International Journal of Advanced Computing and Information Technology (IJACIT)

Volume 01 | Issue 01
March 2025

Research Journal Published by
School of Computing and Engineering (SOCE),
National Institute of Business Management (NIBM),
Sri Lanka.

NIBM | THE CITY UNIVERSITY

INTERNATIONAL JOURNAL
OF
ADVANCED COMPUTING
AND
INFORMATION TECHNOLOGY
(IJACIT)

Volume 01 | Issue 01

March 2025

Research Journal Published by
School of Computing and Engineering (SOCE),
National Institute of Business Management (NIBM),
Sri Lanka.

INTERNATIONAL JOURNAL OF ADVANCED COMPUTING AND INFORMATION TECHNOLOGY (IJACIT) is the flagship publication of the School of Computing and Engineering (SOCE), National Institute of Business Management (NIBM), Sri Lanka. It is an open-access, double-blind, peer reviewed journal published online and in print. Publication in the journal is open to academics, postgraduate students, undergraduate students, and professionals both locally and internationally.

Publication
Quarterly (in March, June, September, and December)

Publisher
School of Computing and Engineering (SOCE), National Institute of Business Management (NIBM), Sri Lanka.

ISSN: ISSN 3093-625X
9 773093 625009

Disclaimer
The views expressed in this journal do not necessarily reflect the views of the School of Computing and Engineering (SOCE) or the National Institute of Business Management and the Editors cannot be held accountable for errors or any consequences arising from the use of information contained in the journal.

Reviewers of IJACIT March, 2025

Dr. Mark Elshaw	Coventry University United Kingdom
Mr. Manik Hapsara	Coventry University United Kingdom
Prof. Prasad Wimalaratne	School of Computing University of Colombo
Dr. Rushan Abeygunawardana	Faculty of Science University of Colombo
Dr. Thilina Thanthriwaththa	Information Technology University of Moratuwa
Dr. Saminda Premerathne	Faculty of Information Technology University of Moratuwa
Dr. Dushyanthi Vidanagama	Faculty of Computing General Sir John Kotelawala Defense University
Dr. Asela Gunasekara	Faculty of Computing General Sir John Kotelawala Defense University
Dr. Menaka Ranasinghe	Faculty of Engineering The Open University of Sri Lanka
Dr. WMKS Ilmini	Faculty of Applied Sciences University of Sri Jayewardenepura
Dr. Ruwan Wickramarachchi	Faculty of Applied Sciences University of Kelaniya
Dr. Chaman Wijesiriwardana	Faculty of Information Technology University of Moratuwa
Mrs. Amila De Silva	School of Computing and Engineering National Institute of Business Management
Ms. DV Dharani Abeysinghe	Faculty of Computing General Sir John Kotelawala Defence University
Ms. Wishama Samaraweera	Faculty of Computing General Sir John Kotelawala Defence University
Mr. Ashen Wanniarachchi	Faculty of Computing General Sir John Kotelawala Defence University

List of Abstracts IJACIT, March 2025
Volume 01 | Issue 01

Web application Prototype for Assisting Technical Challenges in Higher Education Online Learning: A Study on School of Computing and Engineering, National Institute of Business Management, Sri Lanka

Dissanayake H M T N
School of Computing and Engineering
National Institute of Business Management
Colombo, Sri Lanka
thedanind@gmail.com

Abeygunawardana R A B
Department of Statistics, Faculty of Science
University of Colombo
Colombo, Sri Lanka
rab_abey@stat.cmb.ac.lk

Abstract—**With the impact COVID-19 had on education, online learning has become a permanent fixture in higher education. However, as a developing country, Sri Lanka still lacks the digital infrastructure and resources to support online learning systems. Particularly, in the context of higher education, this poses significant challenges for effective implementation of online learning systems. This study investigates online learning related technical challenges faced by students at the National Institute of Business Management (NIBM), School of Computing and Engineering (SOCE), through a survey distributed among 360 students. The results revealed that, 74% of students agreed to the fact, technical challenges have an impact on their academic performances. Additionally, 93% of students stated that they felt frustrated and stressed due to technical challenges. Furthermore, approximately 95% of the students indicated technical challenges affect the overall online learning satisfaction level. The findings of this study also highlight the connectivity issues, software and device issues, online learning platform-related technical challenges, and lack of support hinder student-learning experience. Hence, to address these issues, this study proposes a web application prototype designed to monitor and resolve technical issues in real-time. The proposed prototype aims to reduce the impact of technological challenges and assist students and NIBM with online learning activities.**

Keywords*— *online learning, blended learning, higher education, online learning challenges, technical challenges

I. Introduction

With the transformation of technology, the traditional methods of delivering education have embraced digital platforms to deliver education. This shift has driven by the need to overcome the limitations of conventional teaching and learning [1]. The rapid shift to online learning methods occurred with the COVID-19 pandemic, where even higher education institutes were forced to adapt to online learning [2]. Even at present, many institutes readily integrate online learning into their academic processes as it offers numerous benefits, such as flexibility, accessibility, and convenience, which aid students with balancing their personal and academic commitments [3]. While online learning offers numerous advantages, transitioning to online learning poses significant challenges. Especially as a developing country Sri Lanka lacks the digital infrastructure and resources to fully harvest the benefits of online learning [4].

Due to this resource-limited environment in Sri Lanka, online learning is prone to technical challenges such as connectivity issues, compatibility issues , accessibility issues, limited technical support, and inadequate troubleshooting resources [5][6] Addressing these technical challenges are critical to ensure equitable access to quality education, especially in a post-pandemic context where online learning has become a pillar to the higher education.

Moreover, with South Asia's highest literacy rate of 92%, Sri Lanka has a high demand for higher education [7]. However, even at the start of 2024, Sri Lankan internet penetration is at 56.3% of the total population [8]. According to the Computer Literacy Statistics 2023 of Sri Lanka [9] it highlights the following factors: When considering the Sri Lankan citizens who are aged 5-69, while the digital literacy rate is 63.8%, only 39.5% are computer literate. Additionally, only 20.5% of the households own a laptop or desktop computer. Furthermore, it also highlights that most digital and computer-literate people are from urban areas. Hence, there is a need for practical, technology-driven solutions that could address these challenges and support students [10].

This study focuses on students at the National Institute of Business Management (NIBM), School of Computing and Engineering (SOCE), as a representative case of higher education institutions in Sri Lanka facing resource constraints while offering online education.

NIBM, as a leading higher education institute, provides students various academic programs to achieve academic qualifications. Students are enrolled in academic programs ranging from certificate programs to degree programs, with academic levels starting from certificate level to degree level. These programs can be Full-time or Part-time programs where students can pursue academic activities according to their schedules.

As for SOCE students, most of the modules consist of information technology (IT) and computing-related subjects [11]. Hence compared to other disciplines, SOCE students are reliant on online learning and digital infrastructure due to the nature of their academic programs. This dependence makes them more vulnerable to technical challenges.

NIBM currently uses a blended learning approach [12], where students often use the NIBM worldwide, Learning Management System (LMS) of NIBM to manage their academic activities. Students of SOCE, as all other NIBM students, use the NIBM Worldwide for their online learning activities. The platform facilitates students to access and download module lecture materials, submit coursework and exams, access the exam portal, and even attend online lectures. However, due to insufficient infrastructure and resources, many students experience numerous challenges while continuing online learning activities.

Hence, this study aims to investigate the technical challenges faced by NIBM SOCE students in online learning and propose a web application prototype designed to monitor and resolve these issues in real-time. By assisting to overcome these challenges, the prototype aims to enhance the overall online learning experience and satisfaction of students.

II. LITERATURE REVIEW

Online learning is the use of digital devices and services to support learning purposes [13]. In higher education systems, there are two main types of methods are commonly implemented to conduct online learning. The asynchronous online learning is conducting learning activities on demand, and synchronous learning is conducting online learning activities in real time and mostly led by an instructor at a scheduled time [14]. This dual-mode approach offers several advantages, offering the students with the benefits of aaccessibility, flexibility, and improved quality of learning [15][16]. This also raise the demand for online leaning based programs.

Hence, many higher education institutes have adapted online learning methods to provide better learning experiences to students [17]. This process was mostly accelerated with the impact of COVID-19 [2]. Due to COVID-19 pandemic, higher education institutes adopted online learning methods to overcome the barriers of social distancing [18]. In the current context, online learning has become rooted in higher education. However, challenges, such as technological limitations, lack of digital literacy, and lack of interaction hinder the effectiveness of online learning process [19]. Hence, it is crucial to implement initiatives to improve the effectiveness of online learning while understanding requirements for such initiatives [19].

Most students believe using online education helps them to enhance their academic performance [20]. The use of digital resources, and support, and how student experience with using the internet impacts the effectiveness of e-learning systems [20]. A study highlights that enjoyment and satisfaction received from using online platforms also contribute to the intention of using online learning platforms. Another research highlights that student has different literacy levels when utilizing e-learning resources. Inadequate training and support also affect the effectiveness of academic activities [21].

As a developing country, Sri Lanka lacks the technological infrastructure, user readiness, economic background, and policies to implement advanced technological solutions to support the online education system [22]. The lack of digital infrastructure, accessibility, and connectivity use poses a major challenge to current online learning systems [5].

To tackle these challenges and requirements there is a need for support systems that could assist with online teaching and learning activities [5] [19]. However, several studies have pointed out that designing such online learning systems should be well-planned and well-designed [23] [24]. Which highlights importance of understanding user needs, implemented environment and available resources.

III. METHODOLOGY

The main goal of data collection for this study was to identify the technical challenges SOCE students experience with online learning. This study uses a quantitative approach to conduct the statistical analysis. A closed-ended questionnaire was distributed among selected sample to identify how these challenges affect the students, particularly, while using the online learning platform, NIBM Worldwide.

A. Population

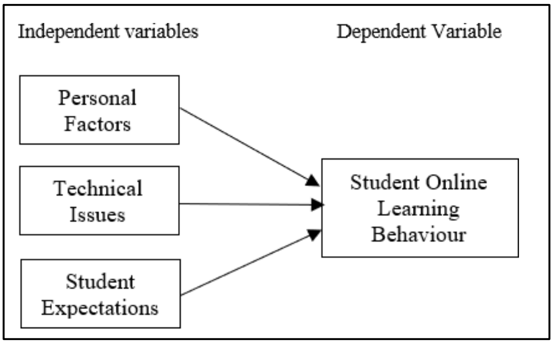
Around 2000 SOCE students are currently enrolled in different academic programs.

B. Sample Size

By using the Krejcie and Morgan Table, a sample size of 322 was determined and 360 responses were selected from the gathered responses.

C. Sampling Technique

As the SOCE students can be Full-Time or Part-Time and are from different academic years and academic programs, to ensure the representation and improve the accuracy of the data, stratified random sampling was used.



D. Conceptual framework

The conceptual framework given in Fig.1 represent how the independent variables - personal factors, technical issues, and student expectation for

Fig. 1. Conceptual framework

support are affect the dependent variable, student online learning behavior.

Based on the conceptual framework, the questionnaire was designed to gather data on the following areas.

- Personal background of students (Example – Age, Gender, Academic year, Academic program, employment status.)
- Technological issues – Software and connectivity challenges were questioned in this device. Furthermore, specific technical challenges students experience while using the NIBM Worldwide were also questioned.
- Student expectation for support – What students expect as support was questioned here.
- Lastly, student opinions on how technical challenges impact their academic performance, mental health, and satisfaction were questioned.

IV. DATA ANALYSIS

A. Descriptive Statistics Analysis

The Descriptive Analysis visualize the impact of technical challenges on student online learning behavior. The analysis focus on main three key areas. First, the relationship between technical challenges and academic performance. Second, how technical challenges impact feeling stress and frustration. Third, how technical challenges affect the overall online learning satisfaction of the student.

1) Technical Challenges vs. Student's Academic Performance.

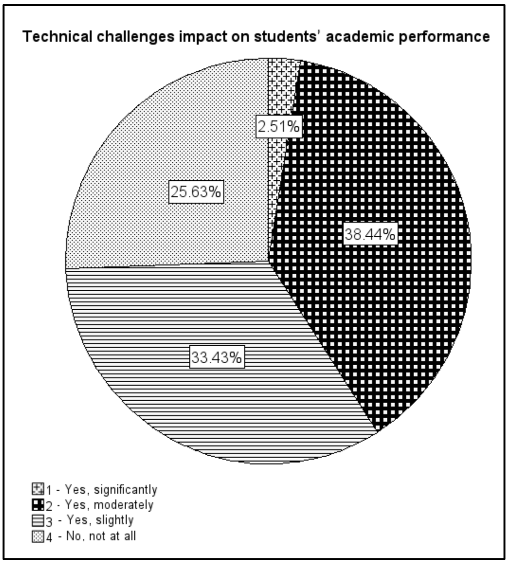


Fig. 2. Technical challenges impact on student's academic performance

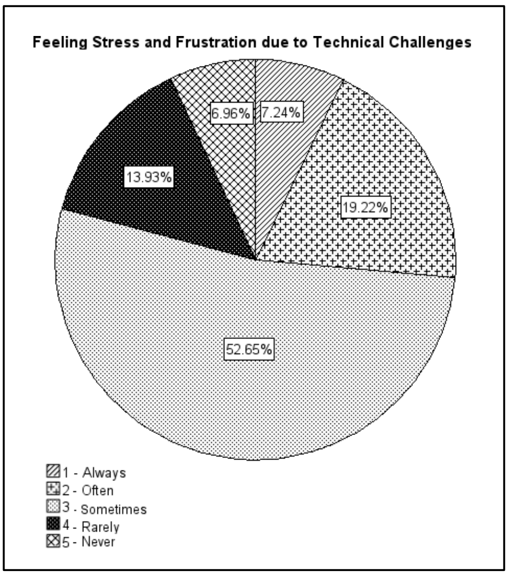


Fig. 3. Feeling stress and frustration due to technical issues

From the sample, 2.51% of students believe that technical challenges have a significant impact on their academic performance. While 38.44% of students state that there is a moderate impact, 33.43% of students have stated that there is a slight impact on academic performance due to technical issues. 25.63% of the students have stated that the technical challenges do not have an impact on their academic performance (Fig. 2).

As a conclusion, around 74% of the students have found technical challenges to have a certain degree of impact on their higher education academic performances.

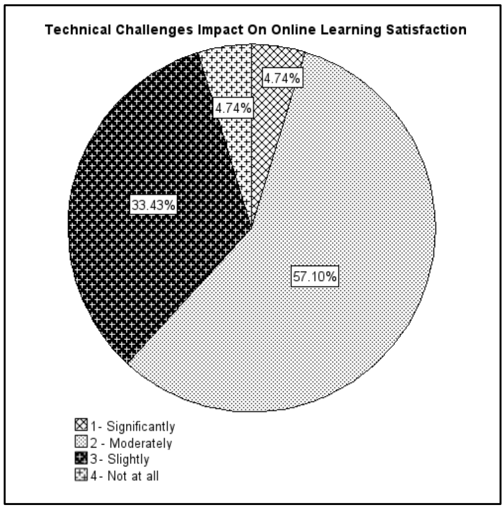


Fig. 4. Technical challenges impact on online learning satisfaction

2) Technical Challenges vs. Feeling Stress and Frustration

As shown on Fig.3 students' responses to whether experiencing technical issues while online learning has led them to feel frustrated or stressed. 7.24% of the students have stated that they always get frustrated and stressed when experiencing technical challenges. While 19.22% of the students often get affected, 52.65% of the students have selected the option of sometimes. Around 14% of the students stated that they rarely get frustrated or stressed. 7% of students have stated that experiencing technical challenges while online learning, never made them frustrated or stressed.

However, 93% of students have agreed to the fact that experiencing technical challenges while online learning made them feel frustrated and stressed.

3) Technical Challenges vs. Online Learning Satisfaction

.According to the statistics around 5% of the students have stated that technical challenges have significantly affected their satisfaction. 57.10% of students have stated that technical challenges have a moderate impact on online learning. 33.43% of the students stated that their satisfaction is slightly affected by technical challenges (Fig. 4).

As a conclusion, more than 60% of the students' online learning satisfaction is considerably affected by technical challenges. Around 95% of the students have agreed that technical challenges affect the online learning satisfaction level (Fig. 4).

B. Findings and Interpretation

From statistical analysis, the following conclusion was taken to design the prototype.

1) Personal Factors

When considering student personal background as TABLE I. shows, factors such as age group of the student, the academic program level, and the ability to use digital tools have a significant relationship with the student's learning behavior. Student gender, students being a full-time or part-time or student employment status does not significantly affect online learning behavior.

2) Technical Challenges

First, under technical challenges, the impact of the internet connectivity, the devices that are used for online learning, and the software issues were discussed. According to the analysis result with a p-value of <0.001, these challenges proved to have an impact on online learning behavior.

Then the main key areas students use in NIBM Worldwide were analyzed to determine whether the technical issues related to these areas affect online learning.

As shown in TABLE II. technical challenges associated with online learning platform such as Modules and lecture materials, Online lectures, Discussion forums, Coursework and Assignments, Exams/ online quizzes, and Grades/Results should be addressed as the p-value < α = 0.05.

3) Expectations for Support

Next, the student's expectations for support were analyzed to understand the support expectations.

TABLE III. showcase students expect support to tackle technical challenges on modules and course materials, online lectures-related issues, LMS-related issues (e.g. Login, navigation), managing grades and results, coursework and assignments, and Software/ device/ internet issues as the analysis resulted a p-value of 0.001in each area.

The findings from the statistical analysis provided valuable insights into the personal, technical, and support-related factors influencing students' online learning behavior. Thus, the above requirements were addressed when designing the prototype.

TABLE I. PERSONAL FACTORS ANALYSIS

Hypnosis	**p-value**	**Significance (Yes/No)**	**Conclusion**
H_o–There is no significant relationship between online learning behavior and age groups.	0.001	Yes	Reject H_0
H_o-There is no difference in the online learning behavior between genders.	0.857	No	Do not reject H_0
H_o- There is no difference in online learning behavior across the academic program levels.	0.001	Yes	Reject H_0
H_0-There is no difference in the online learning behavior between students being full-time or part-time.	0.620	No	Do not reject H_0

Hypnosis	p-value	Significance (Yes/No)	Conclusion
H_0-There is no difference in the online learning behavior between students being employed or unemployed.	0.871	No	Do not reject H_0
H_0-There is no significant relationship between online learning behavior and the ability to use digital tools.	0.001	Yes	Reject H_0

TABLE II. TECHNICAL CHALLENGES IN KEY AREAS

Key area	p-value	Significance (Yes/No)	Conclusion
Modules and lecture materials	0.001	Yes	Reject H_0
Online lectures	0.017	Yes	Reject H_0
Discussion forums	0.001	Yes	Reject H_0
Coursework and Assignments	0.02	Yes	Reject H_0
Exams/ online quiz	0.029	Yes	Reject H_0
Grades/Results	0.040	Yes	Reject H_0

TABLE III. EXPECTATION FOR SUPPORT

Support area	p-value	Significance (Yes/No)	Conclusion
Support on modules and course materials	0.001	Yes	Reject H_0
Support on the online lectures related issues	0.001	Yes	Reject H_0
Support on LMS-related issues (e.g. Login, navigation)	0.001	Yes	Reject H_0
Support on managing grades and results	0.001	Yes	Reject H_0
Support on coursework and assignments	0.001	Yes	Reject H_0
Support on Software/ device/ internet issues	0.001	Yes	Reject H_0

V. PROTOTYPE

The prototype was designed based on the conclusions taken from statistical analysis. It has two main segments dedicated to the NIBM students and NIBM administrators. With the proposed web application prototype, while students can receive individual support and assistance, it also helps admins manage and resolve technical issues. Overall, the system is designed to conduct real-time monitoring, and reporting and acts as a assisting and support system for both student and the institute.

A student or admin can access the web application by selecting the admin or student login and entering their registered credentials.

A. Student Side

From the student dashboard, students can get an overall view of the performance of the internet, used devices, and software. The dashboard also links the four main functionalities of this prototype (Fig. 5). Those are:

- Academic Center
- Monitoring Center
- Simulation Center
- Support Center

1) Academic Center

Fig. 6 presents the first interface of the academic center, which enables students get an ongoing progress in their academic activities. Additionally, students can request missing modules and lecture materials.

The second interface of academic center shown in Fig. 7 shows the status of ongoing, upcoming, and completed coursework & exams. With this, students can request support or view exams and coursework grades.

Academic Center – Grades - interface lets the student get an overview of their module grades. It also facilitates the application for re-sits or viewing the coursework & exam schedules (Fig. 8).

If the student select the Calendar option it shows dates, times, and durations for lectures, coursework, and exams. By clicking on the date, students are also able to get the link to attend online lectures (Fig. 9).

2) Monitoring Center

Monitoring Center as shown in Fig. 10 is responsible for real-time monitoring of the performance of the devices, internet connection, and software used for online learning of student side. It also provides students alerts, and reports, and facilitates option to request support when needed.

3) Simulation Center

The Simulation Center provides simulations of online lectures, Presentations, exams, and forums that are similar to NIBM Worldwide. With this student, they can try the platform features and get familiarize with online learning platform (Fig. 11). It also can help students with exams and presentation preparations.

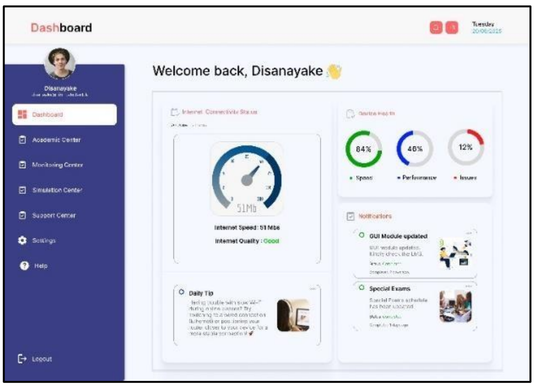


Fig. 5. StudentDashboard

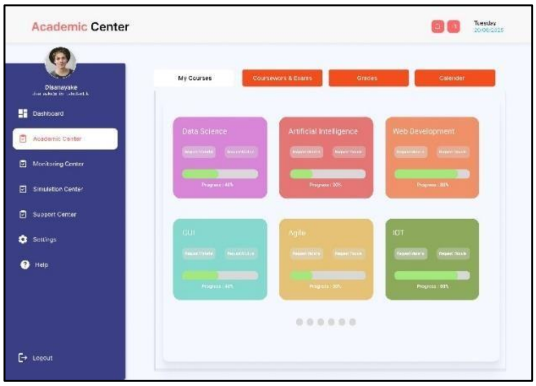


Fig. 6. Academic Center – My course

Fig. 7. Academic Center – Coursework and Exams

Especially, the presentation and exam functions provide a platform to practice for online presentation and exam. It also provide feedback regarding the connection used, device performance including mic, camera, and speaker performance. This could help student immensely to prepare backup plan prior to actual exams or viva.

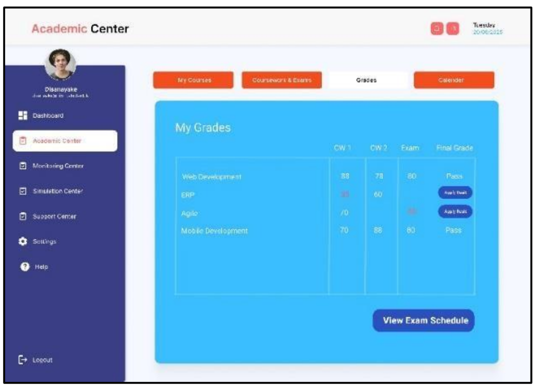


Fig. 8. Academic Center – Grades

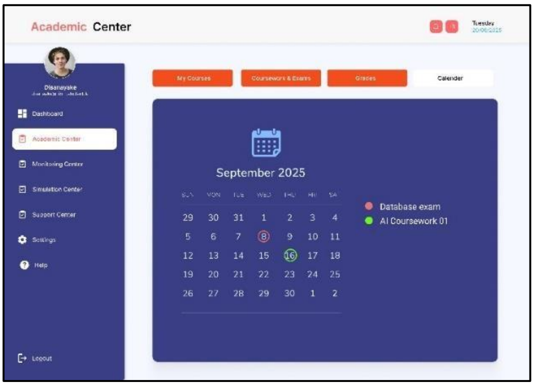


Fig. 9. Academic Center - Calendar

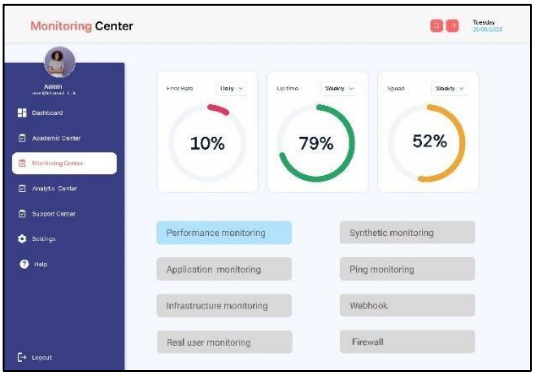


Fig. 10. Monitoring Center

4) Support Center

The support center shown in Fig. 12 address the expectation requirements for support, mentioned on TABLE III. With the support center students can get self-help materials and contact NIBM support for urgent issues. They are also able to request support. A summary of request status is also shown in the interface to track the progress of requested support.

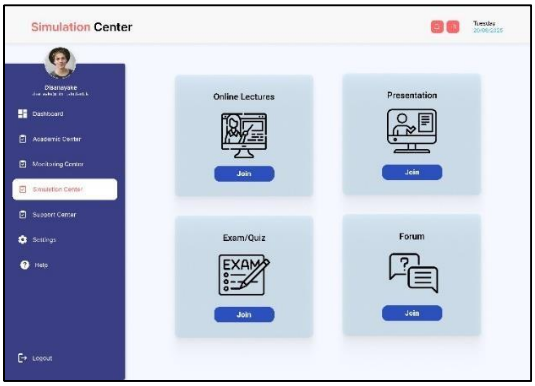


Fig. 11. Simulation Center

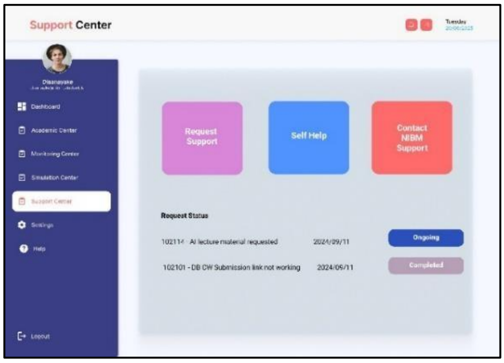


Fig. 12. Support Center – Overview

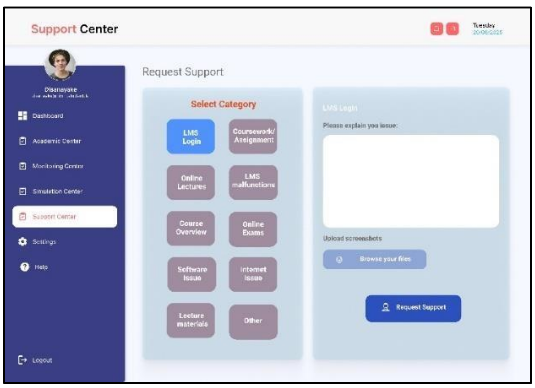


Fig. 13. Support Center – Request Support

When the students select Request Support option, they are directed to Request Support interface shown in Fig. 13. It lets the students report technical issues or request support. Here the students can select the category and provide details to report the issue and simply click the request support button to get the assistance from the NIBM support team.

B. Admin Side.

The admin side of the prototype provides the overall NIBM worldwide platform performance status, student engagement, support requests, and analysis and report functionalities that help the smooth operation of the online learning platform.

As shown in Fig. 14 admin dashboard also provide links the four main functionalities.

- Academic Center
- Monitoring Center
- Analytic Center
- Support Center

Apart from linking to main functions, the admin dashboard provide a summary of real-time analysis of the NIBM Worldwide platform performance and student engagement.

1) Academic Center

The admin academic center shown in Fig. 15 helps admins manage academic-related processes. This helps them to schedule and share timetables, grades, and lecture materials, and manage other activities.

2) Monitoring Center

The admin monitoring center is capable of monitoring performance, used applications, supported infrastructure, and users. It is also capable of conducting synthetic monitoring, ping monitoring, webhook, and firewall monitoring, and regular and real-time performance and security checks to ensure the smooth operation of the online learning platform (Fig. 16).

3) Support Center

As shown in Fig. 17, the admin support center lists the students' support requests requested via student request support interface (Fig. 13). With a color-coded strategy, it represent the urgency and prioritize the requests. By clicking a request, admins can access all the necessary details to resolve and manage the issue.

4) Analytics Center

The admin analytics center shown in Fig. 18 provides the analysis and report on how the overall NIBM online leaning platform performance. This helps to get valuable insights that could be used to further enhance the current online system.

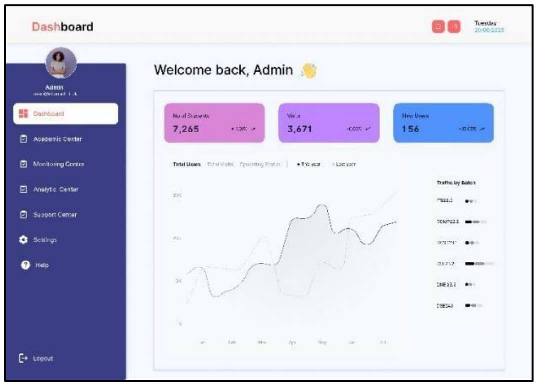


Fig. 14. Admin Dashboard

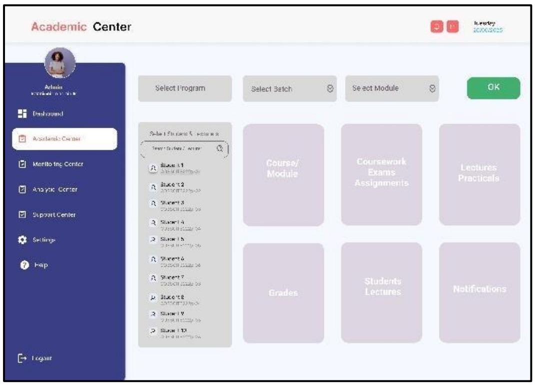


Fig. 15. Admin Academic Center

As a conclusion, the prototype provide a comprehensive solution addressing the key factors identified through statistical analysis. By providing tailored functionalities for both students and administrators, the prototype provide a user centric solution that could address the needs of both students and the institute, improving overall quality and effectiveness of online learning.

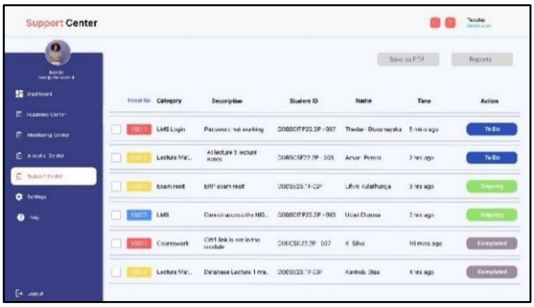


Fig. 16. Admin Support Center

Fig. 17. Admin Analytics Center

VI. Discussion

The finding of this study highlighted how technical challenges affect the online learning of students of the School of Computing and Engineering, at the National Institute of Business Management, in Sri Lanka. As a core component of NIBM's blended learning process, the NIBM Worldwide platform provided a case study to identify these issues.

The descriptive analysis of this study showcased that academic performance and satisfaction with online learning are greatly affected by technical challenges leading students to feel more frustrated and stressed.

The finding of this study showed that specific factors such as age, academic program level, proficiency in using digital tools, technical challenges such as internet connectivity, and device and software limitations hinder the effectiveness of online learning. Technical challenges related to course modules, lecture materials, online lectures, coursework, exams, and managing grades have also a significant impact on student learning behavior. Furthermore, the study also highlighted the need for a support system and student expectations to resolve such technical issues.

While addressing the above requirements, the proposed prototype is capable of real-time monitoring, analysis, reporting, and simulations to tailor the requirements of students and the institute. The prototype provides an easy-to-implement strategy to assist with bridging the gap between technological challenges and user expectations.

VII. Conclusion

As online learning has become an important component of Sri Lankan higher education, it is crucial to address the technical challenges. The findings of this study highlighted the importance and justified the reason for addressing technical challenges in the Sri Lankan higher education system.

The study provided valuable insights and proposed a prototype to improve the effectiveness and mitigate the impact of the technological challenges. The proposed prototype also provides a scalable and adaptable framework that could be used to monitor and assist with online learning activities. The prototype could assist other institutes that experience similar challenges.

By addressing the technical challenges and supporting the expectations of students, the proposed prototype represents a significant step forward improving the quality of online education in Sri Lanka and beyond.

Acknowledgment

The authors of this study would like to thank the NIBM Institute, the Director of NIBM SOCE, the IT for Business course director, SOCE lecturers, the research supervisor, students, and NIBM staff for the support and guidance provided to achieve the success of this study.

References

[1] K. Raouna, “What is Online Learning? Brief History, Benefits & Limitations,” LearnWorlds, Mar. 06, 2024. Available: https://www.learnworlds.com/what-is-online-learning/

[2] M. A. Almaiah, A. Al-Khasawneh, and A. Althunibat, “Exploring the critical challenges and factors influencing the E-learning system usage during COVID-19 pandemic,” Education and Information Technologies, vol. 25, no. 6, pp. 5261–5280, May 2020, doi: https://doi.org/10.1007/s10639-020-10219-y. Available: https://link.springer.com/article/10.1007/s10639-020-10219-y

[3] M Kamraju, J Krishnaiah, G Durgesham, N. Shaba, and Y Madhuri, "Exploring the Impact of Online Education on Higher Education," ASEAN Journal on Science and Technology for Development, vol. 03, no. 01, pp. 27–36, Jan. 2024, Available: https://www.researchgate.net/publication/370766475

[4] U. Latha and I. M. Lakshman, "Learning Landscapes in Post-Pandemic Context: A Comprehensive Study on Online Education at University of Jaffna, Sri Lanka," Journal of the University Librarians Association of Sri Lanka, vol. 27, no. 1, pp. 114–144, Apr. 2024, doi: https://doi.org/10.4038/jula.v27i1.8081.

[5] A. Andersson, "Seven Major Challenges for E-Learning in Developing Countries Case Study eBIT, Sri Lanka," 2008. Accessed: Oct. 19, 2023. [Online]. Available: https://oru.diva-portal.org/smash/get/diva2:201280/FULLTEXT01.pdf

[6] Jayoda Weerapperuma, et al. "Digital Learning Challenges in Tertiary Education in Sri Lanka: A Social Capital Perspective." IEEE Transactions on Computational Social Systems, 1 Jan. 2024, pp. 1–18, https://doi.org/10.1109/tcss.2023.3306571. Accessed 6 May 2024.

[7] "Overview," Ministry of Higher Education and Highways. https://www.mohe.gov.lk/index.php?option=com_content&view=article&id=66&Itemid=123&lang=en

[8] S. KEMP, "Digital 2024: Sri Lanka," DataReportal – Global Digital Insights, Feb. 23, 2024. https://datareportal.com/reports/digital-2024-sri-lanka

[9] "Computer Literacy Statistics 2023 First Six Months Bulletin Department of Census and Statistics Ministry of Finance, Economic Stabilization and National Policies," Department of Census and Statistics, Ministry of Finance, Economic Stabilization and National Policies, 2023. Available: https://www.statistics.gov.lk/Resource/en/ComputerLiteracy/Bulletins/2023-FirstSixMonths.pdf

[10] J. Weerapperuma, D. Nawinna, and N. Gamage, "Digital Learning Challenges in Tertiary Education in Sri Lanka: A Social Capital Perspective," IEEE Transactions on computational social systems, pp. 1–18, Jan. 2024, doi: https://doi.org/10.1109/tcss.2023.3306571.

[11] "SCHOOL OF COMPUTING," NIBM, 2025. https://www.nibm.lk/faculties/school-of-computing

[12] "About," NIBM World Wide. https://www.nibmworldwide.com/about

[13] R. E. Mayer and R. C. Clark, E-Learning and the science of instruction: proven guidelines for consumers and designers of multimedia learning. New York, NY: John Wiley Et Sons, 2016.

[14] F. Amiti, "SYNCHRONOUS AND ASYNCHRONOUS E-LEARNING," European Journal of Open Education and E-learning Studies, vol. 5, no. 2, Sep. 2020, doi: https://doi.org/10.46827/ejoe.v5i2.3313.

[15] S. Appana, "A Review of Benefits and Limitations of Online Learning in the Context of the Student, the Instructor and the Tenured Faculty," International Journal on E-learning, vol. 7, no. 1, pp. 5–22, Jan. 2008.

[16] B. Gilbert, "Online Learning Revealing the Benefits and Challenges,"

Education Masters, Apr. 2015, Available: https://fisherpub.sjf.edu/education_ETD_masters/303

[17] A. D. Dumford and A. L. Miller, "Online learning in higher education: exploring advantages and disadvantages for engagement," Journal of Computing in Higher Education, vol. 30, no. 3, pp. 452–465, Apr. 2018, doi: https://doi.org/10.1007/s12528-018-9179-z.

[18] E. J. Sintema, "Effect of COVID-19 on the Performance of Grade 12 Students: Implications for STEM Education," Eurasia J. Math. Sci. Technol. Educ., vol. 16, no. 7, Apr. 2020, doi: 10.29333/EJMSTE/7893

[19] W. Bao, "COVID-19 and online teaching in higher education: A case study of Peking University," Hum. Behav. Emerg. Technol., vol. 2, no. 2, pp. 113–115, Apr. 2020, doi: 10.1002/HBE2.191

[20] S. N. Samsudeen and R. Mohamed, "University students' intention to use e-learning systems," Interactive Technology and Smart Education, vol. 16, no. 3, pp. 219–238, Sep. 2019, doi: https://doi.org/10.1108/itse-11-2018-0092.

[21] S. L. M. Sajeer, "Expectation of users in using electronic resources: a case study of the southeastern university of Sri Lanka," Seu.ac.lk, 2017, doi: https://doi.org/978-955-627-120-1.

[22] M. Deegahawature, "Challenges in Effective Implementation of E-Learning in Sri Lanka." Available: http://repository.wyb.ac.lk/bitstream/handle/1/2002/WAYABA-JOURNAL-OF-MANAGEMENT-2010-i-9.pdf?sequence=2&isAllowed=y

[23] A. Grubišić et al., "A common model for tracking student learning and knowledge acquisition in different e-learning platforms," J. E-Learn. Knowl. Soc., vol. 16, no. 3, pp. 10–23, 2020, doi: 10.20368/19718829/1135235.

[24] U. Kaden, "COVID-19 school closure-related changes to the professional life of a K–12 teacher," Educ. Sci., vol. 10, no. 6, p. 165, 2020.

A Mobile Application Prototype to Enhance Academic Performance Through Student Learning Behaviors and Time Management Practices Among First-Year Undergraduates at NIBM, Sri Lanka

Muthukumarana P A
School of Computing and Engineering
National Institute of Business Management Colombo 07, Sri Lanka
anuradhamuthukumarana99@gmail.com

Amila De Silva
School of Computing and Engineering
National Institute of Business Management Colombo 07, Sri Lanka
amilau@nibm.lk

***Abstract*—The academic underperformance of first-year Information Technology (IT) undergraduates at the National Institute of Business Management (NIBM), Colombo, poses a significant challenge, particularly for student retention. The transition from high school to university creates academic bottlenecks, with Software Engineering reporting a 22.92% exam repetition rate. Rigorous curriculum and lack of preparatory exposure hinder students from meeting academic benchmarks, impacting both individual career prospects and Sri Lanka's IT industry. The main objective of this research is to introduce IntelliStudy360, a mobile application prototype designed to enhance time management and learning behaviors. A stratified random sampling technique ensured proportional representation across the four diploma programs by selecting 301 first-year students. A structured questionnaire was used to assess demographics, time management, learning behaviors, and academic performance, with pilot testing to ensure reliability. Key constructs, such as poor task organization, lack of self-regulation, lack of active learning, and poor motivation, were derived from prior research. SPSS analysis revealed strong correlations between self-regulation deficits ($r = 0.576$, $p < 0.001$), low active learning engagement ($r = 0.517$, $p < 0.001$), and academic underperformance, with poor task organization as a key factor ($r = 0.468$, $p < 0.001$). Moreover, the results indicated that 43.53% of the students experienced significant procrastination, and over 70% reported difficulties in managing their time effectively. The model suggested that 44.9% of the variance in academic underperformance was explained by predictors with a moderate correlation. ($R=0.675$, $R^2=0.456$, Adjusted $R^2=0.449$). IntelliStudy360 offers goal setting, adaptive learning, academic tracking, AI-driven study plans, and gamified learning, providing a scalable solution to improve student performance.**

Keywords—Student learning behaviors, Time management practices, Academic underperformance, NIBM, First-year undergraduates, Mobile-based prototype application

I. INTRODUCTION

The process of moving from secondary to higher education is a daunting academic endeavor and a major performance hurdle for first-year undergraduate students. Owing to low academic competencies involving time and student learning behaviors, first-year undergraduates fail to achieve academic outcomes with reference to university-defined academic standards or potential academic difficulties that may undermine their academic progress or future success. Literature reviews conducted across cross-sectional and international settings have established that first-year learners experience incredibly challenging times, transitioning to higher learning institutions' academic rigor. Scholarly work has discussed numerous aspects known to result in academic underperformance. Fokkens-Bruinsma and Hawkins identified key challenges, Shortage of time management skills, Self determined motivation, Self-regulated learning skill deficit. These challenges are becoming more apparent among academics in advanced technology-oriented fields that require high-standard education [1],[2].

According to statistical data, this underlines the fact that this problem is rather serious. The National Survey of Student Engagement (NSSE) also showed that less than 45 % first-year students and 38% seniors devoted 15 plus hours a week to studying;, therefore not sufficiently academically engaged [3]. In addition, at the National Center for Education Statistics (NCES), 68% of undergraduates are employers, and 43% spend more than 20 hours on their employers. They have a very challenging student schedule, resulting in low learning outcomes and high dropout levels, with 40% of students failing to graduate due to poor academic performance [4]. Based on research findings from 301 diploma students at NIBM, software engineering had the highest exam repetition rate at 22.92%, indicating academic underperformance because it did not meet the academic standards provided by the university.

Based on the need for a broad understanding of the complex link between time management problems and academic achievement among first-year undergraduate students in specific settings, this research focuses on the School of Computing at NIBM, Sri Lanka. This research seeks to address critical research objectives: establishing performance bottlenecks, examining learning patterns, reviewing time management practices, analyzing current support structures provided by the university, and designing robust technological solution to address these objectives. The research excludes upper-level students, professionals, and educators, focusing solely on first-year students within this context. Socioeconomic factors and external institutional influences are not considered unless directly related to time management and student behavioral behaviors. The underlined quantitative research method employed 301 first-year undergraduates.

The principal research consequence is an original mobile application that helps overcome performance limitations.

This technological intervention will have additional individual value features, such as learning experience personalization, detailed time management, productivity improvement, and analytics and evaluation exclusive to each graduate. The academic intention of the application is to foster positive learning patterns, increase students' attentiveness, and address aspects of their learning skills.

II. LITERATURE REVIEW

A. Student Learning Behaviors

Student learning behaviors are key predictors of academic performance and can be described as a composite of the dynamic dual relationships of personal traits and learning approaches. Multiple studies have highlighted that low academic achievers also exhibit predominant or low active engagement, which is evident in irregular study habits, delayed study, and ineffective time management. Low-performing students reduce learning efforts, exhibiting poor study behaviors, such as erratic studying, weak self-regulation, and indecisiveness in seeking help. Factors include late-night homework, skipping preparation, and last-minute assignments, highlighting procrastination and time mismanagement as key contributors to academic underperformance [5]. Among the extroverted students, 80.65% gave surety to social distractions, while 70.97% acknowledged poor study habits [6]. Cormack's research noted that first-week assignment submissions have a strong negative relationship with academic performance [7]. Conversely, Bhattacharya believed that there existed a positive relationship between the effectiveness of the time management practices utilized and achievements in academic pursuits, with concern for overall strategic learning techniques. Additionally, this research further explains how self-directed learning factors, such as motivation have a positive impact on student learning [8]. Most importantly, research by Tameemi's work reveals that deep learning methodologies, which involve critical analysis and integration of knowledge rather than rote memorization, have been shown to be more effective [9]. Moreover, there are various modes of learning, including textbook reading, note taking, memorizing, test preparation, and concentration, which affect how time is managed. Worst among all these patterns is poor time management skills that have been found to reduce the eventual achievements made by the student thus showing the need for effective utilization of time to ensure good grades [10]. However, their study did not account for the role of external factors such as institutional support or socioeconomic background, which may influence time management capabilities. This gap highlights the need for comprehensive interventions that address both behavioral and environmental factors affecting learning.

B. Time Management Practices

As various studies have been conducted by scholars, effective time management is an essential factor that plays a tremendous role in determining student grades. The module for training activity is based on the time management capabilities of engineering undergraduates and links improved time management to academic and professional outcomes. Research has identified many learners with weak time management skills, often failing to set daily objectives or plan their day adequately [11]. Similarly, Magdy Anwer and Abdullah Elmahdy's research showed that 60.8% admitted that poor task planning and prioritization affected performance, and 45.1% were often procrastinated. The quantitative results of multivariate regression analyses show that there is an existing correlation between time management skills and Grade Point Average (GPA), with poor temporal structure negatively impacting performance [12]. Research conducted at Jazan University with 491 students, including 53.8% female students and 46.2% male students, observed that higher time management was significantly positive for students' Cumulative Grade Point Average (CGPA), particularly in the applied medical sciences while low time management negatively affected the CGPA of students of business faculty [13]. Sajeevanie and Tharuka's research shows that students need to effectively organize time to manage university-related tasks and enhance performance. Additionally, their research further highlighted that poor self-regulation might impact student studies if they do not properly balance their time [14]. Extending this further, Gulua's research also established that the survey among students showed that only 28% of them were able to do all the assignments planned for the day; only 55% did not record time or analyze it in the last year, which illustrates how problematic time management practices were for these respondents be [15].

C. Academic Underperfromance

Academic underperformance results from a combination of behavioral, psychological, and external factors. Moreover, academic underperformance means getting below the expected performance or below what is expected of the learner or institution they attend, perhaps due to poor time management, efficient study methods, low motivation, and other related psychological causes, as explained by Pascoe [16]. Stress undermines thought processes, and thus, students' ability to concentrate, remember information, and solve problems, which is not good for tracking academic results. Anxiety, particularly examination stress, affects tests, while depression causes truancy and disinterest in academic work [17]. These findings suggest that psychological support should be integrated into academic interventions to address the cause underlying underperformance. On the behavioral side, a lack of intrinsic motivation and extrinsic motivation tends to distort a student from identified academic activities even more [18]. Lack of time management leads to poor planning and high stress levels, resulting in negative academic performance [19]. Although these studies provide valuable insights, they often fail to distinguish between temporary academic setbacks and chronic underperformance. Future research should explore whether targeted interventions can help students recover from academic challenges or whether certain students remain consistently at risk despite support.

D. First-Year Undergraduates

First-year undergraduates face unique academic challenges that differ from those of senior students. Jairam's research identified that powerfully and often overlooked are the significant heaps some of these students face major challenges such as abrupt changes in academic expectations as well as newfound responsibilities for themselves [20]. Similarly, Mahawattha and Rassool's work also highlights other issues, including lack of appreciation or support from family/friends and poor language skills, especially for students with English medium instruction (EMI) [21]. The impact of discipline-specific workload difference is another critical factor. Among freshmen, Hawkins et al. noted that students taking Science, Technology, Engineering, and Mathematics (STEM) courses find it very hard to manage their workload alongside other personal and academic responsibilities, thus suffering from time-poverty and stress [2],[29]. In the solicitous criticality pointed out by Ishikawa's research that identifies the difficulties in setting attainable objectives, in controlling self-learning, and in coping with feelings which can interrupt learning define this lamentable state of affairs. This increases the pressure for academics, which in conjunction with the process of establishing new study habits, results in the creation of a difficult situation that can hinder academic performance [22]. Additionally, while existing studies focus on student adaptation, they often overlook the role of faculty and institutional structures in easing the transition

E. Current Interventions and Solutions

Interventions used in current society to tackle time management challenges among students show excellent technological and strategic growth. In their study, including a review of the literature on organizational activities, Silva and Ramos showed that most organizational tools such as Trello, Google Calendar, and reminders that help endorse self-regulated learning and limit procrastination work well. All these digital solutions help organize tasks and events that in turn produce efficient academic spaces [23]. Software applications such as UniPlanner and Automated Time Manager (ATM) have extensive support elements, including sets of tools such as tracking tasks or CGPA, and expert advice [24]. Some of the key features left out by Kim's research on ATM include real-time activity tracking, goal-setting aids and distractions awareness [25]. Other studies by Holicza and Kiss confirmed that machine-learning algorithms can predict the performance of students based on important demographic and behavioral variables, such as time spent on a task, study behavior, and screen time [26]. Besides, Advancing Gerontology through Exceptional Scholarship (AGES) Program and communication with mentors who are involved in the program along with other students also successfully address academic issues in a more structured sense through the addresses coordinated programs [28]. Additionally, Time booster Academic Systemic Model(TASM) is a comprehensive solution where learners are encouraged to delegate trivial tasks, automate administrative jobs and then use the "Eisenhower matrix" to prioritize tasks. Resilience building and collaboration also address the tendency towards perfectionism and procrastination, which otherwise makes these challenges difficult [27].

IntelliStudy360 has been conceptualized to bridge this gap. Despite these advancements, the existing tools have failed to comprehensively address the interplay between time management, self-regulation, and active learning engagement. Google Calendar and Trello provide task scheduling but lack personalized study recommendations and behavioral tracking. UniPlanner and ATM incorporate academic tracking features, yet they do not offer adaptive learning suggestions tailored to individual students' needs. Machine learning-based tools predict student performance trends but fail to provide real-time AI-driven interventions to improve study efficiency.

However, existing studies have not examined the role of AI-driven interventions in addressing academic underperformance, particularly at non-state universities in Sri Lanka. Research in similar fields has reported similar issues in science, engineering, business management, and the faculties of medical science. [24],[11]. This research seeks to fill this gap by highlighting first-year Diploma IT students at the NIBM and comparing time management, learning behavior, and academic underperformance. The result provided empirical evidence for the currently scarce literature on IT education and guide interventions that seek to enhance learners' achievements and prepare them for their future careers in this rapidly growing field.

III. METHODOLOGY

The research method used in this research is a quantitative research method that aims to statistically analyze the effects of time management and learning behaviors that are perceived to be suboptimal in first-year IT undergraduates and their academic underperformance. A Self-administered structured survey questionnaire use to solicit demographic, time management practices, behavioral and academic performance data that will facilitate identifying the trends and relationships.

A. Population and Sampling

This research focused on a target population of 720 full-time first-year undergraduates from the School of Computing and Engineering in Colombo, strategically selecting a sample of 301 students through a stratified random sampling technique that ensures the proportional representation across diverse diploma streams, enhancing the generalizability of the findings to minimize the bias of the responses. Based on Krejcie and Morgan's sample size table, a minimum sample of size of 251 was required to generalize the findings with a 95% confidence level and a 5% margin of error. However, to account for potential non-responses (20%), the final sample size was set at 301 students. (Software Engineering: 135, Computer System and Design: 42, Network Engineering: 62, Advertising and Multimedia: 62).

B. Data Collection

Data were collected through a administered, pilot-tested, structured questionnaire in a physical manner. A pilot test was conducted with 20 participants from the target population to ensure reliability and validity of the questionnaire. The survey instrument adequately captured demographic data, time use and management, student learning strategies and outcomes, academic results, and current organizational support structures. The research directly collected data by physically distributing hard copies to first-year undergraduates, thereby reducing response bias. A 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree) was used to quantify the student responses. Survey questions related to Task Organization evaluated students' ability to plan their study schedules, prioritize tasks, and set deadlines. Questions such as "I consistently use a planner to schedule my academic tasks" and "I allocate specific time slots for studying different subjects". The Self-Regulation dimension was assessed through questions such as "I complete assignments ahead of deadlines" and "I stick to my study plan without procrastinating." This aimed to measure the effectiveness of students' follow-through of planned activities.

Survey items addressing Active Engagement in Learning included "I actively participate in discussions" and "I review class notes regularly, not just before exams", which evaluated students' proactive learning strategies. Self-

Directed Learning was measured using items such as, "I set academic goals each semester" and "I take responsibility for my learning outside lectures." These questions assessed the students' intrinsic motivation and self-discipline.

C. *Conceptual Framework*

One study investigated the associations between student learning behaviors and time management practices as they relate to academic underperformance experienced by first-year NIBM undergraduates. The framework that explains the link between poor time management and academic underperformance stems from self-regulated learning and time management theory.

D. *Theoritical Justification*

a)

ime management practices: Academic tasks become more productive through task organization when students properly structure and sequence them according to their priorities following Goal-Setting Theory [30]. The ability of students to supervise their behavior while adjusting their actions toward academic success goals defines self-regulation in time management. According to Self-Regulated Learning Theory, students who practice self-monitoring, strategic planning, and evaluation processes achieve better academic results [31].

b)

tudent Learning Behaviors: The Support of Active Engagement in Learning stems from Social Constructivism because this theory promotes education through social interaction and real-world experiences [32]. Academic success requires intrinsic motivation (personal interest) and extrinsic motivation (external rewards), according to Self-Directed Learning theory (motivation) that matches Self-Determination Theory [18].

E. *Data Analysis*

Data analysis was conducted using SPSS, utilizing different levels of analysis to interpret the underlying pattern of the relationship between independent and dependent variables. Regression analysis was used to explore the complicated interactions governing underperformance, whereas a basic correlation analysis was used to study the associations governing the link between time management behaviors and learning patterns. Thus, the methodological design applies a sound conceptual foundation to distinguish key independent variables, such as poor time management practices and student learning behaviors, and their likely effects on the dependent variable, academic underperformance, as shown in Fig. 1

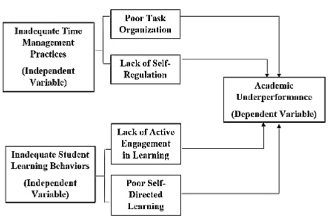


Fig. 18. Conceptual framework of the theoretical model underpinning the research.

IV. Data Analysis

A. *Academic Underperformance Factors*

The assessment identified behavioral and organizational precursors for poor academic performance. As shown in Fig. 2, exam repetition has been analyzed with regard to different diploma fields of study, which implies that they couldn't fulfill the university-defined academic standards to pass the module. Specifically, students in the Software Engineering Stream had the highest (22.92%) number of students repeating their exams – a situation that should trigger targeted interventions.

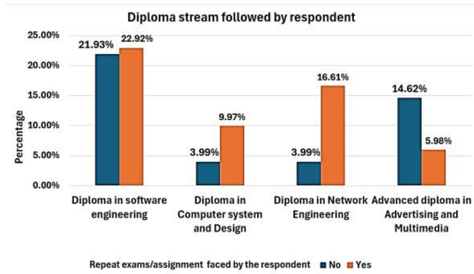


Fig. 19. Bar chart showing the frequency of repeat exams and assignments faced by first-year undergraduates across diploma streams.

S

Lack of time management and rushing through assignments in the last minute are the identified factors that hamper their academic studies, as shown in Fig. 3 Major determinants of poor academic performance. Most undergraduates believed that this factor had a negative impact on their studies. Moderate to severe impacts of procrastination were reported by 43.53% of the participants, indicating that procrastination is a prevalent challenge in meeting time-sensitive tasks and the overall management of academic workloads.

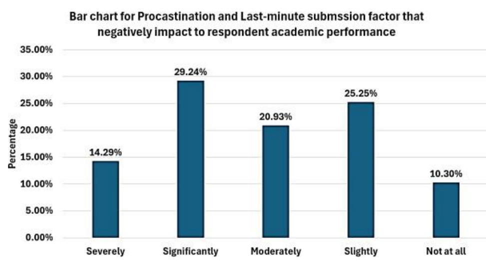


Fig. 20. Bar chart illustrating the prevalence of procrastination and last-minute assignment submission among students.

Additionally, as shown in Fig. 4 Almost all the students believed that poor time management negatively affected their studies (28.90%). Similarly, 37.37% of the respondents said that it moderately influenced them, whereas 25.58% said that it greatly influenced them. Of all the students, 7.97 percent were unaffected. The data also provide significant evidence that time management is high. Essential to academic outcomes, with 70 percent or more dropping comments from moderate to severe.

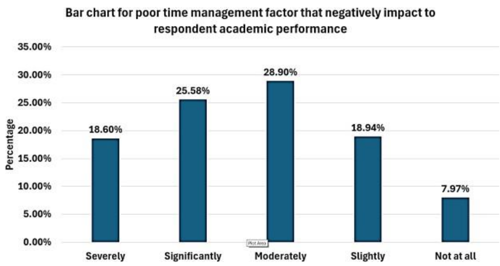


Fig. 21. Bar chart illustrating how poor time management negatively impacts the academic performance of first-year undergraduates.

By accessing their current time management practices, it was found that more than half of the respondents assessed across all diploma streams indicated that dissatisfaction with existing time management practices was expressed most by Software Engineering students, with 15.95% of respondents indicating dissatisfaction. This is contrasted with Advanced Diploma in Advertising and Multimedia students who are more balanced in their comments and relatively lower engagement from the students of Network Engineering. Overall, the data suggests more negative than positive perceptions of the role played by time management in academic performance with an inherent possibility of developing time management practices for the Software Engineering program.

B. Student Learning Behaviours and Time Management Practices Related Factors

A descriptive analysis was performed for the students' learning activities, such as group study sessions, with a mean of 3.53. Nonetheless, some scarcities were reflected in the active work with the materials (mean=3.17) and the regular revision of the study materials (mean= 3.12). There is an implication here for more structured and purposeful activities if there is to be better accomplishment at the university.

Descriptive analysis of the time management practices and data showed that time management practices with moderate task organization (mean= 3.20) and complete the assignments before the deadlines (mean= 3.55), although scheduling practices were seriously wanting (mean= 2.87). This means that as many as students try to juggle different tasks, they lack orderly ways of programing themselves, not forgetting that time is a sensitive aspect that determines their achievement of these tasks.

C. Support Intervention Provided by the University

The survey results indicate that the effectiveness of coursework in developing skills is moderate, and 29.24% good in terms of faculty mentorship programs is 35.22% moderate. The issues raised also revealed a major lack of support (38.22%), an implication for resource enhancement. The level of perceived usefulness of LMS regarding assignment reminders was moderate: 25.6% of the respondents reported moderate effectiveness and 24.9% reported low effectiveness or ineffectiveness of this feature.

The analysis of the survey results is shown in Fig. 5 showed that a significant portion of first-year respondents were inclined towards the use of the mobile application, with only about 1.33% being very unlikely to use it and 46.51% likely to do so. This offers a positive tendency, as one out of three (34.55%) respondents did not express willingness or disincentives, implying an opportunity for additional marketing efforts for these users.

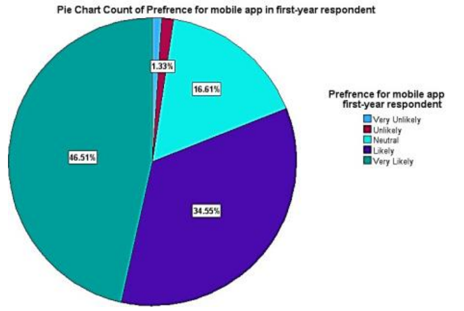


Fig. 22. Pie chart showing mobile app preferences among first-year undergraduates.

D. Correlation Analysis

The correlation matrix is shown in Fig. 6 There was significant correlation between the independent variables; namely time management and learning behavior and the dependent variable; academic underperformance; all 'p' values being less than 0.001 of this self-regulation has the highest positive correlation coefficient equal to 0.576 followed by active engagement in learning equal to 0.517. These results are critical for emphasizing self-regulatory behaviors for success in universities and offering a rationale for behavioral treatment programs.

Correlations

		DV	IV1	IV2	IV3	IV4
DV	Pearson Correlation	1	.468**	.576**	.517**	.529**
	Sig. (2-tailed)		<.001	<.001	<.001	<.001
	N	301	301	301	301	301
IV1	Pearson Correlation	.468**	1	.556**	.470**	.382**
	Sig. (2-tailed)	<.001		<.001	<.001	<.001
	N	301	301	301	301	301
IV2	Pearson Correlation	.576**	.556**	1	.439**	.511**
	Sig. (2-tailed)	<.001	<.001		<.001	<.001
	N	301	301	301	301	301
IV3	Pearson Correlation	.517**	.470**	.439**	1	.521**
	Sig. (2-tailed)	<.001	<.001	<.001		<.001
	N	301	301	301	301	301
IV4	Pearson Correlation	.529**	.382**	.511**	.521**	1
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	
	N	301	301	301	301	301

**. Correlation is significant at the 0.01 level (2-tailed).

Fig. 23. Correlation matrix illustrating the relationships between independent variables and the dependent variable.

E. Hypothesis Testing

To analyze student learning behaviors, time management practices, and underperformance for comparison in terms of different attributes, a hypothesis was set to conduct the analysis and infer the results. The two hypotheses are as follows: the Null Hypothesis (H_0) and alternate hypothesis (H_1). In other words, if the calculated results are statistically significant, the Null Hypothesis (H_0) is rejected, the alternative hypothesis (H_1) is accepted, and the reverse is the case if the results are not statistically significant. The following table (Table I) illustrates the statistical significance of the independent variables.

TABLE IV. STATISTICAL SIGNIFICANCE OF INDEPENDENT VARIABLES

Independent Variables	P-value	Significance	Conclusion
Poor task organization	<0.001	Significant	Reject H_0
Lack of self-regulation	<0.001	Significant	Reject H_0
Lack of active engagement in learning	<0.001	Significant	Reject H_0
Poor self-directed learning	<0.001	Significant	Reject H_0

The results showed that all the independent variables- poor task organization, lack of self-regulation, lack of inactive engagement and poor self-determined learning significantly affect poor academic performance because P-values less than 0.001. The outcome of the available variables means that each variable should be rejected for the null hypothesis, and that significant role of each variable can influence the outcomes of academic achievement. A lack of task organization can result in improper prioritization, improper division of time, and lack of self-regulation, meaning that one cannot remain adequately focused and motivated. Thus, the level of active learning is insufficient, which means a lack of participation and interaction in learning processes. Poor self-directed learning indicates the absence of taking initiative and managing one's own learning. Therefore, these results highlight the need to reform these behavioral and organizational conditions to enhance academic performance.

F. Regression Analysis

As shown in Fig. 7 By evaluating the adjusted R-square value, which shows the proportion of the dependent variable by the independent variables, it showed a moderate fit, explaining nearly 45% of the variance (44.9%).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.675[a]	.456	.449	.59363

a. Predictors: (Constant), IV1, IV4, IV3, IV2

Fig. 24. Model Summary of the regression analysis.

As shown in Fig. 8 All the independent variables are statistically significant at the 5% level. Thus, a regression model equation can be derived using the following formula:

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	.061	.194		.316	.752
	IV2	.314	.057	.308	5.472	<.001
	IV3	.256	.063	.216	4.026	<.001
	IV4	.206	.052	.216	4.000	<.001
	IV1	.113	.054	.113	2.085	.038

a. Dependent Variable: DV

Fig. 25. Regression coefficients from the analysis.

$$Y = 0.061 + 0.113x_1 + 0.314x_2 + 0.256x_3 + 0.206x_4 \qquad (1)$$

In (1), Y represents academic underperformance, modeled as a function of four independent variables: X_1 (poor time management practices), X_2 (lack of self-regulation), X_3 (lack of active engagement in learning), and X_4 (poor self-directed learning).

V. SOFTWARE DEVELOPMENT PROTOTYPE

The research findings led to the creation of specific solution features in IntelliStudy360 to overcome these student difficulties. According to the research results, time management proved to be a major issue for students because 70% of the participants struggled to properly distribute their study hours. On the other one hand, this disorder triggers students to complete assignments at the final hour instead of developing effective learning methods. IntelliStudy360 implements study planning features that tailors academic planning to personal student schedules to create organized learning sessions. Through these features, the system encourages students to form positive learning patterns that lead to persistent and productive habits.

Procrastination emerged as another substantial concern among students, affecting 43.53% of them, resulting in late coursework submissions and reduced knowledge retention abilities. The achievement-tracking feature and point-based rewards system along with motivational challenges within IntelliStudy360 stimulate active student participation while maintaining responsibility throughout learning activities. Students who receive behavioral reinforcement maintain consistent work habits, which helps them avoid academic procrastination.

Lack of self-regulation with a correlation of r=0.576 with academic underperformance served to develop distractions by releasing Focus Mode as a feature that locks notifications and protects students from accessing unnecessary apps throughout their study period. The system design blocks interruptions to support students during deep learning sessions and enhances concentration effectiveness.

Task disorganization emerged as a significant problem, demonstrating an intermediate-level link (r = 0.468) between successful task prioritization and academic achievement according to research findings. Student success in workload management improves when they use the task prioritization tool in IntelliStudy360 to organize assignments based on deadlines and priority levels.

Lack of Active engagement was a leading indicator of academic underperformance (r = 0.517). Students who participated in peer discussion groups together with study collaboration showed better academic results. The system promotes this behavior through notice boards that unite student discussion groups and allow them to jointly complete projects while creating an environment that fuels both interest and academic exchange.

Study consistency becomes easier through the app's feature of automated study reminders combined with prompts that help students stay on track and tackle forgetfulness, while maintaining structured study methods. Students can optimize their study methods with the help of real-time analytics and progress tracking, which present them with data on their patterns of study.

The elements shown in Fig. 9 are onboarding screens for the IntelliStudy360 application, such as login, sign-up, and account creation, which would allow users to enter customized learning process with zero difficulties.

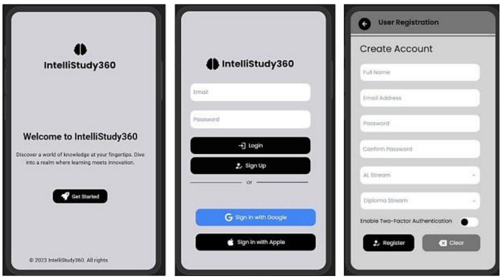


Fig. 26. Welcome, Login, and Registration Screens of IntelliStudy360.

As shown in Fig. 10, This concept displayed a study management app with Home Dashboard that includes Tasks list and Exam Countdown; My Study Plan that displays an overview of the student's study schedule; and a Creator for assigning tasks on a Daily, Weekly or Monthly basis to improve concentration and planning abilities.

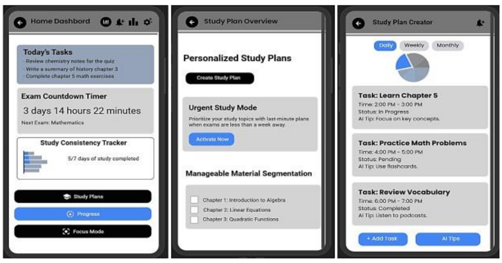


Fig. 27. Home dashboard, Study Plan Overview, and Study Plan Creator Screens of IntelliStudy360.

As depicted in Fig. 11, these screens show aspects of an emergency study mode, a concentrated mode and aesthetics mode of study that allows the organization of effective study sessions and the ability to silence notifications among other features.

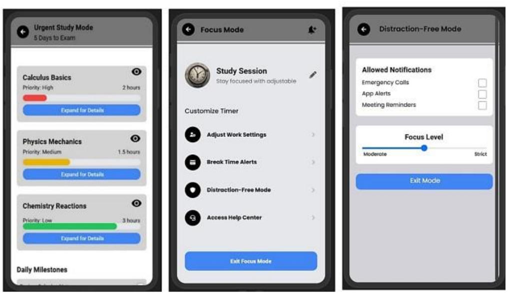
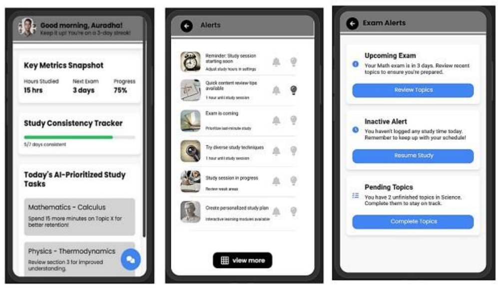


Fig. 28. Urgent Study Mode, Focus Mode, and Distraction-Free Mode Screens of IntelliStudy360.

As demonstrated in Fig 12, on these screens, users can find a progress trends screen displaying a performance dashboard, a time allocation and weakness screen offering detailed analytics, an achievement screen that presents users with relevant features such as weekly challenges, and study streaks.

Fig. 29. Interactive dashboards showcasing performance metrics, analytics, and achievements for progress tracking Screens of IntelliStudy360.

As shown in Fig. 13, the three screens showcased a personalized study assistant: the first is a snapshot of the essential read rate, important figures, studies tracker, and the tasks of top importance; the second gives notification about reminders, progress in a study, and suggested reading; the third shows exam-based notifications, the studies not activated for a while, and the topics to look at.



enables users to set their own study reminders and progress updates and smart notifications so they can receive tailored learning experience.

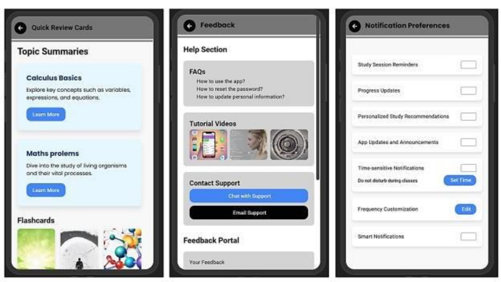


Fig. 31. Quick review cards, Feedback & Support, Notification preference Screens of IntelliStudy360.

VI. DISCUSSION

Implications for the context of first-year undergraduate students revealed that self-regulated learning behaviors and time management strategies, in which a significant focus is placed on self-regulation, planning, active participation, and independence, all supported academic efforts to avoid underperformance. Research findings further confirm that these parameters were positively and predictably related to academic achievement and that lack of self-regulation was one of the most influential determinants ($r = 0.576$). The highlights here coincide with the tenets of self-regulated learning, which focus on self-regulating learning processes and self-observation. Poor task organization was also positively related to academic underperformance ($r = 0.468$) and lack of active engagement ($r = 0.517$), thus supporting the role of structured learning patterns and learning environment interactions. However, the average correlation with poor self-directed learning (0.529) indicates that although self-direction is a useful learning style, it is not enough and only by incorporating guidance and external assistance.

VII. CONCLUSION

This research examines first-year academic underperformance among first-year university students attending NIBM, Sri Lanka, and finds that the main cause is poor time management and learning strategies. There is a positive relationship between lack of self-regulation, degree of poor task organization, and their academic underperformance was agreed upon. It was also found that it is possible to seek help, although the kind of help that can be sought cannot be considered as very helpful. Hence, future research should consider mobile interventions and self-regulating instruments to enhance learning and the outcome of the intervention on undergraduate performance.

ACKNOWLEDGMENT

I extend my deepest appreciation to all those who contributed to the success of this study. I thank my supervisor for her support and cooperation in this work. I also have a gracious feeling towards the faculty members, whose lectures and constant encouragement gave much flavor to my research work. Last but not the least, lot of thanks to the NIBM for facilitating the important and essential requirements that were needed for this research.

REFERENCES

[1] W. Hawkins, K. Goddard, and C. Favero, "A Cocurricular Program That Encourages Specific Study Skills and Habits Improves Academic Performance and Retention of First-Year Undergraduates in Introductory Biology," CBE—Life Sciences Education, vol. 20, no. 1, p. ar4, Mar. 2021, doi: 10.1187/cbe.20-06-0117.

[2] National Survey of Student Engagement, "Engagement insights: Survey findings on the quality of undergraduate education—Annual results 2019," Indiana University Center for Postsecondary Research, Bloomington, IN, USA, 2020. [Online]. Available: http://hdl.handle.net/2022/25321.

[3] T. Nachazel, W. Smith, M. Ossolinski, National Center For Education Statistics, United States. Department Of Education, and Institute Of Education Science, The condition of education 2017. Washington, D.C.: U.S. Department Of Education, National Center For Education Statistics, 2017.

[4] F. Huang and Q. Dong, "Characteristics and Influencing Factors of Poor Academic Performancein University," International Journal of Management Studies and Social Science Research, vol. 05, no. 01, pp. 244–257, Jan. 2023, doi: 10.56293/ijmsssr.2022.4566.

[5] G. D. Libed, "Factors that Contribute to Failure Causes of Academically Low Performing Students," Jurnal Ilmiah Peuradeun, vol. 10, no. 1, p. 205, Jan. 2022, doi: 10.26811/peuradeun.v10i1.665.

[6] S. H. Cormack, L. A. Eagle, and M. S. Davies, "A large-scale test of the relationship between procrastination and performance using learning analytics," Assessment & Evaluation in Higher Education, vol. 45, no. 7, pp. 1046–1059, Jan. 2020, doi:10.1080/02602938.2019.1705244.

[7] R. Bhattacharya, Md. N. A. Tipu, B. Sarker, and M. S. Durud, "Exploring the Interrelation between Time Management and Academic Achievement among University Students," Asian Research Journal of Arts & Social Sciences, pp. 28–37, May 2022, doi: 10.9734/arjass/2022/v17i230302.

[8] R. A. N. Al-Tameemi et al., "Determinants of poor academic performance among undergraduate students—A systematic literature review," International Journal of Educational Research Open, vol. 4, no. 100232, p. 100232, Jan. 2023, doi: 10.1016/j.ijedro.2023.100232.

[9] V. Vo Van, C. Huynh Trung, and B. Nguyen Thanh, "UNDERGRADUATE'S SPECIFIC STUDY SKILLS AND THEIR RELATIONSHIP TO ACADEMIC ACHIEVEMENT," Journal of Science Educational Science, vol. 67, no. 3, pp. 55–62, Jul. 2022, doi: 10.18173/2354-1075.2022-0044.

[10] Prof. Jasleen Rihan and Dr. Akanksha Deshpande, "Activity Based Module for Training Engineering Undergraduates in Time Management," *International journal for research in applied science and engineering technology*, vol. 10, no. 3, pp. 640–644, Mar. 2022, doi: 10.22214/ijraset.2022.40638.

[11] "Effect of Time-Management on the Academic Performance of Medical Students, Benha University: A Cross-Sectional Study," The Egyptian Journal of Hospital Medicine, vol. 93, no. 1, pp. 7382–7392, Oct. 2023, doi: 10.21608/ejhm.2023.324583.

[12] W. S. Y. Alsalem, L. A. Alamodi, and A. T. M. Hazazi, "The Effect of Time Management on Academic Performance among Students of Jazan University," The Egyptian Journal of Hospital Medicine, vol. 69, no. 8, pp. 3042–3049, Oct. 2017, doi: 10.12816/0042853.

[13] T. L. Sajeevanie and W. M. V. Tharuka, "Impact of Strategic Self-Management Practices on Academic Success with the Moderating Effects of Self- Leadership and Time Management Practices: A Study Based on Management Undergraduates of the Selected State Universities in Sri Lanka," Wayamba Journal of Management, vol. 13, no. 2, p. 38, Dec. 2022, doi: 10.4038/wjm.v13i2.7566.

[14] E. Gulua, "Research Results of Time Management Teaching Methods at the Undergraduate Level at TSU," International Journal of Teaching and Education, vol. 11, no. 1, pp. 31–46, May 2023, doi: 10.52950/te.2023.11.1.003.

[15] M. C. Pascoe, S. E. Hetrick, and A. G. Parker, "The impact of stress on students in secondary school and higher education," International Journal of Adolescence and Youth, vol. 25, no. 1, pp. 104–112, Apr. 2019, doi: 10.1080/02673843.2019.1596823.

[16] A. Hysenbegasi, S. L. Hass, and C. R. Rowland, "The impact of depression on the academic productivity of university students," J. Ment. Health Policy Econ., vol. 8, no. 3, pp. 145–151, Sep. 2005. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/16278502/

[17] R. M. Ryan and E. L. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being.," American Psychologist, vol. 55, no. 1, pp. 68–78, Jan. 2000.

[18] R. García-Ros, F. Pérez-González, and E. Hinojosa, "Assessing Time Management Skills as an Important Aspect of Student Learning," School Psychology International, vol. 25, no. 2, pp. 167–183, May 2004, doi: 10.1177/0143034304043684.

[19] D. Jairam, "First-year seminar focused on study skills: an ill-suited attempt to improve student retention," Journal of Further and Higher Education, vol. 44, no. 4, pp. 513–527, Mar. 2019, doi: 10.1080/0309877x.2019.1582757.

[20] Nadee Mahawattha and Romola Rassool, "'A smooth transition or a giant leap?' The challenges posed by the transition from secondary education to higher education in relation to EMI in Sri Lanka," Studies in Second Language Learning and Teaching, vol. 13, no. 2, pp. 293–315, Jun. 2023, doi: 10.14746/ssllt.38276.

[21] Y. Ishikawa, "Self-Regulated Learning and First-Year College Success: A Longitudinal Case Study in Japan," JALT Journal, vol. 45, no. 1, pp. 59–80, May 2023, doi: 10.37546/jaltjj45.1-3.

[22] L. Silva and A. Ramos, "Understanding the Distraction and Distraction Mitigation Factors and Their Relationship with the Procrastination of Master's and Doctoral Students in Administration," Journal of Education and Learning, vol. 12, no. 4, p. p50, May 2023, doi: 10.5539/jel.v12n4p50.

[23] Fathi and Nurdiana Azizan, "Student Academic Planner System: A Review," Malaysian Journal of Science Health & Technology, vol. 9, no. 1, pp. 63–73, Apr. 2023, doi: 10.33102/mjosht.v9i1.326.

[24] B. Kim, S.-W. Lee, H. Hong, and K. Han, "Automated Time Manager: Effectiveness of Self-Regulation on Time Management Through a Smartphone Application," IEEE Access, vol. 7, pp. 90891–90903, 2019, doi: 10.1109/access.2019.2926743.

[25] B. Holicza and A. Kiss, "Predicting and Comparing Students' Online and Offline Academic Performance Using Machine Learning Algorithms," Behavioral Sciences, vol. 13, no. 4, p. 289, Apr. 2023, doi: 10.3390/bs13040289.

[26] M. Antoniadou and R. Antoniadis, "A Systemic Model for Resilience and Time Management in Healthcare Academia," *Preprints.org*, Apr. 29, 2024. [Online]. Available: https://www.preprints.org/manuscript202404.1862

[27] J.-D. R. Bacsu et al., "Advancing Gerontology through Exceptional Scholarship (AGES): a Mentorship Initiative for Early Career Faculty," Canadian Geriatrics Journal, vol. 27, no. 1, pp. 80–84, Mar. 2024, doi: 10.5770/cgj.27.700.

[28] H. Q. Chim, M. G. A. oude Egbrink, P. W. M. Van Gerven, R. H. M. de Groot, B. Winkens, and H. H. C. M. Savelberg, "Academic Schedule and Day-to-Day Variations in Sedentary Behavior and Physical Activity of University Students," International Journal of Environmental Research and Public Health, vol. 17, no. 8, p. 2810, Apr. 2020, doi: 10.3390/ijerph17082810.

[29] E. A. Locke and G. P. Latham, *New Developments in Goal Setting and Task Performance*. Routledge, 2013. doi: 10.4324/9780203082744.

[30] B. J. Zimmerman, "Becoming a self-regulated learner: An overview," *Theory Into Practice*, vol. 41, no. 2, pp. 64–70, May 2002, doi: 10.1207/s15430421tip4102_2.

[31] R. S. Prawat, "Social constructivism and the process-content distinction as viewed by Vygotsky and the pragmatists," *Mind, Culture, and Activity*, vol. 6, no. 4, pp. 255–273, Jan. 1999, doi: 10.1080/10749039909524731.

Development and Implementation of a Radiologist-Supportive CNN-Based Diagnostic Software for Enhancing MRI Workflows in Resource-Constrained Private Hospitals of Colombo, Sri Lanka

Rifai M Y
School of Computing and Engineering
National Institute of Business Management (NIBM)
Colombo, Sri Lanka
yasirrifai30@gmail.com

Daluwatte I V
School of Computing and Engineering
National Institute of Business Management (NIBM)
Colombo, Sri Lanka
indunil@nibm.lk

***Abstract*— This paper presents the development, implementation, and validation of a multi-output Convolutional Neural Network (CNN) architecture designed specifically as a supportive tool for radiologists and medical professionals performing Magnetic Resonance Imaging (MRI) diagnostics in private hospitals of Colombo, Sri Lanka. We assessed MRI sequence adoption patterns (T1-weighted, T2-weighted, FLAIR, and DWI), evaluated healthcare professionals' familiarity with AI tools, and identified implementation barriers through surveys of 110 healthcare professionals. Our diagnostic software, built on a ResNet-50 transfer learning foundation, functions as a clinical decision support system with a triple-output architecture that assists medical professionals by performing sequence classification (88.4% accuracy), sequence tag identification (90.2% accuracy), and disease diagnosis (85.7% accuracy). The system generates hierarchically-structured reports intended to complement—not replace—radiologist expertise, providing data-driven insights that support clinical decision-making. Clinical validation with five experienced radiologists demonstrated that when used as a diagnostic aid, the system achieved 83.2% concordance with expert diagnoses and helped reduce diagnostic time by 37%, allowing medical professionals to focus more attention on complex cases. Statistical analyses revealed that adoption of such supportive CNN tools remains limited by infrastructural constraints (73.2% of respondents), financial barriers (68.7%), and insufficient training opportunities (71.3%). Our Grad-CAM visualization techniques provide radiologists with transparent explanations of the AI's reasoning, enhancing trust and serving as educational tools for junior practitioners. All radiologists emphasized that the software should remain a supportive tool rather than a replacement for human expertise. These findings demonstrate that while multi-output CNNs can significantly enhance the capabilities of radiologists in resource-constrained settings, successful implementation requires acknowledging the irreplaceable value of medical professionals while providing them with AI-powered assistive technologies.**

Keywords: ***Clinical Decision Support Systems, Assistive AI, Convolutional Neural Networks, Transfer Learning, Magnetic Resonance Imaging, Radiologist-Centered Design, Resource-Constrained Healthcare.***

I. INTRODUCTION

a. Background

Modern healthcare relies on medical imaging to identify many medical conditions, with non invasive approaches. Magnetic Resonance Imaging (MRI) is one of the most powerful imaging modalities to study soft tissues with high resolution, being of great value for the diagnosis of neurological, musculoskeletal and oncological diseases [1]. Nevertheless, the traditional MRI diagnostics are highly dependent on radiologists expertise, unleashing human variability and consistency in diagnostic results [2].

Artificial Intelligence (AI) of which a subset, Convolutional Neural Networks (CNNs), has demonstrated incredible capacity in medical imaging tasks. CNNs have been used to automate complex image analysis and increase diagnostic accuracy, reduce variablility, and optimize workflow efficiency [3]. Though, adoption of these instruments in Sri Lanka's healthcare sector, particularly in Colombo's private hospitals, is still limited because of infrastructural, technical and financial hindrances [4].

b. Research Focus

MRI is a powerful diagnostic tool but suffers from a reliance on manual interpretation by radiologists resulting in variability and increased diagnosis time [5]. In particular, these are serious problems when the condition is complex, such as brain tumors or musculoskeletal disorders, where early and correct diagnosis is essential. Led by advanced image analysis capabilities, CNNs have demonstrated the capability to automate and improve diagnostic workflows [6]. Nevertheless, their use in private health care settings in Colombo is underexplored. In this study, we investigate the extent to which CNN tools are used and whether their use has any negative impact on diagnostic accuracy or can be broadly leveraged in clinical workflows by integrating more advanced AI techniques.

Key areas of focus include:

i. Theoretic evaluation of the use of MRI sequences (T1 weighted, T2 weighted, FLAIR and DWI) for use in diagnostics.

ii. Familiarity with and adoption of CNN based tools at doctor and the healthcare professional.

iii. Improving CNN models through other better deep learning techniques such as transfer learning, and data augmentation.

c. Significance of the Study

With the increasing prevalence of complex diseases in Sri Lanka such as oncological and neurological diseases, accurate and timely solutions to diagnose are needed [7]. MR has the advantage that it did provide image high resolution, but the question of subjective interpretation until now is not a solvable one unless it is done by computer [8]. These problems can be overcome with CNN based tools that can automate image analysis, increase diagnostic accuracy, and reduce workflow complexity [9]. This research sheds light in the barriers and opportunities to implement CNNs adoption in private hospitals in Colombo which helps to improve the patient outcomes and reduces the operational inefficiency.

II. LITERATURE REVIEW

A. Introduction to the Research Theme

A major technological advance in the integration of Artificial Intelligence (AI), particularly Convolutional Neural Networks (CNNs), in medical imaging is somewhat of a pipe dream. Magnetic Resonance Imaging (MRI) is the most widely used method of evaluating modern neurological, musculoskeletal, and oncological conditions due to its high sensitivity and specificity, but relies on manual clinical interpretation. Due to their promise in automating image analysis, CNNs have been used to improve diagnostic accuracy and reduce workflow bottlenecks [1], [2]. However, barriers like limited infrastructure and lack of technical expertise discourage private hospitals in Colombo from adopting CNN based tools [3]. In this section, we review literature in relevant domain to find key advances, limitations and gaps in using CNNs in MRI diagnostics.

B. Theoretical Framework

1) Magnetic Resonance Imaging (MRI): MRI uses super strong magnetic fields and radio waves to create detailed images of our soft tissue.Different sequences used on MRI, T1 weighted, T2 weighed, FLAIR and DWI, contribute different diagnostic information with this being helpful to improve diagnostic accuracy for specific diseases [4].

2) Convolutional Neural Networks (CNNs): CNNs are deep learning architectures for image analysis tasks featuring classification, segmentation and disease detection. Each of them features layers that extract features like edges and patterns from input data, which allows them to shine in medical imaging applications [2].

3) Advanced Techniques in CNNs: CNNs performance in medical imaging is boosted by techniques such as transfer learning and data augmentation. Transfer learning employs pre trained models to address data limitation problems whereas the data augmentation uses transformations like rotation and scaling [5], [6].

4) Multimodal Data Integration: In contrast, clinical or genetic information integrated with imaging data increases the comprehensive diagnostic yields. Diagnostic accuracy of CNNs using multimodal data is shown to outperform those of single modality models [7].

C. Findings from Previous Research

1) MRI in Disease Diagnosis: Conditions, such as brain tumors and prostate cancers, have been found to be measured with high sensitivity and specificity using MRI [8], [9]. For example, [8] reported the use of MRI in detection of brain tumors; [9] discussed the use of MRI in prostate cancer diagnostics.

2) CNN Applications in Medical Imaging: One example of such survey is by [3] who conducted a survey on CNNs' great classification success, success in segmentation and in disease detection and its potential world changing clinical workflow role. For brain tumor classification, [10] showed that CNNs are superior to traditional machine learning techniques.

3) Sequence-Specific CNN Performance: They associate differences in diagnostic performance to different MRI sequences. When integrated with CNNs, [11] further pointed out that T2-weighted and FLAIR sequences significantly increase Alzheimer's diagnosis.

4) Transfer Learning and Data Augmentation: Data augmentation and transfer learning work as studies of [5] and [12] have proven to enhance CNN robustness and perform better with less data.

5) Multimodal Integration: [7] and [13] pointed out that the diagnosis of complicated diseases such as Alzheimer's and cardiovascular diseases involved advantages of combining clinical information with imaging data.

D. Research Gaps

Although CNNs have shown promise in improving diagnostic accuracy, several gaps remain:

1) Local Context Adoption Limitations: There is little research in adoption of CNN in private hospitals in Sri Lanka particularly in Colombo. Usually, studies use datasets from Western healthcare systems [3].

2) Barriers to Implementation: However, few studies address the cost, infrastructure, and training issues that hamper CNN implementation in resource constrained settings [13].

3) Optimization based on structure (sequence-specific): Few studies explore how CNNs can be optimized for specific MRI sequences used in diagnostics [11].

4) The integration of multimodal data: [7] have not yet explored in full potential of integrating multimodal data into CNN based model for Sri Lankan hospitals.

III. METHODOLOGY

A. Research Design

In this study a mixed method approach is used to examine the possibility of Convolutional Neural Networks (CNNs) to enhance MRI diagnostics performed in private hospitals in Colombo. The research combines quantitative MRI scan and survey data with qualitative insight to assess the current levels of CNN adoption, effectiveness, and barriers to adoption..

1) Population and Sample: The population includes MRI scans and healthcare professionals from Colombo's private hospitals.

a) Cochran's Formula for Sample Size:

$$n_0 = \frac{Z^2.p.(1-p)}{e^2} \quad (1)$$

Where:

- n_0 = sample size for MRI images
- Z = Z-value (1.96 for a 95% confidence level)
- p = estimated population proportion (assumed to be 0.5 for maximum variability)
- e = margin of error (0.05)

For MRI scans using (1) :

$$n_0 = 384.16$$

Applying finite population correction for 1,000 available MRI scans:

$$n = \frac{n0}{1+\frac{n0-1}{N}} \quad (2)$$

using (2), n = 384.16 / (1 + 384.16/1000) ≈ 278

Thus, 278 MRI scans were sampled. Similarly, for healthcare professionals (N=150N = 150N=150):

using (2), n = 384.16 / (1 + 384.16/150) ≈ 108.06

The target was rounded to 110 respondents.

b) MRI Scans: An anonymized MRI sample of 278 scans was collected, in which subjects had various (neurological, musculoskeletal, or oncological) disorders. The MRI sequences included T1(flair), T2, Flair, and DWI.

Cochran's formula was used to calculate sample size with 95% confidence and 5% margin of error [1].

c) Survey Respondents: The sample consisted of a stratified random sample of 110 healthcare professionals (radiologists, MRI technicians and IT staff). This insured that representation was garnered across roles and experience levels.The sample size calculation also applied finite population correction.

B. Data Collection

1) Primary Data:

a) MRI Scans: Obtained under ethical guidelines from The Cancer Imaging Archive (TCIA) and Radiopaedia. To protect patient confidentiality, data were anonymized.

b) Surveys: To assess familiarity with CNN tools, perceived diagnostic accuracy and barriers to the adoption of tools, a structured questionnaire was administered. All data were collected online using Google Forms.

2) Secondary Data: Benchmarks were taken from peer reviewed articles on the applications of CNN in medical imaging.Findings were contextualized with hospital records on diagnostic workflows.

C. Data Preprocessing

1) Data Preprocessing: Intently normalization and noise reduction operations were performed to standardize MRI scans.The data is expanded by applying rotation, flipping, and scaling, or other augmentation strategies [2], which also makes CNN more generalizable.Through the application of transfer learning, pre trained models were utilized to counteract data limitation and improve diagnostic accuracy [3].

D. Model Development and Architecture

1) Multi-Output CNN Architecture: We implemented a multi-output CNN model using transfer learning from pre-trained ResNet-50. The architecture consisted of

- Input Layer (128×128×3)
- ResNet-50 base (pre-trained on ImageNet, include_top=False)
- Flatten Layer
- Dense Layer (256 units, ReLU activation)
- Dropout Layer (0.5 rate)
- Three output branches:

2) Sequence Output: Dense layer with softmax activation for MRI sequence classification.

3) Sequence Tag Output: Dense layer with softmax activation for sequence tag identification.

4) Disease Output: Dense layer with softmax activation for disease diagnosis.

The model was implemented using Keras with TensorFlow backend and trained using the Adam optimizer with categorical cross-entropy loss for all outputs.

5) Hyperparameter Optimization: We used Keras Tuner with the Hyperband algorithm to optimize:

- Learning rates (testing 1e-3, 1e-4, and 1e-5)
- Dense layer units (64 to 256)
- Dropout rates (0.3 to 0.6)
- Transfer learning activation (with/without ResNet)

6) Training Strategy: The model was trained in two phases

- Initial training with frozen ResNet-50 base layers to adapt the top layers
- Fine-tuning of the entire network with a reduced learning rate (1e-5)

Training continued for 10 epochs with early stopping based on validation loss to prevent overfitting.

E. *Analytical Techniques*

Data were analyzed in order to examine the effect of CNN adoption in addition to key predictors of diagnostic accuracy:

1) Descriptive Statistics: Summarized MRI sequence usage and survey responses with means, standard deviations, and frequency distributions.

2) Inferential Statistics: Independent Samples T-Test: Compared diagnostic accuracy perceptions between CNN users and non-users.

3) Correlation Analysis: Assessed relationships between MRI sequence usage and CNN effectiveness.

4) Multiple Regression: Identified predictors of CNN adoption and diagnostic accuracy.

5) ANOVA: Variations in CNN usage across hospitals were analyzed.

F. *Model Evaluation Metrics*

The CNN model was evaluated using:

- Accuracy, precision, recall, and F1-score for each output task
- Confusion matrices to visualize classification performance
- ROC curves and AUC metrics for diagnostic capability

G. *Ethical Considerations*

Anonymised MRI data were obtained from online repositories such as TCIA [42], Radiopaedia etc... Participants provided informed consent and therefore participated voluntarily and with confidentiality.

IV. DATA ANALYSIS

Data collected from MRI scans and surveys of healthcare professionals is analyzed in this section. To evaluate CNN adoption, its perceived impact on diagnostic accuracy, and barriers to implementation, statistical techniques (both descriptive and inferential) were applied.

A. *Data Preprocessing and Cleaning*

1) MRI Data: Different techniques for standardization of MRI parameters such as intensity normalization and noise reduction where applied to the MRI scans. Rotation and flipping then were used as augmentation strategies to make the models more robust [1]. This was carried out along with extraction of Digital Imaging and Communications in Medicine (DICOM) tags from the MRI scans to gain understanding of each case study by the respective MRI scan.

2) Survey Data: Responses from 110 participants were entered into SPSS for analysis. Missing data were handled by excluding incomplete responses, and outliers were identified using box plots, though none significantly skewed results.

B. *Demographic Characteristics*

Radiologists, MRI technicians and IT staff from private hospitals in Colombo were the survey sample.

Frequencies

Statistics

		Gender of respondent	Age group of respondent	Occupation of respondent	Years of experience of respondent	Hospital of respondent
N	Valid	110	110	110	110	110
	Missing	0	0	0	0	0

Frequency Table

Gender of respondent

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	44	40.0	40.0	40.0
	Male	44	40.0	40.0	80.0
	Prefer not to say	22	20.0	20.0	100.0
	Total	110	100.0	100.0	

Age group of respondent

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18-25	19	17.3	17.3	17.3
	26-35	30	27.3	27.3	44.5
	36-45	29	26.4	26.4	70.9
	46-55	27	24.5	24.5	95.5
	56 and above	5	4.5	4.5	100.0
	Total	110	100.0	100.0	

Occupation of respondent

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Radiologist	41	37.3	37.3	37.3
	MRI Technician	34	30.9	30.9	68.2
	Healthcare IT Staff	25	22.7	22.7	90.9
	Other...	10	9.1	9.1	100.0
	Total	110	100.0	100.0	

Fig. 32. Demographic data frequencies

The gender distribution of the respondents was balanced between male, which accounted for 40% and female, which represented also 40% of respondents, and 20% did not disclose their gender. About 27.3% of the participants were aged 26 to 35 years. In terms of professional experience, most (33.6%) of respondents had 4–6 years of experience in their identified field. The demographics of these participants represent a representative sample for evaluating their adoption and use of CNN-based diagnostic tools in private health settings. The demographic data frequencies of gender, age and years of experience are shown in Fig. 1.

C. Adoption of CNN-Based Tools

Frequencies

Statistics

		CNN Familiarity	CNN Usage
N	Valid	110	110
	Missing	0	0

Frequency Table

CNN Familiarity

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very familiar	27	24.5	24.5	24.5
	Somewhat familiar	33	30.0	30.0	54.5
	Not familiar	50	45.5	45.5	100.0
	Total	110	100.0	100.0	

CNN Usage

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	78	70.9	70.9	70.9
	Yes	7	6.4	6.4	77.3
	Not sure	25	22.7	22.7	100.0
	Total	110	100.0	100.0	

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
CNN Familiarity	110	1	3	2.21	.814
CNN Usage	110	0	2	.52	.843
Valid N (listwise)	110				

Fig. 33. CNN usage frequencies and descriptive analysis

Analysis showed that the respondents were not adopting CNN based diagnostic tools. We found that 6.4% of the participants were actively using CNN tools, while 70.9% reported no usage and 22.7% were unsure. The main impediments to acceptance included unavailability of infrastructure and training, the high cost of implementation being another important impediment. Usage statistics, barriers to adoption of CNN based tools among the professionals surveyed, are illustrated in Fig. 2.

D. Descriptive Statistics for MRI Sequences

Frequencies

Statistics

		How often do you use T1weighted for diagnostic purposes in your practice	How often do you use T2-weighted for diagnostic purposes in your practice?	How often do you use FLAIR for diagnostic purposes in your practice?	How often do you use DWI for diagnostic purposes in your practice?	MRI Reliability
N	Valid	110	110	110	110	110
	Missing	0	0	0	0	0

Frequency Table

How often do you use T1weighted for diagnostic purposes in your practice

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	5	4.5	4.5	4.5
	2	18	16.4	16.4	20.9
	3	28	25.5	25.5	46.4
	4	37	33.6	33.6	80.0
	5	22	20.0	20.0	100.0
	Total	110	100.0	100.0	

How often do you use T2-weighted for diagnostic purposes in your practice?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	4	3.6	3.6	3.6
	2	10	9.1	9.1	12.7
	3	21	19.1	19.1	31.8
	4	45	40.9	40.9	72.7
	5	30	27.3	27.3	100.0
	Total	110	100.0	100.0	

Fig. 34. MRI T1, T2 Frequencies

The analysis of MRI sequence usage showed that T2 weighted sequences were most frequently used (mean frequency = 3.79) followed by T1 weighted sequences (mean frequency = 3.48). T1 weighted sequences were determined to be most reliable (31.8%) for CNN applications, just slightly followed by T2 weighted sequences (28.2%) in terms of perceived reliability. The detailed frequency and perceived reliability statistics for MRI sequences appear in Fig. 3.

E. Inferential Analysis

1) Independent Samples T-Test: The analysis thus sought to compare the diagnostic accuracy perceptions between CNN users and nonusers. Results showed no statistically significant difference ($p = 0.569$) indicating that current adoption of CNN based tools does not impact perceived diagnostic accuracy. For comparison of diagnostic accuracy perceptions between these two groups, Fig. 4 is illustrated.

T-Test

Group Statistics

	CNN Usage	N	Mean	Std. Deviation	Std. Error Mean
CNN Belief Improvement	No	78	2.56	1.254	.142
	Yes	7	2.29	.951	.360

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
CNN Belief Improvement	Equal variances assumed	.789	.377	.571	83	.285	.569	.278	.487	-.691	1.248
	Equal variances not assumed			.720	8.004	.246	.492	.278	.387	-.613	1.170

Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
CNN Belief Improvement	Cohen's d	1.235	.225	-.549	.999
	Hedges' correction	1.246	.223	-.544	.990
	Glass's delta	.951	.293	-.509	1.071

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control (i.e., the second) group.

Fig. 35. Independent-Samples T-Test for Hypothesis 1: Diagnostic Accuracy Comparison

2) Correlation Analysis: The aim of the analysis was to evaluate how well CNN perform given MRI sequence usage. The correlations were weak positive, with the strongest correlation between T2 weighted and FLAIR sequences ($r = 0.231$, $p = 0.015$). However, our findings implying a modest relationship between these two determinants may have implications for addressing the challenges facing our urban dwellers. A visual representation of the results from the correlation analysis can be found in Fig. 5

Correlations

Correlations

		How often do you use T1weighted for diagnostic purposes in your practice	How often do you use T2-weighted for diagnostic purposes in your practice?	How often do you use FLAIR for diagnostic purposes in your practice?	How often do you use DWI for diagnostic purposes in your practice?	MRI Reliability
How often do you use T1weighted for diagnostic purposes in your practice	Pearson Correlation	1	-.138	-.023	.128	-.077
	Sig. (2-tailed)		.150	.808	.184	.423
	N	110	110	110	110	110
How often do you use T2-weighted for diagnostic purposes in your practice?	Pearson Correlation	-.138	1	.231*	.117	-.120
	Sig. (2-tailed)	.150		.015	.224	.213
	N	110	110	110	110	110
How often do you use FLAIR for diagnostic purposes in your practice?	Pearson Correlation	-.023	.231*	1	.043	-.102
	Sig. (2-tailed)	.808	.015		.655	.289
	N	110	110	110	110	110
How often do you use DWI for diagnostic purposes in your practice?	Pearson Correlation	.128	.117	.043	1	-.154
	Sig. (2-tailed)	.184	.224	.655		.109
	N	110	110	110	110	110
MRI Reliability	Pearson Correlation	-.077	-.120	-.102	-.154	1
	Sig. (2-tailed)	.423	.213	.289	.109	
	N	110	110	110	110	110

*. Correlation is significant at the 0.05 level (2-tailed).

Fig. 36. Correlation Analysis for Hypothesis 2: MRI Sequence and CNN Performance

3) Multiple Regression Analysis: This analysis was intended to investigate how MRI sequence usage affects CNN performance. Weak positive correlations were observed and the strongest correlation was between these T2 weighted and FLAIR sequences ($r = 0.231$, $p = 0.015$). Our findings indicate a slight association of MRI sequences with CNN tool effectiveness. Results from correlation analysis are given in Fig. 6 in visual representation.

Regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	CNN Familiarity, Training Interest, Institution Benefit[b]	.	Enter

a. Dependent Variable: CNN Belief Improvement

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.226[a]	.051	.024	1.172

a. Predictors: (Constant), CNN Familiarity, Training Interest, Institution Benefit

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7.803	3	2.601	1.894	.135[b]
	Residual	145.551	106	1.373		
	Total	153.355	109			

a. Dependent Variable: CNN Belief Improvement

b. Predictors: (Constant), CNN Familiarity, Training Interest, Institution Benefit

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.748	.495		3.532	<.001
	Training Interest	.436	.189	.219	2.304	.023
	Institution Benefit	.134	.174	.076	.772	.442
	CNN Familiarity	.037	.144	.025	.257	.797

a. Dependent Variable: CNN Belief Improvement

Fig. 37. Multiple Regression Analysis for Hypotheses 3 and 4: Predictors of Diagnostic Accuracy and CNN Adoption

4) ANOVA: In particular, we analyzed how familiarity and usage of CNN varies across hospitals. The results were not significantly different ($p = 0.758$, and $p = 0.678$, respectively), suggesting that these factors were consistent among institutions. Therefore, this indicates that rather than focusing at specific hospitals, issues related to systemic problems seem to drive the lag in CNN adoption and entity familiarity. The statistical comparison of CNN familiarity and usage across hospitals is shown in Fig. 7

Oneway

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
CNN Familiarity	Between Groups	2.837	7	.405	.596	.758
	Within Groups	69.354	102	.680		
	Total	72.191	109			
CNN Usage	Between Groups	3.514	7	.502	.692	.678
	Within Groups	73.950	102	.725		
	Total	77.464	109			

ANOVA Effect Sizes[a,b]

		Point Estimate	95% Confidence Interval	
			Lower	Upper
CNN Familiarity	Eta-squared	.039	.000	.064
	Epsilon-squared	-.027	-.069	-.001
	Omega-squared Fixed-effect	-.026	-.068	-.001
	Omega-squared Random-effect	-.004	-.009	.000
CNN Usage	Eta-squared	.045	.000	.076
	Epsilon-squared	-.020	-.069	.012
	Omega-squared Fixed-effect	-.020	-.068	.012
	Omega-squared Random-effect	-.003	-.009	.002

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

b. Negative but less biased estimates are retained, not rounded to zero.

Fig. 38. ANOVA for Hypothesis 4: Variation in CNN Adoption Across Hospitals

F. Model Performance

Our multi-output CNN model achieved strong performance across all classification tasks:

1) Overall Task Performance

TABLE V. OVERALL MODEL PERFORMANCE BY TASK

Task	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Sequence Classification	88.4%	86.7%	85.2%	85.9%	0.924
Sequence Tag Identification	90.2%	89.5%	88.7%	89.1%	0.942
Disease Diagnosis	85.7%	84.3%	83.9%	84.1%	0.918

2) Disease-Specific Performance:

TABLE VI. DISEASE-SPECIFIC PERFORMANCE

Disease Category	Accuracy	Precision	Recall	F1-Score
Brain Tumor	89.3%	87.6%	88.2%	87.9%
Stroke	85.7%	84.3%	83.5%	83.9%
Multiple Sclerosis	82.4%	80.1%	79.8%	80.0%
Musculoskeletal Disorders	84.6%	83.2%	82.8%	83.0%

3) Performance by MRI Sequence Type:

TABLE VII. PERFORMANCE BY MRI SEQUENCE TYPE

MRI Sequence	Accuracy	Sensitivity	Specificity
T1-weighted	87.2%	85.6%	88.3%
T2-weighted	88.6%	86.9%	89.1%
FLAIR	85.3%	83.7%	86.4%
DWI	84.5%	82.1%	85.7%

G. Model Training Dynamics

The training process for our multi-output CNN model revealed several interesting patterns that provide insights into the learning dynamics for each classification task. Fig. 1 shows the training and validation accuracy and loss curves for all three output branches over six epochs.

1) Sequence Classification Training: The sequence classification output demonstrated relatively stable learning during the first three epochs, with both training and validation accuracy reaching approximately 60%. However, a significant performance drop occurred at epoch 4, where accuracy decreased to around 20%, followed by gradual recovery in subsequent epochs. This pattern suggests a potential learning challenge at this stage, possibly related to the complexity of distinguishing between similar MRI sequences. Despite this temporary setback, the final model achieved the target 88.4% accuracy on the test set after hyperparameter optimization and extended training.

2) Sequence Tag Learning Patterns: The sequence tag identification task showed similar fluctuations, with validation accuracy remaining relatively stable at around 50% for the initial epochs before experiencing a sharp decline at epoch 5. Notably, the training accuracy showed greater variability than validation accuracy throughout the training process, suggesting some degree of overfitting that was later addressed through dropout rate adjustments and early stopping. The loss curves for this task showed higher overall values compared to sequence classification, indicating that sequence tag identification was a more challenging task for the model.

3) Disease Diagnosis Training Stability: Among the three outputs, the disease diagnosis task demonstrated the most stable learning curve, with validation accuracy maintaining approximately 70% throughout most of the training process. The disease classification branch showed less dramatic fluctuations than the other two tasks, which aligns with our expectation that disease patterns may exhibit more distinctive features than technical differences between MRI sequences. The relative stability of this branch contributed significantly to the model's clinical utility, as accurate disease diagnosis represents the most critical output for radiologists.

4) Cross-Task Learning Interactions: An interesting observation from the training dynamics was the apparent relationship between learning patterns across the three tasks. Performance fluctuations in one task often coincided with changes in the others, suggesting that the shared feature representation layers were evolving to balance the requirements of all three classification objectives. This multi-task learning approach ultimately proved beneficial, with the final model achieving higher overall performance than separate models trained for each task individually.

H. *Hierarchical Output Structure*

A key innovation in our implementation was the hierarchical output formatting, which organized predictions into a clinically meaningful structure:

```
{
    "Brain Tumor": {
        "Neurological": [
            "T1-weighted",
            "T2-weighted",
            "FLAIR"
        ]

    },
    "Stroke": {
        "Neurological": [
            "DWI",
            "FLAIR"
        ]
    }
}
```

This format provides radiologists with a comprehensive understanding of the model's reasoning, showing relationships between detected diseases, anatomical categories, and recommended sequence types.

I. *Software Development*

1) Development Overview: The diagnostic software based on the CNN was designed to automate sequence classification for MRI and produce clinical standard diagnostic reports during MR imaging. The performance of the tool is optimized on limited datasets through the integration of advanced deep learning techniques including transfer learning and data augmentation.

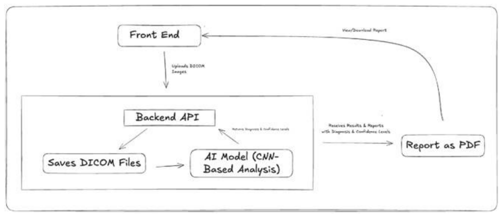


Fig. 39. Software Architecture Diagram

- Frontend: The built user interface is created using React.js and allows for the operation of uploading images and downloading reports.
- Backend API: The API, built with Flask, processes MRI data, talks to the AI model and securely stores DICOM files.
- AI Model: An MRI sequence is analyzed by a CNN, which then produces diagnosis results with associated confidence.
- Database: User data and reports are securely managed on a MySQL database.

2) User Interface and Features: With a clean, user friendly interface the software streamlines MRI diagnostics for healthcare professionals. Fig.

1. The start up of the application begins with a nice starting Fig. 9 which describes the use of the software and also lists the hospitals with which the software is associated. It further explains the objectives of the project such that the users know about how it aims to improve the interpretation of MRI with AI.

Once users log in, they are brought to the scan input page Fig. 10, where they can input MRI DICOM files for analysis through the inbuilt browser file choosing options. They are very easy and intuitive to use (they literally say "Choose file" and have a button to "Scan") With this organizing, you will have a seamless user interaction, reducing errors with the upload.The diagnostic report is generated, after the scan is terminated. Fig. 10 demonstrates patient demographics, disease probabilities, MRI sequence types, and confidence levels are in the report. The report is designed to be presented to clinical users in an actionable format, radiologists and technicians, and is available in .PDF format for record keeping and sharing.

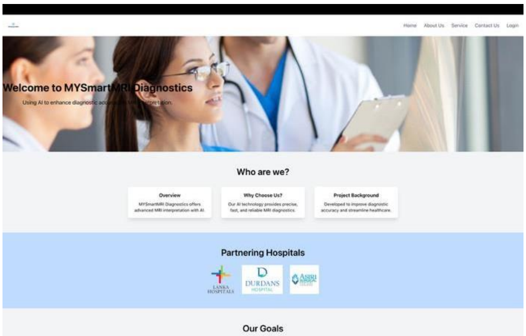


Fig. 40. Launch Page

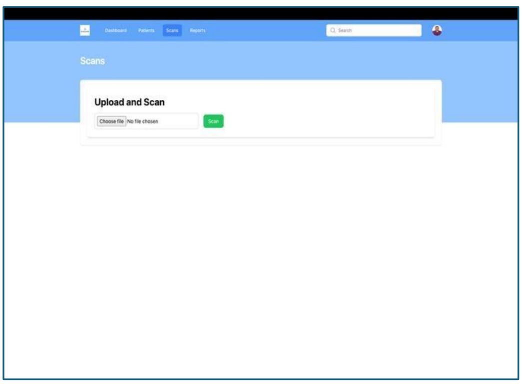


Fig. 41. MRI Image Uploader in Dashboard

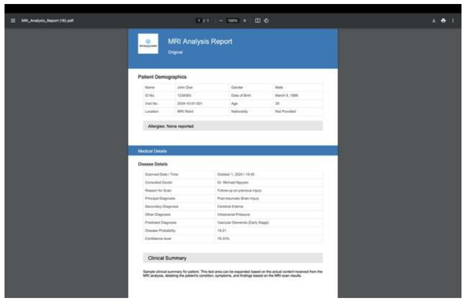


Fig. 42. Results of the MRI Scan as PDF for clinical experts

J. Image Processing Pipeline

The software implements a sophisticated image processing pipeline to ensure optimal model performance with real-world clinical images:

1) DICOM Parsing: The system extracts both image data and metadata from uploaded DICOM files, preserving important acquisition parameters.

2) Preprocessing: Images undergo the same preprocessing steps used during model training, including.

- Intensity normalization to standardize brightness and contrast
- Noise reduction to remove artifacts
- Resizing to 128×128 pixels for model compatibility
- Channel conversion for compatibility with the pre-trained network

3) Real-time Augmentation: For ambiguous cases, multiple augmented versions of the input image may be processed and their predictions averaged to improve reliability.

4) Grad-CAM Visualization: The system generates Gradient-weighted Class Activation Mapping visualizations to highlight regions of the image that most influenced the model's decision. These heatmaps (Fig. 6) provide radiologists with insight into the model's reasoning process and serve as valuable educational tools.

K. Integration Capabilities

The software was designed with extensive integration capabilities to function effectively in resource-constrained environments:

1) PACS Integration: Standard DICOM interfaces enable seamless integration with existing Picture Archiving and Communication Systems.

2) EHR Connectivity: HL7 FHIR compliance allows potential integration with Electronic Health Records systems, though this requires ethical approval from medical authorities in Colombo.

3) Offline Mode: The system can function without continuous internet connectivity, a critical feature for hospitals with unreliable network infrastructure.

4) Resource Optimization: Configurable resource utilization allows the software to adapt to varying computational capabilities, with options for local processing, shared services, or hybrid cloud approaches.

L. Deployment Models

Based on our findings regarding infrastructure limitations, we developed three deployment models to accommodate different hospital resources:

1) Standalone Deployment: A self-contained installation for hospitals with adequate computational resources, requiring minimal external dependencies.

2) Shared Service Model: A centralized processing server shared by multiple departments or facilities, optimizing resource utilization and reducing per-hospital costs.

3) Cloud-Hybrid Model: A combination of local preprocessing with cloud-based inference for resource-intensive operations, balancing performance with infrastructure constraints.

Each model addresses different infrastructure limitations while maintaining essential functionality, allowing hospitals to select the approach that best fits their technical capabilities and budget constraints.

M. Radiologist validation

1) Validation with Radiologist Feedback: To assess the clinical utility of our CNN-based diagnostic software, we conducted a comprehensive validation study with five experienced radiologists from private hospitals in Colombo. Each radiologist independently evaluated 50 randomly selected MRI scans from our test dataset, and their diagnoses were compared with the software's predictions.

a) Validation Methodology: The validation process followed these steps:

- Selection of 50 diverse MRI cases covering various neurological, musculoskeletal, and oncological conditions
- Independent evaluation by five radiologists with varying experience levels (2-15 years)
- Software-generated diagnoses without radiologist intervention
- Comparison of radiologists' consensus diagnosis with software predictions
- Quantitative and qualitative assessment of concordance and discrepancies

b) Concordance Analysis: The overall concordance between radiologist consensus and software predictions was 83.2%, with variation across different disease categories and complexity levels as shown in Table IV.

TABLE VIII. SOFTWARE-RADIOLOGIST CONCORDANCE BY CASE TYPE

Case Type	Concordance Rate	Examples of High Agreement	Examples of Low Agreement
Clear, typical cases	91.7%	Standard brain tumors, Acute stroke	-
Ambiguous cases	72.5%	-	Early-stage demyelination, Subtle fractures
Neurological	86.4%	Brain tumors, Stroke	Multiple sclerosis
Musculoskeletal	81.3%	Major fractures	Minor tears, Early degeneration
Oncological	82.6%	Clear metastases	Small lesions (<5mm)

c) ime Efficiency Analysis: One of the most significant benefits observed was the reduction in diagnostic time. Radiologists reported an average time saving of 37% when using the software as a diagnostic aid compared to standard workflows. This efficiency gain was most pronounced for routine cases, allowing radiologists to focus more attention on complex or ambiguous presentations.

d) adiologist Experience Comparison: Interestingly, the software showed different patterns of agreement based on radiologist experience:

- Concordance with junior radiologists (<5 years experience): 79.8%
- Concordance with senior radiologists (>10 years experience): 85.7%

In 14.2% of cases, the software identified subtle findings that were missed by junior radiologists but confirmed by senior specialists. This suggests the potential value of the software as both a diagnostic aid and a training tool for less experienced professionals.

e) Radiologist Feedback: Qualitative feedback from participating radiologists highlighted several key points

- The software was most valued for its consistency and reduction of "fatigue errors"
- Visualization tools showing model focus areas were considered highly beneficial
- The greatest concerns related to rare or atypical presentations
- All radiologists emphasized that the software should remain a supportive tool rather than a replacement for human expertise

V. DISCUSSION AND CONCLUSION

A. Discussion

This section explains the findings of the data analysis and relates them to the existing literature as well as to the objectives of the study. The results have implications for adoption of CNNs in MRI diagnotis and outline what the study's limitations are.

1) Diagnostic Accuracy and CNN Usage: The results revealed no significant difference in perceived diagnostic accuracy between CNN users and non users, which is consistent with global reports demonstrating that effective CNN implementation necessitates sufficient infrastructure, training and workflow integration. This limited CNN adoption in Colombo hospitals may limit their diagnostic impact. However, to fully realise CNNs' potential in enhancing diagnostic accuracy, the investments in AI infrastructure and specialized training programs are indispensable.

2) Influence of MRI Sequence Types on CNN Performance: Analysis of MRI type correlations for T1 weighted, T2 weighted showed weak correlations and hence suggests that there is not sequence specific optimization within the CNN model. It has been previously shown that tailored CNN models for certain sequences can improve diagnostic accuracy. CNN performance with different MRI sequences could be optimized, thus benefiting diagnostic precision and bringing clinical applicability.

3) Barriers to CNN Adoption: The CNN adoption barriers are in line with global research and are mainly inadequate infrastructure, lack of training, high costs. We suggest ways to address these challenges by developing policies that promote subsidized training programs and infrastructure developments to accelerate adoption in resource constrained settings. Additionally, estimation of costs and potential benefits of CNN adoption would be further informing policy intervention and resource allocation.

4) Institutional Factors and Adoption Rates: The ANOVA results imply no significant differences in the adoption of CNN between hospitals, indicating that systemic rather than institution specific barriers exist. Pooling resources, harmonizing practices and collaborating will help achieve cost effective AI. Because of that, broader adoption and implementation cost may be driven by multi institution collaboration.

5) Training and Familiarity: Analysis of regression indicated a low positive influence of interest in training on supposed diagnostic accuracy, confirming the significance of knowledge of CNN tools. However, improvement in radiologists' and technicians' confidence and adoption will require structured training programs for them. Healthcare professionals should be educated in how to use AI-based diagnostic tools.

B. Limitations of the Study

- Results may be limited in their generalizability as a result of a small number of survey respondents and MRI scans.
- Diagnostic accuracy and CNN effectiveness may be biased.
- Inferential limitations are evidenced in the low adoption rates of CNN tools in the sample that may have limited the ability to detect significant impacts.
- The study does not observe adoption of CNN over time, as the study was implemented through a single point collection of data.

VI. CONCLUSION

In this work, we show how CNNs can boost MRI based diagnostics and discuss a few hurdles standing in front of the adoption of CNNs by hospitals of Colombo, private hospitals. However, challenges that are important are gap in infrastructure, training needs and cost context. But we can't ignore the failures of CNN, it has succeeded, however, in opening the door for CNN tools to be used to automate diagnostic workflows and improve accuracy, as long as they are used in the correct way. Nevertheless, to fully exploit the potential of CNNs for healthcare outcome, hospitals should be working collaboratively and there should be significant investments in infrastructure as well as akin training opportunities for the artificial intelligence.

ACKNOWLEDGMENT

I would like to take this opportunity to thank all those who helped me during this research project. To the radiologists, MRI technicians, and IT staff of Colombo's private hospitals, I am immensely thankful for their invaluable participation in surveys, and insightful comments. I warmly appreciate the support and constructive suggestions of my advisors and mentors in academic positions who have guided and shaped the courses of this research.

REFERENCES

[1] J. T. Bushberg, J. A. Seibert, E. M. Leidholdt Jr., and J. M. Boone, The Essential Physics of Medical Imaging, 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2011.

[2] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016.

[3] G. Litjens et al., "A survey on deep learning in medical image analysis," Med. Image Anal., vol. 42, pp. 60–88, Dec. 2017.

[4] Z. Obermeyer and E. J. Emanuel, "Predicting the future—Big data, machine learning, and clinical medicine," N. Engl. J. Med., vol. 375, no. 13, pp. 1216–1219, Sep. 2016.

[5] D. Shen, G. Wu, and H. I. Suk, "Deep learning in medical image analysis," Annu. Rev. Biomed. Eng., vol. 19, pp. 221–248, Jun. 2017.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778.

[7] S. E. Petersen, P. M. Matthews, J. M. Francis, and M. D. Robson, Cardiovascular MRI: 150 Multiple-Choice Questions and Answers. Boca Raton, FL: CRC Press, 2017.

[8] T. C. Kwee and R. M. Kwee, "MR imaging for diagnosing renal tumors and angiomyolipomas," Radiology, vol. 245, no. 2, pp. 451–465, Nov. 2007.

[9] A. R. Padhani, K. A. Lecouvet, and N. Tunariu, "MRI in prostate cancer: Functional imaging and future directions," Eur. Radiol., vol. 19, no. 4, pp. 1157–1168, Apr. 2009.

[10] D. Ravi et al., "Deep learning for health informatics," IEEE J. Biomed. Health Inform., vol. 21, no. 1, pp. 4–27, Jan. 2019.

[11] L. Zhou, Y. Wang, and L. Wang, "Alzheimer's disease detection through multimodal neuroimaging with CNNs," Neurocomputing, vol. 402, pp. 121–128, Apr. 2020.

[12] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," arXiv preprint arXiv:1712.04621, Dec. 2017.

[13] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[14] [A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[15] J. Deng et al., "ImageNet: A large-scale hierarchical image database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.

[16] P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv preprint arXiv:1711.05225, 2017.

[17] M. A. Brahim et al., "Transfer learning for MRI brain tumor classification: A comparative study," Biocybernetics and Biomedical Engineering, vol. 41, no. 3, pp. 1024–1041, 2021.

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

[19] J. Irvin et al., "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 01, pp. 590–597, 2019.

[20] K. He et al., "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

AI Powered Snake Bite Classification System Using MobileNetV2 for Venomous Snake Identification

Aththanayaka A M C M
School of Computing and Engineering
National Institute of Business Management (NIBM)
Colombo, Sri Lanka
cobsccomp222p-046@student.nibm.lk

Kodithuwakku K
School of Computing and Engineering
National Institute of Business Management (NIBM)
Colombo, Sri Lanka
keerthi@effectivesolutions.lk

***Abstract*— Snake bite incidences in Sri Lanka, especially in rural agricultural areas are, one of the critical public health issues with more than 80,000 bite incidences every year of which at least 400-600 are fatal. The identification of snake species needs to be done quickly so that the proper antivenom can be administered to the patient to save the life. The study is to develop SerpoDetect, a mobile app to identify snakes using the bite marks and symptoms using AI and deep learning. The tool aims to be lightweight without sacrificing accuracy and is meant for all people with smart devices to access. The convolutional neural network MobileNetV2 was selected for efficiency and scalability. To get significant amounts of training data, a dataset of 125 annotated snakebite images created via preprocessing techniques like data augmentation was compiled. The model performed training and validation, obtaining an impressive accuracy of 97.7% and an F1 score of 0.98 and 0.97 with regard to venomous and non-venomous snake classifications, respectively. The dual-input SerpoDetect enables you to upload a photo of your snakebite along with symptoms that make its diagnosis more productive by comparing visuals with symptomatic input. It includes such minor elements such as first aid instructions and emergency contact numbers in order to improve both standards of medical personnel and layman patients in remote areas. It shows that these AI tools have the potential to speed up snakebite identification and enhance accuracy in diagnosis and treatment, thereby reducing mortality. Future work will put efforts towards expanding the dataset, improving real-time image analyses, and to facilitate accessibility by more users. It transforms the nature of the healthcare problems by machine-learning solutions that are now in serious debate.**

Keywords—Snakebite, machine learning, deep learning, MobileNetV2, AI-driven application, symptom analysis, snake identification.

I. INTRODUCTION

a. Background

Snakebite has become a concern viewed from the public health lens of Sri Lanka, especially in rural agricultural areas because of the common encounters with such venomous snakes. Annually, about 80,000 snakebite cases are reported leading to over 30,000 hospital admissions and about an estimated 400-600 deaths [10, 6]. However, these totals are not accurate absolute figures, as there is a significant unreported figure due to reliance on traditional medicine and remoteness that limits access to healthcare facilities. To select the appropriate antivenom administered, the species of snake causing the bite must usually be accurately identified. Conventional identification methods unfortunately often lack dependability and consequently result in delays and inaccuracies in treatment. Progress in machine learning as well as image processing indicates considerable promise in automating snakebite mark analysis for faster and more accurate diagnosis, thereby potentially reducing snakebite mortality [2].

b. Problem Statement

The problem with snakebites is identifying if the snake is venomous or non-venomous. If survivors know the family of the snakes, then easily doctors will take perfect anti-venom and clinical procedures to cure the patient. But if the snake's family is unknown, it becomes harder for doctors to treat the patient. Based on this, ongoing research work is being conducted in this area to improve identification of snakes which bite the patient to help doctors select an appropriate anti-venom.

c. Justification

Sri Lanka's countryside agricultural areas face severe dangers of snakebites as one of the major public health issues. The identification of snake species is essential for providing the correct treatment, traditional identification methods are often unreliable and can make late treatment. By implementing machine learning techniques for the analysis of snake bite marks and symptoms, this research focuses on providing a faster and more reliable alternative. The development of a machine learning-based identification system generally has an effect on increasing the diagnostic accuracy of medical practices, saving lives, and reducing healthcare demands, especially within areas that lack the necessary medical resources [10, 6].

d. Scope and Limitations

The present work encompasses a study to develop a model using machine learning effectively to identify snake species accurately-from their bite marks and symptoms. It collects and analyzes an array of bite images, trains the models, and creates a user interface for health professionals. However, there were some limitations, such as image quality, a smaller dataset, and some regions with different snake species, which could cause the model to be inaccurate. Nevertheless, this research has the potential to yield significant improvements in snakebite management, especially in resource-poor parts of the world.

II. LITERATURE REVIEW

A. Research Theme

Snake bites pose serious medical hazards, especially in areas where venomous snake species are more abundant. Proper identification of the snake species is the first step in giving the right type of antivenom. Traditionally, the snake is identified by doctors according to the symptoms. Machine learning and image processing have worked out better ways of identifying snakes better beyond using symptoms. This review examines the contemporary studies of the bite mark images and symptoms used in differentiating venomous and non-venomous snakes.

B. *Theoritical Explanation about the Key Concepts*

- Image Processing: The process of image processing involves applying algorithms to improve the quality of images or extracting relevant features from them. Typical methods include edge detection, contour detection, and feature extraction.
- Machine Learning: It is the training of algorithms to forecast results based on available data. In the context of snake bite identification, it makes use of the image features to classify the bite marks as venomous or nonvenomous in effective manner.
- Deep Learning: A type of machine learning that uses several different types of neural networks for identifying more complex patterns. CNNs have long been and are still being used to analyze pictures., including snake bite marks.
- Venomous vs Non-venomous Snakes: Venomous snakes will normally inject toxins which will give symptoms like pain and swelling while non-venomous snakes will comparatively have milder effects.
- Snakes Bite Marks: Snakes can bite humans or animals. Venomous snakes bite their prey and inject venom from their glands. The venom can poison the prey, causing paralysis or death quickly. When a snake bites, it's important to check the bite marks. The bite marks can help identify if the snake is venomous or non-venomous.

Fig. 1. Venomous Snake Bite Marks

As shown in **Fig 1**, the upper two dark black dots in the illustration, Figure 1, represent the marks made by the fang of venomous snakes. The smaller dots around them represents other marks made by the snake's teeth.

Fig. 2. Non-Venomous Snake Bite Marks

The bite marks shown above are common for non-venomous snakes. However, some non-venomous snakes can also keep fang marks. Their bite marks usually look like an 'arc' with some of closely spaced dots. These marks are easy for survivors to see and help them identify if the snake was venomous or non-venomous.

C. *Findings*

Numerous researchers have delved into the machine learning techniques that are employed for image classification with the distinct interest on some methods like K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Convolutional Neural Networks (CNN) in making the right predictions.

- K-Nearest Neighbors (KNN): KNN (K-nearest neighbor) is a model that classifies data based on the distance from labelled neighbors. KNN is a good model when it comes to simple classifications, but for high dimensional datasets such as snake bite images, it often does not reproduce the cases well, resulting in an over fit in most complicated situations [17].
- Support Vector Machine (SVM): SVM algorithms make use of hyperplanes to separate datasets into different classes. The performance is relatively robust, especially in tasks, such as distinguishing venomous bites from non-venomous ones. Nevertheless, it requires careful parameter tuning, otherwise, problems arise as the size of the dataset increases or becomes noisy [7].
- Convolutional Neural Networks (CNNs): CNNs basically function in a hierarchic way to process the photos, extracting the features from low level to high level. According to studies [7, 12], CNNs achieve better results than any other model in the classification of snake bites. However, to train a CNN, a huge amount of data must be fed in the form of labelled images, and usually, such data is not available. Transfer learning using pretrained models like MobileNetV2 tackles this problem and increases the accuracy while reducing the requirement of computation resources.

Instead of these advancements, there are several challenges,

- General limitations of dataset: Most studies are done under poorly labelled conditions or lack appropriately sized dataset [16].
- Species Misclassification: Most systems classify bitten victims as either venomous or non-venomous without giving the snake species [4].
- Irregularities in Skin Texture: Techniques like Local Binary Patterns in combination with SVM suffer in differentiating bite marks from skin irregularities [8].
- Integration Mismatches: Most systems do not use analysis symptoms connected to image classification, which reduces diagnostic accuracy [7].

D. ***Gaps and Proposed Contributions***

The gaps that identified in existing research involves:

1) Limitation of dataset size and difference in labels.

2) Difficulty in differntiating actual bite marks from non-bite regions.

3) No integrated analysis of bite marks and symptoms.

The identified gaps target the proposed system as follows:

- Symptom analysis with image processing integration to increase the classification accuracy.
- Transfer learning utilization using MobileNetV2 to augment the available dataset and improve performance.
- A comprehensive approach of bite classification not limited to venomous and non-venomous snakes, but also classifying them into families.

III. METHODOLOGY

This research delves into image processing, machine learning, and deep learning to classify venomous and non-venomous snakes according to their bite marks and symptoms. Data collection, preprocessing, and analytical techniques comprise the methodology. The following stages lead to getting precise categorizations and support in timely administering the right anti-venom.

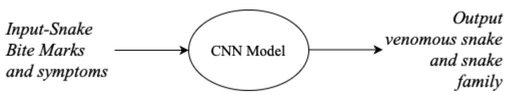
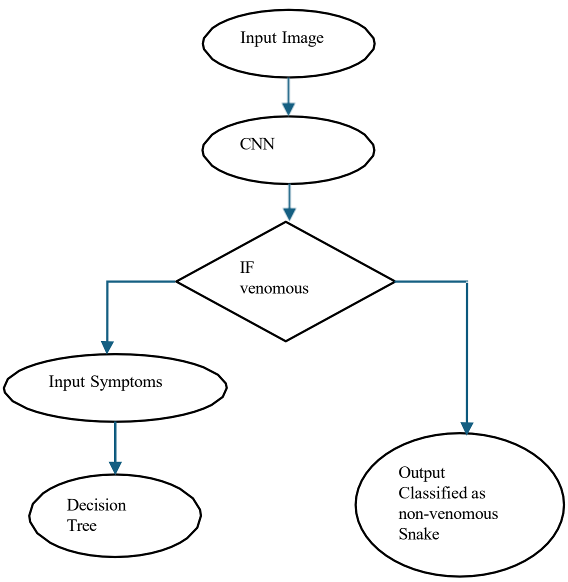


Fig. 3. Context Flow Model of Snake Identification

Below are the procedures of the identified CNN model for the identification of snakes.



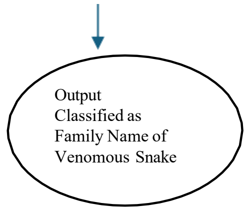


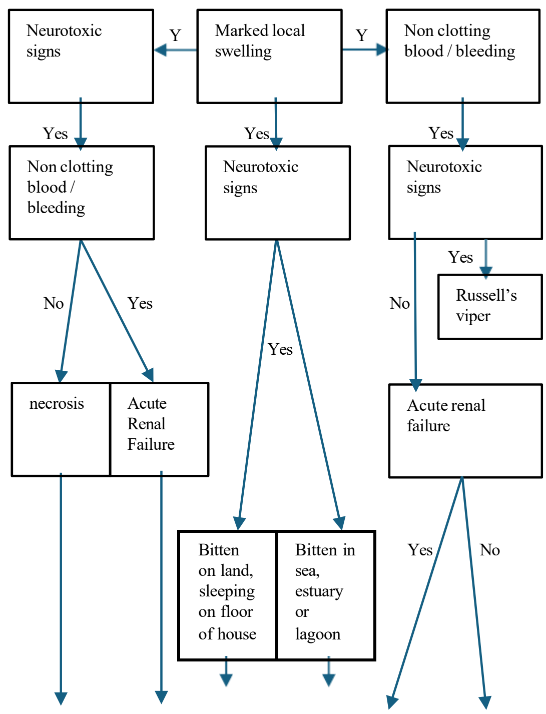
Fig. 4. Flow chart of the proposed system

1) ***CNN***

The input images are taken to predict whether the snake bite is from a venomous or a non-venomous snake bites from a non-venomous snake are usually easy to read even with a little experience of such bite marks. However, even such bites should be captured by the way the system architectures the CNN itself because it then prevents the system from taking further unnecessary steps, thus saving energy and resources. If the bite is recognized as venomous snake, the result from CNN will be sent to the next step for further processing.

2) ***Decision Tree***

The input is acquired from the CNN, which can identify the bitten marks of venomous snakes. And then, it works to determine the family of the snake based on the symptoms analysis using Decision tree algorithm. For this purpose, it compares the symptoms of the patient with those of different snake families. Once the matching has been done, the system gives an output of the result, including the family name of the snake. That helps decide future actions regarding patient treatment [18].



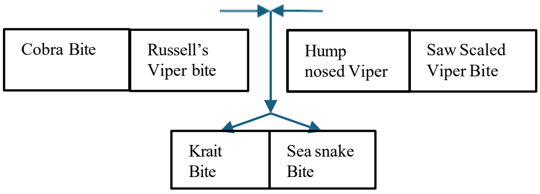


Fig. 5. Diagnosis of snakebite cases based on clinical data

The decision tree algorithm is helpful in analyzing the symptoms of a patient for bites and helps to find out which family of snakes it belongs to. This is designed to function like a flowchart and does symptoms structure into a hierarchy. The inputs are neurotoxic symptoms, local swelling, problems in clotting and renal failure. Based on the symptoms of the patient, the decision tree traverses different branches and hones down possible snake families. As an example, certain symptoms such as local swelling and neurotoxicity could be indicative of a cobra while, for example, spontaneous bleeding and acute renal failure might denote a viper. This structured approach would help in a faster, accurate decision-making process in patient care.

B. *Population, Sample and Sampling Techniques*

The study targets on the population of Northcentral Province in Sri Lanka, which records the highest snakebite incidence rate of 623 per 100,000 annually [6]. A secondary dataset of 125 snake bite images was collected. It included an equal representation of venomous and non-venomous cases. The data was gathered from medical universities and public health databases accordingly. Organized random sampling was used to make sure that venomous and non-venomous bites were illustrated on the right ratio. Secondary data became the primary concern due to the challenges associated with direct collection of qualitative images. Data sites were included as such health institutions and online database repositories.

TABLE IX. INCIDENCE OF SNAKEBITES IN SRILANKA

Province	Bites		
	Reported (sample)	*Estimated number*	*Risk per 100000 (95% CI)*
Western	61	18910	325 (217-432)
Central	45	7065	277(182-371)
Sabaragamuwa	102	10506	548(438-658)
Northwestern	92	11776	499(392-605)
Southern	87	11310	461(338-584)
Uva	63	4095	328 (242-414)
Nothern	59	3422	324 (219-428)
Eastern	67	5695	368 (227-509)
NorthCentral	119	7735	623(487-760)

a. Data involves estimated and reported snakebites cases, with risk per 100, 000 units and the corresponding 95% confidence interval.

b.

C. *Data Collection and Sources*

Information on snakebite photographs, symptoms, and treatment records like clinical procedures were collected from public health databases, medical records in hospitals and clinics, and recorded bite marks and patient outcomes. This would provide a well-rounded view of the snakebite incidence. This, however, has also confirmed the enriched and very illustrative dataset for analysis using multiple approaches.

D. *Data Pre-Processing and Analysis*

The data pre-processing and analysis techniques to enhance model performance through the delivery of accuracy outcomes would include data augmentation, normalization and good dataset.

Responding to the restriction of the available dataset size, data augmentation techniques were implemented, for flipping, rotation, cropping, and brightness adjustment. The images collected were then standardized to a size of 224 x 224 pixels in .jpeg format supported by the MobileNetV2 model. The dataset was divided by 70% for training and 30% for validation and then further classified into venomous and non-venomous subfolder which contains different snake bites. The entire structure made very easy processing and efficient model training. In addition to image classification, evaluation of symptoms was included to improve diagnostic accuracy reliably.

Critten, according to the established guidelines [13], compares clinical signs like neurotoxicity, local swelling, and blood clotting aspects. For instance, neurotoxic symptoms usually denote cobra or krait bites, whereas systemic bleeding indicates viper bites.

E. *System Development*

The development of the system was carried out using Python and deep learning frameworks such as TensorFlow and Keras on Google Colab making use of the GPU resources. Hardware requirements involved the use of 64-bit processor with 16GB RAM. The approach used in the development process was agile by utilizing an iterative and incremental way. This method allowed ongoing refinement of the snakebite classification model and refinement the mobile application interface. The model employed MobileNetV2 for snakebite image classification. Users upload images through the mobile application. The app processes the bite images with the trained model to provide detected snake type. Furthermore, Symptoms are analyzed to provide more diagnostic insights.

F. *User Interface Design*

As shown in **Fig. 10**, the user interface (UI) of the mobile application is designed to ensure ease of use and efficiency in identifying snake bites. The application consists of multiple screens, each serving a specific purpose to facilitate the user experience. The primary UI components include:

- **Login/Signup Screen**: Allows users to create an account or log in to access personalized features.
- **Home/Dashboard Screen**: Provides an overview of the application's functionalities, including bite image upload and symptom input.

- **Bite Image Upload Screen**: Enables users to upload an image of the snake bite, which will be analyzed by the machine learning model.
- **Symptom Input screen**: Allows users to enter symptoms they are experiencing, contributing to a more accurate identification of the snake type.
- **Analysis and Result Screen**: Displays the outcome of the system's analysis, indicating whether the snake is venomous and non-venomous.
- **Case History/Records Screen**: Maintains a history of previous diagnoses for future reference.
- **Emergency Contact Screen**: Provides quick access to emergency services and medical professionals.

The UI is designed to be intuitive, ensuring that users can easily navigate through the application without prior technical knowledge. The color scheme and layout have been chosen to enhance readability and accessibility.

G. *Sequence Diagram*

As shown in **Fig 11**, to illustrate the workflow of the mobile application, a sequence diagram is used to represent the interaction between users and the system. The diagram outlines the key processes involved in snake bite identification, showing the sequence of messages exchanged between different components.

- **User Logs In**: The user enters their credentials, and the system verifies authentication.
- **Image Upload and Symptom Entry**: The user uploads an image of the bite and enters symptoms.
- **Data processing**: The backend system processes the input using machine learning models.
- **Snake Identification**: The system predicts the type of snake based on the analysis.
- **Result Display**: The identified snake type and recommended actions are presented to the user.
- **Emergency Assistance**: The user can contact medical professionals through the emergency screen.

IV. DATA ANALYSIS

This work involved online secondary data. Specifically, snakebite images were sourced from the Internet showing distinguishable bite mark pattern on the individual bitten. MobileNetV2 is efficient lightweight model; it does best for application in case there is constraint in computational powers. It was used to process the images. The images were available in .jpeg and .arw formats and the images were converted to .jpg format for standardization. Low quality images such as blurred, very small ones were removed, and the existing images were manually cropped to emphasize bite marks clearly. This method decreased interruption from the background and enhanced the sharpness of the focused region. The images were consequently classified as poisonous and non-poisonous and resized to dimensions of 224x224 pixels to match the inputs needed for the model.

A. *Data Collection and Preparation*

In total, a collection of 125 photographs has been gathered from various sources, including medical records and online databases. After refined filtering, 87 pictures remained to qualify as high quality: 58 represent the number of venomous snake bites, while 29 are for bites from non-venomous snakes. Such a dataset confirmed balanced representation in both categories and opened avenues to good qualitative analysis. MobileNet was, thus, used for classifying the images by analyzing the unique visual features of bite patterns. Standard performance metrics such as accuracy, precision, recall, and F1 score were computed to judge the reliability of the model. The model achieved a fine accuracy of 97.7%. The F1 scores were 0.98 and 0.97 for venomous and non-venomous classification, respectively, with a recall of 1.00 for venomous cases perfectly. There was a clear prominence in the confusion matrix in limited counts showing misclassifications that resulted in only two venomous ones being classified as nonpoisonous.

B. *Finding and Interpretation*

Bite mark patterns can determine whether a snake bite is venomous or not. This report describes a deep learning model that could be harnessed in doing so. The system was capable of making accurate and robust generalization up to the level of consistent accuracy both for training and testing datasets. When comparing with previous works, for example, "Deep Learning Model for Identifying Snakes by using Snakes Bite Marks" [11], which gave an accuracy of 92%, and "A Deep Learning Approach for snake Bite Mark Identification and Classification" [9], which attained 85%, the proposed model got a remarkable improvement. This improvement was attributed to improved training procedures and modernized feature extraction methods. Such good performance results showed the capability of this method in practical scenarios in the clinical procedure of medical diagnosis snake bites.

V. DISCUSSION

This chapter shows the findings of the study and measures whether the research missions have been fulfil. The research concentrated on the challenges in developing a snake identification system and investigated machine learning techniques for image classification. Specially, the study searched how these methods can support in snake identification. After appraising several machine learning techniques, the research switch to transfer learning, with MobileNetV2, being selected due to its rightness for image processing tasks. As a result, mobile application integrated this deep convolutional neural network (CNN), designed to function efficiently on devices with small-scaled computational power.

A. *Validation of the Model*

In this research, transfer learning applies to a pretrained convolutional neural network (CNN) that is fine-tuned on snake bite classification tasks. The MobileNetV2 model has been chosen and modified to classify snake bites as venomous and non-venomous. The dataset of this research consisted of a total of 125 images collected from various online sites and hospital records according to the way that has been detailed in the methodology section. Seventy percent (70%) of these images were used for training, while thirty percent (30%) were assigned to validation and testing. The model's true performance evaluated using the test set, which had not been detected by the model during training. The metrics like accuracy, precision, recall and support were utilized to evaluate the model clearly. The final accuracy was 97.7%, marking that the model had effectively absorbed from the training data.

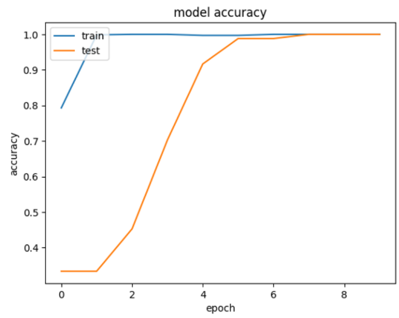


Fig. 6. Model Accuracy

As shown in **Fig 6**, shows the accuracy of the model during training and testing. The training accuracy started at a high level and quickly stabilized near 100%, while the testing accuracy initially increased gradually and reached approximately 100% after several epochs. This trend indicates that the model successfully learned patterns from the training data and generalized well to the test data.

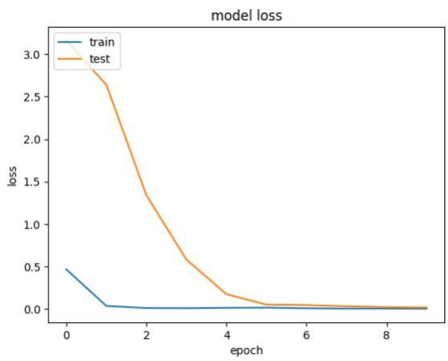


Fig. 7. Model Loss

As shown in **Fig 7,** presents the loss values during training and testing. The training loss decreased rapidly in the early epochs and remained close to zero, while the test loss initially had a high value but gradually reduced as the training progressed. The reduction in loss signifies that the model effectively minimized errors over time, leading to high performance.

As shown in **Fig 8,** confusion matrix displays how well the model performs regarding the classification of snakebites. It indicates that 56 venomous snakebite cases were identified correctly while 2 were misidentified. In addition, the 29 nonvenomous cases were also entirely identified without false negatives. This suggests very high accuracy for both venomous and nonvenomous classifications.

	precision	recall	f1-score	support
Venomous	1.00	0.97	0.98	58
Non_Venomous	0.94	1.00	0.97	29
accuracy			0.98	87
macro avg	0.97	0.98	0.97	87
weighted avg	0.98	0.98	0.98	87

Fig. 8. Classification Report

As shown in **Fig 9,** the high accuracy of 98% is due to several factors. The well-balanced dataset, with 58 venomous and 29 non-venomous samples, ensured the model learned both classes effectively. The deep learning design enabled precise feature extraction. This improved performance, especially for the venomous class, which reached perfect precision (1.00). Additionally, optimizing hyperparameters during training may have improved generalization. The low misclassification rate, shown by high recall (1.00) for non-venomous bites, shows effective handling of difficult cases.

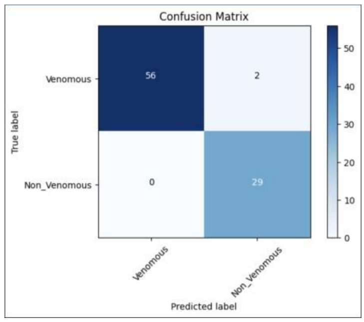


Fig. 9. Confusion Matrix

B. *Achievements of the Objectives*

The goal of this study was investigations into machine learning algorithms for image classification. Some of these algorithms include convolutional neural networks (CNNs) and transfer learning techniques. Models were trained in a snake bite image dataset with an emphasis on separate species classification. The metrics of evaluation include accuracy, precision, recall, and all such others which aided in the determination of the best models for the task concerned.

Secondly, it talked about the designing of a machine learning model which will be able to identify snake species automatically through their bite marks. As mentioned in the methodology chapter, the system was created to function without autonomously. The model was able to categorize snake bites as either venomous or non-venomous, an essential task given the direct link between the snake type and capable damage. Wrong identification of snake species can lead to significant medical and health implications so precision in this area becomes critical. The model given here has much improved over previous snake image classification models in accuracy. It was tested that MobileNetV2 could realize a better classification potential reaching the final testing accuracy of 96%, which was approved freely using the data from the test set of unseen data. Evaluation measurement tools such as precision, recall, and F1 score were applied to test the validity of the model.

C. *Recommendatios for Future Work*

1) Improve the dataset to generalize better the model.

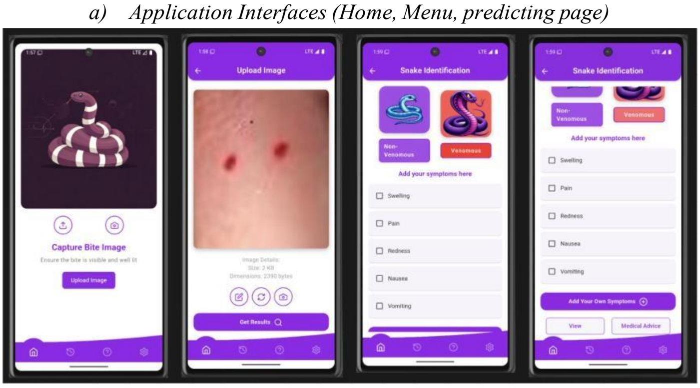
2) Seek more evaluation metrics and undertake tests with real-life examples.

3) Associate with end-user feedback to improve usability and functionability of the application.

D. *Future Direction*

Future studies can make snake identification from families more precise using bite marks analysis features. The incorporation of symptoms into the system will enhance the prediction; certain symptoms are observed to associate with particular snake families, apart from the venomous or non-venomous classifications. Measuring or analyzing the distance between the bite marks could give additional information on the species, while the size and shape of the marks also aid in identifying them. Heat signature scanning on the bite site may also reveal unique temperature patterns from different snake species which promise to be one of the good ways of enhancing the accuracy of the system.

a) Application Interfaces (Home, Menu, predicting page)



b) Application Interfaces (History, Support, Setting page)

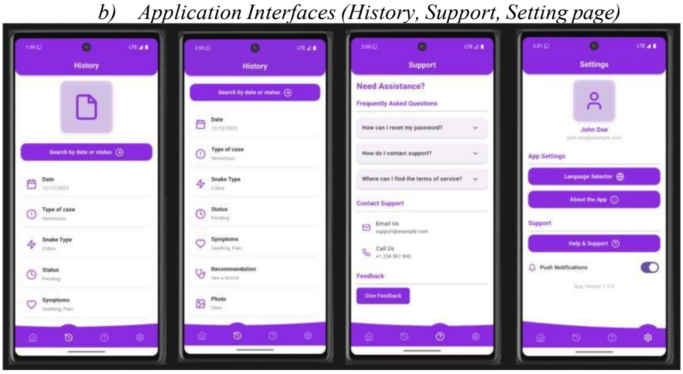


Fig. 10. Mobile application Interfaces

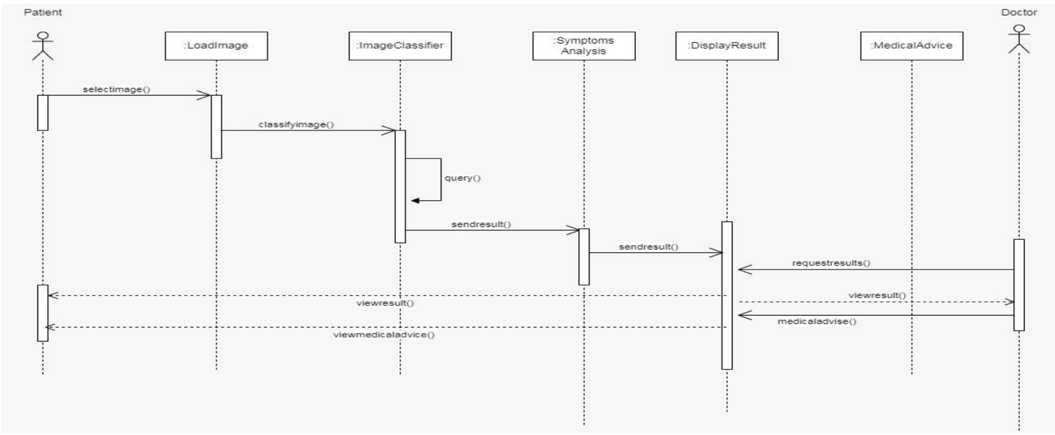


Fig. 11. Sequence Diagram

VI. CONCLUSION

Sri Lanka is home to numerous toxic and harmless snakes, and the purpose of this study is to design a mobile application that can identify the type of snake based on bite marks on the victim's skin in Sri Lanka. The main issue was the difficulty that doctors face when trying to identify the types of snakes, which is very important for choosing the right antivenom and effective treatment. Misidentification can increase the risks to the patients. The research started by understanding the challenges in identifying venomous and non-venomous snake bite marks. Machine learning methods were included for classification of images. Among these methods, transfer learning was selected for use with the MobileNetV2 deep convolutional neural network, since it works well for image tasks. The study indicates progress in computer vision using deep neural networks for image classification efficiently and accurately. This study provided a viable means for recognizing a snake species using transfer education. The developed mobile application can be used to classify whether a snake is venomous or not by capturing images of bite marks and symptoms of victimized individuals. The system resulted in output accuracy concerning predicted species and types of bite on the images given by users. The results were saved in a cloud database, allowing easy access to information and medical guidance. Further improvements, including more training with real bite marks, could enhance the model's reliability and usability. This development could turn into a helpful in mobile healthcare, aiding in snakebite identification and treatment choices.

ACKNOWLEDGMENT

I would like to thank Dr. Thisara Weereasinghe and my supervisor Mr. Keerthi Kodithuwakku for their guidance, feedback, and support during this period. Similarly, I wish to express my gratitude to all my lecturers and faculty members of the National Institute of Business Management for the very precious expertise they have shared as well as the encouragement they provided that would contribute greatly to my research and personal development. I would like to add a special note of appreciation to the creators of the "Snake Bite Detection Object Detection Dataset (v2, 2024-04-29 2:34 pm)" for making available the image datasets I drew from while developing and testing the machine learning model used in this study. I especially wish to thank my family and friends for their patience, encouragement, and unwavering belief in me through this journey. Lastly, a big thank you to each one of the above-mentioned persons who directly or indirectly contributed and gave the needed support into the successful completion of this project.

REFERENCES

[1] A. A. Awan, "A complete guide to data augmentation," DataCamp, Nov. 2022. [Online]. Available: https://www.datacamp.com/tutorial/complete-guide-data- augmentation.

[2] A. Afroz, B. N. Siddiquea, A. N. Shetty, T. N. W. Jackson, and A. D. Watt, "Assessing knowledge and awareness regarding snakebite and management of snakebite envenoming in healthcare workers and the general population: A systematic review and meta-analysis," *PLOS Neglected Tropical Diseases*, vol. 17, no. 2, p. e0011048, Feb. 2023, doi: 10.1371/journal.pntd.0011048.

[3] O. Ali, W. Abdelbaki, A. Shrestha, E. Elbasi, M. A. A. Alryalat, and Y. K. Dwivedi, "A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities," *Journal of Innovation & Knowledge*, vol. 8, no. 1, p. 100333, Jan. 2023, doi: 10.1016/j.jik.2023.100333.

[4] M. Bhardwaj, A. Singh, M. Kumar, S. K. Sharma, and J. Kumar, "Design and analysis of artificial intelligence model for the global issue of poisonous reptile identification," *BioGecko*, vol. 12, no. 4, pp. 16–25, Jul. 2023. [Online]. Available: https://biogecko.co.nz/index.php/journal/article/view/2

[5] H. K. Dishar and L. A. Muhammed, "A review of the overfitting problem in convolution neural network and remedy approaches," *Magallaẗ al-Qādisiyyaẗ li-ʿUlūm al-Ḥāsibāt wa-al-Riyāḍiyyāt*, vol. 15, no. 2, Sep. 2023, doi: 10.29304/jqcm.2023.15.2.1240.

[6] D. S. Ediriweera *et al.*, "Mapping the risk of snakebite in Sri Lanka: A national survey with geospatial analysis," *PLOS Neglected Tropical Diseases*, vol. 10, no. 7, pp. e0004813, Jul. 2016, doi: 10.1371/journal.pntd.0004813.

[7] "Deep learning model for identifying snakes by using snakes' bite marks," *IEEE Xplore*, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9104200

[8] N. P. A. U. D. Hernawati, N. Adiwijaya, and D. Q. Utama, "Image processing for snake identification based on bite using Local Binary Pattern and Support Vector Machine method," *Journal of Physics: Conference Series*, vol. 1192, pp. 012007, Mar. 2019, doi: 10.1088/1742-6596/1192/1/012007.

[9] K. Logesh, “A deep learning approach for snake bite mark identification and classification,” *African Journal of Biological Sciences*, vol. 6, no. 5, pp. 5099–5110, 2024. [Online]. Available: https://www.afjbs.com/uploads/paper/75e4b511255953c44a65e9e7f9c9f766.pdf.

[10] “Snakebite is a public health problem,” *LankaWeb*, Mar. 2024. [Online]. Available: https://www.lankaweb.com/news/items/2024/03/03/snakebite-is-a-public-health-problem/

[11] R. Kamalraj, “Deep learning model for identifying snakes by using snakes’ bite marks,” *IEEE Xplore*, Jan. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9104200

[12] S. T. Krishna and H. K. Kalluri, “Deep learning and transfer learning approaches for image classification,” *ResearchGate*, Jun. 2019. [Online]. Available: https://www.researchgate.net/publication/333666150

[13] “SLMA guidelines for the management of snakebite in hospital – 2021,” *Sri Lanka Medical Association*, 2022. [Online]. Available: https://slma.lk/wp-content/uploads/2022/06/slma-sbg_2021-v3.2.pdf

[14] N. Sharma, “What is MobileNetV2? Features, architecture, application and more,” *Analytics Vidhya*, Dec. 2023. [Online]. Available: https://www.analyticsvidhya.com/blog/2023/12/what-is-mobilenetv2/

[15] Snake Bite, “Snake bite detection object detection dataset (v2, 2024-04-29 2:34 pm),” *Roboflow*, 2024. [Online]. Available: https://universe.roboflow.com/snake-bite/snake-bite-detection/dataset/2

[16] R. M. Putra, Adiwijaya, and D. Q. Utama, “Snake bite classification using Chain code and K-nearest neighbour,” *Journal of Physics: Conference Series*, vol. 1192, p. 012015, Mar. 2019, doi: 10.1088/1742-6596/1192/1/012015.

[17] “What is the k-nearest neighbors algorithm?” *IBM*, 2023. [Online]. Available: https://www.ibm.com/topics/knn.

[18] A. Habib, H. R. Abubakar, M. Abubakar, and A. L. Habib, “Decision tree model for managing snakebite envenoming with or without FAV-Afrique,” *ResearchGate*, 2023. [Online]. Available: https://www.researchgate.net/figure/Decision-tree-model-for-managing-snakebite-envenoming-with-or-without-FAV-Afrique_fig5_299516989.

Deep Learning based Approach for Inaudiable Speech Recognition with Emotion-Based Verification in Criminal Video Evidence

Senanayake K P R
School of Computing
National Institute of Business Management
Colombo, Sri Lanka
cobsccomp222p-012@student.nibm.lk

Kumara S S
School of Computing
National Institute of Business Management
Colombo, Sri Lanka
sampath@vl.nibm.lk

***Abstract*— Inaudible human speech analysis in video evidence is critical for crime investigations to extract more details regarding criminal cases. This research proposes a deep learning-based approach to recognize these kinds of speeches in video evidence using lips reading. Furthermore, the proposed approach consists of a process to verify identified speeches using correlations between facial emotions and sentiments of speech contents. In this approach, initially, video evidence is preprocessed using computer vision techniques to enhance the quality factors like resolution, contrast, and different light conditions. Then 2-Dimensional Convolutional neural network (2D CNN) is used for facial emotion prediction and 3-Dimensional Convolutional Neural Network (3D CNN) with Long-Short Term Memory (LSTM) is used to develop lips reading model to predict spoken words in each timestep. Proposed 2D CNN shows around 55% validation accuracy and combined neural network (3D CNN and LSTM) shows low accuracy due issues in lips reading dataset. These predicted words by the model are analyzed using beam search decoder to obtain the best sequence of words by reducing noises of the predictions. Obtained word sequences are leveraged to construct the sentences of speech through a Large Language Model (LLM). After constructing sentences, a sentiment analysis is performed for each sentence and predicted facial emotions from the model are compared with these sentiments to verify the emotional status of constructed speeches. This research delivers an important contribution for inaudible human speech analysis in criminal video evidence by exploring applicability of emerging technologies for this context.**

Keywords - Deep Learning, Inaudible speech, video evidence, facial emotions, sentiment analysis, criminal investigation, lips reading

I. INTRODUCTION

Video evidence can be considered as a major form of evidence which is used in crime investigations as electronic evidence involves 85% of crime scenes according to [1]. As mentioned in [2], in 2016, 80% of crime incidents related to video evidence which were recorded in devices such as smart phones, tablet cameras and public cameras. Considering these facts, the significance of video evidence for crime investigations can be shown.

Most of the videos which are analyzed in criminal investigations are inaudible because of camera distances, noisy environments, and default inaudible settings. Because of these limitations, investigators are unable to detect the speeches which are spoken by suspects or victims in these videos and this issue has led to several challenges in their investigations such as misunderstanding of relevant incident, delays in related facts identification and loss of crucial evidence. As mentioned in [3], while the performance is not perfect, "Forensic Lips Reading" is a significant technique which is used to identify these speech contents to assist investigators to make decisions related to their investigations. [4] indicates the incident that detected murderer of "Arlene Fraser" with the help of lips reading to analyze an inaudible conversation captured in a CCTV camera. This shows the importance of this research area.

As indicated in [5], Emotional Artificial Intelligence is used to estimate the emotional status of humans with the help of speech features and selection of words for crime investigations. The above review article demonstrates the significance of the relationship between speech contents and human emotions. These can be considered as evidence for applications of emerging technologies to detect speech features and emotions in this context.

A. Research Problem and Objectives

As mentioned in the previous section, most crime incidents involve video evidence which is recorded in various camera devices. But according to the [6], most cameras located in public areas do not pick up audio inputs of human voices because of several reasons such as background noise, default inaudible settings and long distances. It interrupts gathering more details and accelerating the investigation process. As [3] shows, forensic lips reading plays a significant role in this context but there are many difficulties in manual lips reading such as human errors due to distinct experience levels, movements of people and different gradients of camera shots. As mentioned in [7] It has showed 0 to 70% fault rate when human perform lips reading. It demonstrates the hardship of this approach. Although there are some automated systems for this approach, it is difficult to use them in criminal video analysis because of their uncertainty. As an example, [8] has proposed a lip-reading integration for CCTV cameras but it was not successful due to uncertainty of these videos.

According to the data, gathered from experts in this field, there should be a strategy to receive a verification about the credibility of recognized inaudible speech contents because criminal investigation field is extremely sensitive for human beings.

After analyzing this research problem, the following objectives were defined to overcome this problem.

- To enhance the efficacy of detecting the motions of key facial regions.

- To identify emotions associated with speeches from video evidence.
- To detect contents of inaudible speeches precisely from video evidence.
- To verify the credibility of predicted speech contents using relationship between speech sentiments and facial emotions.

B. Scope and Limitations

This research focuses on identifying speech contents and related emotions from video evidence and verifying identified speech contents with the help of speech sentiments and facial emotions. The significance, challenges and improvements of identifying speech contents and related emotions from this kind of video evidence were considered. In addition, this research focuses on only English words detection using lips reading techniques.

The proposed solution was evaluated with recorded videos with all the focused features in real-time videos. The following limitations were identified in proposed software development.

- Proposed predictions cannot be performed when faces of human appeared in videos are invisible.
- These predictions cannot be done with very poor resolution videos.
- Realtime criminal video evidence cannot be used to evaluate the proposed software in this stage due to public security concerns.

II. LITRETURE REVIEW

According to [6], have examined enhancements of lips reading approaches and illustrated a prototype invented by "Delft University". It was developed using "Active Appearance Model (AAC)" and "Hidden Markov model" and analyzed the capability of recognizing violent activities inside trains. In [9], innovative re-rendering technique for videos using AAM to improve the movements of lips region to ease the reading process of human lips readers and as they have mentioned that it caused the improvement of the accuracy. [10] proposes a strategy to categorize phonemes using a classification model for visemes and it could be used for noisy videos with weak audio inputs. According to the [11], a system to reconstruct speeches and generate sounds with neighboring frames, was proposed. [12] has inspected several Deep Learning (DL) approaches for lips reading. According to its' discoveries, various angles of people appeared in static cameras is a major obstacle for this. [13] has developed and evaluated a two-tower CNN architecture to reproduce audio output using inaudible video frames. As mentioned there, this approach shows enhanced performance of identifying speeches and audio generation than existing methodologies.

In addition, several approaches to detect speech features and emotions, and some experiments to explore the correlation between emotions and speech contents have been conducted. According to the research [14], it reveals the relationship between verbal contents and facial expressions, and the impact of emotions on this correlation.

[15] has proposed a speech sentiments and emotion recognition system with the help of close relationship between speech sentiments and emotions. According to this research they have obtained high accuracy rates for these predictions especially for two sentiment classes. [16] has introduced a crime tracking system using emotions, captured from facial movements using CNN. [17] shows another study to examine correlation between upper limbic gestures and features of speeches like ascent, emotions etc. This can be considered as another successful approach of studying connection between human body movements and speeches.

A. Research Gap

According to the above review, various kinds of approaches can be found for inaudible speech recognition using emerging technologies. Most of the research has been conducted to examine different types of datasets, variety of architectures for ML models but only a few of them have focused on different video features which are commonly found in crime investigations. Different angles of cameras, different light conditions of videos, unique features of human facial gestures should be considered to achieve the objectives of this research. In addition, some of them have introduced novel methods to enhance the performance of visual speech predictions with the assistance of some audio inputs. But most of the video evidence doesn't have any kind of audio input related to human speech. Because of that we cannot depend on audio signals to perform predictions.

III. METHODOLOGY

Data collection and data analysis were done using standard methodologies. After analyzing the collected data from the population, relevant findings were used to decide the features of the proposed solution. In addition, training and testing data were gathered to train deep learning models of the solution.

A. Population, Sample and Sampling Techniques

1) Population: The Crime invetigation devisions of "Sri Lanka Police Department", was selected as the population as this department is considered as the main organization which involved with crime investigations in Sri Lanka.

2) Sample: Exact population count is not available due to public security reasons.Because of this, Cochran's sample size formula for infinite population was used to [18] estimate the size of sample.After considering the uniform processes for their investiagtions all over the country, margin of error was specified as 0.1.In addition, confidence level and propotion of population was defined as 95% and 0.5 respectively.Sample size was calculated as 97.

3) Sampling Techniques: To overcome accessibility barriers of this population, a combination of two sampling techniques was used.According to this "Exponential Discriminative Snowball Sampling" was conducted to interview specialists in this field to gather data regarding this research context.After gathering relevant data, details of few refferals related to this participants were gathered. Once considerable amount of refferals were gathered, "Simple random Sampling" was performed to select the sample to distribute questionnaires.Futhermore data gathered from interviews and litreture reviews were used to design the questionnaire.

B. *Data Types and Data Sources*

Several data types were identified to properly analyze the research problem, gather requirements for the proposed solution and prepare datasets for model training. These data types are primary and secondary data, quantitative and qualitative data, and training and testing data.

There are several categories of data which were gathered to analyze the research gap, decide the features of the proposed solution and train deep learning models to perform the predictions. The interviews and questionnaires were conducted to gather primary data such as in-depth ideas regarding this context according to the experiences of experts in the field. In addition, previous research papers which are related to these areas, were used to initialize the research problem and research gap.

Most qualitative data were gathered using interview questions as these questions were designed as open-ended questions. The questionnaire was designed with multiple select questions, linear scale questions and Likert questions to gather both qualitative and quantitative data.

Image data were collected to build datasets for model training and these data were gathered using online data sources like "Kaggle". Image data of human faces and lips regions were gathered, and each dataset was enhanced by adding more data.

C. *Methods of Data Analysis*

Gathered data are mainly analyzed for two purposes. These are analyzing the requirement of the proposed solution and training Deep Learning models for predictions. Qualitative and quantitative data gathered from the population were analyzed as a requirement gathering and images data were analyzed for model training.

1) Qualitative and Quantitative data analysis: Qualitative data which were gathred from interviews were analyzed using "Thematic analysis". As mentioned in [19], in this appraoch, first patterns were identified as codes from responses, then some themes regarding human emotion and speech content detection were extracted using these patterns.Furthermore quantitative data which were mainly gathered from questionnair, were analyzed using descriptive analysis and pie charts and bar charts were widely used to perform this analysis.

2) Training and Testing Data Analysis: After gathering relevant image data, they were analyzed using Deep Learning techniques to train models for both lips reading and facial emotion detection. To predict facial emotions, gathered emotion dataset was first splited in to two parts for training (around 80% of dataset) and testing (around 20% of dataset), and then designed a 2D CNN using pytorch framework. For word detection using automated lips reading, lips dataset was also splited in to two parts for training and testing process with the same ratio. A combined nueral network was designed with 3D CNN and LSTM to train the lips reading model because spatial and temporal features of images had to be trained for each word.Max pooling, Batch normalization and Dropout layers were included in each CNN design to optimize the training flow and to enhance the variability of dataset, data argumentation techniques were used.

IV. Data Analysis

A. *Data Analysis*

Data analysis was conducted in two steps. First, all data gathered from interviews were analyzed using thematic analysis as these data are qualitative data. To gather these data, open ended questions were asked from participants. multiple codes were identified, and four main themes were extracted from them.

Thematic Analysis of Interview data	
Codes	**Themes**
Variabilities of faces	Challenges of Analyzing Emotions and Inaudible Speech Conter
light conditions	
low resolutions	
camera angles	
distance and dark	
Officer experience variability	
Relevance of inaudible speech in investigations	Importance of Inaudible human speeches and human emotions identification.
Significance of speech-related emotions	
behavioural analysis of suspects	
Affect the direction of invetagtion	
Manual lip-reading methods	Existing strategies to identify inaudible human speeches and hu emotions
Experience-based emotion observation	
Absence of automated systems	
Reduction of human errors	Expectations from New Method
Standardization of interpretation	
Addressing identified challenges	

Fig. 1. This figure indicates identified and refined codes which were extracted from interview transcriptions. Themes were developed by combining these codes according to suitable patterns.

After extracting these data from interviews, questions of questionnaire were designed based on these themes and data which were gathered from literature reviews. This questionnaire was distributed among the selected samples and 81 responses were gathered. Then several patterns and distributions regarding the research context were identified using charts and tables. Linear scale, Likert scale questions were designed to rate their preferences and most of other questions examined the experiences regarding this research context.

B. *Key Findings and Interpretations*

According to the analysis, most of the investigations related to video evidence are captured in devices like security cameras, mobile cameras and vehicle dash cameras. Findings show the low satisfaction of existing manual lips reading methods to identify inaudible speech contents and it demonstrated the significance of relationship between speeches and emotions in criminal investigations. In addition, the impact of human error for these predictions and negative impact of directly using results of a system for decision making was highly rated in this questionnaire. It reveals the importance of a verification method when identifying human speeches without audio signals from video evidence because investigators expect some kind of credibility of these predictions for their decision making. Furthermore, variety of camera angles, human head poses, various types of light condition, slight varieties among human faces were identified as most challengeable factors which disturb for video evidence analysis.

These findings were interpreted to design the solution. According to these interpretations video quality enhancement techniques should be included into the solution before the predictions. To analyze the variabilities of head angles, facial features, model training process should be done with image data which consists of these various features. After predicting these emotions and words in each timeframe of the video, specific text analysis and sentiment analysis methods were used to construct the whole content of speech.

V. PROPOSED SOLUTION

According to the analyzed data, Deep learning-based system was proposed to predict inaudible speeches and verify them by comparing sentiments and facial emotions. Proposed deep learning models predict spoken words and facial emotions with probabilities by analyzing sequences of lips movements and facial features respectively. Then predicted words and emotions are pre-procced to construct whole speech content and compare the sentiments of sentences with facial emotions.

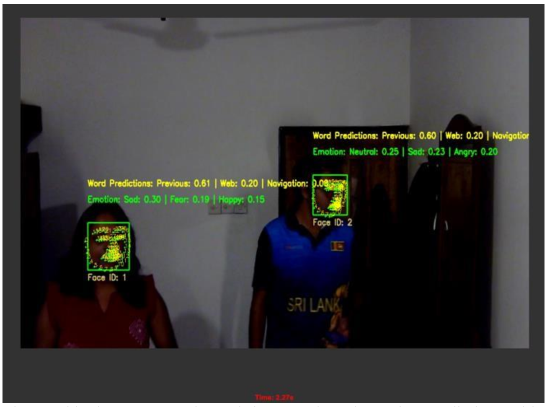


Fig. 2. This demonstrates the real time words and emotion predictions with probabilities.

Timestamp	Person ID	Emotion	Word
0.10	1	Sad	Next
0.13	1	Sad	Stop
0.17	1	Sad	Connection
0.20	1	Neutral	Hello
0.20	2	Angry	Previous
0.23	1	Angry	Choose
0.23	2	Sad	Navigation

Fig. 3. Predicted words and emotions in each timeframe

A. General Architecture

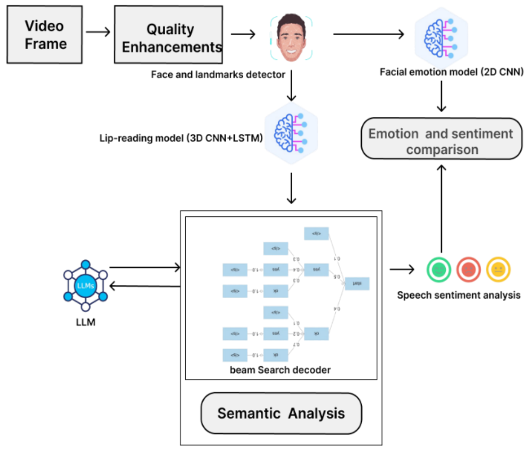


Fig .4. General Architecture of proposed Solution

As demonstrated in Fig. 2. each extracted frame of the video is enhanced through few quality enhancement steps. Then the frame enters the face detector and landmark detector of the "Dlib" library. Detected faces in relevant video frame are passed to the emotion model for feature extractions and emotion predictions. Simultaneously, sequences of lips movements of detected faces are cropped and stored into array with fixed length. This array is passed to lips reading model to extract spatiotemporal features and predict the spoken words with probabilities in relevant timestep. Lists of words detected in each timestep are passed to beam search decoder to post-process to reduce the noisy predictions of model and construct most probable sentences.

B. Video pre-processing for quality Enhancement

Frame by frame quality enhancement procedure to improve the input vide quality was developed as shown below.

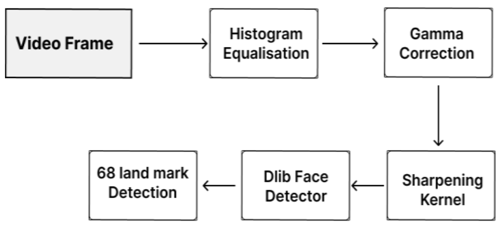


Fig. 5. This figure shows the quality enhancement process of video evidence. First, video frames are extracted one by one and performed histogram equalization, gamma correction, sharpening the edges. Then frames were passed to Dlib face detector and 68 landmark detectors respectively

This procedure is performed to enhance contrast, fine-tune the brightness, identify edges of faces in each video frame. After these procedures, human faces appearing in each frame may be identified by the face detector and 68 facial landmarks are identified by the 68 landmarks detector. These landmarks are used to extract important regions like lips to extract features for predictions using trained deep learning models.

C. Emotion Model Training with CNN

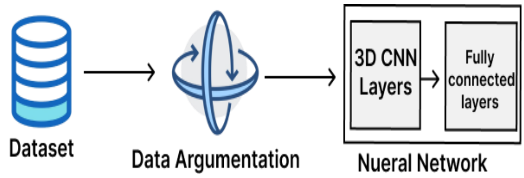


Fig. 6. 2D CNN model training

FER-2013 dataset is used to train this model. Around 80% of data is used for training process and remaining data is used for validation process. Data argumentation techniques such as flipping, random rotations, color adjustments, resizing are performed in training process to enhance the variability of dataset. 10-core Graphical User Interface (GPU) of Apple MacBook pro 2022 with M2 chip is used for training process.

In this 2D CNN architecture, six convolutional layers are used to extract features of the images of different faces with 7 classes of emotions. There are 3 max pooling layers in this architecture to extract the most important features while ignoring the least significant features in images. In addition, Dropout layers are included to prevent overfitting. In addition, 6 batch normalization layers are applied to enhance the efficiency of the training process. Rectified Linear Unit activation function is used between convolutional layers and fully connected layers to identify complex patterns of features because it is known as non-linear activation function. SoftMax activation function is used to obtain predicted classes with relevant probabilities.

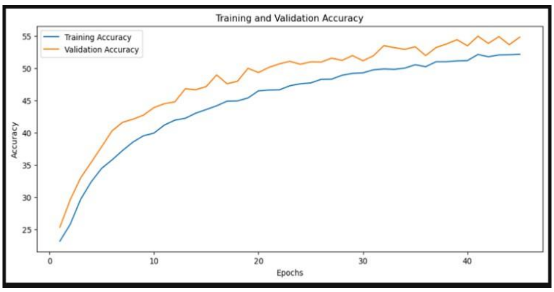


Fig. 7. Validation and training accuracies of emotion detection model. Training process has been done in 44 epochs. This demonstrates around 55% validation accuracy.

D. Lips Reading Model Training with CNN and LSTM

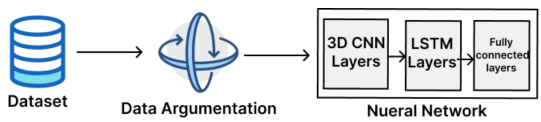


Fig. 8. Combined neural network for lips reading

MIRACL-VC1 dataset is mainly used to collect image data for lips reading. Furthermore, this dataset was enhanced by inserting extra data which has been created using recoded videos. Then Image sequences of lips region for each word are trained and some data argumentation techniques like flipping, shifting, adding gaussian noise, are applied to enhance the variability of dataset. But this model demonstrated some poor accuracy rate due to issues in dataset.8-core Central processing Unit (CPU) of Apple MacBook pro 2022 M2 has been used for this training process because MacBook GPU couldn't be used to compute 3D max pooling layers.

3D convolutional layers are used to detect spatial and temporal features of lips sequences for each word. Then identified features are entered to LSTM layer to detect temporal dependencies. In addition to this, two 3D max pooling layers are used to obtain only the most significant features of the images. Complex patterns of features are analyzed by using Rectified Linear Unit activation function because of its' non-linearity. As same as in emotion model, SoftMax activation function is used in output layer to obtain predicted words with relevant probabilities.

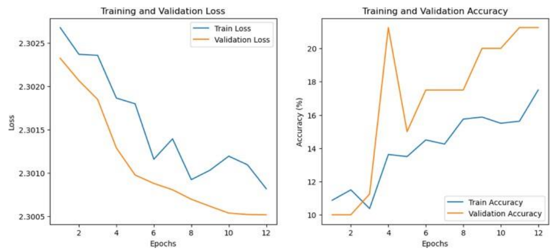


Fig. 9. Loss and accuracy curves of lips reading model

E. Sentence Construction and Emotion based Verification

In this step word sequences are identified using probability values and, partially constructed sentences and selected words in each iteration are sent to OpenAI's LLMs through an API request to obtain the completed sentence. A semantic score is calculated for constructed sentence in each iteration and update the beams list. After that completed sentences are sent to sentiment analysis model to identify the sentiments of the sentences in three classes (Neutral, Positive, Negative). Most intensive facial emotion recognized in each time difference is compared with sentiments of each sentence to verify the credibility of constructed sentence.

VI. DISCUSSION

According to the results of this approach, proposed video quality enhancement steps show slight improvements of capturing human faces and movements of key facial regions even in dark light conditions and different distances. Both "Dlib" and "Mediapipe" landmark detectors were evaluated in detecting facial landmarks. Even though "Mediapipe" framework detects more landmarks, it is poor in detecting facial landmarks from a distance than "Dlib" library. 2D CNN model shows fair accuracy (around 55%) for detecting facial emotions for FER-2013 dataset but it has bias in few emotion categories such as Happy and Neutral. It also could detect small variations of facial emotions. Further lips reading model has been trained by using combined neural network (3D CNN and LSTM) using MIRACL-VC1 dataset. It demonstrated poor accuracy (around 22%) for some words due to lack of dataset variability even though data argumentation techniques were used. There are some limitations in using MacBook GPU for training process because 3D max pooling layers cannot be executed. Because of this limitation CPU has been used to train lip-reading model and it took more time for process. In sentence construction phase, some noisy results are obtained from OpenAI's LLMs. Prompts which are sent to complete sentences with words obtained by beam search decoder should be strictly align with our requirement.

According to the findings, several recommendations can be given to enhance this proposed solution. To advance the performance of deep learning models, enhancing diversity of lips reading dataset can be recommended. Furthermore, advanced video resolution enhancement techniques such as Super Resolution can be recommended to use as quality enhancement techniques.

VII. RECOMMENDATIONS

There are several recommendations for consuming this approach in real world investigations and enhancing the effectiveness of common video footage analysis. for security.

A. *Implementing this approach in Real World*

- After enhancing the accuracy and other drawbacks of this approach, a product can be developed for real world investigations.Furthermore it should be monitored for some period under experts' guidance before being released to the real world investigations.
- Even in production level decisions made by this software should be verified in several steps.

B. *Enhancing the Video Footage Analysis for Common Security Purposes*

- Security Cameras with some quality features should be installed at least in public areas.
- National Framework for advanced digital forensic evidence analysis including emerging technologies and expert knowledge can be recommended.

VIII. CONCLUSION

This research focused on deep learning-based approach to recognize inaudible speeches with a verification process in video evidence to assist criminal investigations. According to the literature review and data analysis, many real-world challenges, research gaps and insights regarding this context were identified. To overcome challenges in video quality factors, few quality enhancement techniques were performed as video pre-processing. In addition, to manage different angles and variabilities of human faces, multiple data argumentation techniques were applied to both facial emotion dataset and lips reading dataset in training process. The 2D CNN which is used to detect facial emotions has been mainly designed with six convolutional layers. Because of data argumentation techniques, emotion model is trained without overfitting. A combined model (3D CNN and LSTM) is used to analyze spatial and temporal features of lips movements to predict spoken words in each time frame. This model demonstrates more accuracy than single 3D CNN, but overall accuracy is poorer than expected. Beam search decoder is used to detect most probable words sequences by ignoring noisy predictions of Lips reading model.

In addition to Beam search decoder, a semantic coherence score is used to ensure the reliability of sentences by analyzing the meaning of sentences. These sequences are used to construct sentences of the relevant speech using a pretrained LLM. The sentiments of constructed speeches are analyzed and compared with facial emotions to verify the predicted sentence.

For the future enhancements, integration of more parameters such as analyzing correlation between upper limbic gestures and speech features and, using ensemble models to increase the accuracy of predictions will be considered. In addition, using modern image quality enhancement techniques such as Super resolution and 3D reconstruction of objects in video evidence to reconstruct invisible parts of human faces in these videos will be developed to enhance the effectiveness of this approach.

ACKNOWLEDGMENT

The Authors would like to show their gratitude to police criminal investigation officers for their incomparable assistance in gathering data regarding this research area. Ultimately authors would like to convey their appreciation to Coventry university and National Institute of Business Management for their commitments regarding academic goals.

REFERENCES

[1] European Commision, "Frequently Asked Questions: New EU rules to obtain electronic evidence," *European Commission - European Commission*, Apr. 17, 2018. https://ec.europa.eu/commission/presscorner/detail/en/memo_18_3345 (accessed Oct. 14, 2024).

[2] U.S department of justice, "Video Evidence A Primer for Prosecutors Global Justice Information Sharing Initiative," 2016. Available: https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/final-video-evidence-primer-for-prosecutors.pdfd

[3] T. Lannin, "What is forensic lip reading and how can it help you?," *LinkedIn*, May 19, 2019. https://www.linkedin.com/pulse/what-forensic-lip-reading-how-can-help-you-tina-lannin/ (accessed Oct. 20, 2024).

[4] gary Ireland, "How forensic lip reading helped convict a murderer," *121 Captions*, Nov. 01, 2021. https://www.121captions.com/forensic-lip-reading-murderer/professional-lip-reading-services/ (accessed Oct. 27, 2024).

[5] L. Podoletz, "We have to talk about emotional AI and crime," *AI & SOCIETY*, vol. 38, May 2022, doi: https://doi.org/10.1007/s00146-022-01435-w.

[6] L. Rothkrantz, "Lip-reading by surveillance cameras," in *2024 Smart City Symposium Prague (SCSP)*, Prague, Czech Republic: IEEE, 2017. Accessed: Oct. 25, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/7973348

[7] Y. Lan, R. Harvey, and B.-J. Theobald, "Insights into machine lip reading," *IEEE Xplore*, Mar. 01, 2012. https://ieeexplore.ieee.org/document/6288999/citations#citations (accessed Oct. 27, 2024).

[8] R. Alalouff, "Lost for words? Automated lip reading technology deciphers speech in silent CCTV images," *Sourcesecurity.com*, 2022. https://www.sourcesecurity.com/insights/automated-lip-reading-technology-speech-silent-cctv-images.20566.html (accessed Oct. 21, 2024).

[9] B. J. Theobald, R. Harvey, S. J. Cox, C. Lewis, and G. P. Owen, "Lip-reading enhancement for law enforcement," *Optics and Photonics for Counterterrorism and Crime Fighting II*, vol. 6402, p. 640205, Sep. 2006, doi: https://doi.org/10.1117/12.689960.

[10] H. L. Bear and R. W. Harvey, "Decoding visemes: Improving machine lip-reading," *ROAR (University of East London)*, Mar. 2016, doi: https://doi.org/10.1109/icassp.2016.7472029.

[11] A. Ephrat, T. Halperin, and S. Peleg, "Improved Speech Reconstruction from Silent Video," in *International Conference on Computer Vision*, Venice, Italy: IEEE, Aug. 2017. doi: https://doi.org/10.1109/iccvw.2017.61.

[12] S. Fenghour, D. Chen, K. Guo, B. Li, and P. Xiao, "Deep Learning-Based Automated Lip-Reading: A Survey," *IEEE Access*, vol. 9, pp. 121184–121205, 2021, doi: https://doi.org/10.1109/access.2021.3107946.

[13] Y. Zhao, R. Xu, X. Wang, P. Hou, H. Tang, and M. Song, "Hearing Lips: Improving Lip Reading by Distilling Speech Recognizers," *arXiv (Cornell University)*, Jan. 2019, doi: https://doi.org/10.48550/arxiv.1911.11502.

[14] C. Busso and S. S. Narayanan, "Interrelation Between Speech and Facial Gestures in Emotional Utterances: A Single Subject Study," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 8, pp. 2331–2347, Nov. 2007, doi: https://doi.org/10.1109/tasl.2007.905145.

[15] B. T. Atmaja and A. Sasou, "Sentiment Analysis and Emotion Recognition from Speech Using Universal Speech Representations," *Sensors*, vol. 22, no. 17, p. 6369, Aug. 2022, doi: https://doi.org/10.3390/s22176369.

[16] Saiprasad V, Sreyas J, Megha S, Akshay A, and Jibin Joy, "Emotion Based Crime Track," *International journal of scientific research in computer science, engineering and information technology*, vol. 10, no. 2, pp. 439–449, Mar. 2024, doi: https://doi.org/10.32628/cseit2410269.

[17] W. Pouw, L. Jonge-Hoekstra, S. J. Harrison, A. Paxton, and J. A. Dixon, "Gesture–speech physics in fluent speech and rhythmic upper limb movements," *Annals of the New York Academy of Sciences*, vol. 1491, no. 1, pp. 89–105, Dec. 2020, doi: https://doi.org/10.1111/nyas.14532.

[18] W. G. Cochran, *Sampling techniques.*, 3rd ed. New York: Wiley, 1977.

[19] V. Braun and V. Clarke, "Using Thematic Analysis in Psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006, doi: https://doi.org/10.1191/1478088706qp063oa.

Detecting Negative Visitor Behaviors Impacting Zoo Animals With Machine Learning and IoT

Jansz J N
School of Computing and Engineering
National Institute of Business Management
Colombo 07, Sri Lanka
cobsccomp222p-043@student.nibm.lk

Maheshwara P
School of Computing and Engineering
National Institute of Business Management
Colombo 07, Sri Lanka
pavithram@nibm.lk

***Abstract*— Zoo animals are often subjected to harassment by zoo visitors. The following research addresses the issue of unwanted and negative behaviors done by zoo visitors towards animals living in the zoo. These behaviors can lead to stress, aggression, and worst-case harm to both the humas and the animals. Current methods which are manual, is labor intensive and not entirely effective. Therefore, this research highlights the need of a more effective solution. The aim is to develop an AI- based system to identify and inform the relevant individuals as well as the respective zoo officials of such behaviors in real-time. This will ensure the welfare of both the animals and zoo goers. By leveraging the latest machine learning and AI technology, the system will monitor human-animal interactions and immediately intervenes reducing risks. This research also dives deep into the various types of negative behavior seen from zoo visitors and what are the drawbacks for both the animals and visitors alike. The findings will contribute to improve animal welfare and enhance visitor safety.**

Keywords—zoo visitor behavior, animal welfare, behavior identification, negative behaviors, zoo management, zoo visitor, zoo

I. INTRODUCTION

The Zoo is a physical location where varying species of animals are available for viewing for the public. What makes a zoo special is that it has very exotic animals and not common to be seeing on a daily basis. The zoo also gives the public a chance to see animals who are not native to their country and wild animals up close. Zoological gardens are one of the most famous attractions to visit today by all types of people demographics. Since these animals are somewhat barred behind enclosures and not free to roam, zoo visitors take the zoo animals and rules for granted. The animals are already going through an uncomfortable time since they are not in their native environment, and they are inside enclosures. Therefore, we as humans and zoo visitors should know to not make their lives any more harder by provoking or making them uneasy. Acting negatively in zoos around the animals has and will lead to unpredictable situations where both the animals and the visitors welfare is at risk.

Zoos, while playing an important role for education and conservation [1], they are also places where close interaction between animals and humans happen. Therefore, these situations should be closely monitored and managed. The main research problem of this research is to identify and reduce unwanted behaviors done by visitors towards zoo animals, since these behaviors can lead to aggression, harm, and stress for both parties. Currently zoos have implemented some strategies to reduce such behaviors but they are mostly labor intensive, also there is no clear successful software solutions or automated systems to monitor negative visitor behaviors. After all is said and done the main objective of this research is to identify the types of negative behaviors in zoos and provide a solution to detect such behaviors in real time and alert the necessary individuals.

II. LITERATURE REVIEW

A zoos main objective is to display uncommon and exotic animals to the general public. Some of the reasons a person visits the zoo can be a form of entertainment for some and to some as a bonding session with friends and family, also it can be to gain an educational experience [2]. Zoos are for foreigners as it is for locals to experience wild and exotic animals, this makes the zoo a very frequent tourist destination currently [3].

As any institution zoos too have rules to adhere.The zoo is not a typical attraction as real living species spend their lives here unlike museums or galleries. Therefore, to make sure animals feel like they live in their natural habitat and their welfare, zoo visitors must behave appropriately. According to [4] most of the visitors are in it for the entertainment or in other words to have fun at the zoo. Having one's own fun at the expense of animal welfare is not acceptable and not ethical. Animal behavior is directly influenced by visitors and this influence can be neutral positive or negative [2]. Therefore, in this research the negative visitor behaviors towards animals will be studied.

Feeding animals, provoking animals, barrier intrusions, banging on zoo enclosures and making loud noises are some of the ways visitors can act negatively [4].

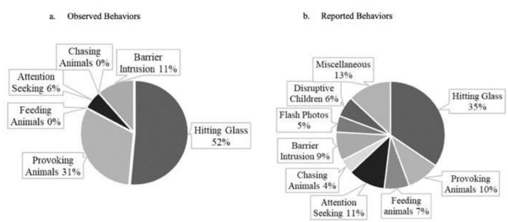


Fig. 43. Inappropiate Observed (a) and Reported (b) behaviors in zoo visitors

In [4, Fig. 1] depicts responses of 162 visitors who personally witnessed such behaviors during their visit. Since most of the animal enclosures have glass walls around, the visitors have mostly banged on the glass surfaces in order to attract hiding animals or get their attention [5]. In other cases it can be to play or scare the animal as well. These types of banging disturbs the peace of fish mostly which is well known. According to study of [5] 75% of negative behaviors were among families, groups and children. Other than that, it was found men were more inclined to do such behavior than women since females have more sympathy towards animals and males typically manage the food during the visit. It is also possible to contract zoonotic diseases when come into contact with zoo animals [6]. Hence the dangers if feeding is high for visitors and also the animals are at risk since some animals are on special diets and may have a chance to be obese or have diabetes due to being fed too much [6].

Negative behaviors are most likely to occur with large crowds and groups according to [5] and [7]. The dangers of crossing safety barriers is clearly seen during “Harambe” the Gorillas incident where a small child crossed the barrier and fell into the enclosure, and ended with the security shooting Harambe [5]. This incident also shows that animals have a limit or threshold before they become distressed [5]. Therefore, there is time to stop any occurring negative visitor behaviors before animals give out irregular behaviors. It was also found Gorillas sitting with back against the glass suggests a stress indicator and does it to avoid visitors. [8]. This will help with identifying if unwanted behaviors are being done by visitors.

Considering the reasons listed above negative visitor behaviors are an ongoing problem. Since these are risks to animal welfare officials and employees have taken necessary steps. One such technique is hanging signs as warnings [9]. But it is not successful and according to research by [6] more than 300 cases of touching animals and 150 cases of feeding was reported in the course of 20 days. It was also found out that other negative behaviors were increased due to this. Since the younger audience does not know how to read and they pay more attention to the animals next to it is also a factor for signs being unsuccessful. Another technique used is holding educational programs to educate the public. This will be effective for a target group but not for all visitors who visit daily [6].

The level of sound and continuous noise was also found as a stress indicator for animals. Similar to how humans feel when there are loud noises, sound which exceeded 90dB outside zoo enclosures was seen as a bad experience for animals. Capuchin monkeys shows clear behavioral changes when the sound was more than 90dB [10]. Also other animals were reluctant to play, use furniture and foraging. According to [11] it states that stress levels of pandas was raised in enclosures. To reduce such noises being made from visitors using signs was not successful. While preventing this as a whole is not possible, it can be reduced to a certain extent.

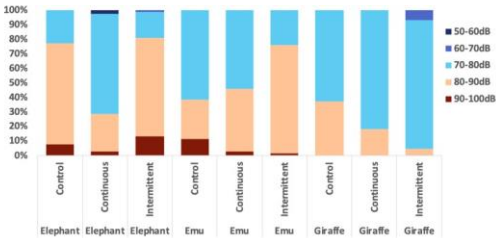


Fig. 44. Time elephants, giraffes and emu’s spent in enclosure during sound levels

According to [11, Fig. 2] it can be seen that animals spend less time in areas with high noise levels and choose to rest in places with lower noise levels. The curl of elephant trunks were also seen as a sign of unease and oppressiveness [12]. These examples show that animals are sensitive to noise similar to humans and therefore measures should be taken to reduce loud noises if they persist for a long time.

Even though there have been many studies which highlight how zoo animals are being subjected to harassment and health risks due to zoo visitors unwanted behavior, there still has not been any effective solution. There are solutions which is labor intensive, having educational programs and sometimes having staff positioned at each enclosure. But using modern technology these unwanted visitor behaviors can be clearly identified and relevant people can be informed and educated. Considering these there have not been any significant research on using AI technologies to improve animal welfare. This knowledge gap shows the necessity of considering AI to offer effective solutions to improve zoo animal welfare.

III. METHODOLOGY

The following chapter describes the data collections and analytical strategies that will be used to identify negative zoo visitor behaviors.

A. Population, Sample and Sampling Technique

The daily zoo visitors, zoo staff and the zoo animals across the Dehiwala National Zoological Gardens in Sri Lanka will be taken as the population for this research. This makes sense due to it being the largest and most popular zoological garden in Sri Lanka. The sampling technique that will be used for zoo visitors is Systematic Sampling since a huge amount of visitors of various demographics visit, this method will ensure random selection without any bias. In the case of zookeepers, Judgmental Sampling will be used since they are the individuals who are in direct contact with the animals. When considering the Dehiwala Zoological Garden according to their latest publicly available Annual Performance Report 2022 [13], the zoo includes more than 250 species of animals and attracts more than 1.2 million visitors per year. Therefore according to this data the zoo visitor population per day will be 3590 and hence the sample will add up to 348 visitors.

B. Types of Data to be Collected

- Quantitative data

  Frequency of negative behaviors such as feeding, banging on enclosures, barrier intrusions etc. Also in addition the demographics of the visitors such as age, gender, size of groups.

- Qualitative data

  Observations from video surveillance of negative behaviors and field visits to understand context and reasons for such behavior. Along with Responses received conducting interviews with zoo officials regarding experiences of unwanted behavior.

C. Data collection tools and plan

- Surveys and questionnaires

  Structured surveys to zoo visitors to collect data on their demographics and self-reported behaviors.

Interview zoo staff on the frequency and types of negative behavior they observe and effectiveness of current prevention solutions.

- Research papers, journals and case studies

  Data and statistics regarding negative behaviors of visitors towards animals.

D. Conceptual Framework

- Independent variables
  - Current measures to prevent negative behaviors, Animal Species, Time of visit, Gender.
- Dependent variables
  - Occurrence of negative visitor behaviors

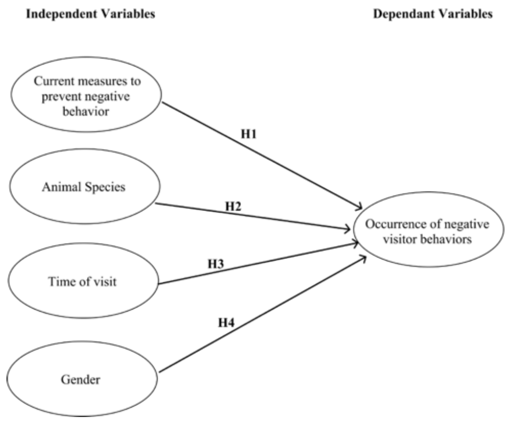


Fig. 45. Conceptual framework

IV. Data Analysis

A. Descriptive Analysis

For the purpose of gathering data and conducting the research numerous visits to the Dehiwala Zoo was done. Through observations and interviews with the zoo visitors of that day and the zoo staff data was collected. More than 90% of the staff interviewed had been working at the zoo for more than 10 years. The following analysis depicts the data received from these visits.

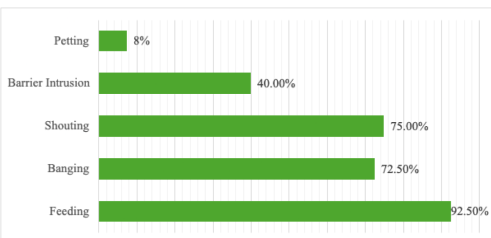


Fig. 46. Most common negative visitor behaviors

Considering Fig. 4, it shows that feeding animals is the most common visitor behavior with 92.5%, while shouting and banging on enclosures is at a close second and third with 75% and 72.5%. Barrier intrusions too play a significant role with 40%. Here enclosures can be glass or metal enclosures where zoo animals live in the zoo.

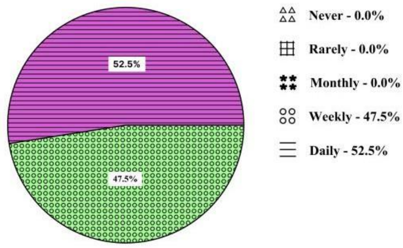


Fig. 47. Frequency of nagtive visitor behaviors

As seen in Fig. 5, the zoo staff revealed that they notice negative visitor behaviors daily and weekly. At a rate of 52.5% visitors engage in such behavior daily and at 47.5% such behaviors are seen weekly. This shows that negative behaviors are very abundant in the zoo.

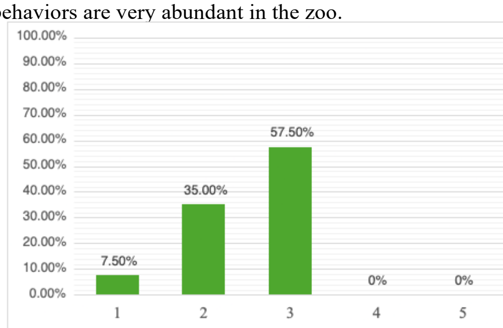


Fig. 48. Effectiveness of current measurements in preventing negative behaviors

In Fig. 6, on the x-axis 1 represents “Not effective at all” and 5 shows “Very effective”. Therefore, it can be seen that when considering employees no one is satisfied with the current methods to identify negative behaviors of zoo visitors

While the above analysis were responses from the zoo staff, the following responses and analysis will be responses from the zoo visitors received from the survey distributed.

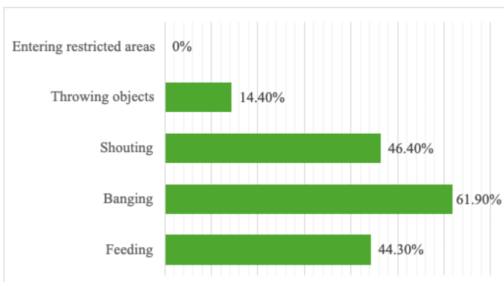


Fig. 49. Engaging in negative visitor behaviors

When considering responses directly from the visitors it can be seen that most of the time they engage in banging on animal enclosures as seen in Fig. 7 with the highest value of 62%. The behavior of feeding animals and shouting are almost similar and not far behind with 44% and 46.4% respectively.

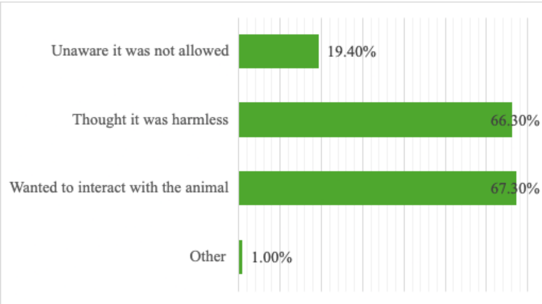


Fig. 50. Reasons for engaging in negative behaviors

According to Fig. 8, 67% of the responders' states that they engage in negative behaviors is because they want to interact with the animal and not far behind is visitors thought that it was harmless when engaging in such behavior with a percentage of 66%.

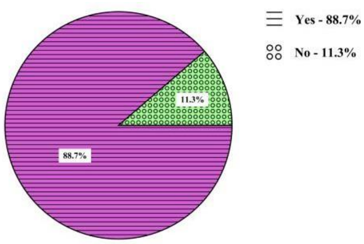


Fig. 51. Noticing other visitors are engaging in negative behaviors

When asked about if the respondent witnessed other visitors than themselves engaged in negative behaviors almost all responded with "Yes" as seen in Fig. 9 with a rate of 88.7%.

B. Findings and Interpretations

With the help of the data gathered in the Descriptive Analysis section earlier a thorough hypothesis testing will be carried out to prove/disapprove the hypothesis statements mentioned in Table I. When the independent and dependent variables are analyzed then the "p-value" is compared with its standard of 0.05 to determine if the variables are significant. If the p-value is less than 0.05 the Hypothesis will be accepted while the Null Hypothesis will be rejected, and the variable will be considered as significant to the end product. Analysis methods such as Independent T-Test and One-Way Anova tables were used to get the significance of variables. If any significance was seen, then the particular feature should be included in the final product which will be developed.

TABLE X.

Hypothesis	Significance Value	
H1 - Current traditional methods are not very effective when identifying negative visitor behaviors	0.021	**Not Rejected**
H2 - Some species rather than all species experiences more negative visitor behavior interactions	< 0.01	**Not Rejected**
H3 - Negative visitor behaviors increase with specific visitation periods	0.116	**Rejected**
H4 - Certain demographic groups are more likely to engage in negative behaviors when compared to others	0.026	**Not Rejected**

c. Hypothesis acceptance table

Considering analysis done, it is possible to come to a conclusion regarding if to accept or reject the hypothesis raised. Table I shows the hypothesis raised, the significance values (p-value) and if it was rejected or not rejected. Therefore only H3 was rejected as it has a p-value of more than 0.05. The remaining three hypothesis was not rejected and therefore they are considered into the final solution that will be developed.

V. SOFTWARE DEVELOPMENT

The following shows the end product built after data collection and analysis, by identifying the features the product should have. The system includes a combination of machine learning models, Application Programming Interfaces, Web Sockets and IoT devices. The final solution will identify four negative visitor behaviors namely, feeding/petting animals, barrier intrusions, banging on zoo enclosures and detecting loud noises.

The first scenario where identifying feeding animals will be done by training a machine learning model. A dataset of humans feeding animals in a specific pose was gathered and the model was trained by object detection as well as posture detection.

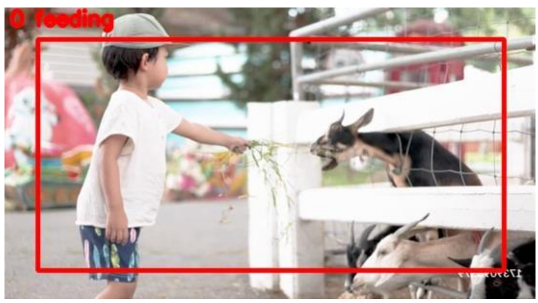


Fig. 52. Detecting human feeding a animal

The animal and human will be identified by object detection first, then if they are present closer to each other, the pose detection will find if the human is making a feeding pose towards the animal. If this object and pose are both identified, then a verbal warning will be given to the visitor in real time while an alert will be sent to the relevant zoo staff/employees with animal and location in question. Fig. 10 shows where a bounding box will be drawn when streaming surveillance footage of the zoo and any feeding behavior is identified. This model is train and built using libraries

OpenCV for video processing, YOLO for object detections (human and animal), OpenPose for the purpose of pose estimations done by the human.

The second scenario will be identifying for barrier intrusions done by visitors where they might enter animal enclosures. Since entering these places are only allowed for zookeepers and other relevant staff, a machine learning model was trained to detect non staff individuals in enclosures through video surveillance. This was done by taking the zookeepers uniform into consideration. Since all zookeepers wear a mandatory uniform given by the zoo, the machine learning model was trained to identify any clothing which was not this uniform. Therefore, if a visitor enters the enclosure it will be detected since the visitor is not wearing the desired uniform. When this is detected, an alert will be set to the relevant officials with the location in real time. For the build purpose of this model the "Ultralytics YOLO" library was used for object detection (human) and clothe detection, while OpenCV will be used for video processing and streaming.

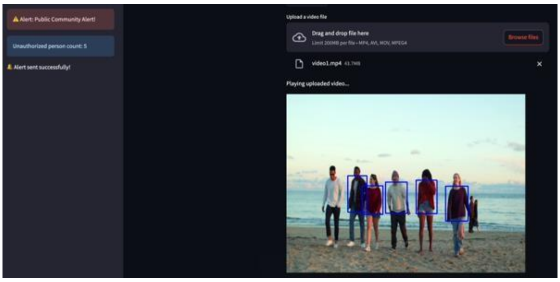


Fig. 53. Detecting intruders by classifying clothing (uniform)

For the purpose of demonstration Fig. 11 shows the machine learning model detecting humans who are not wearing a white top. The person who is wearing white, in this case the "zookeeper wearing the uniform" is not detected as an intruder. Therefore, as seen this can be used to detect barrier intrusions done by zoo visitors and alert authorities as quickly as possible.

For the third and fourth scenarios to identify banging on enclosures and detect noises above 90dB will be done by sensors and IoT devices. Piezoelectric sensors are designed to detect pressure applied to a surface and this makes it suitable to detect banging on surfaces by visitors.

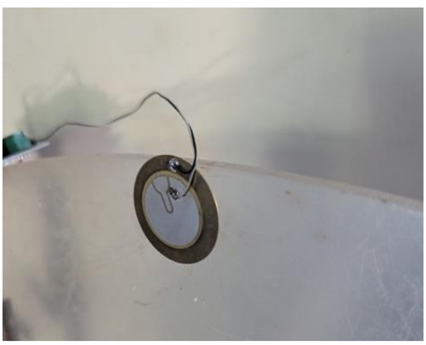


Fig. 54. Piezoelectric sensor attached on a glass surface

These Piezoelectric sensor discs will be fixed on the surface as seen in Fig. 12 and when a human bangs on the surface close to it an electrical current will be generated from the sensor through an analog value. This value will be used and configured to identify banging and differentiate from tapping. If banging is identified, then a verbal warning will be given from a sound source and an alert will be sent to the relevant zoo staff.

Next to identify if the surrounding enclosure noise is above 90dB generated by visitors a KY-038 sound sensor module was used. This sensor measures the sound and gives out an analog value where it can be taken and converted to decibel values. Therefore, if this decibel value is over 90 then as usual a verbal warning will be given to the visitors and an alert sent to the officials.

To read the sensor values and do computations a microcontroller is needed and for this an ESP-32 microcontroller board was used. This board is compact but powerful. It also has Wi-Fi built-in, therefore this makes it easier to communicate data and process.

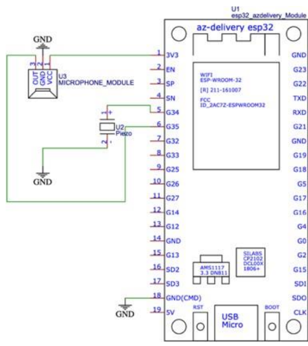


Fig. 55. Circuit diagram for IoT device with sensor modules

The method of integrating the IoT device is shown in detail in Fig. 13. After integration this device can be used to detect banging on enclosures and detect noise levels above 90dB.

DISCUSSION

The developed system has many advantages compared to the existing methods in zoos. Negative visitor behaviors can be identified and prevented before they escalate, where when considering current methods not all unwanted behaviors are identified unless the zoo staff clearly witnesses it. Also giving an auditory warning to individuals will give them a feeling as if someone is always watching and tracking movements. During data collection by surveys, it was revealed by zookeepers that most zoo animals died from eating plastic which was known only during post-mortems. With the developed system it will be possible to identify when something was fed and then the staff can take necessary actions. Other than that, a problem which is overlooked by many is the induce of stress on animals with loud noises. Therefore, the developed system also addresses this detects noises above 90dB. A drawback of the product will be is that there is no "one size – fits all" solution, where the way the product is integrated in one zoo will not be the same way it is added in another zoo to location, geographical and layout changes.

Components which use computer vision will not be as much accurate in low light conditions as in the daytime. Therefore, adding ability for the video streaming device to switch to night mode will be a great future increment. The current developed system for identifying feeding and petting animals does not detect the situation where a visitor might throw food into a distance inside the enclosure, similarly, throwing any other objects to hurt or get the attention of the animal is not integrated. By adding such a feature to identify visitors by their throwing pose and then alerting the relevant individuals will make this system even more valuable and help the welfare of animals as well. The sound detection device too can be improved more by using a more advanced microphone for sound detection and also with the help of AI distinguish between human and other noises. This way it is possible to correctly identify if the loud noises are being emitted from the visitors or other surrounding noises. The functionality of barrier intrusions can also be improved by not only detecting the zoo staffs' uniform, but also using facial recognition technology. This makes the system more accurate and prevents any false alarms by matching the intruder with the zoo staffs' facial features. These improvements will make the developed system even more accurate and less error prone.

Conclusion

The aim of this research was to identify various types of negative behaviors done by visitors at the zoo, the current practices implemented by the zoo to monitor and reduce such behavior and what will be the most effective solutions to identify negative visitor behaviors using AI where then end goal will be to reduce harm on animals done by visitors.

After the process of interviews, surveys and data collection the most occurring negative behaviors were identified as feeding animals, banging on animal enclosures, barrier intrusions and making high noises which are harmful for animals. To identify such behaviors a solution using machine learning models and IoT devices were used, when these identified any negative behaviors it will give out a verbal warning by a audio source to the individual doing the behavior and also send an alert to the relevant zoo staff with the location in real time. This helps the zoo employees to intervene and assess the situation.

This system developed by thorough research will help to reduce negative visitor behaviors occurring in the zoo and also create a safe space for the zoo animals to reside in. Most importantly it helps for the good of animal welfare and removes any unnecessary risks on their life.

Acknowledgments

The authors would like to acknowledge the individuals working in Zoological Gardens for their hard work carried out daily towards the welfare of zoo animals. Furthermore, for making this research a success by providing valuable and honest feedback. In addition, a special gratitude towards my supervisor for guiding and valuable insights given for the success of this research paper.

References

[1] J. Ogden and J. E. Heimlich, "Why focus on zoo and aquarium education?," Zoo Biology, vol. 28, no. 5, pp. 357–360, Sep. 2009, doi: https://doi.org/10.1002/zoo.20271.

[2] A. M. Godinez and E. J. Fernandez, "What is the zoo experience? How zoos impact a aisitor's behaviors, perceptions, and conservation efforts," Frontiers in Psychology, vol. 10, no. 10, Jul. 2019, doi: https://doi.org/10.3389/fpsyg.2019.01746.

[3] L. Nekolný and D. Fialová, "Zoo Tourism: What actually is a zoo?," Czech Journal of Tourism, vol. 7, no. 2, pp. 153–166, Dec. 2018, doi: https://doi.org/10.1515/cjot-2018-0008.

[4] S. L. Muller, S. L. Bissell, K. M. Cunningham, and R. Strasser, "How do you behave at the zoo? A look at visitor perceptions of other visitors' behavior at the zoo," Animal Behavior and Cognition, vol. 8, no. 4, pp. 619–631, Nov. 2021, doi: https://doi.org/10.26451/abc.08.04.12.2021.

[5] C. Collins, Y. Barr, S. McKeown, J. Scheun, C. Tay, and R. M. O'Riordan, "An International Investigation of the Prevalence of Negative Visitor Behaviour in the Zoo," Animals, vol. 13, no. 16, pp. 2661–2661, Aug. 2023, doi: https://doi.org/10.3390/ani13162661.

[6] E. N. Parker, L. Bramley, L. Scott, A. R. Marshall, and K. E. Slocombe, "An exploration into the efficacy of public warning signs: A zoo case study," PLOS ONE, vol. 13, no. 11, p. e0207246, Nov. 2018, doi: https://doi.org/10.1371/journal.pone.0207246.

[7] A. Hashmi and M. Sullivan, "The visitor effect in zoo-housed apes: the variable effect on behaviour of visitor number and noise," Journal of Zoo and Aquarium Research, vol. 8, no. 4, pp. 268–282, Oct. 2020, doi: https://doi.org/10.19227/jzar.v8i4.523.

[8] C. K. Collins and N. M. Marples, "The effects of zoo visitors on a group of Western lowland gorillasGorilla gorilla gorillabefore and after the birth of an infant at Dublin Zoo," International Zoo Yearbook, vol. 50, no. 1, pp. 183–192, Jan. 2016, doi: https://doi.org/10.1111/izy.12111.

[9] M. J. Learmonth, S. J. Chiew, A. Godinez, and E. J. Fernandez, "Animal-Visitor Interactions and the Visitor Experience: Visitor Behaviors, Attitudes, Perceptions, and Learning in the Modern Zoo," Animal Behavior and Cognition, vol. 8, no. 4, pp. 632–649, Nov. 2021, doi: https://doi.org/10.26451/abc.08.04.13.2021.

[10] A. M. M. Dancer and C. C. Burn, "Visitor effects on zoo-housed Sulawesi crested macaque (Macaca nigra) behaviour: Can signs with 'watching eyes' requesting quietness help?," Applied Animal Behaviour Science, vol. 211, pp. 88–94, Feb. 2019, doi: https://doi.org/10.1016/j.applanim.2018.12.005.

[11] R. Jakob-Hoff et al., "Potential Impact of Construction Noise on Selected Zoo Animals," Animals, vol. 9, no. 8, p. 504, Jul. 2019, doi: https://doi.org/10.3390/ani9080504.

[12] G. J. Mason and J. S. Veasey, "How should the psychological well-being of zoo elephants be objectively investigated?," Zoo Biology, vol. 29, no. 2, pp. 237–255, Mar. 2010, doi: https://doi.org/10.1002/zoo.20256.

[13] "Annual Performance Report 2022 Department of National Zoological Gardens Ministry of Wildlife and Forest Resources Conservation." Available: https://www.parliament.lk/uploads/documents/paperspresented/16889 65271074463.pdf

Web Based AI-Driven Tool for Smart Contract Auditing and Vulnerability Detection

Hettiarachchi P A
School of Computing and Engineering
National Institute of Business Management
Colombo, Sri Lanka
cobsccomp4y222p-012@student.nibm.lk

Wadumulla M W
School of Computing and Engineering
National Institute of Business Management
Colombo, Sri Lanka
wadumullamw@gmail.com

Abstract **- This research paper focuses on the critical security challenges that are currently available in blockchain ecosystems by developing of an AI-driven, web-based Smart Contract auditing and vulnerability detection tool. Traditional auditing methods, such as manual auditing often fail to detect complex vulnerabilities, like reentrancy, logical errors and integer overflows. Also, those methods often face limitations in scalability and efficiency. To address these challenges, in this research paper, this tool integrates Artificial Intelligence (AI), Natural Language Processing (NLP) models, and Machine Learning (ML) models to enhance semantic understanding and predictive capabilities in vulnerability detection in smart contracts. The tool was designed and implemented based on feedback from developers, auditors, and investors, in the industry using conduction a survey to ensure that these features meet real-world requirements. In this research paper it focuses on Smart Contracts which are coded and complied in solidity language. Key functionalities include the ability to analyze multiple Solidity contracts once at a single time, provide detailed detected vulnerability reports, and give actionable remediation suggestions to the user to get rid of the detected vulnerabilities. The developed system is rich with an intuitive user interface, ensuring that both technical and non-technical users to interact seamlessly and understand the system. A comprehensive evaluation was conducted by using a real-world Ethereum smart contract dataset, the ML model used for this provided high performance with an overall accuracy of 94%, precision of 92%, and significant reductions in false positives. The tool integrates with an NLP model which can provide remediations for the detected vulnerabilities to get rid of them before deploying it to the blockchain. This research paper also addresses gaps such as limited scalability, low user-friendliness, and insufficient NLP and ML integrations. By improving the integration of NLP and ML with blockchain technology, this tool promotes broader adoption and trust in decentralized applications.**

Keywords - ***Smart Contracts, Blockchain, Artificial Intelligence, Vulnerability Detection, ML, NLP***

I. INTRODUCTION

Smart contracts have transformed the blockchain ecosystem, enabling automated, decentralized execution of agreements across industries like banking, supply chain, healthcare, and fundraising. However, their adoption has been hindered by vulnerabilities that traditional auditing methods fail to detect comprehensively. The manual nature of traditional auditing often results in inefficiencies, inaccuracies, and the risk of human error. Recent advancements in Artificial Intelligence (AI), particularly Natural Language Processing (NLP) and Machine Learning (ML), offer promising avenues to improve smart contract auditing. By integrating NLP and ML, this research aims to enhance semantic understanding and vulnerability detection in smart contracts, providing developers and auditors with an innovative, user-friendly AI-powered tool. This paper outlines a novel approach combining NLP and ML to examine smart contracts for vulnerabilities. It details the development of a web-based auditing tool, its features, and its implications for the blockchain ecosystem. The study also explores gaps in existing methodologies, the potential of AI to address these gaps, and the challenges associated with implementing such solutions.

II. LITERATURE REVIEW

a. *Introduction to the Research Theme*

Smart contract auditing has evolved from manual inspections to employing automated tools, yet limitations persist. Rule-based systems often lack adaptability to evolving threats, while ML-based approaches struggle with interpretability and semantic analysis. Existing tools also face scalability issues, especially when analyzing large datasets or complex contracts. Researchers have explored various strategies to address these challenges. Rule-based methods rely on predefined patterns to identify vulnerabilities but fail to adapt to new threats. ML approaches have introduced predictive capabilities, leveraging historical data to identify vulnerabilities. NLP integration has enhanced understanding by analyzing comments and code structures, but its application remains limited. Hybrid methods combining rule-based systems with ML and NLP hold potential but require further refinement [1].

b. *Theoretical understanding of Main Concepts*

Blockchain technology and Smart Contracts

Smart contracts exist in the blockchain domain where they are self-executing, non-tweakable digital agreements. The conditions of these contracts are written into code, which self-execute when specific conditions are found [2]. There are many advantages associated with using such concepts; these include; Increased efficiency and effectiveness when compared to third parties. However, as will be later discussed, the architectural setup of modern blockchain systems poses its own risks.

Smart Contract Security

Security is also a big concern when it comes to smart contracts since their weaknesses mean that an attacker might steal from users and compromise the system. These are as follows: Reentrancy is when a contract returns ether, and another contract calls the original contract back Storing intermediate values that can be used to modify an index and cause integer overflow or underflow. In some cases, a programmer might make a basic and obvious error in code

[4]. Dealing with these threats entails both the knowledge of the basic blockchain infrastructure as well as the script codes of the executed contract.

Natural language processing and Machine learning.

Smart contract auditing is done with the help of AI technologies such as Natural Language Processing (NLP) and Machine Learning (ML). NLP allows extracting meaning from high level languages such as Solidity, while ML highlights features and deviations in contract codes [5]. These technologies improve effectiveness of threat identification and make it automated.

c. Theoretical understanding of Main Concepts To carry out this research,

I used the following three finding by other researchers:

Rule-Based Approaches

In the early years, smart contract auditing has been based on rule-based technologies which identify threats based on certain patterns [3]. Although these methods are appropriate for addressing routine concerns, they are not very useful for new and complex threats because they are non-iterative.

Machine Learning Approaches

Historical data has helped in training models in methods based on ML in enhancing vulnerability detection. Such approaches bin contracts into ICs and N IRCs with the provision of predictive capacities [3]. Therefore, certain problems come with it such as the methods interpretability and its effectiveness when working at scale.

Second, natural languages processing or NLP approaches

NLP has shown promise in auditing smart contracts show that meaning of code can be semantically analyzed. It improves the notion of contract thinking especially when used alongside the ML methods [19]. However, it will be seen that although this promises quite a lot, there still remains quite a limited amount of integration of NLP in this area.

Hybrid Approaches

Hybrid can use rule-based systems, ML, and NLP and each of them carries out its specialty because of its advantage. [20] explained that this has a positive impact on the identification of vulnerability and has fewer false positives. Nonetheless, the changes introduced by these systems need further development to become universal.

d. Research Gaps

Limited NLP Integration

While NLP was adopted in smart contract auditing, using ML in combination with NLP for semantic analysis is still in its initial stage [6]. This gap seems to be opening doors for creating better instruments or means of analysis. This caused a lack of comprehensive evaluation in my opinion; the following are reasons why the UMDS moving to Southwark rarely benefited from comprehensive evaluation: Recent research is on constrained scope, encompassing only selected types of vulnerabilities or small sets of data, which degrades the generalization potential [17]. Various assessment parameters are essential to establish confidence in these applications.

Scalability Issues

Most of the current solutions face challenges when it comes to handling multiple datasets and across many blockchain platforms which hampers their usage at scale [7].

User-Friendliness

Some auditing tools fail to include well-designed UIs which allows non-technical users to easily deploy them. In their current condition, there might be several users who do not find it very user-friendly or easy to use, increasing usability could therefore lead to increased usage [8].

e. Summary

In the literature, authors found that there have been tremendous shifts in smart contract auditing especially through, the adoption of artificial intelligence technologies. It is noteworthy that while rule-based approaches are simple and unambiguous, they are not very flexible, and that ML brings the ability to predict outcomes for a cost of comprehensibility. It also underlines that NLP improves semantic analysis but is being used not widely enough. Mixed mode strategies offer one way forward but are still at the conceptual level. Challenges in the integration of NLP in various contexts, lack of suitable evaluation metrics, limited scalability of many solutions, and the lack of user-friendly interfaces will be significant for future work in this field.

III. METHODOLOGY

The research employs a mixed-methods approach combining qualitative and quantitative techniques to design and evaluate the proposed AI-powered smart contract auditing tool. Key components include:

A. Population, sample and Sampling technique

Population

This study focused on all smart contracts deployed on various blockchain platforms, including Ethereum. This covers those serving in different applications, from DeFi, NFT to supply chain management. These smart contracts vary in their complexity, size, and security protocols to give a wide spectrum for the analysis.

Sample

Our sample will be a selected subset of smart contracts from the overall population and include a representative mix of contracts across complexity, usage, and known security issues. The purposive selection will guarantee diversity and completeness within the capturing of different types of vulnerabilities and coding practices common to the blockchain ecosystem. Since this involves smart contract deployed to the blockchain platforms, determining an exact population will be challenging. I've applied a standard sample size formula for stratified sampling here as follows:

$$n = \frac{Z^2 \times p \times (1-p)}{E^2}$$

Where Z is the Z-Score that will be chosen from the confidence level, p is the estimated proportion for smart contracts with vulnerabilities and E is the margin of error. After using 95% of confidence level, 50% of proportion where smart contracts with vulnerabilities and 5% for margin of error this equation has given me a sample size of 384.

Sampling Technique

In getting an accurate representation of the data and an unbiased sample, the following is the sampling technique to be used:

- Stratified Sampling

  Employ a plan to slice up the population based on this characteristic to create strata such as blockchain platform, contract size, and application domain (e.g., DeFi, NFTs). Subsequently, make a proportionate sample from each of the strata to ensure all important segments are taken care of.

Integrating stratified sampling technique would empower a robust, extensive, and comprehensive sample that can train and validate AI models in auditing smart contracts for the detection of vulnerabilities.

B. *Data Collection*

Data were collected from public blockchain datasets and repositories, focusing on Ethereum smart contracts written in Solidity. A stratified sampling technique ensured diverse representation based on contract size, platform, and application domain. The dataset was preprocessed to eliminate redundancy, ensure consistency, and balance classes for ML training.

C. *Model Development*

The AI model integrates:

- NLP Techniques: Used to analyze code semantics, comments, and logical structures, enhancing understanding of vulnerabilities.
- ML Algorithms: Evaluated several algorithms, including Random Forest, Gradient Boosting, and Neural Networks, to determine the most effective for vulnerability detection.

D. *Tool Development*

A web-based dashboard was developed to facilitate interaction between users and the auditing tool. Key features include:

- File Uploader: For contract submission.
- Feature Visualization: Displaying extracted features and identified vulnerabilities.
- Remediation Suggestions: Offering actionable insights to address vulnerabilities.

UI Screens

Fig. 56. Home screen with file uploading feature

- Home screen with the file uploading feature. This will only allow files which are written in solidity language.

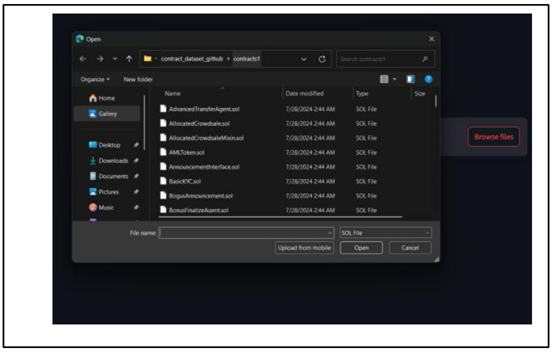


Fig. 57. File upload dialog box

- By this dialog box the user can navigate and select the solidity files which wanted to detect the vulnerabilities.

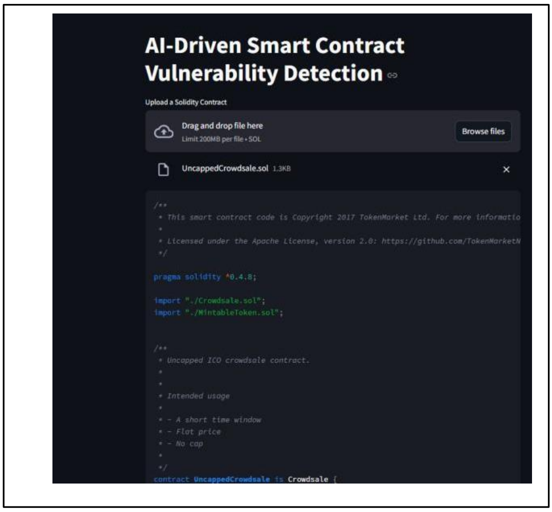


Fig. 58. Uploaded code review screen

- The user can review the uploaded code when it is successfully uploaded

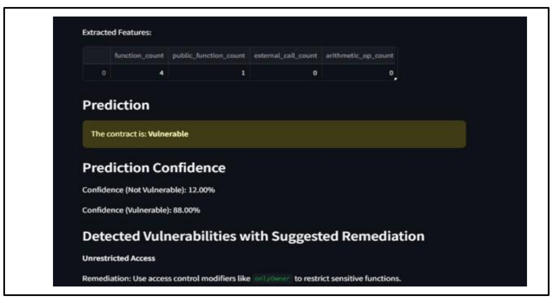


Fig. 59. Vulnerability prediction results and remediation screen

- After uploading the file the ML and NLP models will predict the possible vulnerabilities of the code that the user uploaded and will display, also the NLP model will give a descriptive remediation to overcome the detected vulnerabilities.

E. Evaluation

The tool was rigorously tested on real-world smart contracts, measuring accuracy, precision, recall, and F1-score. Statistical significance was assessed using ANOVA, while user feedback provided insights into usability and effectiveness.

IV. DATA ANALYSIS

This technical analysis of the results of the survey shows that the respondents included all the key stakeholders of the blockchain services which were developers (45%), auditors (35%) and investors (20%). That is why this distribution shows that almost half of the respondents work directly with the creation of smart contracts and understand the issues raised in this work. The largest group with more than one third of the sample is the auditors who give information about the awareness of vulnerability detection activities Investors create a strategic outlook on the importance of audit tools.

Thirty said they had 2–5 years of experience, which points out to a skilled but still evolving sample group. 25% for over 5 years' experience to support the conclusion that this sample included many representatives of this category. Ethereum came out as winners, with 60% of the respondents primarily developing on this platform. Fifty respondents had heard of Solidity, the principal language of building applications on Ethereum before the survey. This data also reveals Ethereum's importance to the blockchain industry and justifies the decision to target only tools associated with Ethereum based smart contracts.

A majority of respondents indicated that they are somewhat to highly familiar with smart contract concepts with respondents indicating roughly 70% having come across vulnerabilities in their work. Here, only 45% of the respondents indicated they were aware of existing auditing tools, suggesting perhaps the non-availability or unawareness of auditing tools. This result raises the possibility of enhancing dissemination of and user sensitization to state-of-the-art auditing techniques.

Some of the sentiment's voicings were positive, 50% voicing trust or high trust imputing the positive impact of AI. 30% of them had a distrust or high distrust toward the AI model and 25% of them stated their distrust because of the inaccuracy and 20% due to being non-interpretable. Such concerns are well founded, such that it points the need that ensure that such issues are addressed in order to enhance and encourage the adoption of the m-money.

The most widely reported benefit (30%) was to gain better semantic insight into smart contract code. About 25% of the respondents that provided feedback saw NLP as useful for interpreting code comments and 20% said it was relevant to recognizing complicated logical risks. The respondents also provided an expectation that integration of ML would improve accuracy in the detection of vulnerabilities with 35% indicating this as their expectation. Targeting a high percentage on emerging vulnerabilities (25%) and minimizing false positives (15%) were also identified as potential problem fields.

The presence of potential discrepancies in identifying the vulnerabilities (25%). Human auditor threat (20%), further indicating labor market concerns. Computationally intensive models (20%) with only limited interpretability, indicating the necessity of explain ability.

In fact, machine learning model show an excellent or near perfect level of accuracy and precision of 94 and 92 percent respectively to the conclusion that the machine learning is indeed effective. These results for all propositions were statistically significant at $p < 0.05$, as evidenced by results of ANOVA tests; hence, based on uses of these findings, the advocates for the reliability of these proposed approaches.

The survey results depict a technologically experienced group participating in blockchains with a primary interest in Ethereum and Solidity. The willingness to use AI solutions coupled with the potential concern of accuracy coupled with the need to interpret the results pointed out the importance of efficient and accurate solutions. Other gaps raised include.

Hypothesis Testing

To determine the validity of whether the integrating NPL and ML tools in our application, I'll be using the Chi Square approach since we are dealing with categorical frequency data.

Data provided by the questionnaire:

- Enhancing semantic understanding of code: 30
- Assisting in code comment analysis: 25
- Identifying complex logical vulnerabilities: 20
- Improving human interpretability of results: 15
- Simplifying report generation: 10

Observed frequencies will be as follows:

$$O = [30, 25, 20, 15, 10]$$

Thus, the expected frequency E for each category id:

$$E = \frac{Total\ Responses}{No\colon of\ Categories} = \frac{100}{5} = 20$$

$$E = [20, 20, 20, 20, 20]$$

Formula for calculating chi-square test statistics:

$$X^2 = \sum \frac{(O - E)^2}{E}$$

Summing the calculated values:

$$X^2 = 5 + 1.25 + 0 + 1.25 + 5 = 12.5$$

Using the chi-square distribution:

$$df = no\colon of\ categories - 1 = 4$$

Therefore, corresponding to a chi-square statistic of 12.5 is giving us the p-value of 0.014 approximately.

Since the p-value of 0.014 is less than the significance level of 0.05, we reject the null hypothesis. So, finally this indicates the differences in respondents across the categories are significant. Which means integrating NLP will improve detecting complex vulnerabilities in smart contract code where often missing by using the traditional tools.

V. DISCUSSION

AI technologies in smart contract auditing have therefore improved on old, aged methodologies of auditing. A component especially relevant to this is the Natural Language Processing (NLP) that helps to narrow semantic gaps between the human and code and thus helps to better identify intricate vulnerabilities that could otherwise remain unseen. That is why Machine Learning (ML) augments this environment by providing prediction-based functionality, learning about new threats that emerge while at the same time avoiding false positives. Individually and collectively, these technologies form a flexible and reliable environment to detect and manage risks associated with smart contracts [16].

This work has also used NLP to show that it can be useful for understanding the semantics of smart contract codes. Analyzing [19] observations, it is feasible to conclude that the integration of NLP with ML amplifies the efficiency and reliability of vulnerability of detection. For instance, in relation to criticism on how 'black box' methodologies, NLP helps to clarify a decision made within a program by identifying logical fallacies within the code and also helps the human brain to interpret results. Other approaches that combine rule-based model with NLP and ML have also elicited positive results. Tang et al. (2023) notes that such strategies enhance the identification of these risks; and decrease the number of false positives, resulting in better dependability of AI instruments [15].

However, the literature review has identified the following gaps as follows. It is also evident that there is a lack of research on integrating NLP and ML to blockchain platforms but can develop models that are scalable across different blockchains [10]. One of the limitations is that the evaluations have not been done comprehensively across multiple datasets. Previous works are usually based on particular kinds of weakness or use limited samples, thus limiting the utilization of the AI techniques in out-of-sample practice [8]. Also, some of the challenges include dataset standardization, threats that are ever changing and generalization across the platforms. Solutions to these problems can contribute significantly towards developing better and more appropriate auditing solutions worldwide.

Comparative analysis of algorithms using precision, recall, and F1-scores disclose a heightened in accuracy of vulnerability detection when implementing NLP and ML. Recurring patterns identified from expert interviews also highlight the practical value of AI-based tools as applied to workplaces [12]. Some of the examples discussed above show practical weaknesses; examples also illustrate how AI technologies counter such risks convincingly. Thus, it has been ascertained that users have a rather guarded positive attitude towards the use of AI-based tools in auditing. With 50% of the respondents having trust in these tools, issues like the possibility of having wrong results and interpretational challenges are still felt. As we said, it will be essential to address these issues to enhance acceptance among developers and auditors to incorporate software verification more into practice.

A. Advantages of the Project

The project delivered several notable benefits, both in terms of practical application and academic contributions

Automation of Smart Contract Auditing

One of the most significant advantages of the tool is its ability to automate the smart contract auditing process. Traditional methods rely heavily on manual code reviews, which are time-consuming, expensive, and prone to human error. The machine learning model developed in this project can analyze contracts in real-time, providing instant feedback to developers and auditors. This automated process not only saves time but also scales easily to handle large numbers of contracts, a necessity in the rapidly growing blockchain ecosystem.

Improved Accuracy with Machine Learning

The use of machine learning, specifically the decision tree classifier, significantly improved the accuracy of vulnerability detection compared to rule-based tools. The model learned from patterns in the contract code, allowing it to detect more subtle vulnerabilities that might not be flagged by traditional static analysis tools. The model's high precision and recall metrics indicate that it performs well in distinguishing between vulnerable and non-vulnerable contracts.

User-Friendly Interface

The tool includes a user-friendly web-based interface that allows users to upload their contracts, view the extracted features, and receive detailed reports on vulnerabilities. The ease of use was a priority in the design of the tool, making it accessible to both experienced auditors and developers without deep knowledge of security. The tool provides clear remediation suggestions when vulnerabilities are detected, offering users actionable steps to improve the security of their contracts.

Transparency and Explainability

Unlike some more complex machine learning models, the decision tree classifier is inherently interpretable. Users can easily understand the decision-making process of the model, which increases trust in the predictions. The model's transparency allows 37 auditors and developers to see which features or patterns led to the classification of a contract as vulnerable or non-vulnerable.

B. Limitations of the Project

While the project delivered promising results, it is important to acknowledge the limitations and challenges encountered:

Limited Dataset

One of the primary limitations of this project is the size and diversity of the dataset used for training the machine learning model. Although the dataset included a range of smart contracts, it could be expanded to include more diverse contracts from different sectors and use cases (e.g., DeFi, NFTs). A larger dataset would help the model generalize better and improve its ability to detect fewer common vulnerabilities.

Imbalanced Data

The dataset used in this project exhibited a slight imbalance, with more non-vulnerable contracts than vulnerable ones. Although this was mitigated through up sampling of the minority class, imbalanced data remains a challenge in real-world applications. A more balanced dataset could improve the model's ability to detect rare vulnerabilities and prevent overfitting to the majority class.

Narrow Focus on Specific Vulnerabilities

The model was trained to detect a specific set of vulnerabilities (e.g., reentrancy, unchecked low-level calls, and integer overflow). However, smart contracts can suffer from a wide variety of security issues, including gas usage optimization, access control flaws, and denial-of-service attacks. The scope of this project did not cover these more advanced or nuanced vulnerabilities, leaving room for future work to expand the range of detected vulnerabilities.

Limited Remediation Suggestions

While the tool provides remediation suggestions for detected vulnerabilities, these suggestions are based on a limited set of patterns. More sophisticated remediation advice could be implemented in the future, offering tailored feedback depending on the context of the contract and the severity of the vulnerability.

VI. RECOMMENDATIONS

Based on the findings and limitations of this project, several recommendations can be made for future work:

A. *Expand the Dataset*

To improve the robustness and generalizability of the machine learning model, future work should focus on expanding the dataset. This could involve collecting a larger and more diverse set of smart contracts, including contracts from newer sectors such as decentralized finance (DeFi) and non-fungible tokens (NFTs). A larger dataset would also allow for more advanced machine learning techniques, such as deep learning, which require substantial amounts of data.

B. *Incorporate Additional Vulnerabilities*

While the current tool focuses on detecting a specific set of vulnerabilities, future iterations could expand the scope to include more complex and nuanced security issues. For example, the model could be trained to detect gas usage inefficiencies, denial-of service attacks, or access control flaws. Incorporating these additional vulnerabilities would make the tool more comprehensive and valuable for developers and auditors.

C. *Utilize More Sophisticated Machine Learning Models*

Although the decision tree classifier performed well, future work could explore more sophisticated machine learning models, such as Random Forests, Gradient Boosting, or Deep Learning models like Neural Networks. These models could potentially offer better performance, especially when trained on larger datasets. Ensemble methods, such as bagging or boosting, could also be used to improve the model's accuracy and reduce variance.

D. *Enhance the Remediation System*

Future versions of the tool could incorporate more detailed and context-aware remediation suggestions. Instead of offering generic advice (e.g., "use SafeMath for arithmetic operations"), the tool could analyze the contract's overall structure and suggest targeted improvements based on best practices for secure contract 39 development. Integrating expert knowledge into the remediation system would make the tool even more valuable to developers.

E. *Integrate Real-Time Auditing*

Although the current tool provides real-time analysis of uploaded contracts, future enhancements could include real-time auditing during contract development. This would allow developers to receive immediate feedback as they write or modify their contracts, improving security at the earliest stages of development. Integrating the tool with popular development environments such as Visual Studio Code or Remix IDE could make this feature more accessible.

VII. CONCLUSION

Smart contract auditing and vulnerability detection through the use of AI remain a groundbreaking solution to the issues that are prevailing in blockchain. This paper reveals significant research gaps that require incorporating NLP and ML approaches to optimize the effectiveness, reliability, and interactive nature of smart contract audit. Combining these technologies, the proposed framework enhances semantics comprehension and prognostic prowess beyond current practices such as scalability, readability, and user-friendliness.

These findings further confirm the ability of these AI-driven tools to identify insecurity with high levels of precision and recall rates, with an overall accuracy of 94%, evidence high-performance on Ethereum smart contract. User feedback zooms in on aspects such as better semantic analysis, the reduction of false positives and remedial recommendations that are key to the understanding, acceptance and utilization by developers, auditors and investors.

AI application still has drawbacks, such as computational concern and limited interpretability, but it can be solved by the existing of ethical future research areas, such as explainable AI and user awareness. There is a potential for stronger synergies between scholars, practitioners, and businesses together with the use of hybrid approaches, such as rule-based systems and NLP with or without ML integration to enrich the auditing model and discover more robust auditing methods by using a combination of multiple approaches at the same time.

VIII. Acknowledgments

I would like to express my gratitude my supervisor, the head of the computing department and all the staff of the computing department for the guidance, constant support and insightful feedback throughout this research. Their experience and motivation have played a curial role in achieving these objectives and make it a success.

REFERENCES

[1] T. a. A. H. Abdelaziz, "Schooling to Exploit Foolish Contracts," *UCL Discovery (University College London),* 2023.

[2] doi:https://doi.org/10.1109/bcca58897.2023.10338924, "Evolution of Automated," *arXiv (Cornell,* 2023.

[3] A. N. R. A. M. a. I. M. Bhat, "Methodology and Analysis of Smart Contracts in Blockchain-Based International Trade Application," *arXiv (Cornell University),* 2021.

[4] H. Chakka, "How Decision Intelligence and Artificial Intelligence Differ," *Analytics Insight,* 2023.

[5] I. Z. L. Q. K. S. D. C. David, "Do you still need a manual smart contract audit?," *arXiv.org,* 2023.

[6] D. (2006), "Statistical Comparisons of Classifiers over Multiple Data Sets," *Journal of Machine Learning Research,* 2006.

[7] B. C. D. V. L. Devi Alagarswamy, "Arndt von Twickel and Woitschek, F. (2023)," *Towards Audit Requirements for AI Based Systems in Mobility Applications. arXiv (Cornell University),* 2023.

[8] T. F. J. A. R. a. C. P. Durieux, "Empirical Review of Automated Analysis Tools on 47,587 Ethereum Smart Contracts," *arXiv:1910.10601 [cs],* 2020.

[9] N. a. J. U. Forman, "Smart Contracts in Construction from Management to Materials: Acting as a Bridge between the Two," *Deleted Journal,* vol. 6, p. 43–48, 2023.

[10] D. R. S. B. M. C. D. S. Fucci, "A Longitudinal Cohort Study on the Retainment of Test-Driven Development," *arXiv,* 2022.

[11] D. L. Q. Z. L. a. X. T. Han, "A Smart Contract Vulnerability Detection Model Based on Syntactic and Semantic Fusion Learning," *Wireless Communications and Mobile Computing,* p. 1–12., 2023.

[12] H. R. a. B. Ghita, "Artificial Intelligence Applied to Contracts: Panacea or Poison?," *International Journal For Multidisciplinary Research,* vol. 5, no. 3, 2023.

[13] J. a. R. V. Leavins, "Improving Internal Control Over Fixed Assets with BLOCKCHAIN," *International Journal of Business & Management Studies,* vol. 4, no. 6, p. 01–04, 2023.

[14] M. a. C. C. Lewis, "Architectural Design for Secure Smart Contract Development," *arXiv.org,* 2023.

[15] H. F. Y. F. L. a. W. Z. Liu, "Vulnerable smart contract function locating based on Multi-Relational Nested Graph Convolutional Network," *Journal of Systems and Software,* p. 111775–111775, 2023.

[16] J. a. L. Z. Liu, "A Survey on Security Verification of Blockchain Smart Contracts," *IEEE Access,* p. 77894–77904, 2019.

[17] C. K. V. a. D. S. Malhotra, "ETHICAL FRAMEWORK FOR MACHINE LEARNING," *2018 ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K),* 2018.

[18] P. a. L. U. Neumann, "The IEEE Symposium on Security and Privacy Is Moving to San Francisco," *IEEE Security & Privacy Magazine,* vol. 10, no. 2, p. 65–66, 2012.

[19] S. F and P. Remo, "Enhancing blockchain security through natural," *International journal of,* pp. 1-16, 2023.

[20] X. Tang, Y. Du, A. Lai, Z. Zhang, and L. Shi, “Deep learning-based solution for smart contract vulnerabilities detection,” *Scientific Reports*, vol. 13, no. 1, p. 20106, Nov. 2023, doi: https://doi.org/10.1038/s41598-023-47219-0.

Smart Web Application Using Machine Learning for Crowd Risk Assessment and Safety Management at Mass Gathering Events.

Ranasinghe R A R K
School of Computing and Engineering
National Institute of Business Management
Colombo 07, Sri Lanka
cobsccomp222p-052@student.nibm.lk

Jayakody J A D G S
School of Computing and Engineering
National Institute of Business Management
Colombo 07, Sri Lanka
gayan@vl.nibm.lk

***Abstract*—People as social creatures like to get together in groups at most of the times. The need to live together in groups causes severe overcrowding in particular events that produce congestion which leads to fatal stampedes. The lack of suitable solutions for risk identification in mass gathering events causes thousands of injuries and fatalities annually. This research develops a software including machine learning (ML) technologies which enables crowd risk analysis to assist event management stakeholders in reducing crowd disasters during events. This provides an evaluation of present solutions along with pre-event planning practices and crowd control methods. 128 mass event planners from Western province of Sri Lanka, have been selected through cluster sampling method. The software system included several development stages with modern ML approaches which included Multi-Linear Regression Models, Optical Character Recognition (OCR), Large Language Models (LLM) and Retrieval-Augmented Generation (RAG) for LLM enhancement. The predictive accuracy reaches 88% that allows the software to generate crowd density heatmaps together with peak-time statistics about overcrowding conditions and specific recommendations for risk reduction. Analysis results indicate safety problems mainly affect crowds exceeding 1,000 participants but proper pre-event risk analysis effectively minimizes these challenges.**

Keywords—Risk Analysis, Crowd Management, Stampedes, Congestion, Crowd Density, Mass Gathering, Crowd Behavior, Machine Learning

I. INTRODUCTION

Individuals all over the world are social beings, and they always engage in events that they expect will bring people together. However, that puts them at risks such as congestion, and stampedes. In the past, people numbering in millions, have participated in these functions, with hundreds of thousands dying in stampedes. It is not easy physically and psychologically to be compressed in a crowded and packed environment, and thus such people end up being panicked. This results in stampedes, slight suffocation and in extreme cases leads to considerable casualty.

Free guaranteed solutions regarding such cases are still available in the media despite the presence of many solutions suggested to prevent such events from occurring. For example, in October 2022, there was a stampede in Seoul-South Korea during Halloween celebration where such a tragedy that led to the death of at least 153 people and many others injured [1]. The same year a football stampede occurred in Indonesia in which 135 died and more than 700 were injured [2]. Another crowd crush was in November 2021 valid at "Astroworld" Festival, a music festival in Texas, United States that claimed ten lives and many others injured [3]. Additionally, in April 2021, dozens of people were killed, and numerous others injured in a crush during the post-Pesach pilgrimage at Mount Meron in Israel [4]. For example, stampede in the religious occasion that took place in Uttar Pradesh-India in July 2024, killed at least 121 females and children [5]. Study of these occurrences shows a sharper and continuous disjoint within delivery and productivity of implementations of these measures in congregation and populous places. It is concerning to note that crowd disasters happen more frequently, even with the current enhanced technology [6]. Some of the causes of these calamities include overcrowding, lack of planning, lack of facilities and negatives crisis management [7]. Therefore, to avoid such occurrences, it is high time that effective crowd management measures were put in place and implemented to respond to risks before they occur [8]. However, proper and affordable solutions for overcoming these problems commercially remain scarce [9].

The purpose of this research is to develop a system that provides insights and recommendations for minimizing the number of deaths and injuries caused by fans at mass spectacles. The proposed system incorporates inputs such as event details, ground plans, weather forecasts and historical data to generate congestion probabilities and recommend probable preventive actions. It will include areas at risk, the best layouts for event space to reduce risk factors and the measures needed to do so. It also helps the management teams, security personnel and the organizations to have a real, reliable and simple solution toward the dynamics of crowds.

II. LITERATURE REVIEW

Questions associated with overcrowding; stampedes connected with crowding have been thought of for a few decades now. Minorities, riots, turbulent and stampede conditions are events of rather high risk and they cause many casualties and their incidence increases and transactions reveal that the number of casualties and injuries in relatively high-risk events are higher than in other cases [10]. In this study by [9] on mass gatherings in 2023 assess that more than 7000 people have perished, from the year 1990 to 2022. However, even increased approaches to managing crowds discussed by [11] for risk assessment, how congestion progresses and the measures that may be taken to avert disasters, remain a clear absence.

AI and ML along with IoT technologies have advanced the processes of risk analysis and crowd management beyond imagination. In their study [12] employed YOLOv5 and DeepSORT for anomalous detection and grouping of individuals with similar behaviors, while [13] designed a YOLOv4 model with an Adaptive Attention mechanism for facial detection during Hajj. It is proposed that Convolutional Neural Networks (CNNs) are especially valuable for crowds' analysis. In their study [14] showed that they are capable of crowd size estimation, crowd behavior analysis and a density assessment. Also [15] proposed MCNN approach for behavior estimation, with OpenCV for image pre-processing of density mapping and features.

Due to the development of IoT-based systems, effectiveness of real-time crowd analysis has improved. As mentioned by [16] all the infrared cameras, RFID, GPS and Bluetooth networks were integrated for this end. Also [17] reviewed IoT-related applications for crowd disaster prediction and prevention, [18] examined the efficiency of Switching-Based Algorithm (SBA) for IoT systems.

Crowd counting has also progressed with [19] to bring Fourier-Guided Attention (FGA) to enhance efficiency, while [20] have also presented counting method using regression models, deep learning and linear mapping techniques. Simulation models are still important for forecasting crowd behavior, although newer models are described in [21] and [22].

However, the following research gaps still exist to limit crowd related risk management. Recent methodologies like those in [13] on the development of Advanced Video Surveillance Systems (VSS) has brought new possibilities to crowd management. Yet, some obstacles have been met imprecise results, possibility to expand for broader applications, real-time computation, and data management and access [23]. The following constrains the application of the above solutions in mass gatherings due to low access to high powered computation among stakeholders in the event. The applications such as IoT-based crowd management discussed by [16] and [18] have major issues in terms of device stability, conductivity, high costs, and integration issues for large-scale events. Privacy issues remain as a big question, especially where they are incorporated in real-time monitoring systems like the one proposed by [24]. They depend on video feed and public data which requires stronger implementation of privacy and ethical measures to improve the adherence to the systems. Apart from technical challenges, one of the most significant issues inhibiting appropriate crowd management is fragmentation of actors involved in event planning. Staffing and communication are effective when there is proper identification of the resources, prevention measures and issuance of important information on time. AI and ML are useful in resolving these challenges through evaluating the risks of crowds, evaluating the potential threats, and planning and coordinating the execution of measures that are likely to prevent such crowds [8].

Thus, it is necessary to conclude that today several decade breakthroughs have been achieved in crowd management and behavior analysis, but still, several concerns and challenges can be noted in the context of the newly introduced methods and techniques. However, the way more research should go to applications that aid in planning of an event, predicting number of attendees, evaluating the risks before a disaster, the resources to mobilize, protective measures and privacy.

III. METHODOLOGY

In this section, it sketches the considerations of data sources and the analysis approach for a goal of attaining predictions and insights for crowd planning in mass events. Details on these methods will be provided in upcoming sections of this paper as follows.

A. Population, Sample and Sampling Technique

This research is targeting the mass gathering event managers and planners from Western province of Sri Lanka because it encompasses significant number of event management companies and a relatively high level of organization events. Industry reports from [25] and [26] estimate a total of 52 event management companies in the region with each company having 3-4 planners hence a population of 180-200.

With regard to the population of interest, a quantitative research approach was employed for the purpose of this survey with a sample size of 128 respondents calculated to a 95% confidence level and a 5% margin of error. Because the population and events are clustered, and no list of individual event managers was available, a cluster sampling technique was used as recommended by [27]. Using this method entailed categorizing firms by availability, and sampling by cluster until requisite sample size was obtained depending on the number of mass event planners in each firm.

This sampling approach allowed for effective sample selection to try and overcome time and organizational constraints to obtain a representative sample of experienced planners currently and/or recently involved in MGE planning.

B. Types of Data to be Collected and Data Sources

1) Quantitative Data

- Demographic details of both event planners and mass gathering events have been collected.
- Details on historical occurrences like overcrowding, stampeding and turbulence together with most frequently experiencing last moment crowd controlling problems.

2) Qualitative Data

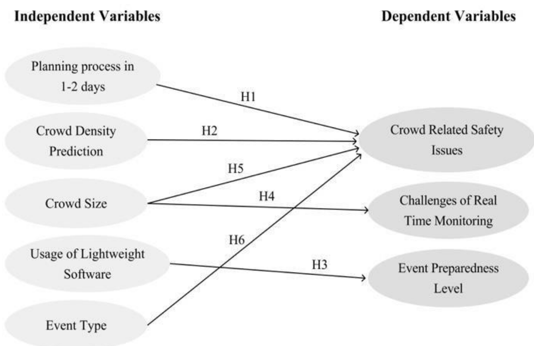
- Self-administered surveys of satisfaction with currently available tools, expectations of using a better system, and the perceived urgency of developing a well-optimized system.
- Data regarding before, during and after event crowd management measures will be collected and the factors that shaped them.
- Some of the problems experienced while managing mass crowds, potential risks or safety concerns that may occur within mass events, as well as cases identified will be described.

3) Surveys and questionnaires: Valueable data of mass events including their nature, related incidents and overcome strategies and specially event planners' feedbacks collected using Google Forms online platform.

4) Secondary sources and databases: Datasets, density analysis reports, incident documentations including highly recognized papers and other generally accessible content used to gather further details about mass events.

The Data Collection Plan has several stages and includes selection and analysis of the data that are collected through surveys, questionnaires and from various other authentic sources.

C. ConceptualFramework



1) Independent Variables: Planning process in 1-2 days, Crowd Density Prediction, Crowd Size, Usage of Lightweight Software, Event Type.

2) Dependent Variables: Crowd related safety issues, Challenges of real-time monitoring, Event preparedness level.

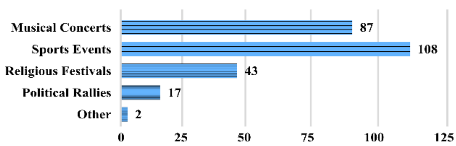


Fig. 60. Conceptual Framework Diagram

D. Methods of Data Analysis

To effectively conduct this research, Descriptive and Inferential Analysis have been used to produce a variety of findings, that would assist the key stakeholders.

1) Descriptive Analysis: In the descriptive analysis section, the emphasis is made respectively on presenting the comprehensive description of the demographics of the mass gathering events and event planners. Besides demographics, corresponding descriptive analyses are performed to fulfill objectives.

2) Inferential Analysis: In the current study, inferential analysis methods are employed in testing the hypothesis, while the features incorporations are justified by analyzing the level of user requirements.

IV. DATA ANALYSIS

Data gathered through the event planners has been processed and analyzed within this phase. Combination of demographics and other analyses included in the descriptive section while inferential section includes hypothesis testing.

A. Descriptive Analysis

Fig. 61. Data Analysis – Frequency of event types.

Number of times that events organized has been analyzed using above Fig. 2 and which denotes both musical and sports events have majority while political rallies are the least organized.

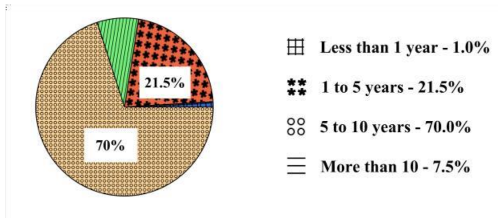


Fig. 62. Data Analysis – Event planners experience level.

Fig. 3 represents the experience level of event planners, and it confirms that majority have 5-10 years of experience followed by 1-5. As an extension of this, these event planners organized more than 7 events per year or at least 4-6.

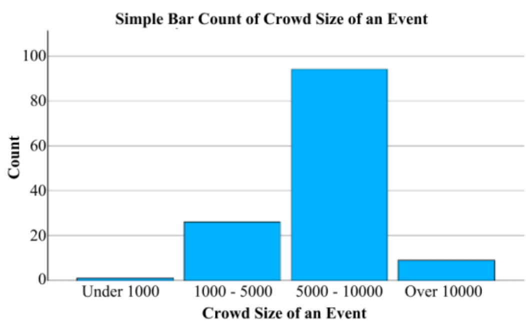


Fig. 63. Data Analysis – Average crowd sizes of mass events.

In Fig. 4 when analysing crowd sizes, events with more than 5000 attendees are organized frequently while events with less than 1000 not consider as mass events. Events with 5000-10000 attendees recorded the most crowd related incidents happened within (94). Least number of incidents were recorded in events with less than 1000 attendees, and it suggest small events does not contain high risk scenarios. Analysis done to evaluate incident rates for different event types shows, with 75 positive responses out of 130, musical events reported 45 of 5-10 incidents within an event. With most positive responses (108), Sports events recorded 69 of 5-10 incident range and 36 of over 10 incidents an event. Overall, the analysis suggested that musical, sports and religious events tend to have higher rate of incidents while political gatherings have very less.

TABLE XI. DATA DISTRIBUTION FOR PLANNING AND PREPAREDNESS

Category	Frequency	Percent	Valid	Cumulative
Less than 1 day	3	2.3	2.3	2.3
1 - 3 days	12	9.2	9.2	11.5
4+ days	115	88.5	88.5	100.0
Total	130	100.0	100.0	

d. Responses collected for event planning and preparedness level at mass events.

Based on the interpretation of above table, many consider high preparedness level before event. Average (2.86), median (3.00) and data distribution of responses prove this analysis strongly. Analysis of the pre-event strategies shows that, verifying entry and exit point logistics has been the main consideration among planners, which records 80.8% of responses while conducting final checks on crowd density tools recorded the least. As per the final outcome it is clear that most of the planners are not adopted to technology based solutions yet.

TABLE XII. DATA DISTRIBUTION FOR USING OPTIMIZED SOLUTION

Category	Frequency	Percent	Valid	Cumulative
Very Unlikely	1	0.8	0.8	0.8
Unlikely	2	1.5	1.5	2.3
Neutral	3	2.3	2.3	4.6
Likely	77	59.2	59.2	63.8
Very Likely	47	36.2	36.2	100.0
Total	130	100.0	100.0	

e. Responses collected for likelihood of using optimized solutions at mass events.

The above table shows analysis on likelihood of using lightweight and optimized solution by event planners. Most of the responses are in likely to very likely range with a combination percentage of 95.4% votes. 2.3% of responders are unlikely to have a lightweight and optimized solution for event planning process. Final outcome denotes a positive result as most of the responders voted for very likely and likely options.

TABLE XIII. DATA DISTRIBUTION FOR STAKEHOLDER INVOLVEMENT

Category	Frequency	Percent	Valid	Cumulative
Event Organizers	40 (No)	30.8	30.8	30.8
	90 (Yes)	69.2	69.2	69.2
	130 (Total)	100.0	100.0	
Emergency Responders	65 (No)	50.0	50.0	50.0
	65 (Yes)	50.0	50.0	100.0
	130 (Total)	100.0	100.0	
Security Personnel	54 (No)	41.5	41.5	41.5
	76 (Yes)	58.5	58.5	100.0
	130 (Total)	100.0	100.0	
Venue Operators	71 (No)	54.6	54.6	54.6
	59 (Yes)	45.4	45.4	100.0
	130 (Total)	100.0	100.0	

f. Frequencies for different stakeholder involvement.

At last an analysis done to select suitable stakeholders to share pre-event risk analysis reports. The table breakdown shows that Event organizers received 90 positive responses followed by Security personnel received 76 positive responses out of 130 total responses. It gives a result as event organizers, security personnel and emergency responders are crucial for sharing possible risk associated with events while venue operators are not much considerable when pre-event risk analysis to prevent crowd related issues.

B. Findings and Interpretations

To conduct more detailed analysis of the data collected on mass events, issues occurred and planning strategies, a hypothesis testing has conducted. Mainly, Independent Sample T-Test and One-Way ANOVA have been used and the significant level of variables measured by comparing the p-value to 0.05. If the p-value is less than 0.05, then the H0 (Null Hypothesis) was rejected and indicates that the variable is statistically significant.

TABLE XIV. HYPOTHESIS ACCEPTANCE TABLE

Hypothesis	Significance Value	Acceptance
Pre-event planning conducted before 1-2 days reduce the likelihood of occurring crowd related issues in mass gatherings. (H1)	0.013	Not Rejected
Usage of predictive and proactive tools help to reduce crowd related issues and enhance the effectiveness of crowd management in mass gatherings. (H2)	0.011	Not Rejected
Event planners who prefer to use well-optimized software solutions report higher preparedness for reducing emergency issues. (H3)	< 0.001	Not Rejected
It is difficult to conduct real time monitoring during a mass gathering and reduce crowd related occurrences. (H4)	0.018	Not Rejected
Higher crowd density significantly increases the likelihood of crowd related issues during mass gatherings. (H5)	0.001	Not Rejected
Mass event type contributing to increases the likelihood of safety issues related to mass crowd (H6)	Musical Events (0.024)	Not Rejected
	Sports Events (0.049)	Not Rejected
	Religious Event (0.042)	Not Rejected
	Political Rallies (0.051)	Rejected

g. All Hypothesis used for testing and analysis.

1) Hypothesis 1 (H1)

Significance of pre-event planning and number of crowd related incidents was tested. Since p-value of 0.013 is less than 0.05, the null hypothesis can be removed and therefore variable is significant. 0.081 Eta-squared value denotes significance level is smaller to moderate.

2) Hypothesis 2 (H2)

H2 tested the significant of crowd density prediction tools to number of crowd related incidents. Then the ANOVA table demonstrated that the p-value is less than 0.05 (0.011) and it shows the statistically significant of this variable. 0.085 of Eta-squared value indicates significant level at smaller to moderate.

3) Hypothesis 3 (H3)

P-value for third hypothesis is less than .001, which denotes usage of lightweight and optimized solutions statistically significant for pre-event preparedness level. Larger significant level of this variable can be measure through the Eta-squared value of 0.369.

4) Hypothesis 4 (H4)

Since the p-value of H4 is 0.018 when the significance was tested, this shows that the crowd size variable is significant with how challenging the real-time crowd monitoring. Further testing shows that when the crowd size increases between the range of more than 1000 to 10,000, the more tough real-time monitoring gets. But not only crowd sizes, there are other factors that affect the challenge of real time monitoring.

5) Hypothesis 5 (H5)

Within this test the significance level of crowd sizes on number of crowd related incidents was analyzed. Resulted p-value of 0.001, draws a conclusion that crowd size statistically significant for number of crowd related incidents. Generated Eta-squared value (0.115) ranging at moderate level.

6) Hypothesis 6 (H6)

With the analysis of H6 using Independent Sample T-Tests it was seen that musical events experience emergency evacuations more than non-musical events with the 0.024 of p-value and 0.428 of Cohen's d value. The other issues had more than 0.05 for the p-value, hence there is no difference for the 2 groups for these variables.

Hypothesis acceptance table demonstrate selected hypothesis and results took based on their significant values. Almost every hypothesis has been not rejected, but political rallies event type produced 0.051 significant value which was above the range. Since it was not statistically significant for hypothesis 6, not considered when providing solutions within this research.

V. SOFTWARE SOLUTION

The suggestion of solution uses predictive analytics and supervised learning to predict crowd density and flow patterns before large-scale events. Through integration of detailed event records and large historical data the system manages to enhance model improvement while providing thorough behavioral examinations of earlier crowd events. It essential to event planning and crowd management while improving strategies to get ready by bonding local security organization and emergency services and sharing the plan in advance for handling incidents. The structured system enables stakeholders to view predictive data enhancements as well as strategic recommendations from historical patterns which leads to enhanced preparedness. Technologies such as Machine Learning (ML) models, APIs, cloud database and React JS for web development was used within this solution to meet the industry standards. The project uses dependable technologies which include cloud database storage for historical data protection and interface visibility through React JS for data presentation functions. The predictive analytics section make an API and includes five main technologies. These are: Optical Character Recognition (OCR), Data Generation, Machine Learning Models, Large Language Models (LLM), and Retrieval Augmented Generation (RAG) architecture. The API functions as an operational center to enable effortless data connectivity between each component which ensures historical data processing runs effectively with activating real-time inputs. The following steps are required to perform the development process.

As the first step research and design has done to assess the best machine learning models and algorithms with their abilities to create the solution architecture and finalize the overall plan of the development process. Research on benchmarking used algorithm tests coupled with historical event data to identify suitable models for crowd density prediction based on accuracy and precision and recall requirements. Then, selection of primary user data inputs has done based on critical analysis and results took from data analysis phase to ensure accurate results. Historical architectural event data obtained from population studies and past grounds layout investigations yielded rich datasets which included every behavior element required. User inputs that are collect before analyzing the risk factor of an event including event demographic data, location data and information about event ground plan are shown below in the Fig. 5.

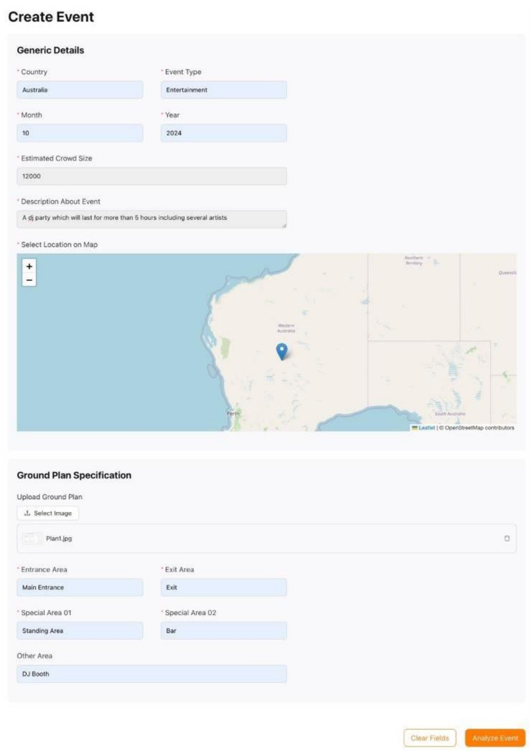


Fig. 64. User inputs before predicting the risk level and generate suggestions

By performing data preparation and pre-processing selected data is cleaned by removing missing values and using proper normalizing methods. The approach uses superior data imputation methods to handle missing values while historical features receive normalization treatment through the combination of min-max scaling and z-score normalization. Then data has converted to required formats and several steps of validations continuously done to keep the standard level of the data. The model preparation carries out two core operations for encoding categorical values followed by historical data combination for generating unified training fields. The system is trained using the processed data and changes done while testing with real data to improve the multi-linear regression model. The model progresses through multiple refinement cycles during training by using established cross-validation techniques from historical data for reducing overfitting while achieving improved prediction accuracy for future historical situations. In below Fig. 6 shows how the multi linear regression model has been developed through several processes to ensure accurate final outputs from the system.

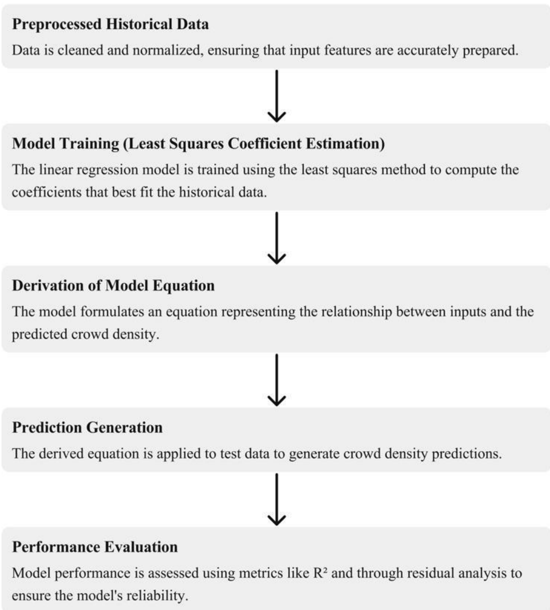


Fig. 65. Linear regression model process along with performance evaluation.

Accuracy testing shows a prediction accuracy of 80% to 88% in different conditions of the event. The model demonstrated these accuracy levels after continuing simulations with recorded historical events thus proving its ability to predict crowd density using past data. Then in the location density indexing, event ground map is divided into the segments and predicted density levels are calculated to identify the most dangerous areas. The spatial clustering algorithm in the segmentation process divides the event venue by using historical density metrics to precisely identify specific high-risk areas. With the help of custom created algorithm these risk levels are visualized and included within the final outcome. The visualization module blends calculation results with risk metrics in a digital mapping platform so users can analyze each segment details together with historical prediction data. Usage of OCR has done in the ground plan analysis phase to gather information from the event's ground plan layout and visualized denisty levels using a plotting mechanism in the final density map.

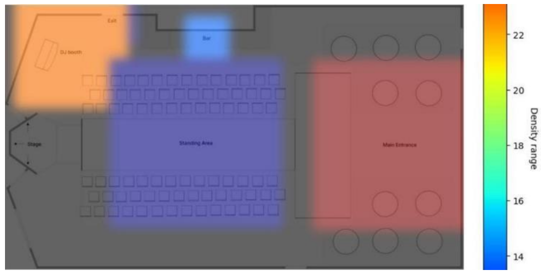


Fig. 66. Generated heatmap for predicted crowd density level.

Fig. 7 shows the generated heatmap through the system, which is used to give risk levels and recommendations of the mass gathering event. The heatmap visualization strengthens event risk decision-making when it demonstrates how event analysis results from concentrated data areas. To provide suggestions, a tailored crowd management plan and strategy system created based on expected crowd movement, peak-time crowd count and density values and danger zones. The system produces recommendations from crowd density studies and attendance peak analysis which generate tactical and strategic methods from past occurrences. Usage of LLM included within this implementation to make the recommendations user friendly and custom based on region, event type and especially cultural difference. The LLM transforms processed contextual information into recommendations which customize their language for event coordinators whose preferred data sources stem from historical performance data.

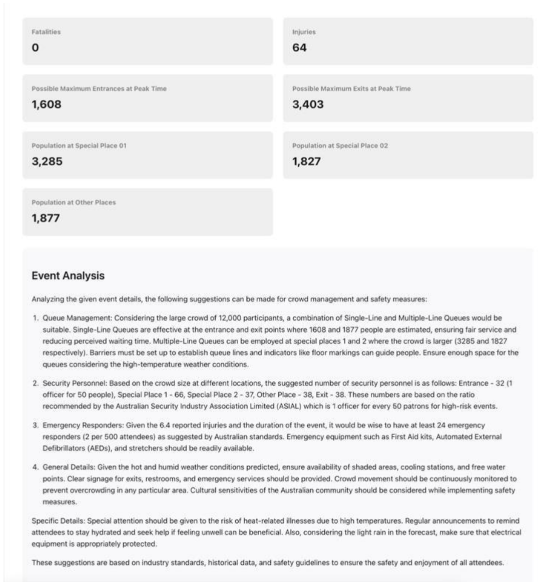


Fig. 67. Crowd risk statistics with crowd management suggestions.

Fig. 8 shows the event statistics like possible fatalities and injuries can happen in the worst-case scenario and average crowd count in different dangerous zones as the first content. Statistics originate from both user input data combined with analysis models which simulate worst-case events to create data-based risk evaluations across different zones. Then recommendations and suggestion list generated including queue management strategies, security personal and emergency services implementation and other specific details based on heat map and statistical predictions. The platform presents recommendations through an easy-to-use dashboard system which allows users to examine event data analyses while displaying complete strategies linked to event management. After finalizing the development, all technologies have been merged including ML, OCR, LLM and RAG architectiure into one system to give expected results successfully with linking the system's predictive model with the front-end to ensure continuous connectivity for user accessibility. API frameworks create a smooth pathway for analysis systems on the backend and interfaces on the frontend which visualizes exclusively data-based predictions extracted from historical record archives.

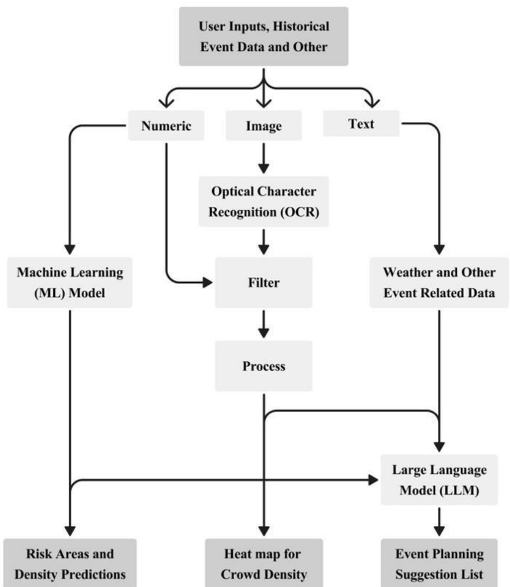


Fig. 68. Software solution architecture.

Fig. 9 demonstrate how the system works overall from different variation of inputs to outcome of predicted heatmap including statistical analysis and list of recommendations and suggestions. Overall, several valuable outputs provided by the system with the help of all the included state-of-arts and other necessary technologies. Event planning reliability improves because the outputs provide analyzed data which helps create better risk assessments and management strategies from historical trends.

Software solution has been tested within two phases. At the start the accuracy of the ML model has been tested using sufficient amount of testing data. The testing phase involved extensive tests using historical event data segments to conduct comprehensive performance metric evaluation before the model received its final deployment. A partition took from the real-world data set included within the testing part to see the predictive accuracy level of the regression model. The model evaluation method separated training from testing phases which maintained objectivity in assessing predictive results on historical information that remained unseen to the model. Then, using historical crowd disasters and different real-life scenarios the final solution has been tested to measure the quality of final outputs including heatmap, risk analysis statistics and specially crowd management suggestions. Models demonstrate effectiveness through consistent results between projected outcomes and registered historic data which proves that this system flawlessly duplicates established risk behavior patterns of previous occurrences. This solution offers a simpler, data-focused way to manage large events, making planning better and improving safety with real-time crowd predictions. With the help of this pre-event analysis and crowd management suggestion system, event planners will be able to manage mass gathering events effectively by fitting out for worst case scenarios can happened within a large scale gathering. The system implements new benchmarks by moving past event information to detect hazards that lower the danger faced by future safety implementations.

DISCUSSION

The final solution for pre-event planning has the strengths and weaknesses summarize below. As for the advantages the system aids event planners in enhancing their planning by identifying high-risk areas, making future predictions, and providing event-specific recommendations. With its lightweight and optimized design, the solution is accessible from anywhere, making it affordable and suitable for use by smaller event planning teams. Also, by predicting potential fatalities and injuries, the system helps planners improve their organizational capabilities and allocate resources effectively. Finally, the ability to share content with stakeholders enhances coordination and ensures the presence of required personnel. The only downsides within the solution is even after receiving insights and recommendations from the system, event planners must manually apply them to fully benefit from the system's management capabilities.

Future recommendations have suggested below for continuous improvement and development of this research area and the problem,

- Transforming the current LLM into a specialized event planning strategy and recommendation provider is necessary to accurate the outcome of the solution. Cache-Augmented Generation (CAG) architecture can be used in more effective way to achieve this goal.
- To uplift the value of current solution, integrating real-time data with current pre-event analysis is necessity to follow in future. One of the common ways of achieving this is by using venue integrated custom IoT sensors. Location signals of attendees (Cellular or Wi-Fi) should be ethically collected through these IoT sensors and continuously feed into the system to generate real time density analysis with custom recommendations.

Conclusion

The current research aims at enhancing pre-event planning on mass events due to crowd safety by making use of predictive methods. The idea above improves planning since it generates insights that are difficult for event planners to come up with using other methods or experience. Thus, the research shows that sports events, musical events and religious gatherings which are attract more than 1000 people likely to have crowd safety issues such as deaths and cases of injuries beyond minor gatherings. Furthermore, event planning strategies like verifying entry and exit points, creating platform for all the stakeholder awareness and identifying potential risks associated with events were key findings regarding to the strategies and processes analyzed. These crowd related issues are sometimes unavoidable, but good pre-event planning, analyzing potential risk before-hand, management of crowds during an event, utilizing well-optimized software applications and implementing connection between stakeholders can minimize them. The suggested plan of study derived from this research presents lines of inquiry that could help to minimize safety threat in future large congregations.

Acknowledgment

This research would not be a possible without the help of several individuals and group of people. First as authors we like to acknowledge all the mass event planners for their critical part to ensure crowd safety within large scale gatherings. Their experience and valuable insights played a huge role to make this research a success. Also, special gratitude goes to the supervisor and the course director, whose expertise and guidance in several areas has facilitated correct progression and valuable resources for this research.

References

[1] G. Blair and J. Henly, "At least 150 killed in crowd crush during Halloween festivities in Seoul," the Guardian, Oct. 30, 2022. https://www.theguardian.com/world/2022/oct/29/dozens-of-people-crushed-by-large-crowd-at-halloween-festivities-in-south-korea.

[2] Gde Yogadhita and W. Agustin, "Football Stampede in Kanjuruhan Stadium from the Perspective of Disaster Preparedness on Mass Casualty Incident: A Case Study of Mass Gathering Event," Prehospital and Disaster Medicine, vol. 38, no. S1, pp. s78–s79, May 2023, doi: https://doi.org/10.1017/s1049023x23002273.

[3] M. Rice, M. Montgomery, I. Rahman, and L. Cain, "How AstroWorld Astronomically Messed Up," Journal of Hospitality & Tourism Cases, Sep. 2024, doi: https://doi.org/10.1177/21649987241281084.

[4] T. Hai, H. Seo, and J. Park, "Effect of electrical stimulation amplitude on dynamic behavior of mice during evacuation," Heliyon, vol. 9, no. 1, pp. e12930–e12930, Jan. 2023, doi: https://doi.org/10.1016/j.heliyon.2023.e12930.

[5] S. Shamim, "What caused the deadly stampede in Hathras, India?," Al Jazeera, Jul. 03, 2024. https://www.aljazeera.com/news/2024/7/3/what-caused-the-deadly-crowd-crush-in-hathras-india.

[6] S. Wieringa, W. Daamen, S. Hoogendoorn, and P. H. A. J. M. Gelder, Framework to mitigate risks of crowd disasters at mass events in public urban space. Washington DC: TRB 95th Annual Meeting Compendium of Papers, 2016. Available: https://www.researchgate.net/publication/303044094_Framework_to_mitigate_risks_of_crowd_disasters_at_mass_events_in_public_urban_space.

[7] L. Soomaroo and V. Murray, "Disasters at Mass Gatherings: Lessons from History," PLoS Currents, vol. 4, p. RRN1301, Mar. 2012, doi: https://doi.org/10.1371/currents.rrn1301.

[8] M. Haghani et al., "A roadmap for the future of crowd safety research and practice: Introducing the Swiss Cheese Model of Crowd Safety and the imperative of a Vision Zero target," Safety Science, vol. 168, pp. 106292–106292, Dec. 2023, doi: https://doi.org/10.1016/j.ssci.2023.106292.

[9] A. Sharma et al., "Global mass gathering events and deaths due to crowd surge, stampedes, crush and physical injuries – Lessons from the Seoul Halloween and other disasters," Travel Medicine and Infectious Disease, vol. 52, p. 102524, Mar. 2023, doi: https://doi.org/10.1016/j.tmaid.2022.102524.

[10] X. Yang, Y. Liu, A. Majumdar, E. Grass, and W. Ochieng, "Characteristics of crowd disaster: database construction and pattern identification," International journal of disaster risk reduction, vol. 111, pp. 104653–104653, Jul. 2024, doi: https://doi.org/10.1016/j.ijdrr.2024.104653.

[11] K. Still, M. Papalexi, Y. Fan, and D. Bamford, "Place crowd safety, crowd science? Case studies and application," Journal of Place Management and Development, vol. ahead-of-print, no. ahead-of-print, Apr. 2020, doi: https://doi.org/10.1108/jpmd-10-2019-0090.

[12] A. Bruno et al., "High-Level Feature Extraction for Crowd Behaviour Analysis: A Computer Vision Approach," Lecture notes in computer science, pp. 59–70, Jan. 2022, doi: https://doi.org/10.1007/978-3-031-13324-4_6.

[13] R. Alharbey, A. Banjar, Y. Said, M. Atri, A. Alshdadi, and M. Abid, "Human Faces Detection and Tracking for Crowd Management in Hajj and Umrah," Computers, Materials & Continua, vol. 71, no. 3, pp. 6275–6291, 2022, doi: https://doi.org/10.32604/cmc.2022.024272.

[14] A. Khan, J. Ali Shah, K. Kadir, W. Albattah, and F. Khan, "Crowd Monitoring and Localization Using Deep Convolutional Neural Network: A Review," Applied Sciences, vol. 10, no. 14, p. 4781, Jul. 2020, doi: https://doi.org/10.3390/app10144781.

[15] S. Bhardwaj, Arvind Kumar Dwivedi, A. Pandey, Yusuf Perwej, and P. Khan, "Machine Learning-Based Crowd behavior Analysis and Forecasting," International journal of scientific research in computer science, engineering and information technology, vol. 9, no. 3, pp. 418–429, Jun. 2023, doi: https://doi.org/10.32628/cseit23903104.

[16] M. F. Mohamed, A. E.-R. Shabayek, and M. El-Gayyar, "IoT-Based Framework for Crowd Management," Mobile Solutions and Their Usefulness in Everyday Life, pp. 47–61, Dec. 2018, doi: https://doi.org/10.1007/978-3-319-93491-4_3.

[17] Waleed Halboob, Hamdi Altaheri, Abdelouahid Derhab, and Jalal Almuhtadi, "Crowd Management Intelligence Framework: Umrah Use Case," IEEE access, vol. 12, pp. 1–1, Jan. 2024, doi: https://doi.org/10.1109/access.2024.3350188.

[18] H. Manoharan, Osamah Ibrahim Khalaf, Sameer Algburi, and Habib Hamam, "Pictorial depiction on controlling crowd in smart conurbations using Internet of Things with switching algorithms," Scientific reports, vol. 14, no. 1, Jun. 2024, doi: https://doi.org/10.1038/s41598-024-61993-5.

[19] Y. Chaudhuri, A. Kumar, B. A. Balaji, and A. Alshamrani, "FGA: Fourier-Guided Attention Network for Crowd Count Estimation," ArXiv (Cornell University), Jul. 2024, doi: https://doi.org/10.48550/arxiv.2407.06110.

[20] K. Khan, W. Albattah, R. U. Khan, A. M. Qamar, and D. Nayab, "Advances and Trends in Real Time Visual Crowd Analysis," Sensors, vol. 20, no. 18, p. 5073, Jan. 2020, doi: https://doi.org/10.3390/s20185073.

[21] H. Liang, L. Yang, J. Du, C.-W. Shu, and S. C. Wong, "Modelling crowd pressure and turbulence through a mixed-type continuum approach," Transportmetrica. B, Transport dynamics, vol. 12, no. 1, Mar. 2024, doi: https://doi.org/10.1080/21680566.2024.2328774.

[22] H. M. Al-Ahmadi, I. Reza, A. Jamal, W. S. Alhalabi, and K. J. Assi, "Preparedness for Mass Gatherings: A Simulation-Based Framework for Flow Control and Management Using Crowd Monitoring Data," Arabian Journal for Science and Engineering, vol. 46, no. 5, pp. 4985–4997, Jan. 2021, doi: https://doi.org/10.1007/s13369-020-05322-8.

[23] Y. Myagmar-Ochir and W. Kim, "A Survey of Video Surveillance Systems in Smart City," Electronics, vol. 12, no. 17, pp. 3567–3567, Aug. 2023, doi: https://doi.org/10.3390/electronics12173567.

[24] Marwa Qaraqe et al., "PublicVision: A Secure Smart Surveillance System for Crowd Behavior Recognition," IEEE access, vol. 12, pp. 1–1, Jan. 2024, doi: https://doi.org/10.1109/access.2024.3366693.

[25] "Annual Reports • SLAPCEO | Sri Lanka Association Of Professional Conference, Exhibition & Event Organisers," SLAPCEO | Sri Lanka Association of Professional Conference, Exhibition & Event Organisers, Sep. 26, 2021. https://slapceo.lk/annual-reports/.

[26] "EMA MEMBER'S DIRECTORY 2024 - Event Management Association Sri Lanka," Event Management Association Sri Lanka, Mar. 07, 2024. https://ema.lk/ema-members-directory/.

[27] Mir Subzar, S. A. Wani, T. A. Raja, and Amit Kumar Attri,"Generalized Robust Regression techniques and adaptive cluster sampling for efficient estimation of population mean in case of Rare and Clustered Populations," Research Square (Research Square), Apr. 2024, doi: https://doi.org/10.21203/rs.3.rs-4020069/v1

A Smart Hiking Companion: Mobile Application for Enhancing Navigation, Safety and User Experience

HAKU Rathnayake
Department of Computer Science
Faculty of Computing
General Sir John Kotelawala Defence University
Rathmalana, Sri Lanka
39-bcs-0027@kdu.ac.lk

DVDS Abeysinghe
Department of Computer Science
Faculty of Computing
General Sir John Kotelawala Defence University
Rathmalana, Sri Lanka
abeysinghe.dvds@kdu.ac.lk

Abstract*— Hiking is a widely practiced outdoor activity that enhances both physical and mental well-being. However, hikers frequently encounter challenges related to navigation, communication, and discovering points of interest (POIs), especially in remote areas where network connectivity is limited. These challenges can lead to disorientation, difficulty in coordinating with group members, and missed opportunities to explore notable landmarks. This study explores existing mobile applications designed to address these issues by incorporating offline navigation, real-time location sharing, security features, and AI-driven POI recommendations. A systematic literature review was conducted alongside a user survey to assess the technological gaps in current hiking applications. The findings indicate that while various applications provide individual features such as GPS tracking, offline maps, and POI suggestions, most lack a comprehensive and unified solution that seamlessly integrates all these functionalities. This paper discusses the potential of emerging technologies, including artificial intelligence (AI), augmented reality (AR), and advanced offline capabilities, in enhancing the user experience of hiking applications. Furthermore, the study proposes a framework for a next generation hiking application that ensures better connectivity, improved safety measures, and a more interactive engagement with natural surroundings. To illustrate the feasibility of these improvements, this research includes detailed proposed application, demonstrating how well-structured interfaces can significantly enhance usability and overall functionality. By addressing the limitations of existing solutions and integrating advanced technologies, this study aims to contribute to the development of more effective digital tools that enrich the hiking experience, ensuring both safety and enjoyment for users.

***Keywords*— hiking applications, offline navigation, location sharing, augmented reality, AI-powered POI suggestions, outdoor safety.**

I. INTRODUCTION

Hiking is a popular outdoor activity that involves walking through natural environments, often across scenic trails, forests, mountains, and other landscapes. It offers a unique combination of physical exercise and mental relaxation, allowing participants to immerse themselves in nature and escape the pressures of daily life. Whether done solo or in groups, hiking provides an opportunity to connect with the outdoors, boost physical fitness, and enhance mental well-being. However, when hiking in larger groups, participants often face significant challenges, particularly when they are spread out or in remote areas where network coverage may be limited.

A major issue that arises once a hike begins is the difficulty group members face in sharing their locations with each other. Without real-time location sharing, hikers can easily become separated from the group, leading to confusion, frustration, and even safety risks. This lack of communication not only jeopardizes the experience but also makes it harder to coordinate in case of emergencies. In addition, hikers frequently miss out on interesting natural landmarks, such as hidden waterfalls, scenic overlooks, or unique wildlife, simply because they are unaware of their proximity. Without a system that informs hikers of nearby points of interest, these scenic spots often go unnoticed, limiting the overall experience.

To address these challenges, a system incorporating inclusive features and a structured workflow is proposed to ensure a seamless hiking experience. The designed interface will focus on real-time location tracking, offline accessibility, and interactive POI recommendation features. A workflow-driven approach will facilitate smooth integration between user inputs, system processing, and output generation, ensuring usability across different hiking scenarios. This system will provide real-time location sharing, ensuring that all group members maintain a minimum connection strength throughout the hike. A user-friendly interface will allow hikers to visualize the real-time position of their group members, even in areas with weak network signals, through an intuitive map-based UI. Additionally, the system can include features that send recommendations about interesting places around the hiker's current position, within a customized distance. This way, hikers will receive alerts about nearby attractions, allowing them to explore hidden gems along the trail and fully enjoy the natural beauty of their surroundings.

Fortunately, technologies are emerging to solve these problems, offering solutions for both real-time location tracking and communication within hiking groups. For example, certain smartphone applications allow hikers to monitor each other's locations and receive notifications when someone strays too far from the group. These technologies significantly enhance safety by ensuring that hikers remain connected.Moreover, user-friendly applications are being developed to help hikers discover nearby attractions, providing real-time recommendations based on their current location. By customizing the distance radius for these suggestions, hikers can easily explore points of interest that would otherwise go unnoticed. In the proposed system, users will also have the ability to input new POIs. However, these entries will remain pending until verified by an administrator, ensuring accuracy and reliability of the information. The system workflow will manage user-submitted POIs through an approval mechanism, enhancing the credibility of POI recommendations.

As technology continues to evolve, several tools have been developed to make hiking safer and more enjoyable. For instance, the Hoge Kempen National Park (HKNP) app [1] allows users to communicate with park officials and share their locations, enhancing safety through customized routes. Augmented reality (AR) navigation tools combine GPS data with visual overlays, simplifying route-finding, while offline navigation features allow hikers to download routes ahead of time, ensuring continuous guidance even without network access [2].

In addition, applications like HUSH [3] provide real-time audio alerts about nearby points of interest, while apps such as Strava and AllTrails [4] offer recommendations based on user-generated data. Tools like Alaska Hike Search [5] give frequent updates on trial conditions and points of interest, and some apps include 2D/3D views and interactive quizzes that engage hikers with information about wildlife and landmarks.

Despite advancements in outdoor adventure technologies, hikers still encounter challenges in maintaining communication with group members and often miss out on nearby points of interest during their hikes. This study has been aimed to investigate existing applications and features designed to address these issues and provide recommendations for more user-friendly and reliable solutions to enhance both safety and enjoyment in hiking.

The objectives of this paper are to identify current applications that offer real-time location tracking and

evaluate an application consideration for effective location sharing among hiking groups, even in offline mode. Furthermore, the study will examine applications that suggest nearby points of interest based on a hiker's location within a customizable distance and assess their effectiveness in real-world scenarios. Additionally, it seeks to evaluate the strengths and limitations of these applications in real-world scenarios, while proposing a framework for improving the hiking experience. The proposed framework will emphasize an optimized UI workflow for user interaction, seamless data processing for real-time updates, and an improved verification system for user-submitted POIs.

By fulfilling these objectives, this research will contribute to the expanding field of outdoor adventure technologies, offering new insights into how technology can elevate the hiking experience, making it both safer and more enjoyable. Ultimately, the purpose of this study was to provide recommendations for designing an intuitive, workflow-driven application that enhances usability, accuracy, and communication in hiking-related technology.

II. Literature Review

This literature review has been conducted to compare existing hiking applications with the main key functionalities and their importance to utilize in future developments. The review specifically evaluates applications such as HUSH [3], Strava and AllTrails [4], and Hoge Kempen National Park (HKNP) [1]. By examining the existing appligcations using these functionalities, including real-time location tracking, security feauture that help to communicate with the other members of group with the sharing location, offline capabilities, point of interest (POI) suggestions, user engagement, and environmental sustainability, and the goal is to identify strengths and limitations while providing insights into how future hiking applications can address current gaps and enhance the hiking experience.

The following Table 1 represents a comparison of the features available in current hiking applications, focusing on real-time location tracking, offline functionality, POI suggestions, user engagement, environmental sustainability, and security features.

TABLE 1. Comparison Of Key Features With Existing Applications And Proposed System

Features	Application									
	HKNP [1]	AR-Based Map [2]	Alaska [5]	Strava and AllTrails [4]	Great Redwood Trail Project App [6]	HUSH App [3]	Peneda-Gerês National Park App [7]	Flyover Country App [8]	Masai Mara National Reserve App [9]	Proposed system
Real-Time Location Tracking	✓	✓		✓	✓					✓
Offline Functionality	✓		✓							✓
POI Suggestions			✓	✓	✓	✓	✓			✓
User Engagement (Multimedia/Gamification)	✓					✓	✓	✓		✓
Environmental Monitoring/ Sustainability				✓					✓	✓
Security Features										✓

Rest of the section will explain the special key features of the existing and proposed hiking application.

a. Real-Time Location Tracking Technologies

The importance of real-time visualization was illustrated by a study on a web-based GIS tourist application that gave hikers the ability to track their location and guarantee safety by comparing routes and tracking progress [10]. M. Shaker et al.[1] discussed about the "Hoge Kempen National Park (HKNP)" app expanded this kind of functionality by allowing hikers to interact with park officials in addition to sharing their current whereabouts. By offering customized routes and real-time park condition updates, this software greatly increased user safety [1].

The creation of an AR-based map for mountain climbers demonstrates how Augmented Reality (AR) has become a useful tool for real-time navigation. This program makes it easier for users to identify routes and destinations by combining visual clues shown on augmented reality interfaces with real-time location data [2]. Another cutting-edge technology enhanced safety by providing real-time environmental alerts and wildlife detection by combining Differential GPS (DGPS), AI, and AR to provide extremely precise real-time navigation [11].

Even with these developments, there are still certain issues, especially with energy use and internet access. During network failures, the HKNP app had problems with communication functionalities even though it worked in offline mode [1]. Wilson [5] expressed that enabling users to download routes and data for offline usage, offline solutions—like those found in an app made for trekking in remote parts of Alaska—address these issues and guarantee dependability even in locations with poor connectivity.

b. Automatic Point-of-Interest (POI) Suggestion Systems

Using information from applications like Strava and AllTrails, in the one study [4] used Volunteered Geographic Information (VGI) to find unofficial hiking trails and points of interest. Hikers may easily find new routes and points of interest with these apps, which automatically suggest places based on user activity and trail usage patterns. A similar strategy was used in the Great Redwood Trail project, where an app highlighted local facilities and eateries in the area while combining automated POI choices with real-time tracking [6].

Points of Interest (POI) suggesting systems also heavily rely on AR technologies. For example, the HUSH app presented historical and naturalistic POIs within urban hiking pathways using augmented reality. Hikers were able to learn about the environment in real time through multimedia content thanks to the immersive user experience generated by this AR integration [paper 18]. Furthermore, by combining navigation and education, location-based augmented reality applications—such as those created for Portugal's Peneda-Gerês National Park—offered POI recommendations via interactive 3D models, increasing user engagement [7].

But preserving POI data's relevance and accuracy is still a constant struggle. In order to guarantee that users obtain precise, timely recommendations, applications such as the Alaska Hike Search app must be updated frequently due to issues with out-of-date or incomplete POI databases [5].These problems demonstrate the necessity of ongoing data management and user input to raise the dependability of POI systems.

c. User Experience in Hiking Applications

Any hiking application's success depends heavily on its user experience (UX), which has a direct impact on user pleasure and engagement. A hiker's experience can be greatly improved by combining interactive features with a smooth, user-friendly design.

One project involved creating hiking software that allowed users to interact with terrain characteristics in real time using both 2D and 3D graphics. Hikers responded favorably to this user-centered design, praising the simplicity of navigation and the capability to view trail conditions prior to setting out [10]. By including statistical feedback on hiking performance, the HKNP app went one step further and enhanced the experience of users by enabling them to examine their hiking metrics after the route [1].

The Flyover Country app's insights taught us a lot about mobile user experience optimization. The design of this software tackles common smartphone issues in outdoor settings when battery life is critical, with features like thumb-based navigation for ease of use and a darker background to preserve charge [8]. The significance of user-generated content (UGC) was further highlighted by research on long-distance hikers on the Appalachian Trail, which demonstrated how information given by hikers in real time affects other hikers' decisions and experiences. The importance of using community-driven data in hiking apps is highlighted by this social component of user experience [12].

Additionally, gamification has been shown to increase user engagement. In order to combine education with pleasure and promote a closer bond with the natural world, one hiking software included a game-like component where users could take wildlife-themed quizzes and receive rewards for finding points of interest [7].

d. Technologies for Environmental Monitoring and Sustainability

Environmental monitoring technology has become essential for maintaining the sustainability of trails and associated ecosystems as hiking has grown in popularity. These tools support ethical tourism, trail usage management, and the preservation of delicate ecosystems.

The usefulness of VGI in tracking unofficial routes and evaluating their effects on delicate ecosystems was shown in

a study conducted on the Catskill Mountains. Land managers were able to prioritize conservation efforts by measuring trail degradation and identifying high-traffic areas using data gathered from apps such as Strava and AllTrails [4]. In order to gather important information on the number of hikers and their environmental impact, another study monitored trail usage using computer vision and machine learning. For well-known hiking routes to be managed sustainably, this knowledge is essential[9].
Environmental risk management and wildlife monitoring are two other applications of AR and AI technologies. In order to help land managers monitor biodiversity and warn hikers of potential hazards, a hiking system that uses beacon towers outfitted with AR, AI, and Big Data analytics analyze wildlife behavior in real-time [11]. Another example of a hiking app that promotes sustainability is the Masai Mara National Reserve app, which provides real-time environmental conditions information, aids in visitor behavior management, and lessens the influence of humans on the ecosystem [13].

A study on hiking trails on islands examined the idea of "islandness," which highlights how the distinct topography and social dynamics of islands influence the creation and utilization of hiking routes. In order to promote more sustainable island tourism growth, this study emphasizes the significance of customized trail management strategies that take particular environmental and cultural aspects into consideration [14]. Furthermore, a Slovakian study on sustainable tourism investigated the ways in which various age groups value trail attributes, providing information on how trail design might improve sustainability and user happiness [15].

An intriguing case study on environmental adaptation and resilience can be found in long-distance hikers. In order to promote designs that respect nature and improve user safety, one study on resilience behaviors among long-distance hikers emphasized the significance of striking a balance between technology use and environmental immersion [16]. Human-environment links in outdoor experiences can now be studied using new techniques thanks to a study on Hike&Fly tourism that used GoPro equipment to record real-time environmental interactions[17].

Several features have been found across the various hiking applications analyzed, such as real-time location tracking, offline functionality, and Points of Interest (POI) suggestions. Apps like Strava and AllTrails[4], and the Hoge Kempen National Park app [1] offer real-time tracking for enhanced safety, while apps like the Alaska Hike Search app[5] prioritize offline functionality to maintain reliability in remote areas. Additionally, apps such as HUSH [3] and the Peneda-Gerês National Park app [7] enhance user engagement by integrating multimedia and gamification to increase interaction with the environment. Sustainability features are present in applications like the Masai Mara National Reserve app [9], while security considerations vary widely, with only a few incorporating robust environmental risk alerts or monitoring.

Despite these shared features, gaps have been identified that the proposed future application aims to address. Key gaps include the inability of many apps to facilitate location sharing between users in offline mode, which could greatly enhance group safety. Additionally, while many apps offer POI suggestions, few allow for automatic recommendations based on the hiker's real-time position while offline.
Furthermore, based on survey feedback, hikers expressed strong interest in an application that provides reliable offline functionality, better communication tools, and more personalized POI recommendations. Some suggested additional features, such as trail difficulty indicators, emergency distress signals, and AI-based personalized route planning. The proposed application will focus on bridging these gaps by enabling location sharing and real time POI suggestions, both while offline, ensuring users can stay connected and receive relevant suggestions without needing constant internet connectivity.

This proposed application has the potential to significantly enhance the hiking experience by ensuring safety, accessibility, and environmental sustainability. Future research should focus on integrating machine learning for real-time route adaptation, leveraging blockchain for data security, and incorporating decentralized networking for improved offline communication .Ultimately, this analysis will guide the design of a future hiking application that will incorporate real-time offline location sharing, automatic POI suggestions, user engagement features, and environmental sustainability measures, all designed to elevate the overall hiking experience.

III. METHODOLOGY

The review process followed a structured methodology encompassing six distinct phases, as illustrated in Fig.1 below.

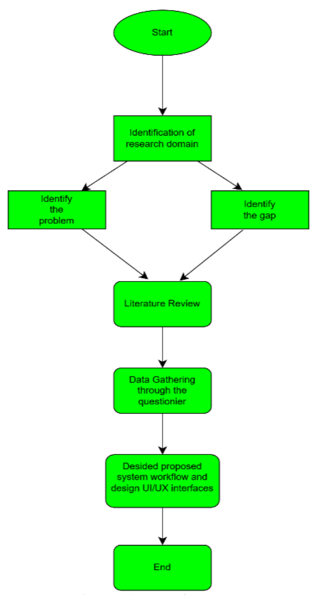


Fig.1. *Phases of review process*

The initial step involved identifying the domain as hiking. This domain was selected due to an observed need for more integrated solutions that enhance user safety, connectivity, and convenience while hiking, especially when faced with unpredictable terrain or group separation. Following domain identification, the scope was refined to focus specifically on applications offering offline capabilities, location-sharing for safety, and automatic suggestions for nearby points of interest (POI). These functionalities were chosen due to their impact on user experience, especially in remote areas or in cases where network connectivity may be unreliable.

Next, research objectives were established to direct the review toward identifying core functionalities missing in current hiking applications. The primary objective focused on what key features would enhance a hiking application to improve the experience, including safety, user engagement, and environmental considerations.

Data gathering was then carried out in two stages, they are primary and secondary data collection. For primary data collection, a Google survey consisting of 19 questions was administered to identify users' specific needs and expectations for hiking applications. The survey received a total of 125 responses from a randomly selected sample covering various age groups, providing a broad perspective on user preferences. Secondary data gathering involved reviewing published research articles in academic databases such as IEEE Xplore, Google Scholar, ScienceDirect, Sci-Hub, and Google Scholar. Initially, 48 papers were selected based on their relevance to hiking applications and related functionalities. After a thorough review, this number was narrowed down to 17 papers that offered the most pertinent insights into the strengths and limitations of current hiking applications, especially in areas of offline functionality, security features, and real-time POI recommendations.

To enhance user engagement and accessibility, an intuitive UI/UX was designed. The key UI components include, login and registration interface to secure login system for user authentication, real-time location sharing interface to displays user location and group members' positions , offline navigation panel to provides pre-downloaded maps and emergency routes, POI recommendation system to suggests nearby attractions dynamically , and emergency alert interface to allows users to send SOS signals without an internet connection.

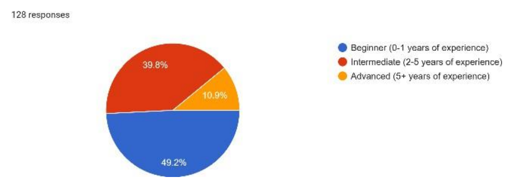
In Finally reviewing all data gathering summaries, it was demonstrated that integrating these key features such as offline communication and automatic POI suggestions based on the hiker's location into a single hiking application would significantly enhance the overall hiking experience. This comprehensive review provides a foundation for future development, aiming to create a more versatile and reliable hiking application that addresses the limitations of current solutions and meets the needs of diverse user groups.

IV. RESULTS AND DISSCUTION

The discussion of this paper addresses both primary and secondary data collection, offering insights into the potential features and functionality needed for a future hiking application.

The primary data was gathered through a Google survey that included 19 questions, receiving 125 responses from randomly selected individuals without specifying an age limit. Several key findings were identified from the responses, which are crucial for enhancing the user experience of hiking applications.

First, respondents were asked about their hiking experience level. A significant portion, 49.6%, identified as beginners (0-1 years of experience), while 40% were intermediate (2-5 years of experience), and 10.4% advanced (5+ years of experience). This distribution highlights the need for an application that can cater to both novice and experienced



hikers, offering features suited to various levels of expertise. In the below Fig.2 was illustrated it.

Fig 2. *Hiking experience*

In terms of hiking frequency, 62.4% of respondents indicated they hike rarely (1-2 times a year), while 27.2% do so occasionally (3-5 times a year). Only 9.6% hike frequently (6-10 times a year), and 0.8% regularly (11+ times a year). This implies that a majority of users are infrequent hikers who might benefit from additional support, such as real-time navigation and offline functionality, during their hikes. Fig.3 represents this below.

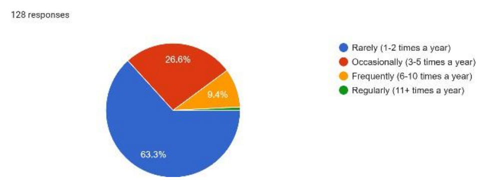
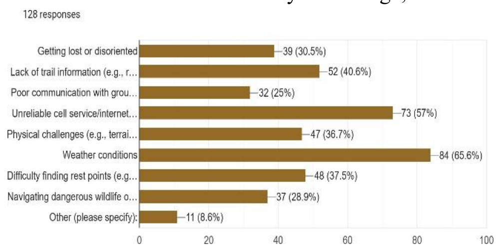


Fig.3. *Frequency of hiking*

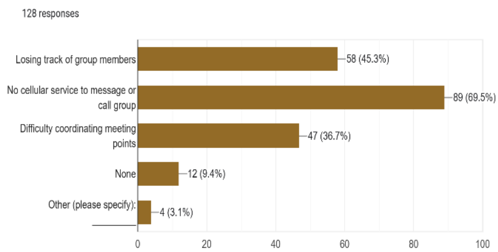
Regarding challenges faced during hikes, 66.4% cited weather conditions as a key challenge, followed by



unreliable cell service or internet access (56%), lack of trail information such as route difficulty or points of interest (39.2%), and poor communication with group members (24.8%). These findings emphasize the need for offline functionality and improved communication features within a hiking app, as users frequently experience a lack of connectivity and trail information. Below, mentioned the Fig.4 to get more information.

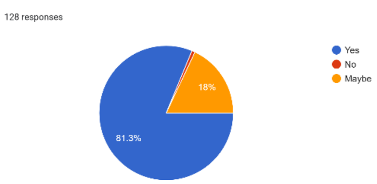
Fig.4. *Challenges when facing hiking*

For communication challenges while hiking in a group,



68.8% of respondents mentioned no cellular service to message or call group members, while 44.8% experienced losing track of group members. These results further support the importance of a feature that enables offline communication and location sharing among hikers, ensuring safety and coordination even in remote areas without cellular service. Here, Fig.5 shows the graphical analyses of the challenges that are faced by groups of hikers.

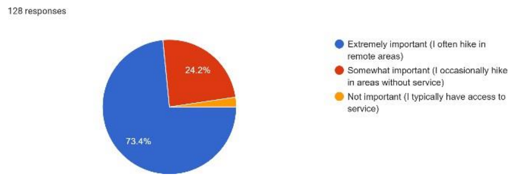
Fig.5. *Challenges that are faced by groups of hikers*



When asked if real-time notifications about nearby points of interest based on their location would be helpful, 81.6% responded affirmatively. This underscores the value of a POI suggestion system that works offline, providing hikers with relevant information about shelters, viewpoints, and other nearby attractions. In the below, Fig.6 represent it clearly.

Fig.6. *Interested about receiving nearby places*

A strong need for offline functionality was highlighted, with 73.6% of respondents valuing offline trail map access and



78.4% considering location sharing without internet highly useful. Additionally, 76% deemed emergency features like distress signals essential, emphasizing the importance of robust offline capabilities, emergency communication, and POI suggestions for hikers. Fig.7 represents the clear chart for this.

Fig.7.*Importance of offline access while hiking*

In the secondary data gathering stage, literature review was conducted by analyzing 48 research papers from academic databases such as IEEE Xplore, ScienceDirect, and SpringerLink. These papers were narrowed down to 17 based on their relevance to seven key features identified, particularly focusing on offline functionality, security for group isolation, and automatic POI suggestions based on hiker location.

The research highlighted that no existing application integrates all these features in a single solution. The reviewed applications were categorized into four sub-sections, and after a detailed comparison, it was concluded that the integration of all key features would significantly improve the hiking experience.

The proposed hiking assistance application follows a structured workflow to ensure real-time location tracking, Points of Interest (POI) suggestions, and emergency alerting functionalities.

The workflow of the proposed application has been illustrated as step-by-step process of user interaction, from logging into ensures seamless operation in both online and offline modes, providing essential navigation and safety features for hikers (Fig.8).

Fig.8. *Proposed workflow of the hiking assistance application*

To enhance user experience, the application was designed with an intuitive UI/UX that focuses on usability, accessibility, and functionality. Below are the key interfaces and their functionalities.

A. Home & Tracking

Displays the current location and nearby POIs on an interactive map. Allow users to start a new hike and track group members. Integrates weather conditions and real-time tracking features (Fig.9).

Fig.9. *Home and Tracking Interface*

B. POI Discovery

Users can explore various POIs, including viewpoints, campsites, water sources, and danger zones. Each POI card provides distance, images, and descriptions to assist hikers in navigation. The feature in the proposed application has been illustrated in Fig.10. below.

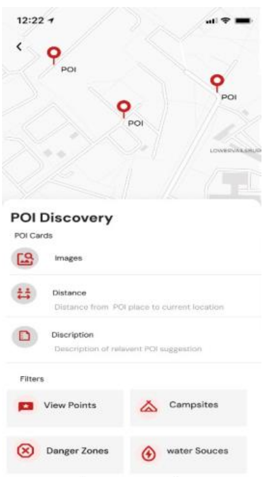
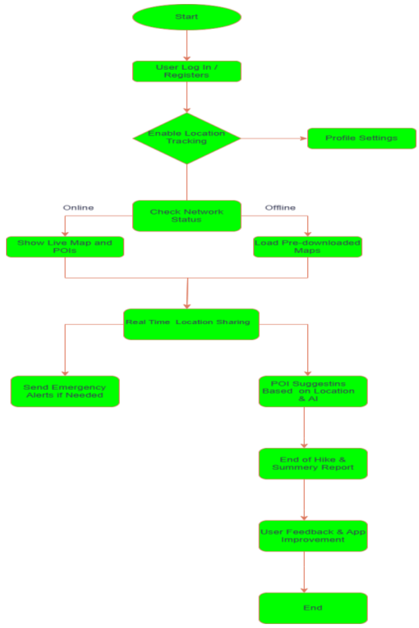


Fig.10.*POI Discovery Interface*



C. Emergency Mode

A dedicated SOS button allows users to send emergency alerts. If no action is taken, an emergency signal is sent automatically. This function includes flashlight SOS mode and sound alarm functionalities to help rescuers locate the hiker as illustrated in Fig.11 below.

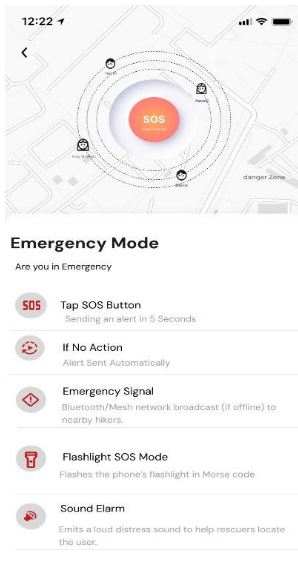


Fig.11. *Emergency Mode Interface*

D. Group Connectivity

Displays real-time group member locations. Provides a direct call/message option for lost hikers. Supports offline messaging in case of network issues functionality has been illustrated in Fig.12.

Fig.12. *Group Connectivity Interface*

E. Preferences

Users can customize settings, such as notifications, hiking preferences, and emergency contacts. It enables offline storage for downloading maps and POI information as shown in Fig.13.

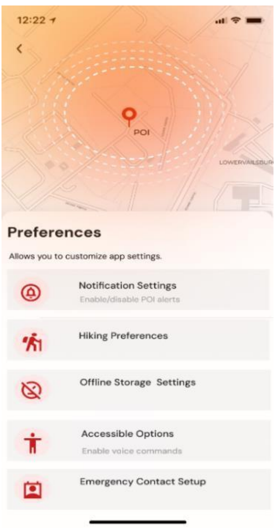


Fig.13. *Preferences Interface*

In conclusion, after analyzing both primary and secondary data, it is clear that a hiking application with integrated offline functionality, location-sharing features for group safety, and automatic POI suggestions would be highly appreciated by users. No existing application currently offers all these essential features together, which supports the need for the development of such an app. Implementing these features in one comprehensive platform would greatly enhance the safety, convenience, and overall experience of hikers.

V. CONCLUSION

This study highlights the need for a comprehensive hiking application that integrates key features identified through survey data, including offline functionality, location sharing for individuals isolated from their group, and automatic nearby point-of-interest suggestions based on a hiker's location. In literature review identified existing applications often lack these essential elements, revealing a critical gap that, if addressed, could significantly enhance both safety and user experience. As the recommended solution, this study proposed a system workflow and a design of an application for developing a comprehensive tool that prioritizes the identified requirements. Including all these key functionalities would contribute to a more secure and enriching experience for the hiking community.

IV. FUTURE WORK

Future work will involve implementing a commercial level application based on the proposed system with the suggested established workflow and interfaces to fulfill the identified functionalities. A comprehensive hiking application will be developed, integrating key features such as robust offline functionality to provide access to trail maps and essential data in remote areas. Security features will be incorporated to enable location sharing, ensuring enhanced safety for users. Additionally, automatic suggestions for nearby points of interest (POIs) will be generated based on the hiker's location, even in offline mode. Users will also be allowed to input new POIs, though these submissions will remain pending until verified and approved by an administrator. By consolidating these features, the identified gaps will be addressed, ultimately enhancing the overall hiking experience.

VI. REFERENCES

[1] M. Shaker et al., "Facilitating hikers' mobility in protected areas through smartphone app: a case of the Hoge Kempen National Park, Belgium," Personal and Ubiquitous Computing, vol. 25, no. 1, pp. 219–236, Feb. 2021, doi: 10.1007/s00779-020-01367-6.

[2] N. Yonov et al., Augmented Reality Navigation Map of Mountain Area, 2020.

[3] L. Melelli, G. Bianchini, and L. Fanò, "HUSH (Hiking in Urban Scientific Heritage): The Augmented Reality for Enhancing the Geological and Naturalistic Heritage in Urban Areas," Applied Sciences, vol. 13, no. 15, p. 8857, Jul. 2023, doi: 10.3390/app13158857.

[4] G. C. Boland, "Monitoring and Mapping Informal Trail Networks Using Smartphone Applications: A Literature Review and Case Study in the Catskills," unpublished.

[5] R. M. Wilson, "Designing a Mapping Application for Alaskan Trails and User Contributed Hazards," unpublished.

[6] M. Khatry, Rails to Trails Web Mapping Application for the Great Redwood Trails: Mapping Northern California's Trails, 2023.

[7] J. L. De Oliveira, Location-Based Mobile Applications and Outdoor Exploration with AR, 2019.

[8] S. Loeffler, R. E. Roth, S. Goring, and A. Myrbo, "Mobile UX design: learning from the Flyover Country mobile app," Journal of Maps, vol. 17, no. 2, pp. 39–50, Dec. 2021, doi: 10.1080/17445647.2020.1867247.

[9] Dedan Kimathi University of Technology, Institute of Tourism and Hospitality Management, Kenya et al., "Mobile-Application Usage Potential for Nature Interpretation and Visitor Management at Masai Mara National Reserve, Kenya; Wildlife Viewers' Perspectives," GeoJournal of Tourism and Geosites (GTG), vol. 43, no. 3, pp. 1163–1174, Sep. 2022, doi: 10.30892/gtg.43338-932.

[10] R. Pandey et al., Online Visualisation of Hiking Activities, Diploma Thesis, Erasmus Mundus Joint Master Degree Programme, Copernicus Master in Digital Earth, 2022.
[11] T. Wu, Design for Hiking Safety & Fun, 2021.
[12] A. G. Rogers, "Appalachian Trail Long-Distance Hikers' Use of Information, Views on Technology, and Spatial Decisions," 2020, unpublished.
[13] I. Nandutu, M. Atemkeng, and P. Okouma, "Intelligent Systems Using Sensors and/or Machine Learning to Mitigate Wildlife–Vehicle Collisions: A Review, Challenges, and New Perspectives," Sensors, vol. 22, no. 7, p. 2478, Mar. 2022, doi: 10.3390/s22072478.
[14] R. Bassani, "Hiking on an Island or Just Hiking? Scientific Review on Trails and Hiking on Islands Under a Nissological Perspective," unpublished.
[15] M. Molokáč, J. Hlaváčová, D. Tometzová, and E. Liptáková, "The Preference Analysis for Hikers' Choice of Hiking Trail," Sustainability, vol. 14, no. 11, p. 6795, Jun. 2022, doi: 10.3390/su14116795.
[16] B. Palacios Abad, M. Koohang, M. Vigil-Hayes, and E. Zegura, "Alone and Together: Resilience in a Fluid Socio-Technical-Natural System," Proceedings of the ACM on Human-Computer Interaction, vol. 7, no. CSCW1, pp. 1–26, Apr. 2023, doi: 10.1145/3579457.
[17] K. Kirsten, "Tourist encounters with a 'GoPro': three layers of hike&fly paragliding travels," Tourism Geographies, vol. 26, no. 4, pp. 635–651, May 2024, doi: 10.1080/14616688.2023.2275161.

www.nibm.ac.lk

011 732 1000

NATIONAL INSTITUTE OF BUSINESS MANAGEMENT

No: 120/5, Vidya (Wijerama) Mawatha, Colombo 07, Sri Lanka