M. S. Islam, "A Fingerprint Based Smart Attendance and Security System Using IoT and Ultrasonic Sensor," 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), Rajshahi, Bangladesh, 2021, pp. 1-5, doi: 10.1109/ACMI53878.2021.9528092

[2] M. S. N. bin Ismail Marzuki, M. A. bin Abdul Razak, A. I. bin Mod Arifin, L. H. bin Zulkornain, A. H. bin Kassim and H. H. bin Abu Bakar, "Arduino-based Fingerprint Scan for Attendance-taking System," 2024 IEEE 6th Symposium on Computers & Informatics (ISCI), Kuala Lumpur, Malaysia, 2024, pp. 133-137, doi: 10.1109/ISCI62787.2024.10668322.

[3] P. Goyal, V. Nigam, S. A. Khan and S. K. Singh, "Automated Attendance Management System," 2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2023, pp. 452-455, doi: 10.1109/ICAC3N60023.2023.10541627

[4] C. Ganjiwale, R. Bachhav, T. Bagal, H. Magar and C. Sukte, "Efficient Attendance Management: A Smart Bluetooth-Based System for Seamless Monitoring and Automation," 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon), Pune, India, 2024, pp. 1-5, doi: 10.1109/MITADTSoCiCon60330.2024.10575910.

[5] M. Ramalingam, S. Vinothkumar, S. Varadhaganapathy and S. Subha, "Multi-User Authentication Using Biometric Sensor Using Parallel Processing Algorithm For Attendance Monitoring," 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2021, pp. 84-89, doi: 10.1109/ICICT50816.2021.9358485.

[6] R. Chauhan, S. Negi, A. Kumar and S. Devliyal, "A Raspberry pi based smart attendance system," 2024 International Conference on Automation and Computation (AUTOCOM), Dehradun, India, 2024, pp. 612-617, doi: 10.1109/AUTOCOM60220.2024.10486161.

[7] Devaprakash, Gowtham, Murali, Muralidharan and V. J. Vijayalakshmi, "Centralized Attendance Monitoring System," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 1288-1291, doi: 10.1109/ICACCS48705.2020.9074162

[8] B. Swain, J. Halder, S. Sahany, P. P. Nayak and S. Bhuyan, "Automated Wireless Biometric Fingerprint Based Student Attendance System," 2021 1st Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology(ODICON), Bhubaneswar, India, 2021, pp. 1-7, doi: 10.1109/ODICON50556.2021.9428983.

Design and Manufacturing of a Control System for a Pneumatic Robotic Arm for Training Purposes

Dinh Hai Vu
Faculty of Mechanical Engineering
Saigon Technology University
Ho Chi Minh City, Vietnam
hai.vudinh@stu.edu.vn

Van Thang Pham
Faculty of Mechanical Engineering
Saigon Technology University
Ho Chi Minh City, Vietnam
thang.phamvan@stu.edu.vn

Thanh Hung Nguyen
Faculty of Mechanical Engineering
Saigon Technology University
Ho Chi Minh City, Vietnam
hung.nguyenthanh@stu.edu.vn

Huu Han Ho
Faculty of Mechanical Engineering
Saigon Technology University
Ho Chi Minh City, Vietnam
han.hohuu@stu.edu.vn

Abstract— This study focuses on designing and manufacturing a new control system for an industrial pneumatic robotic arm, specifically to meet the practical training needs of students in the field of mechatronics. The primary objective of this research is to build an "open" control system, allowing learners to freely connect input and output signals to program the robotic arm according to specific operational requirements. This design enhances students' autonomy and understanding of control systems. It supports flexible and easy adjustments to the robot's operation, helping students gain a stronger grasp of robotic arm control and operation in a real industrial setting. The research methodology consists of three main steps. First, the study applies a programmable logic controller (PLC) as the system's foundation. The PLC was chosen for its flexibility and practical relevance, making it easier for students to engage with industrial applications. Second, the study explores the structure of the pneumatic drive system, control factors, and necessary conditions for stable operation in the robotic arm's pneumatic system. Finally, the research identifies the essential actuators and piston position sensors to ensure effective operation of the robotic arm. The findings demonstrate that the new control system, developed using PLC technology, can flexibly meet practical training requirements and holds great potential for application in technical education. This system not only fulfills basic control needs but also provides extended functionality, allowing students to easily modify the system to practice programming skills for controlling pneumatic robotic arms. As a result, it lays a solid foundation for training in the field of mechatronics.

I. INTRODUCTION

Pneumatic robotic arms are widely used in various industries, particularly in the plastics industry, due to their advantages such as high speed, operational precision, and durability in harsh or hazardous environments where human presence is undesirable. However, most industrial pneumatic robotic arms from major manufacturers are designed with built-in controllers or integrated into automated production systems, making it challenging to intervene or adjust them for specific applications, such as in plastic molding production lines. This design limitation can restrict companies' ability to adapt and improve their production processes to meet specialized requirements.

In today's diverse production landscape, with various scales and types of manufacturing, the ability to flexibly adjust and program pneumatic robotic arms is essential. Operators must have a solid understanding of automated control to be able to program and adjust the robotic arm according to the specific production needs of each business. This need calls for an open control system that allows operators to customize the robotic arm's functions without being limited by the manufacturer's fixed design.

Therefore, this research focuses on designing a new control system for pneumatic robotic arms, capable of meeting the production requirements of factories, allowing users to adjust control parameters as needed, and flexibly reprogram the robot's operations. This control system is intended not only for production but also for educational and training purposes, enabling students and technicians to practice programming and controlling the robotic arm in a hands-on laboratory environment. Although there have been some studies on the application of microcontrollers and microprocessors in pneumatic robotic arms, this research emphasizes developing a highly applicable control system for training purposes, while also meeting diverse production requirements.

II. DESCRIPTION OF THE PNEUMATIC ROBOTIC ARM

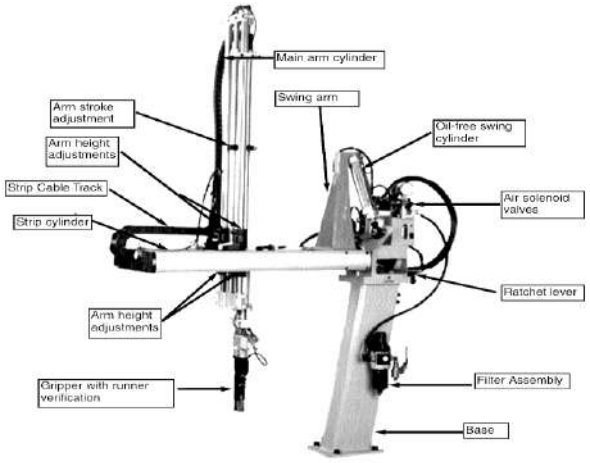


Fig. 1 Functions of Pneumatic Robotic Arm Components

A. Functions of Pneumatic Robotic Arm Components (Fig. 1)

- Main Arm Cylinder: The main cylinder of the arm, providing the primary force and movement to control the up-and-down or vertical motion of the robotic arm.

- Swing Arm: The swing arm allows the robotic arm to rotate or move horizontally, enhancing its reach to various positions.

- Oil-free Swing Cylinder: A swing cylinder that operates without oil, controlling the arm's rotational movement without the need for lubrication, making it suitable for clean environments.

- Air Solenoid Valves: Pneumatic solenoid valves that control the flow of air into the cylinders, thereby regulating the movements of the robotic arm.

- Ratchet Lever: A ratchet lever typically used to lock or adjust the position of the robotic arm.

- Filter Assembly: The air filter assembly removes impurities from the air before it enters the system, protecting the cylinders and valves from damage caused by dust and debris.

- Base: The base of the robotic arm, providing stability to the entire system and anchoring other components on the platform.

- Gripper with Runner Verification: The robotic arm gripper with runner verification functionality, which grips and inspects the item to ensure the product or runner (excess material in plastic molding) is handled correctly.

- Strip Cylinder: The strip cylinder assists in separating parts or pushing items away from the gripper.

- Strip Cable Track: The cable track for the strip cylinder, which secures electrical and pneumatic cables neatly and prevents tangling.

- Arm Stroke Adjustment: Allows for adjustment of the arm's stroke, enabling changes to the arm's range of motion to meet various requirements.

B. Basic Specifications of the Pneumatic Robotic Arm

The dimensions and structural specifications of the robotic arm are shown in Figure 2 and Table 1..

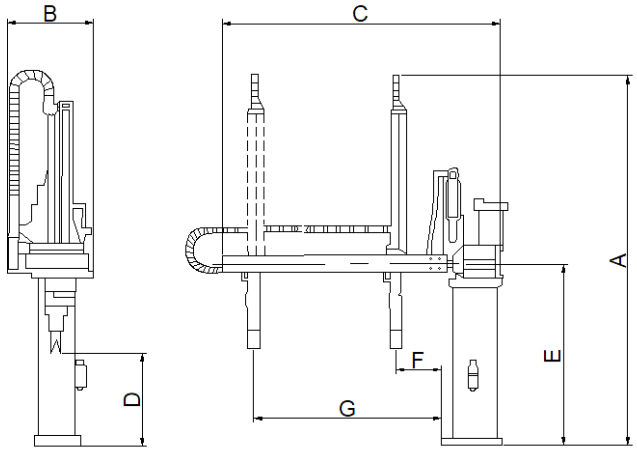


Fig. 2 The dimensions of the robotic arm 1.

Dimensions (mm)	
A - Height	1300
B - Width	290
C - Length	720
D - Clamp Center	120
E - Distance from Axis Center to Base	557
F - Moving Distance	50-430mm
Weight (kg):	
No Load	37kg
With Load	39kg
Performance Characteristics:	
Arm Rotation Angle	45-90°
Clamp Rotation Angle	90°
Minimum Cycle Time	3.2s
Physical Characteristics:	
Air Consumption	0.5kg
Working Air Pressur	11.6MPa
(220V / 1 phase, 50-60 Hz)	2.5W

Table 1. The structural specifications of the robotic arm

C. Pneumatic System Circuit

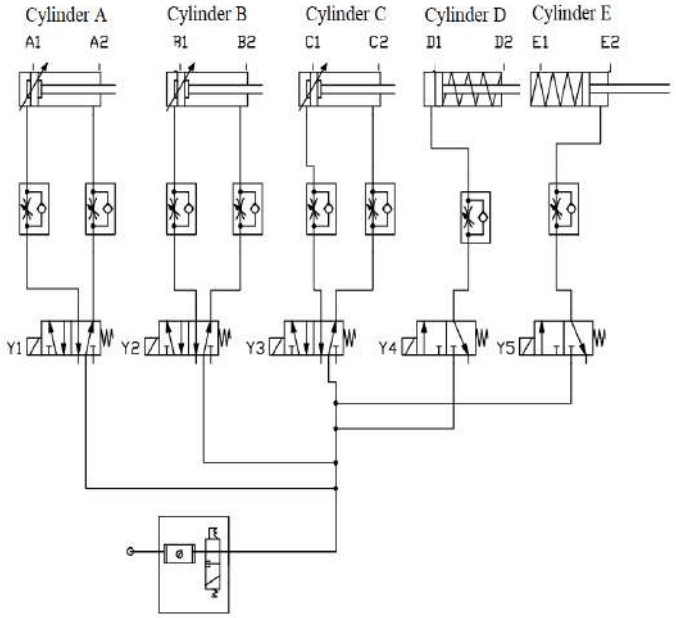


Fig. 3 Pneumatic circuit of the Robotic Arm

The above image (Fig 3) is a diagram of the pneumatic system. The pneumatic robotic arm uses pneumatic energy in the form of pressure and airflow to generate movement. This pneumatic energy is supplied by an air compressor available in the factory. The air pressure is regulated by a pressure control valve, ensuring that the pneumatic pressure supplied to the robotic arm remains constant. The forward and backward movements of the piston are controlled by pneumatic distribution valves operated by a 24 VDC electric control system

III. DESIGN OF CONTROL SYSTEM USING PLC

The objective is to design a flexible robotic arm control system that allows for easy program modification and parameter

adjustments. One of the programs (operating modes) is shown in Figure 4.

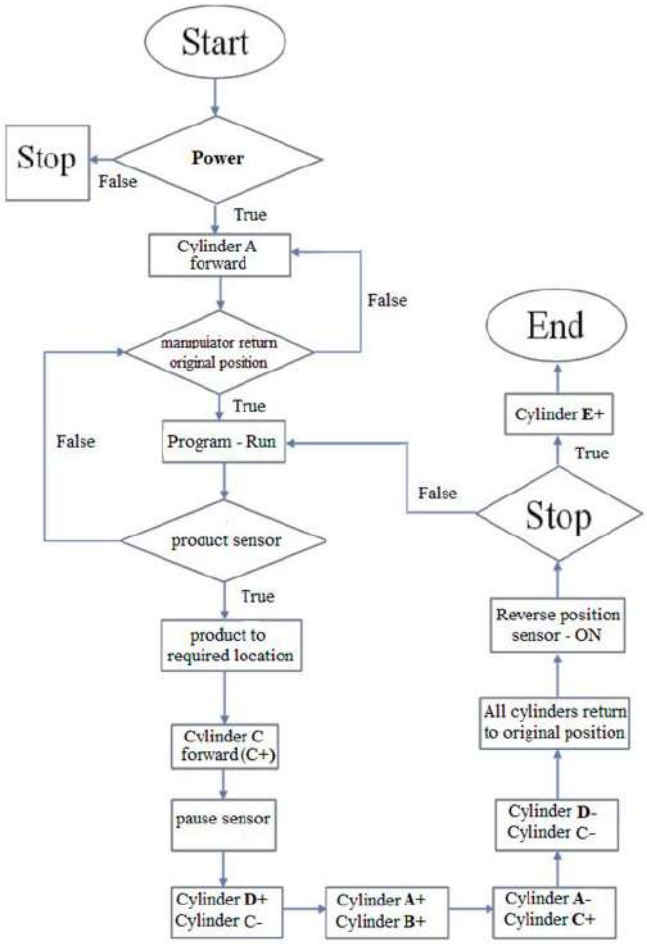


Fig. 4 Flowchart Representing the Operation of the Robotic Arm

The operating procedure of the pneumatic robotic arm begins with checking the start power. If the power is correct, cylinder E is activated, and the robotic arm moves to its initial position. Once the robotic arm is confirmed to be in its starting position, the program is initiated. Next, the sensor detects the presence of the workpiece; if the workpiece is identified at the correct location, it is moved to the designated position, and cylinder C begins operating. At this stage, a temporary stop sensor is activated, and if all conditions are met, cylinder D performs a clamping action while cylinder C moves backward. Following this, cylinder A rotates, and cylinder B moves forward to continue the process. Once completed, cylinder A rotates again to return the workpiece to its initial position along with cylinder C. If the sensor confirms reactivation, the cylinders return to their original positions; otherwise, the process stops. Finally, cylinder E locks to complete the cycle, and the robotic arm returns to its starting state, ready for the next cycle. This diagram describes an automatic sequential control process for the pneumatic robotic arm, with precise steps of inspection and control to ensure efficient and safe operation in an industrial environment.

Figure 5 illustrates the operation of the pneumatic robotic arm using a stroke-step diagram.

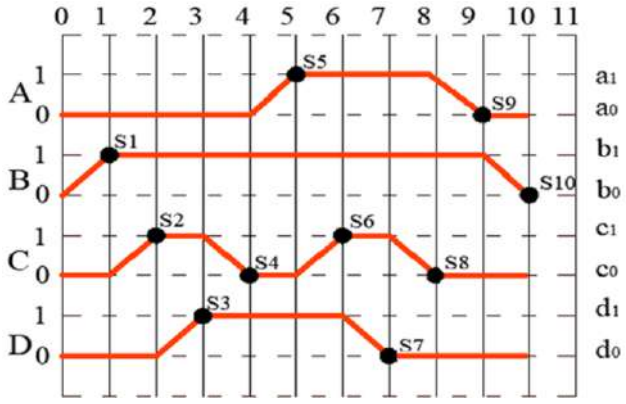


Fig. 5 STROKE-STEP Diagram Representing One Operating Cycle

The control system using a PLC is widely trusted in machinery and production lines due to its many advantages: high stability, excellent noise resistance, and various options for system expansion, such as adding additional modules. For these reasons, in this study, we also chose Mitsubishi's PLC for the pneumatic robotic arm control system (Fig. 6).

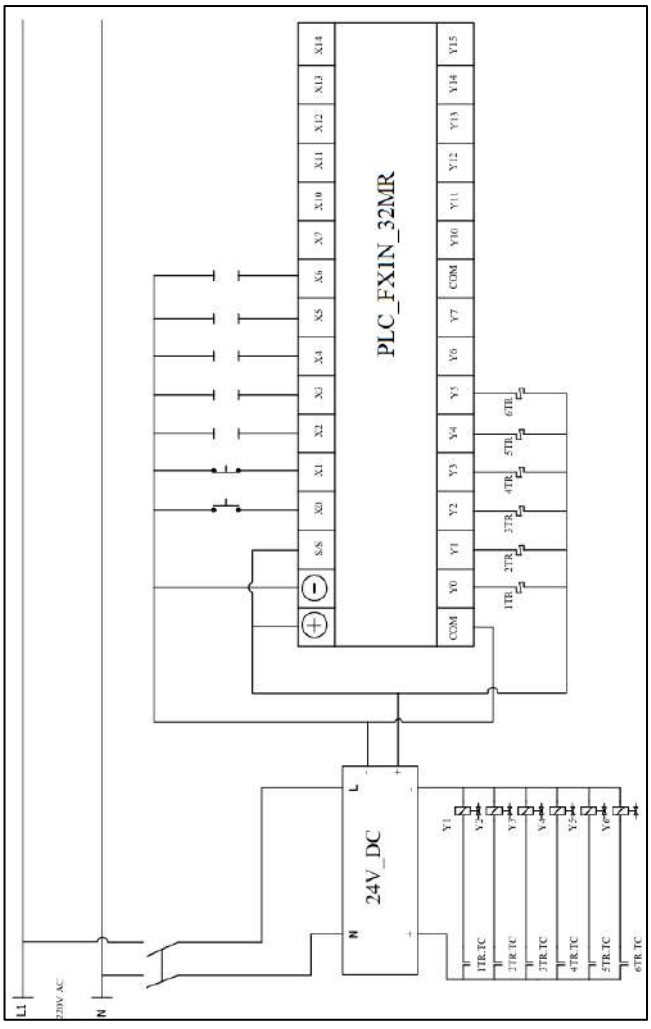


Fig. 6 Electrical System Diagram for PLC Connection

IV. RESULTS OF CONTROL SYSTEM MANUFACTURING



Fig. 7 Control system using the FX1N_32MR PLC



Fig. 8 Pneumatic robotic arm with controller in trial operation

The research team has completed the trial manufacturing of a control system using the FX1N_32MR PLC for a pneumatic robotic arm intended for training purposes, and the trial operation has yielded positive results. Since the system is designed for educational use, the control system was developed with an "open" connection for input signals (from the robotic arm's sensors) and output signals to control the arm's movements through 5/2 and 3/2 distribution valves. Learners programming the robotic arm need to clearly understand these input and output signal addresses and connect them to the controller, providing a foundation for developing control algorithms and programming the robotic arm.

Each different operating mode of the robotic arm corresponds to a different control program. Within the scope of this research, the author proposes four operating modes for the robotic arm, corresponding to four modes as shown in Table 2. Each mode is associated with a specific control program that learners will implement.

MODE 1	***MODE 2***
○ *Arm descend* ○ *Strip forward* ○ *Grip on* ○ *Strip backward* ○ *Arm retract* ○ *Swing outward* ○ *Second arm descent* ○ *Grip off* ○ *Second arm retract* ○ *Swing in*	○ *Arm descend* ○ *Strip backward* ○ *Grip on* ○ *Strip forward* ○ *Arm retract* ○ *Swing outward* ○ *Arm second descent* ○ *Grip off* ○ *Second arm retract* ○ *Swing in*
MODE 3	**MODE 4**
○ *Arm descend* ○ *Grip on* ○ *Strip backward* ○ *Arm retract* ○ *Swing outward* ○ *Second arm descent* ○ *Grip off* ○ *Second arm retract* ○ *Swing inward* ○ *Strip forward*	○ *Arm descend* ○ *Grip on* ○ *Strip forward* ○ *Arm retract* ○ *Swing outward* ○ *Arm second descend* ○ *Grip off* ○ *Arm second retract* ○ *Swing inward* ○ *Strip backward*

Table 2. Operating Mode of the Robotic Arm

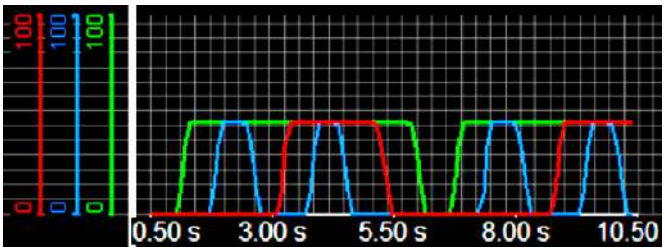


Fig. 9. Experimental results of the stable operation with MODE 1

The experimental results of the stable operation of the pneumatic robotic arm using the MODE 1 program are shown in Figure 9, with a cycle time of 5.3 seconds. The curves in the graph represent the motion of the main cylinders of the pneumatic robotic arm. The stable operating parameters of the pneumatic robotic arm corresponding to the MODES are displayed in Table 3.

MODE	Mode 1	Mode 2	Mode 3	Mode 4
Operating Pressure	6 bar	6 bar	6 bar	6 bar
Cycle Time (s)	5.3	5.6	5.2	5.9

Table 3. Experimental result of the stable operation with the MODE 1

V. CONCLUSION

In this study, a control system using the FX1N_32MR PLC was successfully designed and developed for a pneumatic robotic arm aimed at training and educational purposes. The system's open design allows for flexible connections for input and output signals, making it an effective tool for teaching. This design enables students to connect and program signals from sensors and control the robotic arm's movements through 5/2 and 3/2 distribution valves, providing a hands-on experience in understanding and implementing control algorithms.

Four distinct operating modes were proposed to showcase the versatility of the control system, each with a specific program that learners can implement. This setup allows students to gain practical knowledge in programming, troubleshooting, and adjusting the control system for different tasks, which is essential for developing skills in automation and mechatronics. The number of operating modes of the robotic arm can be easily expanded in the future depending on practical demands. This is also another distinguishing feature compared to previous controllers.

The positive results from the trial operations indicate that this control system can effectively simulate real industrial applications. It offers a robust foundation for educational use, bridging theoretical knowledge and practical application in the field of pneumatic control systems. Future developments could expand on this by integrating additional sensors or components to enhance the system's complexity and functionality, further enriching the learning experience.

REFERENCES

[1] J. Stenerson, *Fundamentals of Programmable Logic Controllers, Sensors, and Communications*. Upper Saddle River, NJ, USA: Pearson, 2004.

[2] Mitsubishi Electric, *FX1N Series Programmable Logic Controller Manual*. Tokyo, Japan: Mitsubishi Electric Corp., 2015.

[3] Festo Corporation, *Pneumatic Components Technical Manual*. Esslingen, Germany: Festo Corp., 2019.

[4] *ISO 10218*: "Robots and robotic devices - Safety requirements for industrial robots," International Organization for Standardization, 2011.

[5] "ANSI/RIA R15.06: Industrial Robots and Robot Systems - Safety Requirements," American National Standards Institute, 2012.

[6] A. Müller and M. Lu, "Application of PLC-controlled pneumatic actuators in industrial automation," *Robotics and Computer-Integrated Manufacturing*, vol. 45, pp. 132–139, Oct. 2017.

[7] J. Smith and R. Johnson, "PLC-based control systems for training in automation and mechatronics," *International Journal of Engineering Education*, vol. 33, no. 4, pp. 1089–1098, Jul. 2017.

[8] D. K. Lim and T. Zhang, "Hands-on learning of PLC and pneumatic systems: A teaching approach," *IEEE Access*, vol. 8, pp. 1040–1049, Jan. 2020.

Section 3:
Civil Engineering (CE)

Description: This field addresses the design, construction, and management of sustainable infrastructure and buildings. Emphasis is placed on efficient project management, material optimization, and environmentally friendly machine design for modern urban development.

Overview of natural ventilation solutions and the potential of applying solar chimneys to enhance energy efficiency and thermal environment in buildings in Ho Chi Minh City, Vietnam

1st Nguyen Cong Huan
Faculty of Environment
School of Technology
Van Lang University
Faculty of Civil Engineering
Saigon Technology University
Ho Chi Minh City, Vietnam
huan.2494403010001@vanlanguni.vn
(Corresponding author)

2nd Huynh Tan Loi
Faculty of Environment
School of Technology
Van Lang University
Ho Chi Minh City, Vietnam
loi.ht@vlu.edu.vn

3rd Nguyen Quoc Y
Faculty of Engineering and Technlogy
Binh Duong Economics
and Technology Univeristy
Binh Duong, Vietnam
y.nq@ktkt.edu.vn

***Abstract*—The paper provides a comprehensive overview of ventilation principles, including wind-driven ventilation, buoyancy-driven ventilation, and combined ventilation systems. Based on these principles, typical ventilation design solutions for buildings are also presented, with the thermal chimney solution emerging as an optimal approach. Various calculation and simulation methods related to ventilation, such as analytical methods, experimental methods, and simulation methods, are also discussed. The characteristics of Vietnam's climate are outlined, with a focus on factors affecting ventilation efficiency. The factors influencing the energy efficiency and thermal environment of ventilation solutions are identified, and the study emphasizes the importance and necessity of seeking ventilation solutions tailored to the specific climatic conditions and architectural features of each building to ensure energy efficiency and thermal environmental performance.**

***Index Terms*—Natural ventilation; Solar chimney; Computational fluid dynamics (CFD) analytical method; Building energy simulation; Cross ventilation**

I. Introduction

In recent years, climate change and the greenhouse effect have become increasingly concerning issues. Humanity is striving to reduce greenhouse gas emissions across various sectors, including the construction industry, which accounts for 30-40% of total energy consumption [1]. As a result, applying energy-saving solutions has become essential. Using passive design strategies has seen significant growth, with natural ventilation receiving substantial attention due to its objectives of energy conservation, improving indoor air quality, and enhancing thermal comfort.

Based on the placement of ventilation openings in buildings, natural ventilation is divided into three main types: Single-Side Ventilation (SSV), Cross Ventilation (CV), and Stack Ventilation (SV) [2]. Among these, Cross Ventilation is regarded as the most effective and has been extensively researched. In-depth studies on the mechanisms and phenomena of CV solutions, such as thermal chimneys, have begun to provide clear direction.

To address the challenges and bring natural ventilation solutions into practical application, alongside using calculation methods such as analytical mathematical equations and experimental approaches, the implementation of simulation solutions is essential. Computational Fluid Dynamics (CFD) [3] and Building Energy Simulation (BES) have become critical tools for evaluating and optimizing the performance of natural ventilation systems. These tools enable designers to predict airflow patterns, analyze thermal comfort, and assess energy performance under various climatic conditions.

II. Objectives and Methods

A. Objectives

The objectives of this review study are to:

- Overview of Location, Temperature, and Solar Radiation in Vietnam
- Overview of principles and application of natural ventilation
- Overview factors affecting natural ventilation
- Overview of Ventilation Performance Calculation and Evaluation Methods

B. Scope

The research will focus on the principles and applications of natural ventilation systems, the computational and simulation models for natural ventilation will be discussed. The factors affecting natural ventilation will also be addressed, through which the potential for applying and improving traditional natural ventilation solutions will be assessed.

C. Methodology

The primary research method used in this paper is analysis and synthesis. Studies published in journals and scientific databases, including Web of Science, ScienceDirect, Scopus, and Google Scholar, were searched and compiled. The research keywords are "ventilation", "natural ventilation", and "natural ventilation simulation", all of which are related to natural ventilation. After compilation, the data will be categorized based on research methods and synthesized to identify existing issues that need to be addressed in the future.

III. Overview of Location, Temperature, and Solar Radiation in Vietnam

A. Location

Vietnam is located in Southeast Asia, stretching from latitude 8°30'N to 23°23'N and longitude 102°08'E to 109°30'E [4]. The country has a tropical monsoon climate, characterized by high temperatures, high humidity, and seasonal temperature variations. This geographical location plays a crucial role in shaping the year-round high average temperatures. Particularly, the central and southern regions, being closer to the equator, experience higher temperatures and solar radiation levels, especially during the dry season, which lasts from December to April [5].

B. Temperature

Vietnam's average annual temperature is high, with notable regional variations. Northern Vietnam, characterized by a humid subtropical climate with four seasons, experiences temperatures between 15°C and 25°C, while summer months (May to August) reach 30°C to 37°C with high humidity. Central Vietnam, influenced by Foehn winds, often sees summer temperatures rise to 35°C to 40°C or higher, particularly from June to August. In Southern Vietnam, the dry season temperatures range from 26°C to 37°C, and the rainy season fluctuates between 25°C and 30°C, with Ho Chi Minh City notably impacted by the Urban Heat Island effect, making it hotter than surrounding areas. Given the country's long summers and frequent temperatures above 35°C, Vietnam's climate is well-suited for natural ventilation solutions, which can improve microclimates and reduce energy consumption, especially in rural and underdeveloped areas that lack modern ventilation systems

C. Solar Radiation

TABLE I
Data solar radiation of Vietnam

Region	**Sunshine (hrs/yr)**	**Solar Intensity (kWh/m^2/day)**	**Solar Power Use**
North	1,600 – 1,750	3.3 – 4.1	Average
Central	2,000 – 2,600	4.9 – 5.7	Very Good
Southern	2,200 – 2,500	4.3 – 4.9	Very Good

The table I shows solar radiation levels in Vietnam range from 3.0 to 5.9 kWh/m² per day, with the highest levels concentrated in the central and southern regions. According to statistics, Vietnam receives approximately 1,460 to 2,500 hours of sunshine annually, generating a significant amount of solar radiation for energy exploitation. This creates favorable conditions for the development of solar energy systems and natural ventilation systems [6].

The simulation-based evaluation of natural ventilation for public buildings such as the Army Broadcasting Center, the Vietnam Military History Museum, and the Headquarters of the Vietnam Telecommunications Department has shown that natural ventilation not only reduces energy consumption but also creates a fresh indoor environment. Studies assessing the impact of solar radiation on thermal chimney ventilation solutions for buildings in Vietnam's climate conditions have demonstrated that higher solar radiation levels increase the effectiveness of thermal chimneys in enhancing ventilation rates [7]. In densely built-up urban areas in Vietnam, the effectiveness of ventilation significantly improves with the application of thermal chimney solutions [8]. Research on thermal chimney models for urban buildings in Vietnam has also been initiated [9], with the results expected to lay the foundation for refining the application of thermal chimney solutions in buildings. These studies have proven that for countries like Vietnam, particularly in Ho Chi Minh City, there is substantial potential for applying natural ventilation solutions, especially thermal chimney solutions, to achieve high efficiency.

IV. Principles and Application of Natural Ventilation

The principles and methods of natural ventilation rely on utilizing natural elements such as wind and temperature differences to circulate air in and out of buildings without the use of mechanical energy. Natural ventilation systems mainly operate based on two primary forces: wind forces, where wind blowing against building surfaces creates pressure that pushes air in and out; and buoyancy forces, where warmer, lighter air rises while cooler, heavier air sinks, generating natural convection currents [2]. The combination of these two forces can optimize the ventilation process, helping to cool and regulate indoor air temperature. The natural ventilation solution includes single-sided ventilation, cross-ventilation, stack ventilation, and hybrid systems that combine natural and mechanical ventilation. These methods not only improve indoor air quality but also contribute to energy savings and environmental protection.

A. Common driving forces

1) Wind-driven forces: Wind-driven ventilation systems shown in figure 1 are a common natural method for ventilating spaces within buildings. When wind blows from the outside, the static pressure on the windward side of the building increases, creating a high-pressure zone [10]. Conversely, on the leeward side, wind vortices can form, creating a low-pressure zone. The pressure difference between these zones

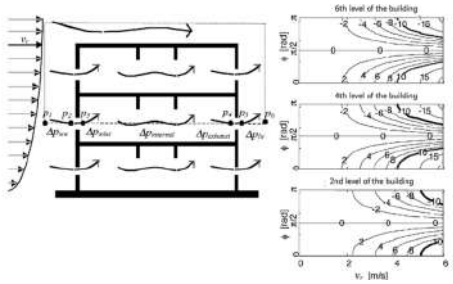


Fig. 1. Wind-driven forces

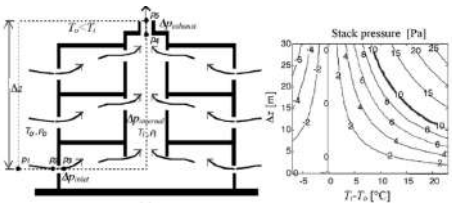


Fig. 2. Buoyancy-driven natural ventilation

causes air to move from the high-pressure area to the low-pressure area, thereby generating ventilation. The intensity of the ventilation depends on the size of the air intake and exhaust openings, as well as wind force.

2) Buoyancy-driven forces: Thermal buoyancy ventilation shown in figure 2, also referred to as ventilation driven by temperature differences, functions based on the disparity in temperature between the indoor and outdoor air [11]. This process is typically characterized by the chimney effect, which relies on the difference in height between the air intake and exhaust points, as well as variations in air density inside and outside the building. When a heat source raises the temperature of the indoor air, it becomes less dense than the cooler outdoor air, causing the warmer air to rise and creating a positive pressure. This pressure difference forces the warm air to escape through exhaust points, facilitating natural ventilation.

Additionally, humidity differences can also contribute to airflow in a manner similar to temperature-driven buoyancy. Since water vapor is lighter than dry air, humidity-driven buoyancy becomes effective in regions where the outdoor air has very low humidity levels, enhancing natural ventilation efficiency in those conditions

3) Combined forces: The two principles of ventilation driven by wind pressure and thermal pressure often operate simultaneously in a building's natural ventilation system shown in figure 3 [12]. Depending on environmental conditions, one of the two forces may become more dominant than the other. For instance, on days with high wind speeds, wind-driven ventilation may prevail, while on calmer days with significant temperature differences between indoor and outdoor air, thermal buoyancy may play a more important role.

These two forces can either complement or oppose each other, depending on wind direction and the temperature difference between indoor and outdoor air. If the wind direction aligns with the airflow generated by thermal buoyancy, these

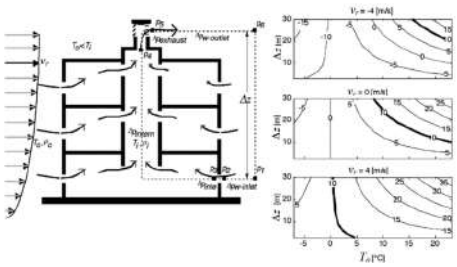


Fig. 3. Combined wind-and buoyancy-driven ventilation

forces will complement each other, enhancing ventilation efficiency. However, if the wind blows in the opposite direction of the thermal buoyancy airflow, these forces may cancel each other out, reducing the overall ventilation rate. Understanding this interaction is crucial for optimizing natural ventilation strategies for different building designs and varying climate conditions.

B. Typical natural ventilation application

The table II **shows** natural ventilation solutions are divided into three main groups: wind-driven forces, buoyancy-driven forces, and combined forces, each with its own advantages and disadvantages. Wind-driven ventilation solutions such as cross ventilation, roof and ridge ventilation, wind-driven turbine ventilators, and wind catchers harness natural wind, providing high efficiency in areas with stable wind flows but largely dependent on weather conditions. On the other hand, buoyancy-driven solutions like solar chimneys, stack ventilation, atrium ventilation, and Trombe walls rely on temperature differences to create continuous airflow even without wind, though they require a significant temperature difference to function effectively. Lastly, combined ventilation systems like mixed-mode ventilation, hybrid ventilation systems, and double-skin façades integrate both natural and mechanical ventilation, offering high flexibility but requiring higher investment and more complex operational management.

In the context of rising energy costs and environmental protection needs, the application of natural ventilation is essential to reduce energy consumption and create sustainable living environments. Notably, the solar chimney stands out as a promising solution due to its ability to generate stable natural airflow, reduce reliance on mechanical systems, and adapt well to HCMC's climate conditions, paving the way for sustainable designs in modern buildings.

V. Overview Factors Affecting Natural Ventilation

A. Building characteristics and thermal mass of building envelope

The building's geometry and the thermal mass of its materials significantly affect the effectiveness of natural ventilation and its impact on energy consumption and the thermal environment [13]. Buildings with taller structures utilize the stack effect - a phenomenon where warm air rises and escapes

TABLE II
VENTILATION SOLUTIONS BASED ON COMMON DRIVING FORCES.

Ventilation Principle	Applied Ventilation Solutions	Description
Wind-driven forces	Cross ventilation	Uses pressure differences to move air through spaces.
	Roof and ridge ventilation	Roof structures assist upward airflow and exhaust.
	Wind-driven turbine ventilators	Turbines drive air in and out using wind power.
	Wind Catcher	Uses wind pressure to bring in fresh air and expel stale air.
Buoyancy-driven forces	Solar chimney	Uses temperature differences for upward airflow.
	Stack ventilation	Hot air escapes, creating pressure-driven airflow.
	Atrium ventilation	Open spaces drive air movement via temperature differences.
Combined forces	Trombe wall	Absorbs heat, creating natural airflow through the building.
	Mixed-mode ventilation	Combines natural and mechanical ventilation.
	Hybrid ventilation system	Uses fans and natural airflow based on needs.
	Double-skin facade	Controls airflow using both buoyancy and wind forces.

through higher openings, pulling cooler air into the building from lower openings. This airflow, driven by temperature differences, is particularly effective in buildings with a high vertical gradient, enhancing natural cooling and reducing the need for mechanical ventilation or air conditioning [14].

Thermal mass, referring to the ability of materials to absorb, store, and release heat, plays a crucial role in stabilizing indoor temperatures. Materials like concrete or brick, which have high thermal mass, can absorb excess heat during the day and release it at night when temperatures drop. This helps maintain a more consistent indoor temperature, reducing the reliance on energy-intensive cooling systems. In regions with large day-to-night temperature variations, this can result in significant energy savings by minimizing the need for active heating or cooling. The use of such materials can also prevent overheating during the day, improving the thermal comfort inside the building while keeping energy consumption at a minimum.

B. Geographical Location and Climate Conditions

The geographical location of a building directly affects its exposure to wind, temperature variations, and solar radiation, all of which influence the potential for natural ventilation. In temperate climates with stable wind patterns and mild temperatures, natural ventilation can be highly effective, reducing or even eliminating the need for mechanical cooling [15]. The mild climate allows for consistent airflow, thus lowering energy consumption. In contrast, extreme climatic conditions—such as very hot, cold, or highly fluctuating temperatures—can diminish the effectiveness of natural ventilation, leading to higher energy demands.

In regions with inconsistent weather conditions, hybrid ventilation systems that combine natural and mechanical ventilation are often necessary to achieve thermal comfort and energy efficiency. These systems utilize natural ventilation when conditions are favorable and switch to mechanical systems during unfavorable weather conditions, ensuring optimal energy performance and maintaining a stable indoor environment.

C. Ambient Temperature

Ambient temperature is a key driver in buoyancy-driven ventilation systems, where airflow is generated by differences between indoor and outdoor temperatures [16]. This phenomenon, known as the stack effect, works effectively in areas with significant day-to-night temperature variations, where warm indoor air can escape through higher openings, and cooler outdoor air can enter through lower ones. In such conditions, buoyancy-driven ventilation can provide free cooling, reducing energy consumption associated with air conditioning.

However, in regions with consistently high ambient temperatures, the effectiveness of natural ventilation is diminished if the outdoor temperature is higher than the indoor temperature. In these cases, mechanical cooling systems are often required, which increases energy consumption [17]. Therefore, designing for buoyancy-driven ventilation requires careful consideration of local climate data to ensure that the system will operate efficiently without leading to increased energy demands for supplemental cooling.

D. Outdoor Wind

Wind-driven ventilation systems harness the pressure differences created by wind passing over a building's surfaces. Strategically positioning openings on the windward and leeward sides of a building allows air to flow naturally through the interior, providing passive cooling [18]. In areas with steady wind patterns, this method can significantly reduce the need for mechanical ventilation, lowering energy consumption.

However, regions with low or variable wind speeds face challenges in wind-driven ventilation, as inconsistent airflow can lead to thermal discomfort and greater reliance on mechanical systems. In such cases, building design should incorporate adjustable openings, wind catchers, or turbines to capture light breezes, optimizing airflow and reducing energy consumption.

E. Solar Radiation

Solar radiation impacts both the thermal environment inside a building and the potential for using natural ventilation [19]. Solar radiation has a significant impact on both energy consumption and the thermal environment inside buildings. Among passive design solutions aimed at minimizing the effects of solar radiation, the solar chimney is an effective solution. Based on the principle of harnessing solar energy to heat the air within a vertical shaft, the solar chimney creates a natural convection flow. When the air inside the chimney is heated, it becomes lighter and moves upward,

escaping through the top of the chimney. This process creates low pressure on the lower floors, drawing in cooler outside air into the building, helping to maintain fresh airflow and natural cooling without the need for electrical power. In areas with high solar radiation intensity, this effect can significantly reduce the need for air conditioning systems, thereby saving energy and lowering operational costs.

F. Relative Humidity

Relative humidity plays a significant role in the effectiveness of buoyancy-driven ventilation systems. In high-humidity environments, the air becomes denser, which can restrict the movement of air within the building, reducing the overall effectiveness of natural ventilation. This can lead to uncomfortable indoor conditions, as the denser, humid air holds more heat, making it more difficult to cool the building naturally.

In contrast, in dry climates, the lower humidity levels make it easier for air to circulate, enhancing the performance of buoyancy-driven ventilation. The lower density of dry air allows for more efficient airflow, improving thermal comfort without the need for mechanical systems. Consequently, buildings in drier climates benefit more from natural ventilation, as the combination of buoyancy and wind forces helps maintain comfortable indoor temperatures with minimal energy use.

VI. Overview of Ventilation Performance Calculation and Evaluation Methods

A. Analytical Methods

Analytical mathematical equations are often used for preliminary evaluations of ventilation performance. These equations typically model airflow using well-established principles like Bernoulli's equation, the continuity equation, and the momentum equation. They consider factors such as pressure differences, wind speeds, and the stack effect to estimate airflow rates [20].

This method is advantageous for its simplicity but also has significant limitations as it does not account for temperature stratification (the layering of different temperatures) or variations in air density. More complex models have been developed to address these factors, but the applicability of these analytical methods remains limited.

B. Experimental methods

Experimental methods involve in-situ measurements or controlled experiments to assess ventilation performance in actual or prototype buildings [21]. Devices like tracer gases (e.g., CO2 or SF6) are frequently used to measure air exchange rates. This technique measures ventilation by releasing a tracer gas and tracking its decay rate over time.

Experimental methods offer the advantage of high accuracy by capturing ventilation performance under real-world conditions, allowing for the precise assessment of dynamic factors such as fluctuating wind speeds and variable internal loads. These methods are particularly effective in accounting for complex interactions that may not be accurately represented in theoretical models. However, the application of experimental methods comes with significant challenges, including high costs and extended time requirements. Despite limited applicability to certain building setups and conditions, experimental methods are valuable for analyzing ventilation systems, offering data that validates and refines computational models and predictions [22].

C. Numerical Methods (CFD)

Computational Fluid Dynamics (CFD) simulation has become one of the most widely used methods for predicting ventilation performance [23]. This approach solves the Navier-Stokes equations to model airflow, temperature, and pressure distributions within a space. CFD provides visual representations of airflow and temperature gradients, helping designers identify issues such as dead zones or thermal discomfort areas. CFD can also simulate interactions among natural, hybrid, and mechanical ventilation strategies, allowing for a comprehensive performance evaluation. This method offers high accuracy, flexibility, and cost-effectiveness when testing design options without needing physical experiments. However, CFD requires substantial computational resources and specialized expertise, and its reliability depends on the quality of input data and boundary condition assumptions, which can lead to inaccuracies if not carefully managed.

D. Energy simulation

Energy simulation methods, such as those implemented in tools like DesignBuilder, EnergyPlus, or TRNSYS, model the performance of ventilation systems in the broader context of a building's energy usage. These methods account for ventilation's impact on heating, cooling, and energy use, enabling designers to optimize it alongside other building systems [24].

In energy simulations, ventilation performance is modeled as part of the building's overall energy balance, enabling a thorough evaluation of its impact on comfort and energy efficiency under various climatic conditions. Ventilation schedules and control strategies for hybrid systems are often included to optimize natural and mechanical ventilation. These computational models are valuable for simulating interactions between ventilation and other systems like HVAC, helping to enhance energy efficiency and reduce costs. However, they require detailed data and careful management of weather and usage assumptions, which can introduce uncertainties if not handled properly.

VII. Discussion—Further Research

A. Discussion

In this paper, we have analyzed various natural ventilation techniques and their applications in building designs, specifically focusing on tropical climates like that of Ho Chi Minh City, Vietnam. Through the integration of Computational Fluid Dynamics (CFD) and Building Energy Simulation (BES) models, we evaluated the performance of natural ventilation strategies, including wind-driven and buoyancy-driven systems, as well as hybrid solutions. These simulations and evaluations provided insights into the optimal placement

of ventilation openings and highlighted the critical factors influencing ventilation effectiveness, such as building height, thermal mass, and climatic conditions.

One key finding of this review is that natural ventilation can significantly reduce energy consumption while improving indoor air quality and thermal comfort when correctly designed and tailored to the local climate. However, the complexity of these designs often requires a multidisciplinary approach involving architects, engineers, and environmental experts.

Despite these promising outcomes, there remain several challenges, particularly in balancing ventilation effectiveness with energy efficiency and comfort levels in buildings where full reliance on natural ventilation may not always be feasible. In highly urbanized environments, where surrounding buildings hinder wind flow, hybrid systems offer a practical solution but require careful optimization to ensure their energy-saving potential is fully realized. Future research should also focus on the long-term performance of these systems, accounting for changing weather patterns and evolving building usage.

B. Further Research

Future research should focus on developing natural ventilation solutions that optimize airflow without being overly dependent on external conditions such as wind and temperature. Additionally, creating comprehensive design guidelines, including simulation and calculation methods for ventilation solutions, will help architects and engineers implement these strategies effectively. The solar chimney stands out as a promising solution to provide stable ventilation, reduce reliance on mechanical systems, and enhance energy efficiency, particularly suitable for climates like Ho Chi Minh City.

C. Conclusion

This study provides an overview of Vietnam's climatic characteristics as well as research on the principles of natural ventilation, emphasizing their important role in reducing energy consumption and improving indoor air quality. The study highlights that strategies such as wind-driven ventilation, buoyancy-driven ventilation, and hybrid systems, when supported by advanced simulation tools like CFD and BES, can optimize ventilation performance and energy efficiency in buildings. However, to fully realize their potential, further research, and innovation are needed to address the current challenges, particularly in tropical climates. The solar chimney solution is seen as a promising option, capable of providing continuous ventilation and reducing reliance on mechanical systems. Therefore, in the future, the development of technical specifications and design guidelines for natural ventilation solutions, particularly solar chimneys, will be essential for evaluating energy efficiency and the thermal environment.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my professors and colleagues for their valuable support and guidance throughout the process of completing this review. Their provision of essential materials and insightful knowledge has been instrumental in shaping the content of this paper.

REFERENCES

[1] L. Pérez-Lombard et al., "A review on buildings energy consumption information," *Energy and Buildings*, vol. 40, no. 3, pp. 394–398, 2008.
[2] H. Y. Zhong et al., "Single-sided natural ventilation in buildings: a critical literature review," *Building and Environment*, vol. 212, p. 108797, 2022.
[3] K. Wijesooriya et al., "A technical review of computational fluid dynamics (CFD) applications on wind design of tall buildings and structures: Past, present and future," *Journal of Building Engineering*, vol. 74, p. 106828, 2023.
[4] T. T. H. Phan and H. A. Nguyen, "Spatial and temporal distributions of temperature and rainfall on tropical islands of Vietnam," J. Water and Climate Change, vol. 14, no. 5, pp. 1395-1412, 2023.
[5] H. T. Hoi, "Potential for solar energy development in Vietnam," International Journal of Environmental Science and Development, vol. 11, no. 7, pp. 358-364, 2020.
[6] World Bank Group, "Global Solar Atlas - Vietnam," *Global Solar Atlas*, Retrieved October 14, 2024, from https://globalsolaratlas.info/download/vietnam.
[7] T. N. Duy et al., "CFD-based investigation on the performance of solar chimney for room natural ventilation," *AIP Conference Proceedings*, vol. 2420, no. 1, Nov. 2021.
[8] N. Anh et al., "The study of ventilation improvement using the solar chimney effect on urban house of Vietnam," *Proceedings of the Korea Housing Association Conference*, pp. 181-184, 2016.
[9] Y. Q. Nguyen et al., "CFD analysis of different ventilation strategies for a room with a heated wall," *Buildings*, vol. 12, no. 9, p. 1300, 2022.
[10] T. Chen et al., "Study of flow characteristics in tunnels induced by canyon wind," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 202, p. 104236, 2020.
[11] T. Kleiven, *Natural ventilation in buildings: architectural concepts, consequences and possibilities*, NTNU, 2003.
[12] M. Fan et al., "A review of different ventilation modes on thermal comfort, air quality and virus spread control," *Building and Environment*, vol. 212, p. 108831, 2022.
[13] D. Tuhus-Dubrow and Mjat Krarti, "Comparative analysis of optimization approaches to design building envelope for residential buildings," *Energy*, vol. 115, pp. 554–563, 2009.
[14] D. Tuhus-Dubrow and Krarti MJB, "Genetic-algorithm based approach to optimize building envelope design for residential buildings," *Building and Environment*, vol. 45, pp. 1574–1581, 2010.
[15] L. Moosavi et al., "Thermal performance of atria: an overview of natural ventilation effective designs," *Renewable and Sustainable Energy Reviews*, vol. 34, pp. 654–670, 2014.
[16] E. Mouriki, *Solar-assisted hybrid ventilation in an institutional building*, Concordia University, 2009.
[17] I. Oropeza-Perez and PAJAe Østergaard, "Energy saving potential of utilizing natural ventilation under warm conditions: A case study of Mexico," *Applied Energy*, vol. 130, pp. 20–32, 2014.
[18] Q. Wang et al., "External wind on the optimum designing parameters of a wall solar chimney in building," *Sustainable Energy Technologies*, vol. 42, p. 100842, 2020.
[19] M. A. Menchaca-Brandan et al., "The influence of radiation heat transfer on the prediction of air flows in rooms under natural ventilation," *Energy and Buildings*, vol. 138, pp. 530–538, 2017.
[20] H. B. Awbi, "Air movement in naturally-ventilated buildings," *Renewable Energy*, vol. 8, pp. 241-247, 1996.
[21] J. C. Phaff et al., "The ventilation of buildings: Investigation of the consequences of opening one window on the internal climate of a room," Report C 448, TNO Institute for Environmental Hygiene and Health Technology (IMG-TNO), Delft, Netherlands, Mar. 1980.
[22] Y. Li et al., "Natural ventilation for office buildings in a hot and humid climate," *Building and Environment*, 2020.
[23] R. Ramponi and Blocken, B., "CFD simulation of cross-ventilation for a generic isolated building: Impact of computational parameters," *Building and Environment*, 2018.
[24] L. Wang et al., "EnergyPlus simulation of hybrid ventilation in office buildings: A case study for climate zones 4A and 5A," *Journal of Building Performance*, 2022.

Deterministic factors governing the load-bearing capacity of concrete-filled steel tube

Tam Nguyen
Faculty of Civil Engineering
Ho Chi Minh City University of Technology
Vietnam National University, HCM city
268 Ly Thuong Kiet Street, District 10, Ho Chi Minh City, Vietnam
nttam.sdh2412@hcmut.edu.vn

Tien Vo
Technical Department of Tien Trung Construction Design and Surveying Co., Ltd
Hoc Mon District, Ho Chi Minh City
vvtien86@gmail.com

Khanh Pham
School of Civil Engineering, International University
Vietnam National University, HCM city
Linh Trung, Thu Duc, Ho Chi Minh
pnlkhanh@hcmiu.edu.vn

***Abstract*— This study examined the load-bearing capacity of concrete-filled steel tube (CFST) columns using distinguished data analysis approaches, including Pearson Coefficient, Mutual Information, and Distance Correlation, aiming to quantify the impacts of steel tube diameter (*D*), column length (*L*), steel wall thickness (*t*), concrete compressive strength (f_{ck}) on the load-bearing capacity (*Nu*). The findings demonstrated the significant influence of *D* and *t* compared to the remaining factors. Moreover, a comparative study was carried out on the performance of the state-of-the-art standards (EC4, AISC, ACI, AIJ) in estimating the load-carrying capacity of CFST utilizing the published data in the literature. The AIJ showed a superb performance, reflected by a high R^2 of 0.928 and low error metrics (RMSE=1312.669, MAE=580.163), among the current empirical approaches.**

I. Introduction

Concrete Filled Steel Tube (CFST) is a composite structure comprising a hollow steel tube and internal concrete, offering superior strength and ductility compared to conventional steel and concrete structures. Thanks to these advantages, CFST has been widely applied in high buildings, bridges [1] and industrial structures since they have high overall structural performance and seismic resistance [2]. The steel and concrete composite has demonstrated remarkable performance in axial forces and lateral forces on the structures used as piers and columns. CFST columns also cut the amount of formwork used, speeding construction and decreasing labor costs. Their intrinsic value of fire resistance and durability make these CFST columns increasingly used in fire and high-temperature risk areas, resulting in even more technologically advanced and safer structures ever built.

Many studies have been conducted to evaluate the characteristics of CFST, particularly the axial compressive strength of CFST columns and the factors influencing this capacity [3]. Despite numerous valuable observations and analytical methods, most previous studies employed numerical approaches to simulate the behavior, which often resulted in outcomes affected by assumptions regarding the interaction between the constituent materials of CFST. The rapid advancement of data-driven approaches has offered new possibilities for estimating the load-bearing capacity of CFST while considering multiple influencing factors.

This study applied various correlation analyses, distinguished by their intuitive concepts, including the Pearson coefficient, Mutual Information, and Distance Correlation, to examine the correlations between input variables and the load-bearing capacity (*Nu*) of CFST columns. The primary goal is to identify the most influential factors and quantify the impacts of relevant variables. The study also evaluates the performance of the state-of-the-art empirical approaches broadly used in practical applications in estimating the load-bearing capacity of CFST columns using regression indicators (i.e., Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and coefficient of determination (R^2)).

II. Research Methodology

This study utilized the published dataset consisting of 390 data points [4] [5] [6] [7] [8] [9] consisting of various CFST designs. The dataset is featured with seven input variables: steel tube diameter (*D*), column length (*L*), steel wall thickness (*t*), concrete thickness (*d*), concrete compressive strength (f_{ck}), steel tensile strength (f_y), and load-bearing capacity of the CFST column (*Nu*).

Three correlation analysis approaches, including Pearson Coefficient, Mutual Information, and Distance Correlation methods, were chosen for their advantages in analyzing correlations and dependencies between input variables.

The Pearson correlation coefficient [10] is a statistical method used to assess the linear relationship between two variables. This method measures the correlation between two variables by observing their simultaneous variations. The Pearson coefficient can range from -1 to 1, where 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no linear correlation between the two variables. The general formula for the Pearson coefficient is as follows:

$$r = \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum(X_i-\bar{X})^2(Y_i-\bar{Y})^2}} \quad (1)$$

where X_i and Y_i are the observed values of the two variables X and Y, $\bar{X}$ and $\bar{Y}$ are the mean values of variables X and Y.

Mutual Information (MI) [11] is an information-theoretic quantification of the amount of nonlinearity between two random variables. MI detects both linear and complex relationships, surpassing the subsidiary methods of linear correlation analyses. A higher MI value shows greater dependence on the two statistical data measures; on the other hand, MI approaches zero in the case of independent variables.

To calculate MI concerning two random continuous variables, X and Y, the following formula is used:

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \quad (2)$$

where $p(x,y)$ is the joint probability distribution of X and Y, $p(x)$ and $p(y)$ are the marginal probability distributions of X and Y.

Distance Correlation [12] (*Dcor*) is the other effective way of examining the linear and nonlinear correlation between two continuous variables. The *Dcor* ranges from 0 to 1, whereby one means that both variables are utterly dependent, whereas 0 means no relationships between the two variables. The formula for Dcor is defined as follows:

$$dCor(X,Y) = \frac{dCov(X,Y)}{\sqrt{dVar(X)dVar(Y)}} \quad (3)$$

where $dCov(X,Y)$ is the distance covariance between X and Y, $dVar(X)$ and $dVar(Y)$ are the distance variances of X and Y.

III. Research Results

Figures 1, 2, and 3 visualize the evaluation results for the correlation between *Nu* and the data features examined. Interestingly, *D* bears the strongest correlation with *Nu* regardless of the analysis approach. The Pearson of this pair of variables (Nu-Dout) is 0.80, implying that this correlation is basically linear. Accompanying the highest mutual information and distance correlation, D is obviously a crucial component of the predictive model for *Nu*. Despite the slight difference in evaluation results, the low r values of the remaining features (r<0.1) (seen in Figure 1), along with the relatively high value of MI and Dcor (shown in Figures 2 and 3), revealed the complex impacts of these features on Nu of CFST. This explained the limitations of those empirical equations, primarily in linear combinations, describing the variability of Nu.

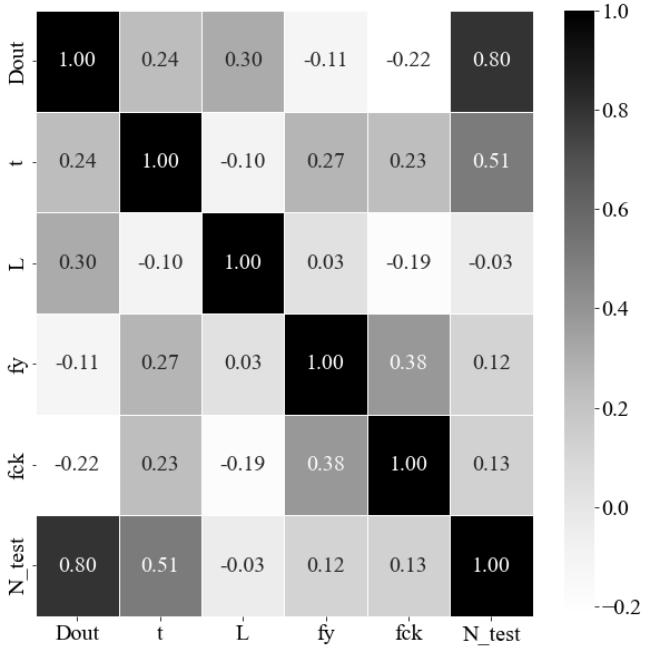


Fig. 1. Peason Correlation model's Heat map

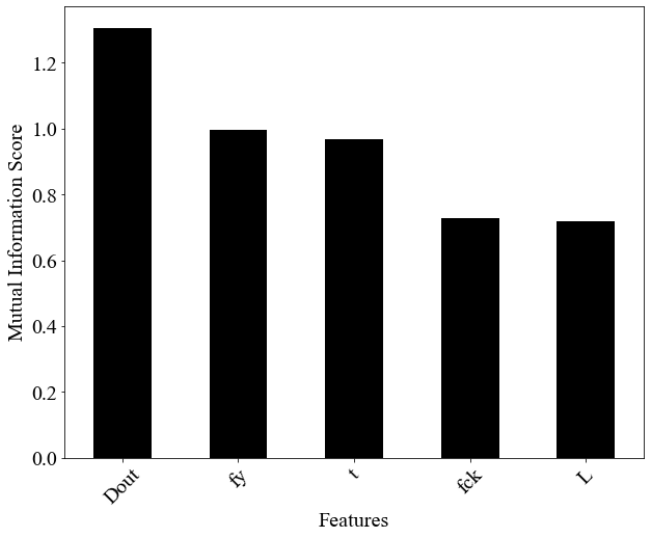


Fig. 2. Mutual Information Scores of Features

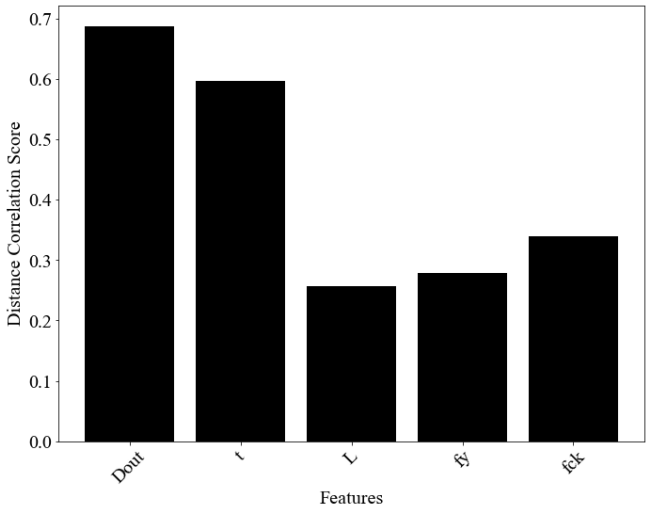


Fig. 3. Distance Correlation Scores of Features

Further elaboration was carried out by examining the performance of current empirical approaches in estimating Nu of CFST from published data in the literature. Table I summarizes those equations from design standards, including EC4, AISC, ACI, and AIJ, broadly used in practical applications. The standard regression metrics (i.e., Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Squared Error (MSE), and R squared (R^2)) were employed to measure the predictability.

The regression analysis results, shown in Figures 4, 5, 6, and 7, indicate that the AIJ model provided superb performance, with the lowest error metrics (RMSE= 1312.669 and MAE = 580.163), explaining 92.8% of the variance in the data. Following that is the AISC model, with relatively low error metrics (RMSE=1577.385 and MAE=837.686), indicating a reasonably accurate prediction. Meanwhile, the EC4 and ACI/AS models show lower accuracy, as indicated by the highest RMSE (1804.227) and MSE (3255237), along with lower R^2 (0.865).

TABLE I. DESIGN STANDARDS FORMULA

Design standard	Shape section	Formula	Eq.
EC4 (2004)	Circular, Rectangular	$N = f_y A_s + f_{ck} A_c$	(4)
AISC 360-10 (2010)s	Rectangular	$N = f_y A_s + 0.85 f_{ck} A_c$	(5)
	Circular	$N = f_y A_s + 0.95 f_{ck} A_c$	(6)
ACI (2005)	Circular, Rectangular	$N = f_y A_s + 0.85 f_{ck} A_c$	(7)
AIJ (2001)	Rectangular	$N = f_y A_s + f_{ck} A_c$	(8)
	Circular	$N = 1.27 f_y A_s + f_{ck} A_c$	(9)

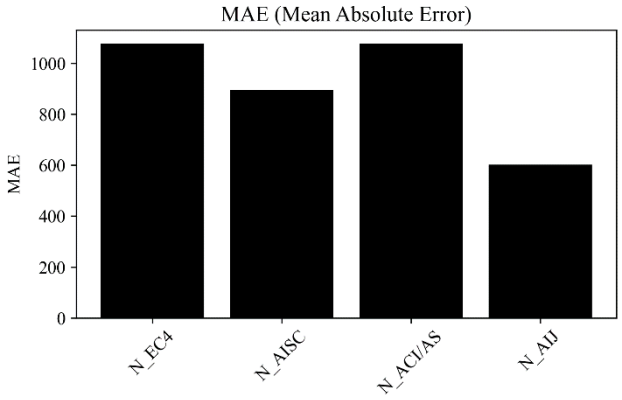


Fig. 4. Comparison of MAE for Different Standards

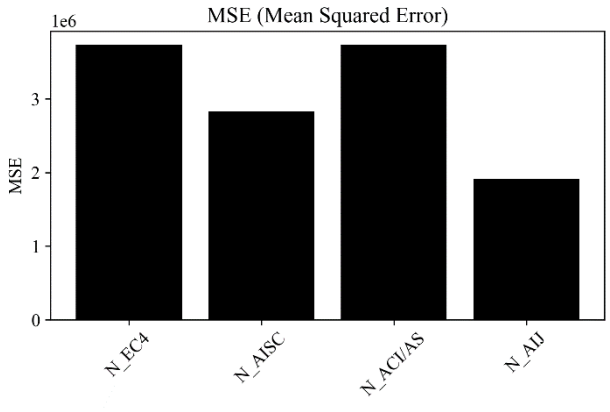


Fig. 5. Comparison of MSE for Different Standards

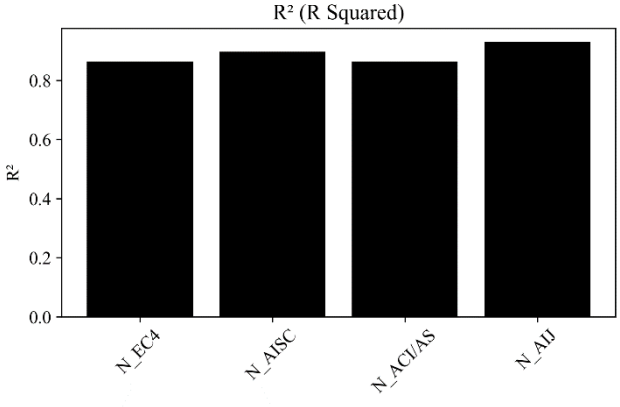


Fig. 6. Comparison of R^2 for Different Standards

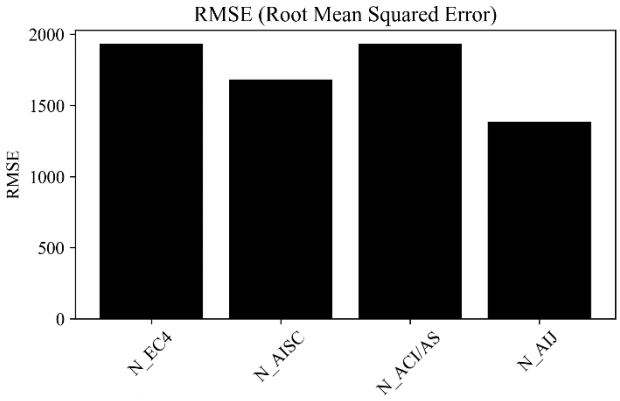


Fig. 7. Comparison of RMSE for Different Standards

IV. Conclusion

According to the research results, the AIJ method is the most effective approach for estimating the load-carrying capacity of CFST columns. Additionally, *Dout* and *t* were the most influential factors in the load capacity. This study offers a scientific justification for the choice of a predictive model and, therefore, enhances the analysis of the load-carrying capacity of CFST.

References

[1] D. E. L. a. M. S. C. W. Roeder, "Concrete-Filled steel tubes for accelerated bridge construction," *Transportation Research Record Journal of the Transportation Research Board,* vol. 2406, no. 1, p. 49–58, 2014.

[2] L.-H. H. a. W. Li, "Seismic performance of Concrete-Filled Steel tubular (CFST) structures," in *Springer eBooks*, 2013, p. 361–368.

[3] M. S. D. A. F. D. a. N. M. A. F. M. F. Shaker, "Parametric Study on the Behavior of Steel Tube Columns with Infilled Concrete—An Analytical Study," *Sustainability,* vol. 14, no. 21, p. 14024, 2022.

[4] X. Z. J. Y. a. C. S. M. Yu, "A unified formulation for hollow and solid concrete-filled steel tube columns under axial compression," *Engineering Structures,* vol. 32, no. 4, p. 1046–1053, 2010.

[5] D.-X. X. a. J. Y. R. L. M.-X. Xiong, "Axial performance of short concrete filled steel tubes with high- and ultra-high- strength materials," *Engineering Structures,* vol. 136, p. 494–510, 2017.

[6] J. Z. a. K. Chaoui, "An experimental behaviour of concrete-filled steel tubular columns," *Journal of Constructional Steel Research,* vol. 61, no. 1, p. 53–66, 2004.

[7] M. D. O. a. R. Q. Bridge, "Design of circular Thin-Walled concrete filled steel tubes," *Journal of Structural Engineering,* vol. 126, no. 11, p. 1295–1303, 2000.

[8] E. A. T. a. E. P. C. A. L. Krishan, "Strength of short centrally loaded Concrete-Filled steel tubular columns," *IFAC-PapersOnLine,* vol. 51, no. 30, p. 150–154, 2018.

[9] C. Avci-Karatas, "Prediction of ultimate load capacity of concrete-filled steel tube columns using multivariate adaptive regression splines (MARS)," *Steel and Composite Structures,* vol. 33, no. 4, p. 583–594, 2019.

[10] E. F. E.-H. a. R. H. A. Shiekh, "A comparison of the Pearson, Spearman rank and Kendall Tau correlation coefficients using quantitative variables," *Asian Journal of Probability and Statistics,* vol. 20, no. 3, p. 36–48, 2022.

[11] T. M. C. a. J. A. Thomas, "ENTROPY, RELATIVE ENTROPY, AND MUTUAL INFORMATION," in *Elements of information theory*, 2001, pp. 13-55.

[12] M. L. R. a. N. K. B. G. J. Székely, "Measuring and testing dependence by correlation of distances," *The Annals of Statistics,* vol. 35, no. 6, pp. 1-26, 2007.

A RESEARCH ON THE COMPRESSIVE RESISTANCE OF CONCRETE USING SEA SAND AND MIXED WITH ECO-CSSB ADDITIVE COMPARED TO CONCRETE USING SEA SAND WITHOUT ADDITIVES AND TRADITIONAL CONCRETE (IN BEN TRE PROVINCE)

Nguyen Quoc Duy
Vietnam National University Ho Chi Minh City - Campus in Ben Tre
Ben Tre City, Vietnam
nqduy211101012.cbt@gmail.com

NguyenDanh Thang
Faculty of Civil Engineering
University of Technology, VNU-HCM
Ho Chi Minh City, Vietnam
ndthang@hcmut.edu.vn (*)
(Coresponding author)

Dao PhuYen
Vietnam National University Ho Chi Minh City - Campus in Ben Tre
Ben Tre City, Vietnam
dpyen@vnuhcm.edu.vn

Abstract - Currently, the demand for sand in construction is very high, but river sand sources are gradually depleting and cannot meet the demand. Meanwhile, sea sand has an abundant supply, and many studies have demonstrated its potential for use in construction. In Vietnam, scientists have successfully developed the ECO-CSSB additive (an additive for mortar and concrete using sea sand and seawater), but the application of this additive in construction remains limited due to the lack of specific standards for salt levels in different regions and the absence of a detailed comparison of its strength and economic efficiency. Therefore, further research is needed to evaluate its effectiveness and practical use, particularly in the Ben Tre area. Through research documents and experimental results, the project team found that concrete made with sea sand and seawater mixed with the ECO-CSSB additive has higher compressive strength than concrete made with sea sand and seawater without the additive in the same mix design. Economically, compared to traditional river sand, ECO-CSSB mixed with sea sand and seawater also offers higher economic efficiency.

Keywords: concrete, eco-cssb additive, compressive resistance, sea sand.

I. INTRODUCTION

Vietnam is one of the countries with high growth rate, needing large sources of raw materials to serve the construction process such as sand, stone and cement. Currently, construction sand is mainly exploited from river sand mines, but with such a large mining output, river sand is not enough to meet the demand.

The special geographical location of Vietnam has a long coastline with very large sand reserves. Taking advantage of this resource helps reduce pressure on river sand resources, while improving efficiency and reducing construction costs, especially in large projects because sea sand is available in large reserves in coastal areas. However, the characteristics of saline sand, with high salt content and other impurities, have a negative impact on the strength of concrete, so it cannot be used in the construction industry yet.

However, due to some properties of sea sand, especially the characteristic of containing high salt content, there is no suitable treatment direction for mixing concrete, especially reinforced concrete. To overcome the limitations of salinity in sea sand, researchers in Vietnam have developed the additive ECO-CSSB (additive for mortar and concrete using sea sand and sea water) with the aim of improving the quality of concrete when using sand and saline water for mixing. However, this is a relatively new product on the market, and management agencies have not yet issued standards and usage norms. In addition, the effectiveness of this additive when applied in actual conditions in Ben Tre has not yet been studied.

II. RESEARCH METHODS

A. Research documents

- Search, read and analyze many documents related to sea water and sea sand.
- Refer to documents and regulations of the state, specifically:
+ TCVN 13754:2023 Saline sand for concrete and mortar.
+ TCVN 7570:2006 Aggregates for concrete and mortar - Technical requirements
+ TCVN 6260:2020 Blended Portland cement
+ TCVN 4506:2012 Water for concrete and mortar - Technical requirements
+ TCVN 3105:2022 Concrete mixtures and concrete - Sampling, preparation and curing of test specimens
+ TCVN 3118: 2022 Concrete - Method for determining compressive strength

B. Experimental research

Conduct casting of concrete samples with durability level B15 (M200) including 5 groups of samples (each group of samples includes 3 concrete samples):
- Group 1: concrete sample mixed with sea sand concrete and sea water without compression additives at 10 days old.
- Group 2: concrete sample mixed with sea sand concrete and sea water with compression additive at 10 days old.
- Group 3: concrete sample mixed with sea sand concrete and sea water without compression additives at 28 days old.
- Group 4: concrete sample mixed with sea sand concrete and sea water with compression additive at 28 days old.
- Group 5: traditional concrete sample compressed at 28 days old.

After casting the sample, wait for the sample to harden for 24 to 72 hours, then remove the mold, take the sample for maintenance and compress the concrete sample when the concrete is 10 days old and 28 days old.

1) Implementation process

According to TCVN 3105:2022 Concrete mixtures and concrete - sampling, fabrication and maintenance of test samples, to determine the properties of concrete and study

its compressive strength the research team chose to cast block samples with pellet size (a=150 mm).

Prepare materials to mix M200 (B15) concrete according to the manufacturer's mixing ratio, including the ingredients as shown in Table 1.

Table 1. Concrete mixing ratio M200 (B15)

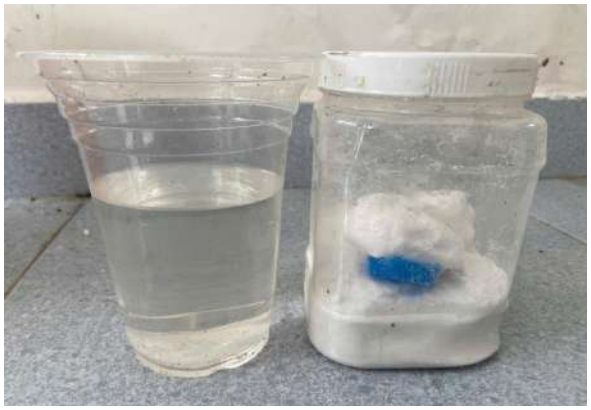
Material	$1m^3$	Model 0.15x0.15x0.15 (m)	Calculation unit
Cement	293	0.989	kg
Sand	466	1.573	liter
Stone	847	2.859	liter
Water	185	0.624	liter



a. Small aggregates (river sand & sea sand)



b. Large aggregate (1x2 cm crushed stone)



c. ECO-CSSB additive



d. Clean water and sea water



e. Cement PCB40

Figure 1. Five groups of samples (a, b, c, d, e)

2) Model casting process:

- Prepare all ingredients according to the ratio as shown in Table 2, proceed to cast concrete samples, concrete samples include: 1 set of traditional concrete samples (1 set of 3 concrete samples), 2 sets of concrete samples cardboard. One set of concrete samples uses sea sand and sea water, two sets of concrete samples use sea sand and sea water mixed with ECO-CSSB additive. There are 15 concrete samples in total.
- Pour concrete samples.
- Maintenance of concrete samples.
- Perform compression tests when the concrete is old enough.



a. Casting concrete samples



b. Concrete maintenance



c.Perform compression test when concrete reaches age

Figure 2. Testing Processing

- Samples for testing compressive strength of concrete are kept in the mold after 28 hours (TCVN-3105), then take the mold out and soak the sample in water for curing. When 10 days old, compress 1 concrete sample using sea sand and sea water, 1 concrete sample using sea sand, sea water mixed with ECO-CSSB additive. And when the remaining 3 sample groups are 28 days old, compress the samples to test their strength and compare.

C. Analyze the results

After compressing the sample, record the results to calculate the compressive strength. Analyze the compressive strength to compare, evaluate the results and economic efficiency then draw conclusions.

The compressive strength of the specimen (R), in megapascals (MPa), accurate to 0.1 MPa, is calculated by the formula:

In there:

$$R = \alpha \times \frac{p}{A}$$

P is the load at which the specimen breaks down, in newtons (N);

A is the bearing area of the specimen, in square millimeters (mm^2);

α is the conversion factor of the test result determined on a sample other than the standard sample to the compressive strength of the standard sample (cubic sample of size 150mm x 150mm x 150mm). The factor is determined experimentally according to the procedure in Appendix B (TCVN 3118:2022).

III. Results, Discussion

A. Research results

After performing the compression test, the results are as following Table 2.

From the Table 2 of results obtained it can be seen that:

- Traditional concrete achieves the highest strength in compression samples.

- Concrete using sea sand and sea water has higher compressive strength than concrete using sea sand and sea water mixed with ECO-CSSB additive at the beginning, but at 28 days old, concrete using sea sand and sea water mixed with ECO-CSSB additive has higher strength than concrete using sea sand and sea water .

B. Compare economic efficiency

At the time of project implementation (September 2024), the price of construction river sand was about 230,000 VND/m3, the price of ECO-CSSB additive was about 50,000 VND/liter and sea sand had large reserves, so the price would be low. In addition, if using ECO-CSSB additive mixed with sea sand concrete and sea water, the standard of concrete mixing water would also be lower because the additive has the ability to limit the negative impact of silt and impurities mixed in concrete. Therefore, it can be seen that the option of using sea sand and sea water mixed with ECO-CSSB additive is economically effective.

IV. Conclusion and Discussion

A. Conclusion

Concrete using sea sand and sea water mixed with ECO-CSSB additive has higher compressive strength than concrete using sea sand and sea water at the same mix ratio without using additives. This shows that ECO-CSSB additive is effective in increasing concrete strength and can be applied to marine projects in the Ben Tre area.

In terms of economic efficiency, compared to the cost of exploiting, transporting and using traditional river sand, the use of ECO-CSSB additive mixed with sea sand and seawater has lower cost and higher economic efficiency.

B. Discussion

The application of ECO-CSSB additive in the production of concrete using sea sand and sea water brings significant socio-economic efficiency. By significantly improving the strength of concrete using sea sand and sea water, this additive helps increase the life of the project, reduce maintenance and repair costs , opening up the prospect of wide application in the construction industry, especially in coastal areas. With the ability to increase compressive strength, this type of concrete can be flexibly applied to many types of projects, from civil works to large infrastructure projects. Thanks to that, this solution not only reduces pressure on traditional river sand sources but also contributes to promoting sustainable development of coastal areas, opening up a new direction for the construction industry.

Table 2: Compressive strength of concrete

Concrete type	Compression day	Age (days)	Compressive force to destroy the specimen (kN)			Average (kN)	Compressive strength (MPa)
			1	2	3		
M1	01/08/2024	10	340.5	358.2	376.5	358.4	15. 9
	August 19, 2024	28	403.8	395.7	428.2	409.2	18.2
M2	01/08/2024	10	319.1	300.1	259.3	292.8	13.0
	August 19, 2024	28	455.3	421.5	439.3	438.7	19.5
M3	August 19, 2024	28	529.1	516.8	540.9	528.9	23.5

In there:
M1 – Sea sand and sea water concrete
M2 – Concrete made of sea sand and sea water mixed with ECO-CSSB additive
M3 – Traditional Concrete

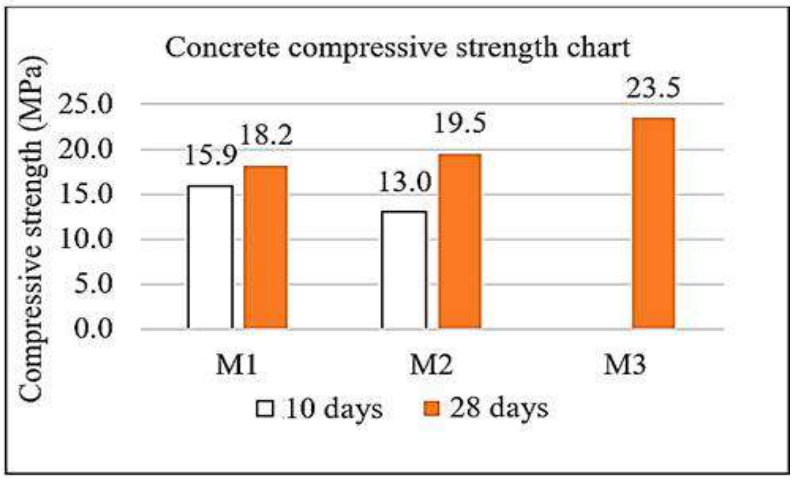


Figure 2: Compressive strength chart of concrete

Due to limited time and budget, the study only stopped at comparing the compressive strength of concrete using saline sand mixed with ECO-CSSB additive compared to concrete using saline sand without additives and traditional concrete. Although the strength of concrete using sea sand and sea water is lower than that of traditional concrete, it can be seen that the additive has a positive impact in limiting the influence of water and sea sand on concrete strength.

It is recommended to conduct further studies to determine the mixing ratio of ECO-CSSB additives, find the appropriate mixing ratio for the characteristics of sea sand and sea water in each locality to achieve the highest efficiency. Conduct more experiments on compressive strength and impact resistance at longer time points to consider the working ability of ECO-CSSB additives in concrete over time.

ACKNOWLEDGMENT

I would like to express my gratitude to the Vietnam National University Ho Chi Minh City - Campus in Ben Tre and my teachers for supporting the funding and research equipment necessary for me to complete this article.

REFERENCES

[1] Keisaburo KATANO, Nobufumi TAKEDA , Yoshikazu ISHIZEKI , and Keishiro IRIYA, "Properties and Application of Concrete Made with Sea Water and Un-washed Sea Sand", *Production Engineering Department, Obayashi Corporation, Japan* , 2013.

[2] Zhen Hai Zhang, Zhen Qun Sang, Le Yu Zhang, Zhen Xiang Ma, Yu Zhang, "Experimental Research on Durability of Concrete Made by Seawater and Sea-Sand", *Advanced Materials Research Vols. 641-642* , 2013.

[3] Jianzhuang Xiao , Chengbing Qiang, Antonio Nanni, Kaijian Zhang, “Use of sea-sand and seawater in concrete construction: Current status and future opportunities”, *Construction and Building Materials* , 2017.

[4] Dhondy, T., Remennikov, A. & Sheikh, M, “Benefits of using sea sand and seawater in concrete: a comprehensive review”, *Australian Journal of Structural Engineering* , 2019.

[5] Jianzhuang Xiao, Qingtian Zhang, Peng Zhang, Luming Shen, "Mechanical behavior of concrete using seawater and sea-sand with recycled coarse aggregates", *Structural Concrete Journal of the fib* 2019.

[6] Menghuan Guo, Biao Hu, Feng Xing, Xiaoqing Zhou, Meng Sun, Lili Sui, Yingwu Zhou, "Characterization of the mechanical properties of eco-friendly concrete made with untreated sea sand and seawater based on statistical analysis", *Construction and Building Materials* , 2020.

[7] Kartheek Thunga, T. Venkat Das, "An experimental investigation on concrete with replacement of sea treated sand as fine aggregate", *Materials today: PROCEEDINGS* , 2020.

[8] Bo-Tao Huang, Jing Yu, Jia-Qi Wu, Jian-Guo Dai, Christopher KY. Leung, "Seawater sea-sand Engineered Cementitious Composites (SS-ECC) for marine and coastal applications", *Cementitious and Concrete Composites* , 2020.

[9] D. Pan, S. Yaseen, Ke Chen, D. Niu, C. Leung, Zongjin Li, "Study of the influence of seawater and sea sand on the mechanical and microstructural properties of concrete", *Journal of Building Engineering* , 2021.

[10] Truong Hoai Chinh, Tran Van Chau, “Research on concrete production from sea sand and sea water in Nha Trang – Khanh Hoa area,” *Journal of Science and Technology - Danang University* , 2013.

[11] Nguyen Van Tam, “Research on using saline sand from Co Chien River, Tra Vinh province to produce concrete,” *University of Technology - Danang University* , 2019.

[12] Dang Thuy Chi, “Experimental evaluation of Geopolymer concrete strength using sea sand and seawater mixture,” *Civil Engineering Journal* , 2022.

SAIGON TECHNOLOGY UNIVERSITY
AIOT LAB VN
Kha Tu Huynh, Cao Hao Thi, Tan Duy Le, Thinh Vo Xuan, Khanh Pham, Vi Chi Thanh

PROCEEDINGS THE 2024 CONFERENCE ON APPLICATIONS OF TECHNOLOGY, AUTOMATION AND CIVIL ENGINEERING

Responsible for publication
Director – Editorial in Chief
BUI MINH CUONG

Responsible for the content
TRUONG THANH SON

Editor : TRUONG THANH SON
PHAM THI MAI
Layout : PHAM THI MAI
Cover design : NGO HOANG QUAN
Copy Editor : NGUYEN TRONG HOANG

SCIENCE AND TECHNICS PUBLISHING HOUSE
70 Tran Hung Dao St., Cua Nam Ward, Ha Noi, Viet Nam
Tel: (84)(24) 3942 2443 Fax: (84)(24) 3822 0658
Website: http://www.nxbkhkt.com.vn Email: nxbkhkt@hn.vnn.vn

BRANCH IN HO CHI MINH CITY
28 Dong Khoi St., Sai Gon Ward, HCMC, Viet Nam
Tel: (84)(28) 38225062 Fax: (84)(28) 38296628
Email: chinhanhnxbkhkt@yahoo.com.vn

Quantity: 10 copies, size: 21^{cm} × 29^{cm} at Nhat Nam Printing and Trading Joint Stock Company.
Address: 007 Lo I, Tan Binh Industrial, Tay Thanh Ward, HCMC, Viet Nam.
Publishing license No: 1150-2025/CXBIPH/01-59/KHKT.
Publishing decision No: 31/QĐXB-NXBKHKT, 10/04/2025.
ISBN: 978-604-67-3280-8.
Printing completed and copyright deposited in the 4th quarter of 2025.

THE 2024 CONFERENCE ON APPLICATIONS OF TECHNOLOGY AUTOMATON AND CIIVIL ENGINEERING

225076H00
ISBN: 978-604-67-3280-8



Not For Sale