SAICSIT2024

SOUTH AFRICAN INSTITUTE OF COMPUTER SCIENTISTS AND INFORMATION TECHNOLOGISTS

The 45th Annual Conference of the South African Institute of Computer Scientists and Information Technologists

Work-in-Progress Papers

The South African Institute of Computer Scientists and Information Technologists (SAICSIT[1]) 2024 conference was hosted by the Computer Sciences Department and the School of Information Technology at Nelson Mandela University. The conference took place in Gqeberha from Monday the 15th to Wednesday the 17th of July 2024.

SAICSIT has been holding this annual conference since 1985. The Institute focuses on research and development in computing and information technology (IT) in Southern Africa. These Proceedings contain the Work-in-Progress papers presented at the SAICSIT 2024 Postgraduate symposium. The purpose of the Postgraduate Symposium is to encourage emerging researchers and prepare them for an academic career. The symposium provides an opportunity for the researchers to present their preliminary findings without needing to go through the rigorous review process associated with the main conference tracks.

SAICSIT Postgraduate Symposium Editor

Janet Wesson   Nelson Mandela University

[1] https://saicsit.org/

Table of Contents

Part I

Postgraduate Symposium

Determinants of employee innovative work behaviour: the role of decent work, social media use and social capital in higher education institutions

Siphe Mhlana[1]

[1] University of Fort Hare, East London, SouthAfrica
litsitsi@gmail.com

1 Introduction

The modern day dynamism in higher education institutions forces universities to increase the innovative work behaviour of employees [1]. Employees are the driving force that integrate innovative ideas into work processes, which is how any organisation's ability for innovation is demonstrated and the competitive advantage achieved [2–4]. Employee innovation in performing their job duties is generally referred to as innovative work behaviour (IWB) [5]. IWB refers to an individual's ability to discover, come up with, support, put into practice, and maintain novel ideas to improve their performance in a certain job function [6]. For higher education to successfully establish and maintain collaborations with industrial sectors and to ensure that graduates' skills are in line with market demands, it must be innovative. However, the research on determinants that affect IWB in higher education institutions has been neglected more especially looking at the role of social media, decent work and social capital.

Social media can promote IWB in a variety of ways including, collaborations, communication, and idea generation. Decent work and IWB are related and re-enforce each other. Decent work refers to opportunities for individuals to secure work that is productive, offers fair compensation, and provides social protection, as well as respect for fundamental rights and values [7].

2 Problem Statement

IWB is crucial for the success and competitive advantage of organizations, particularly in higher education institutions where fostering innovation can lead to improved educational practices, enhanced research outcomes, and overall institutional growth [8]. Despite its importance, the determinants of IWB among employees in higher education institutions remain underexplored. Existing literature suggests that factors such as decent work, social media usage, and social capital may significantly influence employees' innovative behaviors [9–12]. However, there is a lack of comprehensive understanding of how these factors interact and contribute to IWB in the context of higher education. The purpose of this study is to develop a model to enhance IWB through the use of social media, decent work and social capital in the higher education institutions.

2.1 Literature review and conceptual framework

Figure 1 below provides the conceptual framework of this paper and the discussion of each constructs followed.

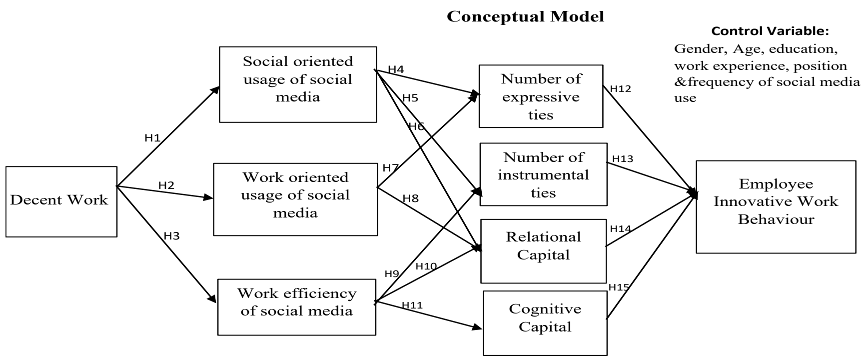

Fig 1. Proposed conceptual framework

As more academic professionals strive to find fair and equitable job conditions, the issue of decent work at higher education institutions is becoming more of a concern [13]. These difficulties include inadequate wages, little work stability, poor benefits and support, and a growing reliance on contract or part-time labour [14]. These concerned lead the International Labour Organisation (ILO) to define the term decent work that guides organisations to implement basic human rights and better incomes for their employees. Based on the fig 1 above there is a positive relationship between decent work and social media usage.

Social media has revolutionized the way higher education institutions interact with students, faculty, and employees. In recent years, social-oriented usage of social media in higher education has become a popular way to engage with stakeholders and foster a sense of community [15]. Social media, which has become a key tool for human interaction in the twenty-first century, is one of the ICT platforms [16].

Universities are embracing social media in growing numbers to promote academic and professional opportunities and to improve the learning process as a whole [17]. Work-oriented usage of social media in higher education includes, recruitment, networking, course promotion, collaborative learning and professional development [18]. Social media is utilised in higher education to promote brands, advertise, and focus on customers, which aids in student retention and recruitment.

In higher education institutions, social media can increase efficiency by facilitating real-time communication, online meetings and webinars, event management, virtual learning, and resource sharing [19]. Efficiency use of social media can save time increase productivity and improve collaboration and knowledge sharing. The secret to properly using social media to boost work efficiency is to use it for what it was designed to do that is, as a tool to enhance collaboration and communication between colleagues, clients, and employees [20].

The term "expressive ties" in higher education refers to the quality and strength of individuals or groups connections based on common goals, ideologies, and feelings that are communicated online [21]. In other words, the number of expressive ties focuses more on the social usage of social media. In higher education, expressive ties are defined by a high level of trust and closeness between faculty and students, as well as the sharing of friendship and social support [22].

Instrumental ties in higher education refer to relationships formed for practical purposes, such as providing information, resources and support [23]. Social media can facilitate the formation and maintenance of these ties among faculty, employees and students. Employees develop trust towards each other by sharing knowledge which is work related and employees who are connected by instrumental ties may be able to recognize each other's knowledge and be able to communicate more successfully with one another to get the information they require [24].

Relational capital refers to the value and benefits that individuals and organisations derive from their social connections and relationships established through social media platforms [25]. In higher education, social media provides a unique opportunity for faculty, employee and students to connect and share information, resources and support [26].

There is no doubt that social media now forms an essential part of modern society and the impart on higher education through teaching and learning. One area where social media may have a tremendous effect is on cognitive capital, which refer to the information, expertise, and abilities of people that contribute to their success in work and personal lives [27].

IWB in universities is essential to drive academic excellence by bringing new ideas and practices to the forefront of education. In the framework of institutes of higher education, IWB can involve developing new curriculum, teaching methods, research approaches and administration procedures [2]. The main drivers of IWB in higher education institutions is the need to stay competitive in a rapidly changing and evolving global environment [8].

3 Methodology

This research adopted quantitative approach for data collection with a survey design. The targeted population of this research is higher education employees at the University of Fort Hare (UFH) chosen through a convenience sampling method. Software package, which is Statistical Package for Social Sciences (SPSS), and Analysis of Moment Structure (AMOS) were adopted as a tool to be used for data analysis. The correlation between the determinants will be measured using correlation and regression. Structural Equation Modelling (SEM) will be used to assess connections between dependent variables. The correlation and regression in this study are going to be used in line with the research hypothesis.

4 Conclusion

This paper provided a progress report based on the postgraduate study. The purpose of the study is to investigate determinants of employee innovative work behavior focusing on the role of decent work, social media usage and social capital in higher education. The conceptual framework was developed showing the relationships among the constructs. The data will be collected to test the hypothesis and confirm the relationship between the constructs. Survey questionnaire will be used for data collection and the data will be analysed.

References

1. Winarsih, Etikariena, A.: Encouraging innovative work behavior among college students through quality education to succeed the SDGs. E3S Web of Conferences. 211, 1–11 (2020). https://doi.org/10.1051/e3sconf/202021101016
2. Namono, R., Obanda, P.W., Ayebale, D., Isiagi, E., Wofuma, G.: Strategizing for innovative work behavior in higher education institutions: the role of creative self-efficacy. Continuity & Resilience Review. 4, 249–266 (2022). https://doi.org/10.1108/crr-03-2022-0005
3. Azeem, M., Ahmed, M., Haider, S., Sajjad, M.: Expanding competitive advantage through organizational culture, knowledge sharing and organizational innovation. Technology in Society. 66, 101635 (2021). https://doi.org/10.1016/j.techsoc.2021.101635
4. Tajeddini, K., Martin, E.: The importance of human-related factors on service innovation and performance. International Journal of Hospitality Management. 85, 102431 (2020). https://doi.org/10.1016/j.ijhm.2019.102431
5. Hewko, S.J.: Individual-Level Factors are Significantly More Predictive of Employee Innovativeness Than Job-Specific or Organization-Level Factors: Results From a Quantitative Study of Health Professionals. Health Services Insights. 15, (2022). https://doi.org/10.1177/11786329221080039
6. Almahamid, S.M., Ayoub, A.E.A.: A predictive structural model of new ways of working on innovative work behaviour: Higher education perspective in the Gulf Cooperation Council. Creativity and Innovation Management. 31, 410–429 (2022). https://doi.org/10.1111/caim.12510
7. Piasna, A., Sehnbruch, K., Burchell, B.: Decent Work: Conceptualization and Policy Impact. 215–224 (2021). https://doi.org/10.1007/978-3-319-95867-5_107
8. Al-Husseini, S., El Beltagi, I., Moizer, J.: Transformational leadership and innovation: the mediating role of knowledge sharing amongst higher education faculty. International Journal of Leadership in Education. 24, 670–693 (2021). https://doi.org/10.1080/13603124.2019.1588381
9. Yan, Y., Deng, D., Geng, Y., Gao, J., Lin, E.: The dual influence path of decent work perception on employee innovative behavior. Frontiers in Psychology. 14, (2023). https://doi.org/10.3389/fpsyg.2023.1302945
10. Sanhokwe, H., Chinyamurindi, W.T., Muzurura, J.: Decent Work and Innovative Work Behaviour: the Mediating Roles of Organisational Learning and Work Engagement. International Journal of Innovation Management. 27, (2023). https://doi.org/10.1142/S1363919623500214
11. Khan, A.N., Moin, M.F., Khan, N.A., Zhang, C.: A multistudy analysis of abusive supervision and social network service addiction on employee's job engagement and innovative

work behaviour. Creativity and Innovation Management. 31, 77–92 (2022). https://doi.org/10.1111/caim.12481
12. Zhao, X., Yi, C., Chen, C.: How to stimulate employees' innovative behavior: Internal social capital, workplace friendship and innovative identity. Frontiers in Psychology. 13, 1–12 (2022). https://doi.org/10.3389/fpsyg.2022.1000332
13. Sălceanu, C.: Higher Education Challenges During Covid-19 Pandemic. a Case Study. Revista Universitara de Sociologie. 16, 104–114 (2020)
14. Heeks, R., Graham, M., Mungai, P., Van Belle, J.P., Woodcock, J.: Systematic evaluation of gig work against decent work standards: The development and application of the Fairwork framework. Information Society. 37, 267–286 (2021). https://doi.org/10.1080/01972243.2021.1942356
15. Asim, M.M., Rasul, A.: Networked Identities: Exploring the Role of Social Networking to Optimize Event Marketing by Higher Education Institutions in the Middle East. Asia Pacific Media Educator. 30, 180–199 (2020). https://doi.org/10.1177/1326365X20970422
16. Nguyen, T.H., Le, H.A., Cu, L.X., Hang, V.T.T.: The Impact of Social Media on the Efficiency of Online Business Performance. Indian Journal of Finance. 16, 46–63 (2022). https://doi.org/10.17010/ijf/2022/v16i3/168703
17. Sarwar, A., Imran, M.K., Akhtar, N., Fatima, T.: Does social media usage boost career prospects of women: an exploratory study in the academia. Kybernetes. 52, 2061–2091 (2022). https://doi.org/10.1108/K-04-2021-0294
18. Chugh, R., Grose, R., Macht, S.A.: Social media usage by higher education academics: A scoping review of the literature. Education and Information Technologies. 26, 983–999 (2021). https://doi.org/10.1007/s10639-020-10288-z
19. Nayak, B.C., Nayak, G.K., Jena, D.: Social recognition and employee engagement: The effect of social media in organizations. International Journal of Engineering Business Management. 12, 1–12 (2020). https://doi.org/10.1177/1847979020975109
20. Louati, H., Hadoussa, S.: Study of Social media impacts on social capital and employee performance–evidence from Tunisia Telecom. Journal of Decision Systems. 30, 118–149 (2021). https://doi.org/10.1080/12460125.2021.1872142
21. Wang, C., Yuan, T., Feng, J.: Instrumental ties or expressive ties? Impact mechanism of supervisor–subordinate ties based on enterprise social media on employee performance. Journal of Enterprise Information Management. 35, 866–884 (2022). https://doi.org/10.1108/JEIM-06-2021-0238
22. Yuan, C.T., Lai, A.Y., Benishek, L.E., Marsteller, J.A., Mahabare, D., Kharrazi, H., Dy, S.M.: A double-edged sword: The effects of social network ties on job satisfaction in primary care organizations. Health care management review. 47, 180–187 (2022). https://doi.org/10.1097/HMR.0000000000000314
23. Chang, Y.Y., Wannamakok, W., Lin, Y.H.: Work conformity as a double-edged sword: Disentangling intra-firm social dynamics and employees' innovative performance in technology-intensive firms. Asia Pacific Management Review. (2023). https://doi.org/10.1016/j.apmrv.2023.01.003
24. Chen, X., Wei, S., Davison, R.M., Rice, R.E.: How do enterprise social media affordances affect social network ties and job performance? Information Technology and People. 33, 361–388 (2020). https://doi.org/10.1108/ITP-11-2017-0408
25. Baima, G., Forliano, C., Santoro, G., Vrontis, D.: Intellectual capital and business model: a systematic literature review to explore their linkages. Journal of Intellectual Capital. 22, 653–679 (2020). https://doi.org/10.1108/JIC-02-2020-0055

26. Ryu, D., Baek, K.H., Yoon, J.: Open innovation with relational capital, technological innovation capital, and international performance in SMEs. Sustainability (Switzerland). 13, 1–13 (2021). https://doi.org/10.3390/su13063418
27. Jia, X., Chowdhury, M., Prayag, G., Hossan Chowdhury, M.M.: The role of social capital on proactive and reactive resilience of organizations post-disaster. International Journal of Disaster Risk Reduction. 48, 101614 (2020). https://doi.org/10.1016/j.ijdrr.2020.101614

The Use of Digital Mental Health Platforms Among Men with Mental Health Conditions in South Africa

Ronaldo Nombakuse[1[0000-0003-4732-3707]] and Pitso Tsibolane[2[0000-0002-7888-1181]]

[1] University of Cape Town, Cape Town, South Africa
nmbxol002@myuct.ac.za, pitso.tsibolane@uct.ac.za

Abstract. At least one billion people globally live with mental health conditions (MHCs), with men in the global South often facing barriers to seeking support due to social stigma, cultural beliefs, traditional masculine norms, and lack of psychoeducation. Despite the rise of digital mental health platforms (DMHPs) that support help-seeking behaviors, their use among men remains underexplored, though studies on women's mental health show positive results. This study aims to explore South African men's perceptions of DMHPs, understand the reasons for their use or non-use, and investigate alternative strategies men adopt to manage MHCs. Employing subjectivism and interpretivism, the study will use an inductive approach informed by social cognitive theory (SCT). Data will be collected through semi-structured interviews and an online qualitative survey using Qualtrics, targeting men aged 18 and above with digital access and MHC awareness. A purposive sampling technique will be used. Thematic data analysis will be performed using NVivo software. The study aims to contribute to existing knowledge on DMHPs, providing insights for South African mental health organizations on men's help-seeking behaviors and outcomes of DMHP use. Additionally, it will inform strategies for digital suicide prevention, mental health education, and tailored DMHP support. Ethical considerations will include distress procedures and trauma-informed approaches, with participants allowed to withdraw and referred to accredited organizations if re-traumatized. An experienced MMH researcher and professional mental health provider will review the research instruments to ensure participants' well-being.

Keywords: Men’s mental health, societal stigma, digital mental health platforms, South Africa.

1 Research Background

Men with mental health conditions (MHCs) in the global South continue to suffer in silence due to masculine standards, patriarchal norms, and societal stigma, and many are reluctant to seek help [1]. For men who live in deep rural contexts where there is inadequate mental health literacy, MHCs are often associated with witchcraft [2]. The increase of MHCs in developing countries is alarming [14]. Men hesitate to be vulnerable as they do not want to be perceived as weak by society [15]. Thus, they barely seek help [20]. Men who are reluctant to seek help often succumb to alcohol and drug abuse as a coping strategy [1-4]. For those who do not succumb to alcohol and drug abuse, suicide is often the only solution [5,16].

Electronic mental health (e-mental health) is an umbrella term associated with digital tools that facilitate the delivery of mental health remotely [6-7]. Scholars in IS often use e-mental health interchangeably with digital mental health platforms (DMHPs), digital mental health interventions (DMHIs), and technology-based mental health interventions (TMHIs) [3, 7-13]. For the context of this study, 'DMHPs' will be used since this term has been used in various digital mental health studies [6-11]. DMHPs include (i) telehealth, (ii) wearable devices, (iii) mobile health applications, (iv) online resources, (v) virtual reality therapy, (vi) online support groups, (vii) digital assessment, (viii) diagnostic tools, (ix) therapeutic games, (x) digital treatment programs, and (xi) social media.

2 Research Problem

The perceptions of South African men with mental health conditions towards DMHPs remain unexplored, although studies on women's mental health (WMH) found positive results [12, 13]. This knowledge gap limits the implementation of suitable DMHPs tailored for South African men with MHCs [1]. To deliver a gender-sensitised psychological treatment or mental well-being support for men with MHCs and those with inclinations [4]. Thus, there is a need for a study that focuses specifically on men's mental health (MMH) and DMHPs within the South African context [14-17]. To address, among other things, the reluctance in help-seeking, the bridge mental health treatment gap, increase the uptake of DMHPs, and investigate DMHPs that are tailored for South African men [16-18].

3 Proposed Research Questions

- **RQ1:** How do South African men perceive digital mental health platforms to manage their mental health conditions?
- **RQ2:** What are the reasons for the use or non-use of digital mental health platforms among men in South Africa?
- **RQ3:** What alternative strategies are being adopted by men in South Africa to manage their mental health conditions?

4 Research Methodology

This study will adopt subjectivism and will employ qualitative research methodology techniques [20]. Interpretive research philosophy will be adopted, assuming that the most suitable way to understand a phenomenon depends on social actors [19]. The study will follow an exploratory approach. The study will follow an inductive approach sensitized by social cognitive theory (SCT) constructs to understand the factors that shape men's perceptions and willingness to use DMHPs [21]. An instrumental single case study design will be employed, while purposive sampling will be used to recruit suitable participants, refer to the inclusion criteria table in Appendix A. The DSM will be used as a guide also to sample self-identified men with inclinations about MHCs but did not receive any formal diagnosis [22].

4.1 Data Collection

Semi-structured interviews will be employed to explore the perceptions of South African men towards DMHPs [23]. A purposively selected sample of South African men will be interviewed until the data saturation point is reached [24]. All participants will be allowed to respond in English or any South African official language, and the interviews conducted in non-English will be translated manually on the interview transcript. An online qualitative survey will be the second primary data source due to the sensitivity of this study. To support men who are reluctant to meet with the researcher face-to-face. The qualitative survey will be created and published on a suitable platform that participants can access, Qualtrics. This study's targeted number of responses is greater or equal to forty men. This number is deemed sufficient particularity when studying a sensitive topic with a hard-to-reach population [23, 25-26].

4.2 Data Analysis

Thematic data analysis will be employed to analyse the data collected with semi-structured interviews and online qualitative surveys [26]. The data will be analysed inductively using NVivo software, and themes and theories will be grounded in the data [23]. The researcher will follow the six steps of thematic data analysis to allow for the inductive emergence of new insights [26].

4.3 Ethical Considerations

All ethical considerations will be observed; the research instruments will first be submitted to the UCT Information Systems Ethics Committee for review and approval. The researcher will use the recommendations of the distress procedure and trauma-informed approach (TIA) on sensitive topics to safeguard the participants' mental well-being [25, 26]. In case of re-traumatisation, participants will be referred to an accredited national mental health organisation. A pre-interview communication pack will be shared with participants, voluntary withdrawal will be allowed, and the data will be removed from the study. Participants will be given pseudonyms to protect their privacy, and any sensitive questions that may cause re-traumatisation will not be asked. The researcher will do post-interview support with the participants, and the researcher will also seek guidance from an experienced MMH researcher to review the quality of the primary data sources.

5 Intended Contribution

This study will add knowledge to the existing digital mental health studies with an emphasis on DMHPs among men with MHCs in South Africa. This study will provide explanations to the developers of DMHPs to understand the preferences of men. The study will educate men who are uninformed about digital mental health literacy and potentially save the lives of those with suicide ideation behaviour (SIB). The study will provide insights to South African mental health organisations and mental health providers about the reasons for reluctance to help-seeking. This study will also be useful to the South African Government and mental health organisations. In providing various strategies for raising awareness about MMH annually in June, October, and November.

6 Delimitations of Research

This study will focus specifically on DMHPs and MMH in South Africa; therefore, WMH will be excluded. Men with severe mental issues (SMI) will be excluded from participating on this study due to their unique vulnerability [7, 26]. This study will not focus on all the MHCs listed in the Diagnostic and Statistical Manual of Mental Disorders (DSM) book. The focus will only be given to MHCs that are prevalent among men [1, 14, 16, 18-20]. In closing, this study will be observed from the perspective of Information Systems (IS), emphasising DMHPs on MMH [6, 15].

7 References

[1] H. C. Masemola, S. V. Moodley, and J. Shirinde, "Perceptions and attitudes of black men in a rural district of South Africa towards depression and its treatment," *South African Family Practice,* vol. 64, no. 1, p. 5557, 2022.

[2] N. J. Bila and C. L. Carbonatto, "Culture and help-seeking behaviour in the rural communities of Limpopo, South Africa: unearthing beliefs of mental health care users and caregivers," *Mental Health, Religion & Culture,* vol. 25, no. 6, pp. 543-562, 2022.

[3] A. Kelemen, E. Minarcik, C. Steets, and Y. Liang, "Telehealth interventions for alcohol use disorder: a systematic review," *Liver Research,* 2022.

[4] E. Schramm *et al.*, "An online program with individualized vs automated support for significant others of depressed individuals–study protocol of a randomized controlled trial," *BMC psychiatry,* vol. 22, no. 1, pp. 1-12, 2022.

[5] K. L. Lovero, P. F. Dos Santos, A. X. Come, M. L. Wainberg, and M. A. Oquendo, "Suicide in global mental health," *Current psychiatry reports,* pp. 1-8, 2023.

[6] L. A. Ellis *et al.*, "The application of e-mental health in response to COVID-19: scoping review and bibliometric analysis," *JMIR mental health,* vol. 8, no. 12, p. e32948, 2021.

[7] T. J. Philippe *et al.*, "Digital health interventions for delivery of mental health care: systematic and comprehensive meta-review," *JMIR mental health,* vol. 9, no. 5, p. e35159, 2022.

[9] J. Strauss, J. Zhang, M. L. Jarrett, B. Patterson, and M. Van Ameringen, "Apps for mental health," in *Mental health in a digital world*: Elsevier, 2022, pp. 395-433.

[10] A. Venning *et al.*, "Exploring the acceptability of a digital mental health platform incorporating a virtual coach: The good, the bad, and the opportunities," *Health informatics journal*, vol. 27, no. 1, p. 1460458221994873, 2021.

[11] A. Mackey and P. Petrucka, "Technology as the key to women's empowerment: a scoping review," *BMC Women's Health,* vol. 21, pp. 1-12, 2021.

[12] T. McCall, M. Foster, and T. A. Schwartz, "Attitudes toward seeking mental health services and mobile technology to support the management of depression among black American women: cross-sectional survey study," *Journal of Medical Internet Research,* vol. 25, p. e45766, 2023.

[13] M. W. Ngwenya and G. O. Sumbane, "The Urgency of Access to Men-Centered Mental Healthcare Services to Address Men's Sensitive Issues in the Communities of South Africa," in *Healthcare Access-New Threats, New Approaches*: IntechOpen, 2022.

[14] SASOP. "Suicide Among Men - Stigma Prevents Seeking Help." The South African Society of Psychiatrist https://www.sasop.co.za/mens-health (accessed 15 Mar 2024).

[15] B. S. Serge, Sy Jadice, Madimba Florence, Baingana Jean, Marie, Dangou Humphrey, Cyprian and E. L. Karamagi, Makubalo, "Suicide in Africa, a neglected reality." [Online]. Available: https://files.aho.afro.who.int/afahobckpcontainer/production/files/iAHO_Suicide_Regional_Fact_sheet_August2022.pdf

[16] T. Mindu, I. T. Mutero, W. B. Ngcobo, R. Musesengwa, and M. J. Chimbari, "Digital mental health interventions for young people in rural South Africa: Prospects and challenges for implementation," *International Journal of Environmental Research and Public Health,* vol. 20, no. 2, p. 1453, 2023.

[17] T. Timakum, Q. Xie, and M. Song, "Analysis of E-mental health research: mapping the relationship between information technology and mental healthcare," *BMC psychiatry,* vol. 22, no. 1, p. 57, 2022.

[18] H. A. Willis and E. W. Neblett, "Developing culturally-adapted mobile mental health interventions: a mixed methods approach," *Mhealth,* vol. 9, 2023.

[19] M. Saunders, "Research Methods for Business Students (6th edn," 2014.

[20] G. Walsham, "Doing interpretive research," *European journal of information systems,* vol. 15, no. 3, pp. 320-330, 2006.

[21] A. Bandura, "Social cognitive theory of mass communication," in *Media effects*: Routledge, 2009, pp. 110-140.

[22] APA, *Diagnostic and Statistical Manual of Mental Disorders (DSM-5).* , 5th ed. American Psychiatric Publishing, 2013.

[23] G. Guest, A. Bunce, and L. Johnson, "How many interviews are enough? An experiment with data saturation and variability," *Field methods,* vol. 18, no. 1, pp. 59-82, 2006.

[24] V. Braun and V. Clarke, *Thematic analysis*. American Psychological Association, 2012.

[25] C. B. Draucker, D. S. Martsolf, and C. Poole, "Developing distress protocols for research on sensitive topics," Archives of psychiatric nursing, vol. 23, no. 5, pp. 343-350, 2009.

[26] L. Dolezal and M. Gibson, "Beyond a trauma-informed approach and towards shame-sensitive practice," Humanities and Social Sciences Communications, vol. 9, no. 1, pp. 1-10, 2022.

Guidelines for reducing perception errors in Autonomous Vehicle System Architecture

Itumeleng Ramala and Monelo Nxozi

Rhodes University, Makhanda, 6139, South Africa

Abstract: Autonomous vehicles (AVs) are expected to reshape the future of transportation, and the autonomous driving system impacts the overall safety of the driving system. AVs require a vast deployment of sensor technologies for environmental perception and to gather broad information on road conditions. However, the advancement from traditional vehicles to AVs faces significant challenges. Perception errors—which form part of the technical challenges (viz. perception, decision, and action errors)— related to the successful commercialization of AVs precipitate safety issues and are deemed to be a key challenge. This study provides guidelines on how to minimise perception errors in AVs for improved perception capabilities and a safer driving experience. This research-in-progress paper provides the preliminary guidelines, which will further be refined following expert evaluation. The preliminary guidelines include sensor fusion algorithms, sensor calibration and maintenance, fault tolerance, real-time data processing, ethical, legal, and societal considerations, testing and validation, and artificial intelligence as key requirements. This study uses the Design Science Research Methodology (DSRM) process to guide the development of the guidelines.

Keywords: Autonomous Vehicles, Perception Errors, Autonomous Vehicle System Architecture.

1 Introduction

1.1 Background

The automotive industry is undergoing a drastic transformation. It is rapidly transitioning from the traditional concept of assembled mechanical parts controlled by humans towards innovative, connected, autonomous, and electric vehicles that offer contextual, intelligent, and personalized customer experience [30, 20]. The concept of autonomous vehicles (AVs) in transportation is evolving and anticipated to revolutionize road traffic by reducing accidents and congestion [24, 22]. [44] note that autonomy is the capability of an intelligent system to accomplish tasks in uncertain conditions in its surroundings. Therefore, AVs are vehicles that can perceive, navigate, and carry out automotive functions independently, as described by [35, 1]. These vehicles, also referred to as autonomous or driverless cars, utilise sensors and software to operate and drive the vehicle [35]. The introduction of AVs includes a vast deployment of sensors. They reduce human assistance and rely on sensors and systems to conduct driving tasks depending on the level of automation [33].

The automation levels of the vehicles depend on the complexity of the autonomous technology applied, the perception range of the environment they drive in, and the degree to which the human driver or vehicle system gets involved in the driving decision, which is closely related to the safety of AVs [44]. This study uses the definition of automated levels published by the Society of Automotive Engineers (SAE), which is widely adopted by automobile manufacturers, regulators, and policymakers [44]. The SAE definition includes six automation levels (viz. Level 0 – No Automation, Level 1 – Driver Assistance, Level 2 – Partial Driving Automation, Level 3 – Conditional Driving Automation, Level 4 – High Driving Automation, Level 5 – Fully Driving Automation) [17, 26, 44, 20, 1, 33]. Most current vehicles are at either level 1 or level 2 due to sensor limitations and high costs, with some classified as level 3 [44, 1]. There are a few level 4 vehicles released as of 2023 [16]. The complexities and challenges encountered in the development of levels 0 to 4 also hinder the development of level 5 vehicles.

Technical challenges are deemed to be the most difficult challenges that AV innovations must overcome [22]. Therefore, this study focuses on addressing these challenges, as they also precipitate safety issues. The technical challenges can be classified as perception, decision, and action errors [44, 22]. The autonomous vehicle system architecture comprises a sensor-based perception layer, an algorithm-based decision layer, and an actuator-based action layer. The failure of the perception layer may mislead the decision and action layers. The inability to resolve perception errors may cause the vehicle to either fail the mission or cause safety problems [44]. Perception errors may be a result of sensor degradation or failure, software malfunction, and errors resulting from vehicle-to-everything communications [44].

1.2 Research Problem

Reliable perception is crucial for the success of AVs, and perception errors pose a significant challenge, leading to inaccurate decisions and unsafe driving actions, ultimately impacting mission outcomes and causing safety issues [23, 44]. AVs need to be able to detect and predict trajectories of other objects in the environment to reduce perception errors [40]. The leading cause of AV fatalities in 2018 was determined to be perception errors [31]. The underdeveloped state of environmental perception technology remains the main obstacle to enhancing the overall performance of AVs and hindering their wide-scale commercialization [44]. Continuous research and testing are crucial to improving AV perception capabilities [32, 41, 21].

1.3 Research Questions

The main question that this study seeks to answer is: *How can perception errors be reduced in Autonomous Vehicle System Architecture to improve safety and limit technical challenges?*

1.4 Research Objectives

The main objective of this study is to develop a set of guidelines to minimise the perception errors of AVs, leading to a safer and more reliable autonomous driving experience. These guidelines are further intended to add to the body of knowledge in relation to the AV system architecture.

2 Research Methodology

This study subscribes to the pragmatism research philosophy, seeking to contribute an innovative artifact in the form of guidelines to solve a real-life problem [16, 6]. The guidelines were developed using the Design Science Research (DSR) as the theoretical lens. DSR follows an incremental approach, emphasising the constant refinement of an artifact from its initial conceptual stage to a final output design as informed by multiple contributions [29]. To fulfill the DSR requirements, the Design Science Research Methodology defined by [29] guided this study, and it will be conducted using a two-phased approach. Phase 1 explores the knowledge base related to this study for the initial design of the guidelines, and Phase 2 focuses on validating and evaluating the initial Guidelines by experts and developing the final set of Guidelines. The state of the research presented in this paper is still in the first phase.

This study is primarily qualitative as it explores various themes related to perception errors in AVs [38]. This approach will help to generate preliminary guidelines from the literature and present them to industry experts within the automotive industry to evaluate and critique. Therefore, the approach to theory development that seems appropriate in this study is the abductive reasoning approach, which combines inductive and deductive reasoning [38]. Expert reviews will be done through in-depth interviews with a limited number of experts [15]. This study employs the purposive sampling technique; 6 to 12 industry experts will be interviewed depending on when the point of saturation is reached [12]. For data analysis, Thematic Analysis will be done through NVIVO, which is a Computer-Assisted Qualitative Data Analysis Software (CAQDAS) [37, 3]. Thematic analysis is appropriate to analyse the interview findings by identifying the themes and patterns from the responses [38].

3 Preliminary Guidelines

Table 1 below presents the preliminary set of guidelines based on the application of the first phase of the Design Science Research Methodology (DSRM) process in this study. There were seven key requirements that guided the development of these guidelines, namely: (a) Sensor fusion algorithms, (b) sensor calibration and maintenance, (c) fault tolerance, (d) real-time data processing, (e) legal and moral consideration, (f) testing and validation, and (g) artificial intelligence (AI). These requirements were identified from a rigorous literature review and elicitation of well-established publicly available standards to ensure the artifact of this study conforms to the well-established guidelines, such as [18, 19], in the automotive industry and safety standards. These guidelines are constructed and described in terms of the following properties: design context, title, description, rationale, and references borrowed from [10, 43, 8]. Experts will be required to rate each guideline using the "Strength of Evidence" rating scale during the second phase of the DSRM process to give them credibility and to determine their level of authority. The level of authority indicates whether or not the guideline must be followed or whether it is only suggested [10, 8]. The application of these guidelines must be in compliance with relevant, well-established safety standards such as [18, 19].

Table 1. Preliminary guidelines

Guideline 1 (Sensor Fusion): Establish algorithms for sensor fusion in AV perception					
Level of Authority	**1**	**2**	**3**	**4**	**5**
Description: Develop and implement more advanced, robust, and precise sensor fusion algorithms that combine the large amount of data generated by individual sensors and related information in a unified format. The algorithms must synchronize all sensor data acquisitions based on a common timestamp to guarantee accurate interpretations.					
Rationale: The sensor fusion algorithms leverage the strengths of different sensors to overcome the limitations of individual sensors and enhance perception accuracy. The algorithms improve accuracy, reliability, fault tolerance and decrease uncertainty in redundant data measurements. The integration of sensors could help address the challenge of object and lane detection in adverse weather conditions.					
References: Realpe, Vintimilla and Vlacic, 2016; Van Brummelen et al., 2018; Rosique et al., 2019; PAS 1880:2020; Wang, 2021; Dauptain et al., 2022; Hasanujjaman, Chowdhury and Jang, 2023; Nahata and Othman, 2023; Oudeif, Mohsen and Alasry, 2024					
Guideline 2 (Sensor Calibration and Maintenance): Regularly calibrate and maintain the vehicle sensors					
Level of Authority	**1**	**2**	**3**	**4**	**5**
Description: Constantly improve and update sensors to help reduce errors caused by sensor drift and degradation.					
Rationale: Sensor calibration and maintenance requirements are crucial for sensor fusion and implementing algorithms for obstacle detection, localization, and mapping. Regular maintenance and updates may help to address any potential issues and keep the perception system up to date.					
References: ISO/PAS 21448:2019; Yeong et al., 2021; Dauptain et al., 2022; Kumar et al., 2022; Su, 2023					
Guideline 3 (Fault tolerance): Ensure the perception system is able to recover from errors					
Level of Authority	**1**	**2**	**3**	**4**	**5**
Description: To ensure safe and secure AV operations, additional sensor coverage should be considered. This can be achieved by installing additional sensors. The implementation of a fault-tolerant system must be able to detect the error, discover the error location, and recover from the detected error.					
Rationale: The degradation or loss of a single sensor can make it necessary to bring the vehicle to a stop. It is necessary to have measures in place for AVs to learn to operate in the presence of faults to avoid unplanned behaviours, more especially behaviours resulting from sensor faults.					
References: Realpe, Vintimilla and Vlacic, 2016; Pas 1880:2020; Zhao et al., 2024					
Guideline 4 (Real-time data processing): Ensure robust V2X communication					
Level of Authority	**1**	**2**	**3**	**4**	**5**
Description: Integrate edge computing and 5G for low latency and higher throughput instead of cloud computing for real-time data processing, as the integration offers more advantages. The large amount of data collected through sensors and communication networks may increase the workload on the cloud and network delays, which hamper cloud computing and pose damage to the system.					

Rationale: Edge computing caters for applications with wireless communication requirements. It stores and processes massive amounts of data where it is generated to overcome any possible network delays. 5G supports highly interactive applications that are computationally intensive and have high quality of service (QoS) requirements.					
References: Hassan, Yau and Wu, 2019; Sittón-Candanedo et al., 2019; Biswas and Wang, 2023; Pandharipande et al., 2023; Trapani and Longo, 2023					
Guideline 5 (Ethical, Legal, and societal consideration): Address Ethical, Legal, and Societal Considerations					
Level of Authority	**1**	**2**	**3**	**4**	**5**
Description: Strengthen research on ethical, legal, and societal issues for improvement of safety and reliability. Engage with ethicists, legal experts, and policymakers to develop relevant standards and legislation to address legal and moral issues tailored to the South African region. The standards and legislation must account for real-world testing and AV commercialization.					
Rationale: AVs require a predetermined way of dealing with specific ethical, legal, and societal issues. The standards will mainly help in making control decisions. However, AVs need ethical guidance on the interaction and communication with other vehicles and pedestrians.					
References: Martínez-Díaz and Soriguera, 2018; Szűcs and Hézer, 2022; Su, 2023					
Guideline 6 (Testing and Validation): Develop and test perception algorithms under conditions very close to reality					
Level of Authority	**1**	**2**	**3**	**4**	**5**
Description: Rigorously test perception algorithms in real-world conditions or conditions that are very close to reality as they enable the vehicle to learn more about different environments and scenarios. Precautionary measures must be taken to ensure human safety. Advanced algorithms must be used for environment mapping, and redundant systems must be employed to cross-verify environmental data.					
Rationale: Extensive testing can help identify potential perception errors and refine the system's performance. Real-world testing, visual testing, simulation, and validation against ground truth data can help identify and reduce the risks of perception errors.					
References: ISO 26262:2018; Rosique et al., 2019; Szűcs and Hézer, 2022; Pandharipande et al., 2023; Piazzoni et al., 2023					
Guideline 7 (Artificial Intelligence): Utilize artificial intelligence algorithms in compliance with safety standards					
Level of Authority	**1**	**2**	**3**	**4**	**5**
Description: Artificial intelligence algorithms must comply with relevant safety standards and legislation. The data used in such algorithms needs to be accurate and representative of the real world. Consider the use of AI algorithms to support all other requirements (Guidelines 1-6).					
Rationale: The use of AI systems, such as machine learning and deep learning in AV perception, can help the system to learn and improve on its own.					
References: Betz et al., 2019; ISO/PAS 21448:2019; Rosique et al., 2019; Fayyad et al., 2020; Yurtsever et al., 2020; Pas 1880:2020; Bachute and Subhedar, 2021; Biswas and Wang, 2023; Nahata and Othman, 2023; Sanjay and Yashwanth, 2023; Zakaria et al., 2023; Hurair, Ju and Han, 2024					

4 Conclusion

This research-in-progress paper provides the conceptual basis for our planned research by stating the research problem space, the selected research methodology (DSR), and the application of DSR in the study. At this point, the research progress is in Phase 1: Theoretical Development of the Initial Guidelines. The empirical evaluation of the guidelines will be undertaken in due course. It is expected that this study will contribute to research on the development of AVs to minimize perception errors by providing guidelines as the artifact to strengthen the perception capabilities of AVs.

Disclosure of Interests. The authors have no competing interests.

References

1. Ahmed, H.U., Huang, Y., Lu, P. and Bridgelall, R., 2022. Technology Developments and Impacts of Connected and Autonomous Vehicles: An Overview. Smart Cities, https://doi.org/10.3390/smartcities5010022.
2. Bachute, M.R. and Subhedar, J.M., 2021. Autonomous Driving Architectures: Insights of Machine Learning and Deep Learning Algorithms. Machine Learning with Applications, 6, p.100164. https://doi.org/10.1016/j.mlwa.2021.100164
3. Bell, E., Bryman, A. and Harley, B., 2018. Business Research Methods. 5th Edition. Oxford: Oxford University Press.
4. Betz, J., Wischnewski, A., Heilmeier, A., Nobis, F., Stahl, T., Hermansdorfer, L., Lohmann, B. and Lienkamp, M., 2019. What can we learn from autonomous level-5 motorsport? pp.123–146. https://doi.org/10.1007/978-3-658-22050-1_12.
5. Biswas, A. and Wang, H.C., 2023. Autonomous Vehicles Enabled by the Integration of IoT, Edge Intelligence, 5G, and Blockchain. Sensors, https://doi.org/10.3390/s23041963.
6. vom Brocke, J., Hevner, A. and Maedche, A., 2020. Introduction to Design Science Research. pp.1–13. https://doi.org/10.1007/978-3-030-46781-4_1.
7. Van Brummelen, J., O'Brien, M., Gruyer, D. and Najjaran, H., 2018. Autonomous vehicle perception: The technology of today and tomorrow. Transportation Research Part C: Emerging Technologies, https://doi.org/10.1016/j.trc.2018.02.012.
8. Cowley, N.L.O., 2009. An Investigation into the Use of Guidelines and Patterns in the Interaction Design Process.
9. Dauptain, X., Koné, A., Grolleau, D., Cerezo, V., Gennesseaux, M. and Do, M.T., 2022. Conception of a High-Level Perception and Localization System for Autonomous Driving. Sensors, 22(24). https://doi.org/10.3390/s22249661.
10. Dix, A., Finlay, J., Abowd, G.D. and Beale, R., 2004. Human-Computer Interaction. third edition ed.
11. Fayyad, J., Jaradat, M.A., Gruyer, D. and Najjaran, H., 2020. Deep learning sensor fusion for autonomous vehicle perception and localization: A review. Sensors (Switzerland), https://doi.org/10.3390/s20154220.
12. Guest, G., Bunce, A. and Johnson, L., 2006. How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. Field Methods, 18(1), pp.59–82. https://doi.org/10.1177/1525822X05279903.
13. Hasanujjaman, M., Chowdhury, M.Z. and Jang, Y.M., 2023. Sensor Fusion in Autonomous Vehicle with Traffic Surveillance Camera System: Detection, Localization, and AI Networking. Sensors, 23(6). https://doi.org/10.3390/s23063335.
14. Hassan, N., Yau, K.L.A. and Wu, C., 2019. Edge computing in 5G: A review. IEEE Access, https://doi.org/10.1109/ACCESS.2019.2938534.
15. Hevner, A. and Chatterjee, S., 2010. Design Science Research in Information Systems. pp.9–22. https://doi.org/10.1007/978-1-4419-5653-8_2.
16. Hurair, M., Ju, J. and Han, J., 2024. Environmental-Driven Approach towards Level 5 Self-Driving. Sensors (Basel, Switzerland), 24(2). https://doi.org/10.3390/s24020485.
17. Ionita, S., 2017. Autonomous vehicles: From paradigms to technology. In: IOP Conference Series: Materials Science and Engineering. Institute of Physics Publishing. https://doi.org/10.1088/1757-899X/252/1/012098.
18. ISO 26262:2018 Road vehicles – Functional safety

19. ISO/PAS 21448:2019 Road vehicles — Safety of the intended functionality
20. Karmanska, A., 2021. The benefits of connected vehicles within organizations. In: Procedia Computer Science. Elsevier B.V. pp.4721–4731. https://doi.org/10.1016/j.procs.2021.09.250.
21. Katiyar, Dr.N., Shukla, Dr.A. and Chawla, Dr.N., 2024. AI in Autonomous Vehicles: Opportunities, Challenges, and Regulatory Implications. Educational Administration Theory and Practices. [online] https://doi.org/10.53555/kuey.v30i4.2373.
22. Kumar, G., James, A.T., Choudhary, K., Sahai, R. and Song, W.K., 2022. Investigation and analysis of implementation challenges for autonomous vehicles in developing countries using hybrid structural modeling. Technological Forecasting and Social Change, 185. https://doi.org/10.1016/j.techfore.2022.122080.
23. Liu, L., Lu, S., Zhong, R., Wu, B., Yao, Y., Zhang, Q. and Shi, W., 2020. Computing Systems for Autonomous Driving: State-of-the-Art and Challenges. [online] Available at: .
24. Martínez-Díaz, M. and Soriguera, F., 2018. Autonomous vehicles: Theoretical and practical challenges. In: Transportation Research Procedia. Elsevier B.V. pp.275–282. https://doi.org/10.1016/j.trpro.2018.10.103.
25. Nahata, D. and Othman, K., 2023. Exploring the challenges and opportunities of image processing and sensor fusion in autonomous vehicles: A comprehensive review. AIMS Electronics and Electrical Engineering, https://doi.org/10.3934/electreng.2023016.
26. Ondruš, J., Kolla, E., Vertaľ, P. and Šarić, Ž., 2020. How Do Autonomous Cars Work? In: Transportation Research Procedia. Elsevier B.V. pp.226–233. https://doi.org/10.1016/j.trpro.2020.02.049
27. Oudeif, M., Mohsen, A. and Alasry, M., 2024. Autonomous Vehicles. [online] Available at: <https://www.researchgate.net/publication/378141518>.
28. Pandharipande, A., Cheng, C.H., Dauwels, J., Gurbuz, S.Z., Ibanez-Guzman, J., Li, G., Piazzoni, A., Wang, P. and Santra, A., 2023. Sensing and Machine Learning for Automotive Perception: A Review. IEEE Sensors Journal, 23(11), pp.11097–11115. https://doi.org/10.1109/JSEN.2023.3262134.
29. Peffers, K., Tuunanen, T., Rothenberger, M.A. and Chatterjee, S., 2007. A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), pp.45–77. https://doi.org/10.2753/MIS0742-1222240302.
30. Pelliccione, P., Knauss, E., Heldal, R., Magnus Ågren, S., Mallozzi, P., Alminger, A. and Borgentun, D., 2017. Automotive Architecture Framework: The experience of Volvo Cars. Journal of Systems Architecture, 77, pp.83–100. https://doi.org/10.1016/j.sysarc.2017.02.005.
31. Piazzoni, A., Cherian, J., Dauwels, J. and Chau, L.-P., 2023. On the Simulation of Perception Errors in Autonomous Vehicles. [online] Available at: <http://arxiv.org/abs/2302.11919>.
32. Ramanagopal, M.S., Anderson, C., Vasudevan, R. and Johnson-Roberson, M., 2018. Failing to Learn: Autonomously Identifying Perception Failures for Self-Driving Cars. IEEE Robotics and Automation Letters, 3(4), pp.3860–3867. https://doi.org/10.1109/LRA.2018.2857402.
33. Rana, M.M. and Hossain, K., 2023. Connected and Autonomous Vehicles and Infrastructures: A Literature Review. International Journal of Pavement Research and Technology, https://doi.org/10.1007/s42947-021-00130-1.
34. Realpe, M., Vintimilla, B.X. and Vlacic, L., 2016. A Fault Tolerant Perception system for autonomous vehicles.
35. Rosique, F., Navarro, P.J., Fernández, C. and Padilla, A., 2019. A systematic review of perception system and simulators for autonomous vehicles research. Sensors (Switzerland), https://doi.org/10.3390/s19030648.

36. Sanjay, S.T. and Yashwanth, S.D., 2023. Self-Driving Car for a Smart and Safer Environment – A Review. Journal of Electronics and Informatics, 5(3), pp.290–306. https://doi.org/10.36548/jei.2023.3.004.
37. Saunders, M., Lewis, P. and Thornhill, A., 2016. Research Methods for Business Students. [online] Available at: <www.pearson.com/uk.>.
38. Saunders, M., Lewis, P. and Thornhill, A., 2019. Research Methods for Business Students. 8th ed. [online] Pearson Education. Available at: <www.pearson.com/uk>.
39. Sittón-Candanedo, I., Alonso, R.S., García, Ó., Muñoz, L. and Rodríguez-González, S., 2019. Edge computing, iot and social computing in smart energy scenarios. Sensors (Switzerland), 19(15). https://doi.org/10.3390/s19153353.
40. Su, K., 2023. A study on the key technologies and existing challenges in the development of autonomous vehicles. In: International Conference on Machine Learning and Automation. [online] pp.165–170. https://doi.org/10.54254/2755-2721/32/20230205.
41. Szűcs, H. and Hézer, J., 2022. Road Safety Analysis of Autonomous Vehicles. Periodica Polytechni-ca Transportation Engineering, 50(4), pp.426–434. https://doi.org/10.3311/PPtr.19605.
42. Trapani, N. and Longo, L., 2023. Fault Detection and Diagnosis Methods for Sensors Systems: a Scientific Literature Review. In: IFAC-PapersOnLine. Elsevier B.V. pp.1253–1263. https://doi.org/10.1016/j.ifacol.2023.10.1749.
43. United States. Department of Health and Human Services. and United States. General Services Administration., 2006. Research-based web design & usability guidelines. U.S. Dept. of Health and Human Services.
44. Wang, J., Zhang, L., Huang, Y. and Zhao, J., 2020. Safety of Autonomous Vehicles. Journal of Advanced Transportation, https://doi.org/10.1155/2020/8867757.
45. Wang, P., 2021. Research on comparison of LiDAR and camera in autonomous driving. In: Journal of Physics: Conference Series. Institute of Physics. https://doi.org/10.1088/1742-6596/2093/1/012032.
46. Yeong, D.J., Velasco-hernandez, G., Barry, J. and Walsh, J., 2021. Sensor and sensor fusion technology in autonomous vehicles: A review. Sensors, https://doi.org/10.3390/s21062140.
47. Yurtsever, E., Lambert, J., Carballo, A. and Takeda, K., 2020. A Survey of Autonomous Driving: Common Practices and Emerging Technologies. IEEE Access, 8, pp.58443–58469. https://doi.org/10.1109/ACCESS.2020.2983149.
48. Zakaria, N.J., Shapiai, M.I., Ghani, R.A., Yassin, M.N.M., Ibrahim, M.Z. and Wahid, N., 2023. Lane Detection in Autonomous Vehicles: A Systematic Review. IEEE Access, https://doi.org/10.1109/ACCESS.2023.3234442.
49. Zhao, X., Fang, Y., Min, H., Wu, X., Wang, W. and Teixeira, R., 2024. Potential sources of sensor data anomalies for autonomous vehicles: An overview from road vehicle safety perspective. Expert Systems with Applications, https://doi.org/10.1016/j.eswa.2023.121358.

A Model for Real Time Stress Management and Intervention for Tertiary Students

Vimbai Matanda [1,2[

0000-0001-5375-631X]], Dieter Vogts[1,3[1111-2222-3333-4444]] and Lynette Barnard[1,4[1111-2222-3333-4444]]

[1] Nelson Mandela University, Gqeberha, Eastern Cape, South Africa
[2] s2199692562@mandela.ac.za
[3]dieter.vogts@mandela.ac.za
[4]lynette.barnard@mandela.ac.za

Abstract. Stress is a widespread issue among tertiary students, negatively impacting their academic performance, mental health, and well-being. This study intends to address this problem by developing a model for real-time stress prediction and intervention utilising smart wearables, specifically the Samsung Galaxy Watch, ranging from Watch 4 to Watch 6.

Keywords: Stress Management, Machine Learning, Stress Prediction, Tertiary Students

1 Introduction

1.1 Background

Stress is an increasingly prevalent issue among tertiary students, with negative impacts on their academic performance, mental health, and overall well-being [1]. The capacity to detect and regulate stress in real time can greatly improve students' quality of life and academic performance [2]. Recent improvements in wearable technology have created new opportunities for continuous health monitoring, including the development of real-time stress management systems [3], [4].

Smart wearables, such as the Samsung Galaxy Watch models, include a variety of sensors that can capture physiological data such as heart rate and heart rate variability, which are indicators of stress levels [5]. Leveraging this data with machine learning techniques enables the creation of models that can predict stress in real-time [3], [6]. The stress predictions enabled by machine learning models can then be utilised to provide timely interventions, allowing students to better manage stress. While prior work has made progress in constructing stress prediction models for wearable devices, many of these attempts have been constrained by controlled conditions, small datasets, and a lack of real-time application [7], [8].

1.2 Problem Statement

Despite the potential of smart wearables in stress management, there is a lack of robust, real-time models specifically designed for tertiary students. Existing solutions often fail to provide timely and accurate stress predictions, leading to delayed or ineffective interventions. Moreover, there is a need for a model that can operate seamlessly in the natural environments of students, offering real-time feedback and support.

1.3 Research Aim

The aim of this research is to develop a model for real-time stress prediction and intervention using smart wearables, specifically targeting tertiary students. The model will use data from the Samsung Galaxy Watch to predict stress in a timely and accurate manner, which allows proactive stress management and intervention.

1.4 Research Significance

This study is significant because it aims to address the issue of stress among tertiary students and poses a technological solution, using smart wearables to improve their overall well-being and academic performance. With real-time stress management, the negative impacts of stress can be mitigated by providing a real-time stress management tool, promoting a healthier and more productive student life.

2 Literature Review

This section begins with an overview of stress management, explores real-time stress management applications, and then focuses on their use among tertiary students. Finally, it explores previous study limitations and how this research aims to address these gaps.

2.1 Stress Management

Stress management involves a variety of treatments and psychotherapies aimed at reducing a person's levels of stress, particularly chronic stress, to improve everyday life [9]. Traditional stress management approaches include cognitive-behavioral therapy, relaxation techniques, and lifestyle adjustments including exercise and time management [10]. While these strategies work, they involve active participation from individuals and frequently fail to provide quick comfort in stressful situations. Technology has opened up new options for stress management, particularly through the use of wearable devices and mobile applications enabling real-time monitoring and interventions [2].

2.2 Real-time Stress Management

Real-time stress management is continuous monitoring of physiological signals such as heart rate variability (HRV), electrodermal activity (EDA), and breathing rate to detect stress and deliver timely interventions [11]. Wearable devices with enhanced sensors and wireless connectivity can collect and analyse physiological data, providing immediate feedback and proposing coping techniques [4]. Real-time stress management has applications in occupational settings, where stress monitoring can improve productivity and well-being [12], and in healthcare, where patients can receive ongoing support in treating chronic stress-related diseases [9].

2.3 Real-time Stress Management in Tertiary Students

Tertiary students endure unique pressures, such as academic pressure, financial difficulties, and social dynamics[1]. Real-time stress management models can be useful in this situation, as they provide students with immediate feedback and suggestions for effective stress management. Studies have examined several techniques for adopting these concepts among students. Examples include, mobile applications that remind students to practice mindfulness or breathing exercises when stress is sensed, have shown promise[4]. Wearable devices that monitor physiological signs and send real-time alerts and recommendations are also under investigation [2].

2.4 Limitations of Existing Studies

Some notable limitations include the use of limited sample sizes or particular demographics, which constrains the generalisability to larger groups. For example, research on police cadets or drivers may not be directly applicable to tertiary students [8]. Variable experimental techniques and a lack of standardised procedures might produce contradictory results, making it difficult to compare data from different studies. Many studies are conducted in controlled surroundings rather than naturalistic settings, which might have an impact on the results' validity when applied to real-world events[3]. There is little evidence on the efficacy of the interventions offered by stress management models.

This study intends to mitigate these constraints by involving a varied sample of tertiary students to increase the findings' generalisability, as well as implementing standardised data collection and analytic processes to assure consistency and dependability in the results. The study will be conducted in students' natural environments to capture real-world pressures and test the proposed model's efficacy in everyday scenarios. The effectiveness of the model's real-time interventions will be assessed.

By addressing these limitations, this study aims to develop a robust and reliable model for real-time stress prediction and intervention in a natural environment, tailored specifically for tertiary students, thereby contributing to their mental health and academic success.

3 Research Design

This section discusses the methodology to be used for this study. It also provides a detailed overview of how the experiment will be designed following the methodology.

3.1 Methodology

The stress prediction model is developed and evaluated in this study using the Design Science Research (DSR) methodology. The research method consists of iterative cycles of design, development, and assessment to ensure that the model is enhanced and validated using empirical testing. Data is collected using different methods to obtain real-time data from Samsung Galaxy Watches, which track heart rate, HRV and other key physiological indicators of stress. The data is then processed and analysed to create a machine-learning model for stress prediction. Some of the machine learning algorithms and neural network architectures to be investigated are Feed Forward Neural Networks (FNN), Random Forests (RF), Convolutional Neural Networks (CNN), and Support Vector Machines (SVMs).

3.2 Experiment Design

The study aims to include a diverse sample of tertiary students from Nelson Mandela University. Participants will be recruited using university methods such as email invitations, posters, and social media. To guarantee a representative sample, students will be drawn from a variety of academic disciplines and years of study. Participants must own a Samsung Galaxy Watch (either a Watch 4, Watch 5, or Watch 6 Classic) to monitor their physiological signals, specifically heart rate and HRV. Participants will also utilise a wearable application to be installed on the participant's Galaxy Watch designed for the study. This application will collect contextual information about the participant's environment and feelings to verify if the proposed model's predicted stress levels reflect the participant's actual experience. This contextual data will be combined with the model's stress predictions to suggest appropriate real-time interventions tailored to the participant

4 Conclusion

This paper presents an approach for real-time stress prediction and intervention for tertiary students utilising smart wearables. While the investigation of various machine learning algorithms and neural networks is ongoing, the results are still inconclusive. Future work related to this study will focus on enhancing the model's performance and determining its applicability to various populations and circumstances.

Acknowledgments. We would like to thank the Telkom Center of Excellence for providing funding for this research.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article

References

1. P. Frazier *et al.*, “Understanding stress as an impediment to academic performance,” *J. Am. Coll. Heal.*, vol. 67, no. 6, pp. 562–570, 2019, doi: 10.1080/07448481.2018.1499649.
2. R. Jerath, M. Syam, and S. Ahmed, “The Future of Stress Management: Integration of Smartwatches and HRV Technology,” *Sensors*, vol. 23, no. 17. Multidisciplinary Digital Publishing Institute (MDPI), Sep. 01, 2023. doi: 10.3390/s23177314.
3. Vividha, D. Agarwal, P. Gupta, S. Taneja, P. Nagrath, and B. Gupta, “Stress Prediction Using Machine Learning and IoT,” in *Proceedings of Data Analytics and Management*, 2022, pp. 615–624.
4. A. S. H. George, A. Shahul, and A. S. George, “Wearable Sensors : A New Way to Track Health and Wellness Partners Universal International Innovation Journal (PUIIJ) Wearable Sensors : A New Way to Track Health and Wellness,” no. August, pp. 15–34, 2023, doi: 10.5281/zenodo.8260879.
5. F. V. Lima *et al.*, “At the Crossroads! Time to Start Taking Smartwatches Seriously,” *Am. J. Cardiol.*, vol. 179, no. Table 1, pp. 96–101, 2022, doi: 10.1016/j.amjcard.2022.06.020.
6. A. A. Al-Atawi *et al.*, “Stress Monitoring Using Machine Learning, IoT and Wearable Sensors,” *Sensors*, vol. 23, no. 21, p. 8875, Oct. 2023, doi: 10.3390/s23218875.
7. G. Vos, K. Trinh, Z. Sarnyai, and M. Rahimi Azghadi, “Ensemble machine learning model trained on a new synthesized dataset generalizes well for stress prediction using wearable devices,” *J. Biomed. Inform.*, vol. 148, no. September 2022, p. 104556, 2023, doi: 10.1016/j.jbi.2023.104556.
8. S. H. Sim, T. Paranjpe, N. Roberts, and M. Zhao, “Exploring Edge Machine Learning-based Stress Prediction using Wearable Devices,” *Proc. - 21st IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2022*, pp. 1266–1273, 2022, doi: 10.1109/ICMLA55696.2022.00203.
9. E. C. Gathright *et al.*, “Stress management interventions for adults with heart failure: Systematic review and meta-analysis.,” *Heal. Psychol. Off. J. Div. Heal. Psychol. Am. Psychol. Assoc.*, vol. 40, no. 9, pp. 606–616, Sep. 2021, doi: 10.1037/hea0001084.
10. V. Khitha and S. Tayade, “Application of Deep Brain Stimulation in Refractory Post-Traumatic Stress Disorder,” *Cureus*, vol. 15, no. 1, pp. 1–10, 2023, doi: 10.7759/cureus.33780.
11. G. Vos, K. Trinh, Z. Sarnyai, and M. Rahimi Azghadi, “Generalizable machine learning for stress monitoring from wearable devices: A systematic literature review,” *Int. J. Med. Inform.*, vol. 173, no. February, p. 105026, 2023, doi: 10.1016/j.ijmedinf.2023.105026.
12. A. Morales *et al.*, “Occupational Stress Monitoring Using Biomarkers and Smartwatches: A Systematic Review,” *Sensors*, vol. 22, no. 17, 2022, doi: 10.3390/s22176633.

Work in Progress: A Case Study on Students' and Lectures' Hedonic Motivations to Use Generative Artificial Intelligence in Higher Learning

Lwazisile Mhlambi[1][0000-0002-7882-0944] and Monelo Nxozi[2][0000-0002-7112-9923]

[1] Rhodes University, Makhanda, Eastern Cape 6139, South Africa
lwazisile@gmail.com

Abstract. The rapid advancement of generative artificial intelligence (generative AI) has transformed teaching and learning practices in the higher learning environment. However, the introduction of generative platforms like ChatGPT—despite it being capable of improving production of lecturers and students–has raised concerns in the matters of accuracy, reliability and data privacy. This work-in-progress paper explores this phenomenon and the intrinsic and extrinsic motivations for the use of generative AI by developing a proposed framework called Value-based Hedonic Motivation System Adoption Model (V-HMSAM) to aid the effective integration of Generative AI in higher learning. The V-HMSAM is a combination of the Hedonic Motivation System Adoption Model (HMSAM) and the Value-based Adoption Model (VAM), which together will account for the intrinsic and extrinsic motivations to use generative AI. The proposed framework will be tested through semi-structured interviews, where students and lecturers from Rhodes University will be used as the population of the case study. The findings of this study will provide theoretical insights into the motivations to use generative AI which could be used to guide the effective integration of generative AI within higher learning spaces.

Keywords: Generative AI, Adoption, Hedonic Motivation, Intrinsic Motivation, Extrinsic Motivation, Higher Learning.

1 Introduction

1.1 Research Context

Artificial Intelligence (AI) is a modern technology that is currently playing a massive role in the digital disruption in the sector of higher education—where technology is applied to enhance learning or training performance [20]. The AI disruption introduced itself in the form of Generative Artificial Intelligence (Generative AI). Generative AI is described as an AI platform that consists of a collection of machine learning algorithms that are designed to generate new data samples based on existing datasets [4]. The platform is able to respond to a user-given prompt by generating a response(s) in the form of text, images, audio, video, and synthetic data within seconds [15]. In the case of Web 3.0, AI and machine learning algorithms are given better leading roles in the delivery of relevant content to users, instead of content, which is chosen by third-party outlets, an element which is prevalent in search engines [7]. Generative AI is capable of generating various forms of content as a response to a user prompt. To achieve this, Generative AI makes use of unsupervised machine learning, which is the unguided form of training of a machine learning algorithm, by using information that has not been labelled or classified and allowing the machine to make its own decisions [2, 8]. It surpasses traditional AI systems as such systems rely on discriminative modelling, wherein data is labelled or classified through supervised machine learning [12].

GPTs, were first introduced by OpenAI in 2018 and are a string of transformer-based deep learning language models that are preconditioned on a collection of data and are able to create human-like content [30]. The GPT architecture has two phases namely, pre-training and fine-tuning [22]. Over time, the GPT architectures have evolved, introducing five iterations namely, GPT-1, GPT-2, GPT-3, GPT-3.5 and GPT-4. In the current landscape of technology, GPTs have made their presence known around the world, with the introduction of the now well-known Generative AI platform named ChatGPT. ChatGPT makes use of GPT-3.5 and GPT-4 technologies, and from its launch in November 2022 ChatGPT has amassed over 100 million users in the space of two months [23]. ChatGPT is an artificially intelligent transformer-based model chatbot that uses natural language processing (NLP) and is centered around the use of computers to develop artificial intelligence models that are able to converse with the end user in human-like manner - which involves analysing, interpreting, and generating human language [16, 18]. ChatGPT has presented itself as a valuable tool for students and teachers, changing their methodologies of engaging their studies and teaching [11]. [6] asserts that ChatGPT is used as an alternative method to access information to search engines like Google, as it speeds up the information-seeking process and provides a direct response whereas Google gives a list of links. Building on information searching, it is also used for personalised tutoring and providing feedback which fills gaps in knowledge based on the individual needs and progress of the student [2, 5]. Furthermore, ChatGPT enables interactive learning with the student, where the student can ask direct questions in a conversational manner [2]. Finally, ChatGPT is also used for language translation by translating educational materials into different languages, enabling its access to wider audiences [14].

Considering it from a teaching perspective, ChatGPT is used to automate the planning of lessons for specific subjects or courses [21]. It is also used for setting quizzes and test questions, including setting rubric for marking [27, 29]. ChatGPT can be used to automate marking or grading as it can be trained to assess student essays or tests, including detecting plagiarism [2, 21]. Within the context of research, ChatGPT is used for its NLP properties to aid researchers in analysing and understanding extensive volumes of textual data, such as social media posts and news articles [6]. It is further used to improve grammar inconsistencies and spelling mistakes and create realistic-sounding text [24]. It is also used to summarise published work on a specific topic, enabling the research to understand the work quicker [21]. ChatGPT is not limited to language-related tasks, it is applicable within the sphere of programming where it is able to suggest code, optimise, complete missing or incomplete code and code debugging [21].

1.2 Problem Statement

Despite holding a promise for a positive change within higher learning spaces by increasing the productivity and efficiency of students and lecturers Generative AI platforms are not completely accurate or reliable, and they pose a risk in relation to the privacy of data. Due to the novel nature of the research within the topic of Generative AI, there are noticeable gaps in research pertaining the use of Generative AI for hedonic purposes as there is a dearth of literature that factors in the value maximisation of the consumer. In pressurised and competitive environments such as higher learning institutions, there is need to explore the extrinsic and intrinsic motivations to use Generative AI by lecturers and students. [11] further points a gap in research in the perceptions of lecturers in the use of adoption of ChatGPT. Previous related research further calls for an increase in the scope of the sample sizes across various academic disciplines and demographics to improve the generalisation of findings [3, 28, 11].

1.3 Research Question

This study aims to answer the following research question: *How does intrinsic and extrinsic motivation influence the adoption of Generative AI among lecturers and students within higher learning institutions?*

1.4 Research Objectives

1. To add to the AI adoption body of knowledge by developing a framework, namely, the Value-Based Hedonic-Motivation System Adoption Model (V-HMSAM), which is a combination of the Hedonic Motivation System Adoption Model (HMSAM) by [17] and Value-based Adoption Model (VAM) by [13], based on the relationship between the ease of use of Generative AI and the user's behavioural intention to use.
2. To provide a greater understanding of the intentions of the use of Generative AI by students and lecturers, including the values and ethics associated with the use of Generative AI, serving as an empirical basis to the governance of Generative AI in higher learning institutions.

2 Preliminary Literature Review

2.1 Constructs of Theoretical Frameworks

The constructs that make up the HMSAM are (1) Perceived Ease of Use, (2) Perceived Usefulness, (3) Curiosity, (4) Joy, (5) Control, (6) Behavioural Intention to Use, and (7) Immersion. The constructs that make up the VAM are (1) Perceived Benefits which have 2 subconstructs which are: (1.1) Usefulness and (1.2) Enjoyment. The second construct being: (2) Perceived Sacrifice and its 2 subconstructs (2.1) Technicality (2.2) Perceived Fee. The third and fourth constructs being (3) Perceived Value and (4) Adoption Intention.

2.2 Planned Search Strategy

Given that study is still within its infant stage, it is intended that a thorough literature review will be conducted to understand the current state of knowledge of the factors that influence the adoption of generative AI by students and lecturers in higher learning. The literature review will be conducted using databases including GoogleScholar, Scopus, EbscoHost, and ScienceDirect, using keywords such as "AI Adoption", "Intrinsic Motivation", "Extrinsic Motivation", "Generative AI", and "Higher Learning". Studies that have been published in the past 6 years (2018 - present) will be considered to ensure relevance.

The anticipated themes that will emerge from the literature review include the factors that influence intrinsic and extrinsic motivation in the adoption of technology. It is presumed that the themes will guide the creation of the theoretical framework.

A potential challenge for this research is finding literature relating to "Generative AI Adoption". To overcome this challenge, literature relating to "technology adoption" and "AI adoption" will also be considered and information from those sources will be used as a benchmark in the data collection stage through the interview process.

3 Research Design and Methodology

3.1 Research Approach

This study is exploratory, as it addresses the "How" question in relation to a phenomenon that occurs when there is limited information available [10]. The study will follow an interpretivist approach, as it focuses on variables and elements in a given context, allowing a deeper understanding of the phenomenon and its complexity in its unique context [1, 25].

A qualitative research method is going to be used in this study, as it aligns with the interpretivist approach to explore and provide deeper meaning to real-world problems [25, 26]. The qualitative research method will enable the researcher to gather the behaviours, experiences, and perceptions of the participants and generate themes and patterns [26].

3.2 Data Collection Techniques and Analysis

Qualitative data will be collected through the use of semi-structured interviews as they enable the researcher to ask open-ended theme-based questions which further allow for in-depth discussions [25]. The sampling methods that will be used will be non-probability sampling. Nonprobability sampling is often used in qualitative research, and it is suitable for this study as expert knowledge of key individuals (i.e. students and lecturers) will be used to answer the research question [19]. Rhodes University will be used as the case study. Therefore, the population for this study will be Rhodes University students and lecturers, to which a purposive sampling technique will be applied to allow the researcher to use their judgement in targeting a select group with exposure to AI tools and who will be best positioned to provide the requisite data for this study [25].

Purposive sampling is also suitable for the study as it allows the researcher to use their judgement to choose the suitable candidates with sufficient diverse characteristics to provide variety in the collected data based on observable key themes [25]. [10] indicate that 12 interviews are typically needed to reach high levels of saturation, which refers to the point at which new data that is collected yields no further insight. Therefore, this study will aim to interview up to 12 students and lecturers, respectively.

The data will be analysed using Computer Assisted Qualitative Data Analysis Software (CAQDAS), in specific, NVIVO, to transcribe the interviews and analyse the data using thematic analysis.

4 Preliminary Thoughts

4.1 Initial Observations and Knowledge and Expected Outcomes

Based on the initial literature search and review, the use of Generative AI within the higher learning context by lecturers and students have been extrinsic or goal-oriented, in a sense that it is used to complete specific tasks, therefore disrupting the traditional norms of higher learning. This observation creates a need to explore the underlying motivations of the use of Generative AI.

The expected outcomes for the research are to explore the interplay between the intrinsic and extrinsic motivations in the use of Generative AI platforms. The findings of this study are expected to aid in the creation of a framework that will aid the effective integration of Generative AI within higher learning environments.

5 Challenges and Next steps

The potential challenge that could be faced is finding research participants (Lecturers), in the interview stage who have the necessary competence and experience in the use of Generative AI.

The next steps in this study are to complete the literature review, to find the existing gaps in research, complete data collection through semi-structured interviews and data analysis.

6 Conclusion

The use of Generative AI in higher learning has proved beneficial to both students and lecturers, however its misuse poses significant risks in the space of higher learning. The study adds to the body knowledge as it aims to explore the intrinsic and extrinsic motivations of the use of Generative AI in higher learning and also develop a framework (V-HMSAM) that factors both motivations. However, more is needed to be done to achieve the objectives of this study such as completing the literature review, collecting the primary data and its analysis.

References

1. Alharahsheh, H.H. and Pius, A. (2020) ‘A Review of key paradigms: positivism VS interpretivism’, Global Academic Journal of Humanities and Social Sciences, 2(3). doi: 10.36348/gajhss.2020.v02i03.001.
2. Baidoo-Anu, D. and Owusu Ansah, L. (2023) ‘Education in the era of Generative Artificial Intelligence (AI): Understanding the potential benefits of CHATGPT in promoting teaching and learning’, SSRN Electronic Journal [Preprint]. doi:10.2139/ssrn.4337484.
3. Bonsu, E.M. and Baffour-Koduah, D. (2023) ‘From the consumers’ side: Determining students’ perception and intention to use chatgpt in Ghanaian higher education’, Journal of Education, Society & Multiculturalism, 4(1), pp. 1–29. doi:10.2478/jesm-2023-0001.
4. Chan, C.K. and Hu, W. (2023) ‘Students’ voices on Generative AI: Perceptions, benefits, and challenges in Higher Education’, International Journal of Educational Technology in Higher Education, 20(1). doi:10.1186/s41239-023-00411-8.
5. Chen, Y., Chen, Y. and Heffernan, N. (2020). Personalized math tutoring with a conversational agent. arXiv preprint arXiv:2012.12121.
6. Chinonso, O. E., Theresa, A. M.-E. and Aduke, T. C. (2023). ChatGPT for Teaching, Learning and Research: Prospects and Challenges. Global Academic Journal of Humanities and Social Sciences, 5(02). https://doi.org/10.36348/gajhss.2023.v05i02.001
7. Essex, D., Kerner, S.M. and Gillis, A.S. (2023) What is web 3.0 (WEB3)? definition, guide and history, WhatIs.com. Available at: https://www.techtarget.com/whatis/definition/Web-30 (Accessed: 01 September 2023).
8. GeeksforGeeks. (2023). Supervised and unsupervised learning. GeeksforGeeks. Available at: https://www.geeksforgeeks.org/supervised-unsupervised-learning/ (Accessed: 29 August 2023).
9. George, T. and Merkus, J. (2021) Explanatory Research | Definition, Guide, & Examples, Scribbr. Available at: https://www.scribbr.com/methodology/explanatory-research/ (Accessed: 12 September 2023).
10. Guest, G., Namey, E. and Chen, M. (2020). A simple method to assess and report thematic saturation in qualitative research. PLOS ONE, [online] 15(5), pp.e0232076–e0232076. doi:https://doi.org/10.1371/journal.pone.0232076.
11. Haglund, J.H. (2023). ‘Students Acceptance and Use of ChatGPT in Academic Settings’
12. Hsu, Y.-C. and Ching, Y.-H. (2023) ‘Generative Artificial Intelligence in Education, part one: The dynamic frontier’, TechTrends, 67(4), pp. 603–607. doi:10.1007/s11528-023-00863-9.
13. Kim, H. W., Chan, H. C. and Gupta, S. (2007). Value-based adoption of mobile internet: An empirical investigation. Decision Support Systems, 43(1), 111 –126.
14. Kohnke, L., Moorhouse, B. L. and Zou, D. (2023). ChatGPT for Language Teaching and Learning. In RELC Journal. https://doi.org/10.1177/00336882231162868
15. Lawton, G. (2023) What is Generative Ai? everything you need to know, Enterprise AI. Available at: https://www.techtarget.com/searchenterpriseai/definition/generative-AI (Accessed: 28 August 2023).
16. Lo, C.K. (2023) ‘What is the impact of chatgpt on education? A Rapid Review of the literature’, Education Sciences, 13(4), p. 410. doi:10.3390/educsci13040410.
17. Lowry, P.B. et al. (2013) ‘Taking “Fun and games” seriously: Proposing the hedonic-motivation system adoption model (HMSAM)’, Journal of the Association for Information Systems, 14(11), pp. 617–671. doi:10.17705/1jais.00347.
18. Lund, B. and Wang, T. (2023) ‘Chatting about chatgpt: How may AI and GPT Impact Academia and libraries?’, SSRN Electronic Journal [Preprint]. doi:10.2139/ssrn.4333415.

19. Nikolopoulou, K. (2022). What Is Non-Probability Sampling? | Types & Examples. [online] Scribbr. Available at: https://www.scribbr.com/methodology/non-probability-sampling/ [Accessed 9 Feb. 2024].
20. Oluwajana, D. et al. (2019) 'The adoption of students' hedonic motivation system model to gamified learning environment', Journal of theoretical and applied electronic commerce research, 14(3), pp. 156–167. doi:10.4067/s0718-18762019000300109.
21. Rahman, M. M. and Watanobe, Y. (2023). ChatGPT for Education and Research: Opportunities, Threats, and Strategies. Applied Sciences (Switzerland), 13(9). https://doi.org/10.3390/app13095783
22. Ray, P.P. (2023) 'CHATGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope', Internet of Things and Cyber-Physical Systems, 3, pp. 121–154. doi:10.1016/j.iotcps.2023.04.003.
23. Ruby, D. (2023) 30+ detailed CHATGPT statistics - users & facts (August 2023), DemandSage. Available at: https://www.demandsage.com/chatgpt-statistics/ (Accessed: 28 August 2023).
24. Sakthi, G. (2023) ChatGPT in education: How students and teachers can use AI to transform learning, YourStory.com. Available at: https://yourstory.com/2023/05/how-chatgpt-change-education-teaching-students-artificial-intelligence#:~:text=ChatGPT%20can%20act%20as%20a,offer%20explanations%20for%20complex%20topics. (Accessed: 19 September 2023).
25. Saunders, M., Lewis, P. and Thornhill, A. (2019). Research methods for business students. Harlow: Pearson.
26. Tenny, S., Brannan, J.M. and Brannan, G.D. (2022). Qualitative Study. [online] Nih.gov. Available at: https://www.ncbi.nlm.nih.gov/books/NBK470395/ [Accessed 2 Feb. 2024].
27. Trust, T., Whalen, J. and Mouza, C. (2023). Editorial: ChatGPT: Challenges, opportunities, and implications for teacher education. Contemporary Issues in Technology and Teacher Education, 23(1), 1-23.
28. Woithe, J. and Filipec, O. (2023) Understanding the adoption, perception, and learning impact of chatgpt in higher education: A qualitative exploratory case study analyzing students' perspectives and experiences with the AI-based large language model, DIVA. Available at: http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1762617 (Accessed: 08 September 2023).
29. Zhai, X. (2023). ChatGPT for next generation science learning. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4331313
30. Zhu, C. et al. (2023). How to harness the potential of ChatGPT in education? Knowledge Management & E-Learning, 15(2), 133–152. https://doi.org/10.34105/j.kmel.2023.15.008

Benefits and Challenges of Robotic Process Automation Adoption in Healthcare

Phillip Sibanda[1[220888914@mandela.ac.za]], Ife Fashoro[2[0000-0002-1791-4744]], and Brenda Scholtz[3[0000-0002-0844-1383]]

Nelson Mandela University, Gqeberha, 6001, South Africa
[1]s220888914@mandela.ac.za
[2]Ife.fashoro@mandela.ac.za
[3] Brenda.scholtz@mandela.ac.za

Abstract. The adoption of Robotic Process Automation (RPA) in healthcare has revolutionised the healthcare industry by automating numerous healthcare tasks. This paper provides an overview of the benefits and challenges of RPA adoption in healthcare institutions, based on a preliminary literature review. Benefits found in the review include reduced administrative burden, increased accuracy, improved efficiency, and lower administrative costs. Challenges found in the review include resistance to change, implementation errors, system integration issues, and lack of expertise. Understanding these challenges is crucial for effective RPA integration. Future research will focus on the adoption of RPA in a South African healthcare institution, with the goal of developing and implementing methods that facilitate the adoption. This approach seeks to maximise the benefits of RPA in a local healthcare institution.

Keywords: Robotic Process Automation, Healthcare.

1 Introduction

The introduction of Robotic Process Automation (RPA) in healthcare institutions has sparked a profound transformation in the industry. RPA has proven to be beneficial in numerous healthcare processes, including patient registration, claims processing, data analysis, and report preparation [5]. RPA automates manual healthcare tasks, leading to extended patient care turnaround times [2]. The market for health-related robotics is projected to grow exponentially over the coming years [6]. RPA not only assists doctors and nurses in performing intricate and accurate tasks but also saves time on performing administrative tasks, thereby enhancing the efficiency of the medical industry [4]. Some studies have reported on the benefits of adopting RPA in healthcare institutions, while others have reported on the respective challenges of adopting RPA in healthcare institutions. No study could be found that summarises these benefits and challenges of RPA adoption in healthcare institutions. This paper provides a brief overview of the benefits and challenges of RPA adoption in healthcare institutions based on the findings of a preliminary literature review. Although the study is exploratory, the findings are crucial in identifying potential research directions.

2 Literature Review Findings

With the help of RPA, healthcare professionals have been able to automate repetitive processes, such as patient registration, data verification, and other related administrative tasks, providing them with more time to focus on patient care and to use their skills more effectively [4],[6]. Despite the significant advantage of reducing the administrative burden for healthcare professionals, the fear of RPA replacing jobs has caused resistance to adopting the technology [8],[15]. Healthcare professionals often hesitate to adopt new processes as they prefer their current system [10]. Resistance to change poses a critical challenge to adopting RPA in healthcare institutions.

By automating tasks such as data entry and record management, RPA has reduced the chances of human error [12],[13]. This automation has enhanced precision**,** data integrity**,** and the reliability of medical information processing. RPA's ability to improve accuracy is one of the key benefits it brings to the healthcare industry [14]. While RPA technology can efficiently process tasks, it cannot check for errors before executing tasks. Thus, software robots can make mistakes quicker than humans without waiting for responses from applications, which can have serious consequences for healthcare tasks [9],[16]. Minor changes to processes can confuse the bots and result in errors. RPA developers have previously reported that forgetting to include the decimal point during the robot configuration process can cause a system failure [9]. Different healthcare institutions may have unique workflows and data entry methods. Adapting RPA bots to handle these variations can be challenging and may require frequent updates to automation scripts.

RPA implementation requires system integration capabilities to enable bots to interact with healthcare systems [17], [18]. Paper-based healthcare system structures have been a cause of adoption challenges. While many healthcare institutions have already transitioned to digital documentation as a more modern and flexible way to store information, other institutions still use paper and unstructured documents [16]. It is difficult for RPA bots to efficiently extract and process relevant information from paper-based systems.

The adoption of RPA in healthcare institutions has resulted in decreased operational costs [5], [6]. Despite its financial benefits, a lack of technical expertise has hindered the effective implementation of RPA. Previous research indicates that integrating emerging technologies such as AI and IoT with RPA is a complex process that demands significant technical knowledge [19], [20]. A limited number of experts are available to assist with developing and implementing RPA for healthcare institutions. Implementing RPA requires technical expertise that healthcare institutions may not have within their organisation, making it challenging to implement RPA effectively [20]. The findings from the review are summarised in Table 1.

Table 1. Benefits and challenges of RPA in healthcare

Benefits	Challenges
Reduced administrative burden [4],[6],[8]	Resistance to change [5],[8],[10]
Increased accuracy and reduced errors [12],[13],[14]	Implementation errors [9],[16]
Improved efficiency [4],[5],[12]	System integration [17],[18]
Lower administrative costs [5],[6]	Lack of expertise [19],[20]

3 Conclusion

The benefits of RPA adoption include, reduced administrative burden, reduced administrative costs, increased accuracy, and improved overall efficiency. However, the transition to RPA is not without its challenges. Resistance to change among healthcare professionals, implementation errors, system integration issues, and a lack of technical expertise present significant challenges. Understanding these challenges is essential for healthcare institutions to effectively navigate the integration process and fully benefit from RPA. Future research will investigate how RPA can be adopted in a South African healthcare institution. The research will use a theoretical and developmental approach to investigate the adoption. The research will investigate effective methods for incorporating RPA into a South African healthcare institution. It will involve tailoring, developing, and implementing these methods to facilitate a seamless transition. Adopting this strategy could provide valuable insights into customising RPA to address specific challenges in the local healthcare sector.

References

1. Shwetha, R., & Kirubanand, V. B. (2021). Remote monitoring of heart patients using robotic process automation (RPA). In ITM Web of Conferences (Vol. 37, p. 01002). EDP Sciences.

2. Palmer, F. (2023). Robotic Process Automation (RPA) in Healthcare: Identifying Suitable Processes for RPA (Doctoral dissertation, Grand Canyon University).

3. Smits, J. (2023). Assessment of Robotic Process Automation technology in Dutch hospitals (Master's thesis, University of Twente).
4. Kaur, J. (2023). Robotic Process Automation in the Healthcare Sector. In E3S Web of Conferences (Vol. 391, p. 01008). EDP Sciences.

5. Administrator. (2024b, April 23). AI & RPA in Healthcare: Benefits, use cases, Trends 2024. AutomationEdge. https://automationedge.com/blogs/ai-and-rpa-in-healthcare/

6. Robotic Process Automation Market Size, Report 2024-2033. (2024b, March 20). https://www.precedenceresearch.com/robotic-process-automation-market

7. Applying the task technology fit model to the healthcare sector: a case study of hospitals' computed tomography patient-referral mechanism. Journal of Medical Systems, 39, 1-14.

8. Fernandez, D., & Aman, A. (2018). Impacts of robotic process automation on global accounting services. Asian Journal of Accounting & Governance, 9.

9. Asatiani, A., Copeland, O., & Penttinen, E. (2023). Deciding on the robotic process automation operating model: A checklist for RPA managers. Business Horizons, 66(1), 109-121.

10. Samhan, B., & Joshi, K. D. (2015, January). Resistance of healthcare information technologies; Literature review, analysis, and gaps. In 2015 48th Hawaii International Conference on System Sciences (pp. 2992-3001). IEEE.

11. Viale, L., & Zouari, D. (2020, July). Impact of digitalisation on procurement: the case of robotic process automation. In Supply Chain Forum: An International Journal (Vol. 21, No. 3, pp. 185-195). Taylor & Francis.

12. Barla, N. H., Almeida, S. M., & Almeida, M. S. (2023). RPA revolution in the healthcare industry during COVID-19. In Confluence of Artificial Intelligence and Robotic Process Automation (pp. 199-229). Singapore: Springer Nature Singapore.

13. Lehtonen, N. (2022). The indirect impacts of robotic process automation: a case study for a Finnish healthcare organisation.

14. Sumathi, N., & Jeyalakshmi, A. (2022). Implementing robotic process automation to improve operational efficiency in the healthcare sector. International Journal of Health Sciences, (I), 9666-9672.

15. Choi, D., R'bigui, H., & Cho, C. (2021). Robotic process automation implementation challenges. In Proceedings of International Conference on Smart Computing and Cyber Security: Strategic Foresight, Security Challenges and Innovation (SMART CYBER 2020) (pp. 297-304). Springer Singapore.

16. Nxumalo, A. (2022). We are exploring the implications of Robotic Process Automation (RPA) on operational efficiency in public health care in South Africa.

17. Lamberton, C., Brigo, D., & Hoy, D. (2017). Impact of Robotics, RPA and AI on the insurance industry: challenges and opportunities. Journal of Financial Perspectives, 4(1).

18. Wewerka, J., & Reichert, M. (2023). Robotic process automation-a systematic mapping study and classification framework. Enterprise Information Systems, 17(2), 1986862.

19. Flechsig, C., Völker, M., Egger, C., & Weske, M. (2022, September). Towards an integrated platform for business process management systems and robotic process automation.

20. Pramod, D. (2022). Robotic process automation for industry: adoption status, benefits, challenges and research agenda. Benchmarking: an international journal, 29(5), 1562-1586.

Stock price prediction using transformer-based models

Akho Ndlodaka

Rhodes University,
Grahamstown, South Africa
akhondlodakaschool@gmail.com

Abstract. This paper proposes integrating FinBERT for sentiment analysis and Temporal Fusion Transformer (TFT) for time series forecasting to enhance stock price prediction accuracy. Diverse datasets, including historical stock prices, financial statements, and market sentiment data from news and social media, will be used. While still in progress, initial findings suggest combining these models can improve prediction accuracy. The paper outlines the methodology for data collection, preprocessing, model construction, and training, and addresses challenges such as computational demands and data alignment. Future work will compare model performances and optimize efficiencies to provide more accurate and robust stock price prediction models for the financial industry.

Keywords: Stock Price Prediction · FinBERT · TFT

1 Introduction

Stock price prediction is essential for informed investment decisions. Traditional methods like ARIMA and GARCH struggle with financial data's complexity [4,6]. Machine learning models, such as SVM and neural networks, offer improved predictions [3, 7].

Transformer-based models, like Temporal Fusion Transformer (TFT) for time series forecasting and FinBERT for sentiment analysis, show great promise [1, 8, 10]. However, integrating diverse data sources remains challenging. This research aims to enhance prediction accuracy by combining FinBERT and TFT, benefiting the financial industry with more robust and reliable models.

2 Literature review

[5] developed a hybrid stock price prediction model combining long short-term memory (LSTM) networks and sentiment analysis. They optimized the LSTM model using NIFTY 50 data (March 2017 - March 2022), finding that 100 epochs and 6 layers performed best. They also analyzed news sentiment for the same period, assessing polarity and subjectivity. The enhanced dataset, integrating sentiment parameters, was used to predict stock prices from March 2022 to

April 2023 using CAT Booster and Random Forest models. Results showed that while LSTM effectively captured past trends, CAT Booster slightly outperformed Random Forest in integrating news sentiment. This research highlights the potential of combining deep learning with sentiment analysis for improved stock price prediction.

FinBERT, a specialized BERT model fine-tuned on financial texts, has been developed to understand and quantify sentiment in financial news and social media [2]. By transforming unstructured text data into sentiment scores, FinBERT enables the incorporation of market sentiment into predictive models, offering a nuanced perspective on market movements.

TFT is a state-of-the-art model designed for multi-horizon forecasting, capable of integrating static and dynamic covariates to capture complex temporal patterns [9]. TFT's architecture includes variable selection networks, gated residual networks, and multi-head attention mechanisms, which together enhance its ability to model intricate time series data effectively.

This collection of studies highlights the effectiveness of integrating advanced deep learning architectures with sentiment analysis, emphasizing the power of transformer-based models in financial sentiment analysis and time series forecasting.

3 Approach

3.1 Data collection

The data will be collected from 50 different companies, ensuring a diverse range of market behaviors and financial conditions.

- *Historical stock prices*: Collected from a reliable financial database, Yahoo Finance. The dataset will include daily stock prices, trading volumes, and technical indicators.
- *Financial statements*: Extracted from company filings and financial reports, providing fundamental financial data.
- *Market sentiment data*: Sourced from news articles (Bloomberg, CNBC, etc.), social media platforms (Twitter, Weibo, etc.), and financial news websites.

3.2 Data pre-processing

- *Data cleaning*: Remove any inconsistencies, missing values, or outliers from the datasets to ensure the quality and reliability of the data. Handle missing data points by either imputing them with statistical methods or removing them if necessary.

- *Normalization*: Normalize numerical data to ensure that all features contribute equally during model training. This will involve scaling the data to a standard range, typically using techniques like min-max scaling or z-score normalization.
- *Financial ratios*: Derive financial ratios from the financial statements, such as price-to-earnings (P/E) ratio, to provide insights into the financial health of the companies.

3.3 Model selection and construction

FinBERT for sentiment analysis

- *Input layer*: Text data from news articles, social media posts, and financial news websites will be pre-processed and tokenized into word pieces suitable for FinBERT.
- *Embedding layer*: The tokenized text data will be converted into dense vector representations (embeddings) that capture the semantic meaning of words and phrases in the financial context.
- *Transformer layers*: The embeddings will be passed through multiple transformer layers that apply self-attention mechanisms to model the contextual relationships between words within each text input.
- *Output layer*: The final layer of FinBERT will output a classification for each text input, assigning sentiment scores (positive, negative, or neutral) based on the contextual understanding of the text.
- *Aggregation*: The individual sentiment scores will be aggregated over defined time periods to generate sentiment indicators that reflect the overall market sentiment for those periods.
- *Integration*: The aggregated sentiment scores will be integrated with historical stock prices and financial statement data, aligning them temporally to be used as features in the stock price prediction model.

Temporal fusion transformer (TFT) for time series forecasting TFT will be employed for time series forecasting in this research. TFT is designed to handle temporal patterns, static covariates, and known future inputs effectively. The methodology will involve several key steps:

- *Input layer*: The model will take in past observed values, known future inputs, and static covariates.
- *Variable selection networks*: Dynamically learn the relevance of each input feature.
- *Gated residual networks (GRN)*: Capture non-linear relationships between features.
- *LSTM layers*: Capture temporal dependencies and patterns in the time series data.
- *Attention mechanism*: Focus on important time steps and features to improve prediction accuracy.
- *Output Layer*: Produce the final stock price predictions.

3.4 Training and validation

- *Data split*: Divide the dataset into training, validation, and testing sets. This will ensure that the model is evaluated on unseen data to prevent overfitting.
- *Loss function*: Use Mean Absolute Error (MAE) or Mean Squared Error (MSE) as the loss function for model optimization.
- *Hyperparameter tuning*: Optimize hyperparameters such as learning rate.
- *Model evaluation*: Assess model performance using evaluation metrics like MAE, RMSE, and MAPE. Cross-validation will be performed to ensure the robustness of the model.

4 Discussion

4.1 Interpretation of results

While this paper is a work in progress, the foundational work on transformer-based models such as TFT for predicting time series and FinBERT for sentiment analysis show promising potential. The findings suggest that integrating FinBERT for sentiment analysis and TFT for time series forecasting can notably enhance the accuracy of stock price predictions.

4.2 Challenges

- *Computational resources*: Training transformer-based models is resource-intensive, requiring high-performance computing infrastructure. This could be a limitation for extended research or real-time application.
- *Temporal alignment*: Accurately synchronizing sentiment data with historical stock prices and financial statements is crucial. Misalignment can lead to inaccurate predictions and misleading insights.

4.3 Model variability

Since this research is still in progress, many textual and time series transformer-based models will be experimented with. If a better model is identified, FinBERT or TFT might be replaced with that model. The study will also compare different model performances and computational complexities to determine the most effective and efficient model. Additionally, different datasets will be incorporated depending on their availability to further enhance the robustness and accuracy of the prediction model.

5 Conclusion

This study proposes integrating advanced transformer-based models, specifically FinBERT for sentiment analysis and TFT for time series forecasting, to enhance the accuracy of stock price predictions. As suggested by foundational research, the promising potential of these models underscores their ability to process complex financial data and extract meaningful insights. While challenges such as computational resource demands and temporal alignment of data remain, ongoing experimentation with various models and datasets aims to address these issues. This study will continue to compare model performances and optimize computational efficiencies to identify the most effective approaches. Ultimately, this research aspires to contribute to the financial industry by providing more accurate and robust stock price prediction models, thereby aiding investors and analysts in making more informed decisions.

References

1. Araci, D.: FinBERT: A Pretrained Language Model for Financial Communications. arXiv preprint arXiv:1908.10063 (2019)
2. Araci, D.: FinBERT: A Pretrained Language Model for Financial Communications. arXiv preprint arXiv:1908.10063 (2019)
3. Atsalakis, G.S., Valavanis, K.P.: Surveying stock market forecasting techniques–Part II: Soft computing methods. Expert Systems with Applications, 36(3), 5932–5941 (2009)
4. Box, G.E.P., Jenkins, G.M.: Time Series Analysis: Forecasting and Control. Holden-Day (1970)
5. Churi, A., Chakraborty, D., Khatwani, R., Pinto, G., Shah, P., Sekhar, R.: Stock price prediction using Deep Learning and Sentiment Analysis. In: 2023 2nd International Conference on Futuristic Technologies (INCOFT), doi:10.1109/INCOFT60753.2023.10425124 (2023)
6. Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the Econometric Society, 987–1007 (1982)
7. Kim, K.I.: Financial time series forecasting using support vector machines. Neurocomputing, 55(1-2), 307–319 (2003)
8. Lim, B., Arik, S.O., Loeff, N., Pfister, T.: Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748–1764 (2021)
9. Lim, B., Arik, S.O., Loeff, N., Pfister, T.: Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748–1764 (2021)
10. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30 (2017)

Young Adults' knowledge of health misinformation on social media

Sivile Manene[1]

[1] University of Fort Hare, East London, SouthAfrica
smanene@ufh.ac.za

1 Introduction

Since the COVID-19 pandemic, the dissemination of health-related information on social media has grown exponentially [1]. An infodemic refers to the overabundance of accurate and inaccurate information during a pandemic, epidemic or endemic, resulting in the prevalence of propaganda, which leads to distrust in public health experts and institutions [2]. Recently, the World Health Organization (WHO) [3] adopted the term to describe the state of health information on social media during the COVID-19 pandemic. Social media was influential in the rise of the COVID-19 infodemic as it allows for ease in sharing and accessing information regardless of the user's geographical location [4].

The McKinsey Health Institute have found that the most prevalent demographic of social media users is 18-39 years old [5]. In an African context, social media serves as a means to stay informed of global health incidents and access relevant health information [6]. This raises concern over the impact misinformation could have on young adults who frequent the social media environment. According to the WHO, the repercussions of misinformation on social media include increased misinterpretation of scientific knowledge, polarization of opinions, escalation of fear, and decreased access to health care [7]. According to Chen et al. [8], over 60% of young adults have shared misinformation at some point in their usage of social media. Given the various challenges faced by the healthcare sectors of various countries in Africa, high awareness of health misinformation on social media could possibly improve engagement between citizens and healthcare experts [9]. This study examines student perception and awareness of health misinformation on social media. It focuses on understanding and addressing misinformation challenges among university young adults.

2 Literature Review

According to Stellefson et al. [10], social media serves as a critical tool in advocating for health information, policies and preventative measures in the event of public health crisis. Social media itself is a digital tool used for sharing ideas and information, making it a convenient means for reaching a wide range of people without having to physically campaign to the masses. It's also less for costly for people or organisations to use, making it a viable option for African governments to use for raising public health awareness in the event of pandemics, epidemics, and endemics. According to Kubheka et al. [11], Africans could benefit from social media's low cost, easy access, and virtual communities as health promotion tools. However, the nature of an infodemic, with widespread dissemination of all information, can impact public health outcomes during a health crisis.

2.1 Impact of Misinformation on Health Outcomes

Due to its ability to remove the physical barriers that often limit access to health information and resources, the use of social media in public health communication and education has significantly increased [10]. On the other hand, for that same reason, social media also serves as a medium for information disorder. Information disorder refers to the cumulation of false information to describe the intentional and unintentional distribution of false or harmful information [12]. Misinformation, on the other hand, refers to false information that is spread with no intent to cause any harm [2]. The nature of information disorder itself is highly deceptive[12]. This is due to claims containing misinformation and disinformation appearing as though they are factual, resulting in the adoption of unhealthy behaviours. A study by Suarez-Lledo & Alvarez-Galvez [13], found that over 87% of social media posts contained health misinformation, with 43% related to vaccines. During the peak of the COVID-19 pandemic in 2020, information disorder was highly prevalent on social media platforms. One particular example was the viral claim made by then president of the U.S, Donald Trump, stating that drinking bleach could mitigate the effects of COVID-19 [14]. Though these claims were made facetiously, the misinformation led to a plethora of poison cases and dire health issues for individuals who attempted to follow this advice.

Within the African context, instances of misinformation about the Ebola epidemic have been widely occurring on social media networks. Misleading claims like: the government using the epidemic to justify lockdowns, the outbreak being a cover to harvest body organs for sale, or the government falsifying cases to gain funding have been prevalent on Uganda's social media ecosystem [15]. This alludes to potential for misinformation to influence the behaviour of individuals of a given population. A study by Harjule et al. [16], which analysed misinformation sharing among young adults, it was found that young adults who spend the majority of time on social media always share misinformation. The main motivation for this is typically the need for self-expression and socialising, hence the need for raising awareness among young adult users. In a study by Lee et al. [17], a correlation was found between people's hesitancy to take the COVID-19 vaccine and misinformation. The study found that lack of expert knowledge, coupled with the prevalence of misinformation about the COVID-19 vaccine on social media had some influence on people's willingness to get vaccinated. In another study conducted by Garett & Young [18], it was found that anti-vaccination accounts have gained over 7.8 million followers on social media since the beginning of the pandemic. The study further found that most individuals who frequently engaged in anti-vaccination clusters on Facebook were vaccine hesitant. According to a study by Aïmeur et al. [19], due to recent COVID-19 infodemic, social media has now became a powerful source for false information dissemination. The same study further found that some of the challenges related to information disorder on social media stem from the deceptiveness of fabricated content, the lack of awareness among users, and the rate in which information can spread due to the nature of social media.

2.2 Challenges and Opportunities in Combating Misinformation

Some the challenges that emanate from combating and mitigating health misinformation on social media stem from the nature of digital platforms. One particular challenge, highlighted by Stellefson et al. [10], is the fact that health experts have no control on what, when, and how health information is disseminated on social media platforms. Given that social media users on popular platforms like Twitter (X) and Facebook can freely engage and interact with health information that may or may not be accurate according to expert consensus, managing information disorder becomes a difficult task. In the forefront of this challenge is the inability for users to distinguish between factual information and false information. According to Shu et al. [20], "social media users are susceptible to disinformation, and they often lack awareness of disinformation". This is due to the content of false information in that it, itself, is often highly sensationalised and written using extreme sentiments to make the user interact with the content more [20]. Harjule et al. [16], also illudes to some of the challenges of misinformation among young adults being lack of information literacy and the ability to detect differentiate between factual and false information on popular platforms.

3 Methods

This study used a quantitative survey approach, and a convenience sampling method to recruit study respondents. Data were analysed using descriptive statistics in SPSS V25. The questionnaire was adopted from Pundir, Devi, and Nath [21] that investigated fake news sharing on social media. The instrument used consisted of a demographic section with 4 items and seven scales with a total of 19 items which were all anchored on a five-point Likert scale. The scales were as followed: Awareness and Knowledge (5 items); Attitude (3 items); Perceived Behavioural Control (2 items); Fear of Missing Out (2 items); Social Networks (2 items); Sadism (3 items) and Intentions (2 items). The questionnaire was piloted amongst 30 young adults that did not participate in the study to ensure content validity. The amendments and recommendations from the pilot study were used to refine the questionnaire. Ethical approval was obtained from the University of Fort Hare Research Ethics Committee, and written consent was obtained from the respondents before the completion of the questionnaire. Data was collected from 255 respondents which were recruited through social media and sent a link to access the questionnaire via Survey Monkey. Respondents were informed of their rights and that participating in the study was voluntary and they could withdraw at any time. Only respondents above the age of 18-year and users fluent in English were asked to participate in the survey. Such criterion are in line with respondent selection used in earlier studies such as Pundir, Devi, and Nath [21]. Data was collected between August and September 2023.

4 Results

The results from the questionnaire are summarised and presented in this section. The reliability of the seven scales discussed in the previous section were tested through the Cronbach Alpha. The Cronbach's alpha score was found to show good internal consistency of 0.887 for the various scales, as suggested by Nunnaly [22]. Most respondents were male (63,4%) and were between the ages of 18-24 years of age (71,7%) followed by the 25-34 years of age group (26.4%). This is typical of the ages expected of young adults at a higher education of learning. The largest part of the respondents was still busy with their undergraduate degree (81.6%), while 12.3% were busy with their honour's qualification. Most respondents were aware of what fake news entails (94,5%) and that they could check suspicious health-related information through trusted sources (87.4%). Most respondents understood that social media sites should be verified (89.0%) and that there are consequences if they share non-verified health-related information on social media (85.8%). However, 18.8% of the respondents were not aware how to spot fake or misleading health-related information on social media.

Almost all the respondents (94,9%) believed that verifying health-related information was a good idea and had a favourable attitude towards verifying health-related information before sharing it on social media (84.7%) and beneficial to other social media users (95,3%). As far as social networks are concerned, 44,7% of respondents did not think that they should share health-related information on social media with their friends while 41.6% did not expect their friends to share health-related information on social media. Half of the respondents indicated that they amuse themselves (52.5%) or have fun (38.1%) when sharing health information on social media. However, most respondents indicated that fake news on social media bothered them (86.7%). Forty percent (40.0%) of respondents felt that their friends were having a more rewarding experience or gratifying experiences (38.4%) when searching for health information on social media. Respondents believed that they could individually contribute to the fight against fake and misleading health-related information on social media (83,9%) through fact-checking and reporting suspicious posts (84.4%). In the future, respondents indicated that they intend to use fact-checking sites (91.0%) and report misleading health-related information (89.4%).

5 Discussion

The aim of the study was to examine student perception and awareness about the spread of health misinformation on social media. Moreover, the researchers sought to identify how seven factors (Awareness and Knowledge; Attitude; Perceived Behavioural Control; Fear of Missing Out; Social Networks; Sadism and Intentions). The results points to a general awareness of misinformation of health-related information on social media and willingness to take specific actions to combat misinformation. Respondents were also intending to fact check and report suspicious information or sites on social media. Shalaby and Solim [23] reported similar results in their study that found harmful habits were relatively less among social media users who had knowledge about the dangers of such practices. To deal with health-related misinformation on social media, awareness and knowledge are the first defence in handling the spread and verification of fake news as they can distinguish between real and fake health information [24–28]. Attitude towards verification of health-related information was found to be very positive. This finding is in line with the results reported by Fishbein and Ajzen [29] that found that a favourable belief about verifying news before sharing on social media will impact positively on intention of the respondents.

Jiang and Beaudoin [30] have reported that there is a weak impact between social networks and intentions to verify health-related information on social media. In line with this study's results, fewer than half of respondents did not think they should share health-related information on social media with friends, nor did they expect their friends to do so. This could be attributed to weak ties or a lack of collective identity among social media users, given the diverse and heterogeneous nature of these networks [30]. Personal Behavioural Control is linked with perceived self-efficacy of social media users who believe they can help fight fake new are more likely to develop verification intention. This study supports this assumption as the respondents did believe they could individually contribute to mitigate the effects of misinformation on social media through activities such as fact-checking and reporting suspicious posts. Most respondents did report that seeing false health-related information on social media bothered them. Half of respondents used social media to amuse themselves and a smaller percentage to have fun with health-related information. Literature suggests that there is no link between sadism and intention to verify health-related information on social media [21, 31]. Previous studies have linked FoMO with various psychological insecurities such as the anxiety of being excluded from the network/group, sense of belongingness and seeking popularity in the context of social media [32, 33]. The results for this study found that a small proportion of the respondents felt that their friends were having a more rewarding experience or gratifying experiences when searching for health information on social media. The study's limitations are a small sample size and recruitment from only one university, limiting generalizability. However, it provides valuable insights into student perception and awareness of health misinformation on social media. The call for respondents was widely circulated to reduce bias. Future research should use a larger sample and include qualitative methods for more detailed analysis.

6 Conclusion

The aim of the study was to examine student perception and awareness about the spread of health misinformation on social media amongst young people. The potential consequences of misinformation about health-related information on social media are serious for the user and once the information is shared, it is very difficult to correct the information. The result from the study is useful to understand the problem of health-related misinformation on social media and how to start developing strategies to mitigate the problem. Moreover, this research has demonstrated that young people that use social media to seek health-related information should be made aware of the problem amongst young people using social media for health information.

References

1. Zielinski, C.: Infodemics and infodemiology: a short history, a long future. Revista Panamericana de Salud Pública. 45, 1 (2021). https://doi.org/10.26633/RPSP.2021.40
2. Manene, S., Hove, C., Cilliers, L.: Mitigating misinformation about the COVID-19 infodemic on social media: A conceptual framework. Jàmbá: Journal of Disaster Risk Studies. 15, (2023). https://doi.org/10.4102/jamba.v15i1.1416
3. World Health Organization: An ad hoc WHO technical consultation managing the COVID-19 infodemic: call for action, (2020)
4. Goel, A., Gupta, L.: Social Media in the Times of COVID-19. J Clin Rheumatol. 26, 220–223 (2020). https://doi.org/10.1097/RHU.0000000000001508
5. Coe, E., Enomot, K., Healy, C.: Gen Z mental health: The impact of tech and social media. McKinsey Health Institute. 1–10 (2023)
6. Omotayo, M., Oyetunji, O.: THE ROLE OF SOCIAL MEDIA IN INFORMATION DISSEMINATION IN THE PANDEMIC ERA. Middlebelt Journal of Library and Information Science. 19, 1596–1595 (2021)
7. World Health Organization: Infodemics and misinformation negatively affect people's health behaviours, new WHO review finds, https://www.who.int/europe/news/item/01-09-2022-infodemics-and-misinformation-negatively-affect-people-s-health-behaviours--new-who-review-finds
8. Chen, X., Sin, S.-C.J., Theng, Y.-L., Lee, C.S.: Why Students Share Misinformation on Social Media: Motivation, Gender, and Study-level Differences. The Journal of Academic Librarianship. 583–592 (2015)
9. Kollamparambil, U., Oyenubi, A., Nwosu, C.: COVID19 vaccine intentions in South Africa: health communication strategy to address vaccine hesitancy. BMC Public Health. 21, 2113 (2021). https://doi.org/10.1186/s12889-021-12196-4
10. Stellefson, M., Paige, S.R., Chaney, B.H., Chaney, J.D.: Evolving Role of Social Media in Health Promotion: Updated Responsibilities for Health Education Specialists. IJERPH. 17, 1153 (2020). https://doi.org/10.3390/ijerph17041153
11. Kubheka, B.Z., Carter, V., Mwaura, J.: Social media health promotion in South Africa: Opportunities and challenges. African Journal of Primary Health Care & Family Medicine. 12, (2020). https://doi.org/10.4102/phcfm.v12i1.2389
12. Wardle, C., Derakhshan, H.: Information Disorder : Toward an interdisciplinary framework for research and policy making. Council of Europe (2017)
13. Suarez-Lledo, V., Alvarez-Galvez, J.: Prevalence of Health Misinformation on Social Media: Systematic Review. J Med Internet Res. 23, e17187 (2021). https://doi.org/10.2196/17187
14. Chiu, A., Shepherd, K., Shammas, B., Itkowitz, C.: Trump claims controversial comment about injecting disinfectants was 'sarcastic', https://www.washingtonpost.com/nation/2020/04/24/disinfectant-injection-coronavirus-trump/, (2020)
15. Mwai, P., Schraer, R.: Ebola in Uganda: The people spreading misinformation online, https://www.bbc.com/news/63741125, (2022)
16. Harjule, P., Manva, Mohd.T., Mehta, T., Gurjar, S., Agarwal, B.: Analysing Misinformation Sharing Amongst College Students in India During COVID-19. Procedia Computer Science. 218, 671–685 (2023). https://doi.org/10.1016/j.procs.2023.01.048

17. Lee, S.K., Sun, J., Jang, S., Connelly, S.: Misinformation of COVID-19 vaccines and vaccine hesitancy. Sci Rep. 12, 13681 (2022). https://doi.org/10.1038/s41598-022-17430-6
18. Garett, R., Young, S.D.: Online misinformation and vaccine hesitancy. Translational Behavioral Medicine. 11, 2194–2199 (2021). https://doi.org/10.1093/tbm/ibab128
19. Aïmeur, E., Amri, S., Brassard, G.: Fake news, disinformation and misinformation in social media: a review. Soc. Netw. Anal. Min. 13, 30 (2023). https://doi.org/10.1007/s13278-023-01028-5
20. Shu, K., Bhattacharjee, A., Alatawi, F., Nazer, T.H., Ding, K., Karami, M., Liu, H.: Combating disinformation in a social media age. WIREs Data Min & Knowl. 10, e1385 (2020). https://doi.org/10.1002/widm.1385
21. Pundir, V., Devi, E.B., Nath, V.: Arresting fake news sharing on social media: a theory of planned behavior approach. Management Research Review. 44, 1108–1138 (2021). https://doi.org/10.1108/mrr-05-2020-0286
22. Nunnally, J.C.: An Overview of Psychological Measurement. 97–146 (1978). https://doi.org/10.1007/978-1-4684-2490-4_4
23. Shalaby, S.F., Soliman, M.A.: Knowledge, attitude, and practice of medical students regarding smoking and substance abuse, Cairo University, Egypt. J. Egypt. Public. Health. Assoc. 94, 11 (2019). https://doi.org/10.1186/s42506-019-0011-z
24. Vraga, E.K., Bode, L.: Using Expert Sources to Correct Health Misinformation in Social Media. Science Communication. 39, 621–645 (2017). https://doi.org/10.1177/1075547017731776
25. Parrott, J.: Finding Truth in the Age of Misinformation: Information Literacy in Islam. Yaqeen Institute for Islamic Research. 1–19 (2018)
26. Khan, M.L., Idris, I.K.: Recognise misinformation and verify before sharing: a reasoned action and information literacy perspective. Behaviour & Information Technology. 38, 1194–1212 (2019). https://doi.org/10.1080/0144929x.2019.1578828
27. Jahng, M.R., Lee, H., Rochadiat, A.M.P.: Public relations practitioners' management of fake news: Exploring key elements and acts of information authentication. Public Relations Review. 46, 101907 (2020). https://doi.org/10.1016/j.pubrev.2020.101907
28. Tseng, Y.-W., Yang, H.-K., Wang, W.-Y., Peng, W.-C.: KAHAN: Knowledge-Aware Hierarchical Attention Network for Fake News detection on Social Media. In: Companion Proceedings of the Web Conference 2022. pp. 868–875. ACM, Virtual Event, Lyon France (2022)
29. Martin Fishbein, Fishbein, M., Ajzen, I.: Predicting and Changing Behavior: The Reasoned Action Approach. (2009)
30. Jiang, S., Beaudoin, C.E.: Smoking Prevention in China: A Content Analysis of an Anti-Smoking Social Media Campaign. Journal of Health Communication. 21, 755–764 (2016). https://doi.org/10.1080/10810730.2016.1157653
31. Ramos-Villagrasa, P.J., del Río, E.F., Fernández-del-Río, E., Barrada, J.R.: Does Evil Prevail? The "Bright" and "Dark" Sides of Personality as Predictors of Adaptive Performance. Sustainability. 12, 474 (2020). https://doi.org/10.3390/su12020474
32. Baumeister, R.F., DeWall, C.N., Ciarocco, N.J., Twenge, J.M.: Social exclusion impairs self-regulation. Journal of Personality and Social Psychology. 88, 589–604 (2005). https://doi.org/10.1037/0022-3514.88.4.589
33. Blackwell, D., Leaman, C., Tramposch, R., Osborne, C., Liss, M.: Extraversion, neuroticism, attachment style and fear of missing out as predictors of social media use and addiction. Personality and Individual Differences. 116, 69–72 (2017). https://doi.org/10.1016/j.paid.2017.04.039

A Comparative Analysis of Autonomous Smartwatch Data Management

Carmen Bekker[1[0000-0002-7400-0041]] and Brenda Scholtz[2[0000-0002-0844-1383]]

Nelson Mandela University, Gqeberha, South Africa
[1] carmen.bekker@mandela.ac.za
[2] brenda.scholtz@mandela.ac.za

Abstract. Smartwatches have gained much popularity in recent years as these devices enhance an individual's well-being and encourage healthy behaviours. These devices detect, track and analyse biological and physiological data, which allow users to gain insights and make decisions regarding their personal data-based insights. However, good data management is crucial to ensure that users gain value from their personal smartwatches. A comparative analysis was conducted of the data management of popular smartwatches. Preliminary results highlighted that Samsung requires less rigour and data preparation before analysis compared to Fitbit when considering step count as a test case. However, Fitbit provided data that was understandable and easy to read.

Keywords: Smartwatches, Data Management.

1 Introduction

Data is a crucial asset for daily activities, can generate future value and assists in decision making [1]. The importance of data and information in connection with continuous capabilities in data collection and processing, as well as progressive technological advancements, have led to the increased usage of data management [2]. The key driver of data management is ensuring that value is derived from data [3].

Smartwatches have biological and physiological data detection, tracking and analysis capabilities [4]. These capabilities interest smartwatch users as they seek to understand and enhance their physical health and fitness, ultimately improving their general well-being [5, 6]. Some studies have shown that smartwatch users gain more value from a larger volume of collected data related to their physical activity and health compared to a smaller volume of data [7].

Smartwatch data collection is autonomous. Smartwatch autonomy refers to the self-managed activities that these devices conduct, which do not require user control, such as measuring step count and heart rate [8]. Various studies have found that smartwatch autonomy benefits users, positively impacting their perceived usefulness and value. Smartwatch users rely on their collected physical activity data so that they can reflect on the insights of their movements and behaviours, which allows them to make data-driven decisions [9].

Several studies have investigated various elements of smartwatch data management, such as the accuracy of autonomous smartwatch data (i.e., step count and heart rate), which is impacted by the method and instrument of data collection [10, 11]. Other studies considered the methods of data analysis when working with raw smartwatch data [12, 13]. However, the steps of smartwatch data management and the availability of different data attributes of the raw data depending on the smartwatch brand are rarely discussed. Users and researchers can gain insights and can identify data correlations if raw smartwatch data is managed and analysed. For example, users and researchers can determine the long-term impact of physical activity (i.e., step count and exercises) on sleep quality by using raw smartwatch data [14].

2 Smartwatch Data Management

From a high-level perspective, data management encompasses many techniques and activities [1]. Often, data management is considered from an organisational perspective rather than a user perspective (i.e., smartwatch manufacturer versus smartwatch user).

A generic data management process proposed by [15] can be considered and applied to smartwatches to assist smartwatch users in managing and analysing their physical activity data. The process consists of five steps: data collection, preparation, storage, analysis, and distribution [15]. Fig. 1 presents the visualisation of these steps.

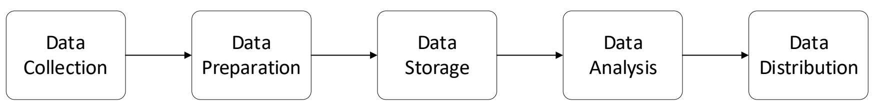

Fig. 1. A Generic DM Process [15]

The smartwatch data collection step is autonomous since the device collects biographical and physiological data using a combination of built-in sensors. Step count data is collected by 3-axis accelerometers and gyroscopes, which determine the wrist displacement of a smartwatch user in different spaces [10, 16]. On the other hand, heart rate data is collected using light-emitting diodes (LEDs) that determine a smartwatch user's blood volume through LED light reflection on the skin [10]. The data preparation step is performed using various algorithms depending on the smartwatch manufacturer [17].

Once the data has been prepared and processed, it is typically stored temporarily in the persistent memory of the smartwatch [18]. However, due to the limited memory capacity available on smartwatches, the data needs to be transmitted from these devices to a wirelessly connected mobile application with a larger memory capacity [12]. Accessing cloud-based services and previously collected smartwatch data via a smartphone is becoming increasingly common [19]. Users can register accounts with cloud-based services and login to these accounts from their smartphones using the companion mobile application. Some studies have highlighted that using a smartwatch with a companion mobile application significantly improves feelings of accomplishment, physical health and emotions in smartwatch users [20].

Often, smartwatch sensor data can be downloaded and accessed in an Extensible Markup Language (XML), JavaScript Object Notation (JSON) or comma-separated values (CSV) format when linked to a cloud services account and companion mobile application [12, 13]. Once the data is in an accessible format, the data analysis step can take place, which depends on the approach of smartwatch users or researchers. One example of an approach to processing CSV files is using a program that extracts and organises the CSV file data into columns. The smartwatch data can be visualised after being obtained from the CSV file data.

After analysis, the smartwatch data can be distributed internally or externally. Internal data distribution is for the smartwatch user who can view their data through visualisations on the device and companion mobile application [21]. Users can also distribute their data externally to friends, family or medical professionals using the available sharing features in commercial smartwatch platforms [22].

3 Comparative Analysis of Autonomous Smartwatch Data

A comparative analysis was conducted where two cases were compared according to their similarities and differences [23]. The authors collected data from two smartwatches, the Fitbit Charge 4 and the Samsung Galaxy Watch 6 as test cases. These devices were selected due to the availability of equipment provided by the Nelson Mandela University's Telkom Centre of Excellence (CoE) Smart Unit.

The authors followed the steps of the generic data management process, as proposed in [15], for April 2024, and then exported the smartwatch data according to the instructions on the smartwatch manufacturer's websites. Step count was used as the case for the comparative analysis and Microsoft Excel was used to store the data.

By default, the autonomously collected smartwatch data was stored according to factors such as physical activities, heart rate, stress, and sleep in a CSV format for both the Fitbit and Samsung smartwatches. The CSV file format is beneficial for smartwatch users and researchers regarding data accessibility as CSV files are machine-readable and can be accessed using various programs such as Microsoft Excel, Google Sheets and LibreOffice [24, 25].

Table 1 summarised the findings from analysing the step count data of the Fitbit Charge 4 and the Samsung Galaxy Watch 6 smartwatches. The general data management techniques were considered for data storage as the folder structure, file structure and data attribute labels remain the same regardless of the Fitbit and Samsung device. The smartwatch device model (i.e. Charge 4 and Galaxy Watch 6) is the factor that determines whether files and attributes contain data (e.g., the Fitbit Charge 4 does not autonomously collect stress data, which results in no stress data being collected and exported). The data storage and analysis of Fitbit and Samsung were considered.

Data storage factors related to the step count file and consisted of the folder structure, location of exported data, and whether the file required conversion from one file format to another for easy readability. Data analysis factors related to the actual step count data, available data attributes and the format of these attributes in the step count files.

Table 1. Fitbit and Samsung Step Count Data Management

Data Management Technique	Aspect	Fitbit Charge 4	Samsung Galaxy Watch 6
Data Storage	Easy Navigation (Folder Structure)		✓
	Requires Conversion to CSV	✓	
Data Analysis	Aggregated Step Count (Daily)		✓
	Readable Date Format	✓	

Fitbit uses a folder structure for grouping its data, which makes locating the step count file was a challenge. Samsung, on the other hand, stores all CSV files in one folder. The step count file for Fitbit was located in the "Global Export Data" folder, which users are not likely to consider as the first folder to search. Unlike most of the other Fitbit data files, the step count file was also not in the expected CSV format. Instead, step count data was in a JSON file format, which is not easy to access compared to a CSV. JSON files are often used for data transmission between a web application and a server and can be opened using text editors (e.g., Notepad) [26]. However, the data within a JSON file does not appear in a tabular format such as CSV files, which makes it difficult to analyse. Therefore, the JSON step count file for Fitbit was converted to a CSV format for analysis.

There was a noticeable difference between the Fitbit converted CSV file data and the default Samsung CSV file. The Fitbit step count file consisted of two unnamed attributes: a date-time timestamp attribute and a step count per timestamp attribute. On the other hand, Samsung presented more data by grouping all daily physical activities into one file. The Samsung CSV file consisted of various data attributes, such as step count, distance, movement type, times when the data was updated, calories and a date-time timestamp attribute. Combining these data attributes into one file makes accessing various types of data convenient for analysis.

Interestingly, regarding step count, Fitbit did not aggregate its daily step count in line with what is usually represented in the visualisations on a user's smartwatch and companion mobile application. The step count data needed to undergo aggregation by date rather than by the date-time attribute. Data aggregation was a tedious process as the steps were measured for each timestamp, even when the step count was zero. By not aggerating daily step count, the Fitbit CSV file contained almost 17,775 rows for a single month of data collection. To aggregate the Fitbit data, the date time attribute needed to be converted to an integer before an Excel function was used to sum all steps from each day in the month.

On the other hand, Samsung already aggregated daily step count data, which aligns with the data that users can visualise on their smartwatch and companion mobile application. Interestingly, the date time attribute was in an unreadable format even when the Excel column was converted to a Custom field for date-time. A Microsoft online forum was used to determine a suitable way to convert the data using an Excel function.

4 Conclusion

The findings of the comparative analysis revealed that Samsung required less rigour and additional data preparation for data analysis. These results were due to Samsung providing step count data in an accessible file format (i.e., CSV), aggregating daily steps on behalf of the user and making the necessary files easy to find for its users and potential researchers. Furthermore, it is convenient for smartwatch users to access various data types in one file, such as steps, distance, calories and exercise types. However, despite requiring additional data preparation before analysis, Fitbit provides simple data such as the step count and timestamp for each day of the month, which could be what certain users and researchers are looking for when analysing the data. Users and researchers alike can determine the accuracy of the step count of the Fitbit smartwatch by analysing step count intervals compared to manual measurements.

It is recommended that future work considers which combinations of data attributes can provide insights into the physical activity and health of smartwatch users beyond what is provided on the smartwatch and the companion mobile applications. The data attributes of other popular smartwatch brands should also be considered in research studies where data is being collected from multiple users with varying devices.

Acknowledgments. The authors would like to thank the Telkom Centre of Excellence (CoE) Smart Unit at Nelson Mandela University for providing the smartwatches for testing the data used in this paper.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Dama International, DAMA-DMBOK : Data Management Body of Knowledge (2nd Edition), New Jersey: Technics Publications, 2017.
2. M. Abu-Elkheir, M. Hayajneh and N. A. Ali, “Data Management for the Internet of Things: Design Primitives and Solution,” Sensors Volume 13(11), pp. 15582 - 15612, 2013.
3. K. Gordon, Principles of Data Management: Facilitating Information Sharing (3rd Edition), BCS, The Chartered Institute for IT, 2022.
4. C. Peng, N. Xi, H. Zhao and J. Hamari, “Acceptance of Wearable Technology: A Meta-Analysis,” Proceedings of the 55th Hawaii International Conference on System Sciences, pp. 5101 - 5110, 2022.
5. E. Karapanos, R. Gouveia, M. Hassenzahl and J. Forlizzi, “Wellbeing in the Making: Peoples’ Experiences with Wearable Activity Trackers,” Psychology of Well-Being volume 6 (4), pp. 1-17, 2016.

6. T. Ha, B. Beijnon and J.-H. Kim, “Examining user perceptions of smartwatch through dynamic topic modeling,” Telematics Informatics 34, pp. 1262-1273, 2017.
7. H. Kang and E. H. Jung, “The smart wearables-privacy paradox: A cluster analysis of smartwatch users,” Behaviour and Information Technology Volume 40 Issue 1, pp. 1 - 14, 2020.
8. W.-C. Cho, K. Y. Lee and S.-B. Yang, “What makes you feel attached to smartwatches? The stimulus–organism–response (S–O–R) perspectives,” Information Technology & People, Vol. 32 No. 2, pp. 319-343, 2019.
9. C. Siepmann and P. Kowalczuk, “Understanding continued smartwatch usage: the role of emotional,” Electronic Markets 31, pp. 795-809, 2021.
10. F. El-Amrawy, B. Pharm and M. I. Nounou, “Are Currently Available Wearable Devices for Activity Tracking and Heart Rate Monitoring Accurate, Precise, and Medically Beneficial?,” Healthcare Informatics Research 2015 Volume 21(4), pp. 315 - 320, 2015.
11. A. Farrokhi, R. Farahbakhsh, J. Rezazadeh and R. Minerva, “Application of Internet of Things and artificial intelligence for smart fitness: A survey,” Computer Networks Volume 189 Issue 4, pp. 1 - 21, 2021.
12. F. De Arriba-Pérez, M. Caeiro-Rodríguez and J. M. Santos-Gago, “Collection and Processing of Data from Wrist Wearable Devices in Heterogeneous and Multiple-User Scenarios,” Sensors (Basel). 16(9), p. doi: 10.3390/s16091538, 2016.
13. M. Buelga Suárez, A. Rebolleda Sánchez , M. Pascual Izco, J. García Montalvo and G. L. Alonso Salinas, “Smartwatch: Looking beyond what you see,” Journal of Electrocardiology 81, pp. 51-59, 2023.
14. T. Semplonius and T. Willoughby, “Long-Term Links between Physical Activity and Sleep Quality,” Medical Science Sports Exercise 50(12), pp. 2418-2424. doi: 10.1249/MSS.0000000000001706, 2018.
15. C. M. Monsanto, “What Is Data Management? A Complete Guide With Examples,” 2022. [Online]. Available: https://blog.hubspot.com/website/data-management.
16. K. Ricchio, P. Lyter-Antonneau and J. M. Palao, “Reliability of fitness trackers at different prices for measuring steps and heart rate: a pilot study,” Central European Journal of Sport Sciences and Medicine 24(4), p. 57–64, 2018.
17. N. Kamal Basha, E. C.-X. Aw and S. H.-W. Chuah, “Are we so over smartwatches? Or can technology, fashion, and psychographic attributes sustain smartwatch usage?,” Technology in Society 69 (2022) issue 101952, pp. 1-11, 2022.
18. M. Boukhechba and L. Barnes, “SWear: Sensing using WEARables. Generalized Human Crowdsensing on Smartwatches,” in 11th International Conference on Applied Human Factors and Ergonomics, San Diego, 2020.
19. W.-H. Lee and R. Lee, “Implicit Sensor-based Authentication of Smartphone Users with Smartwatch,” in HASP 2016: Proceedings of the Hardware and Architectural Support for Security and Privacy 2016, Seoul, 2016.

20. B. Stiglbauer, S. Weber and B. Batinic, “Does your health really benefit from using a self-tracking device? Evidence from a longitudinal randomized control trial,” Computers in Human Behavior 94, pp. 131-139, 2019.
21. T. Blascheck, L. Besançon, A. Bezerianos, B. Lee and P. Isenberg, “Glanceable Visualization: Studies of Data Comparison Performance on Smartwatches,” IEEE Transactions on Visualization and Computer Graphics 25(1), pp. 630-640, 2019.
22. D. A. Epstein, “Everyday Personal Informatics,” 2018.
23. C. G. Pickvance, “Four varieties of comparative analysis,” Journal of Housing and the Built Environment 16, p. 7–28, 2001.
24. T. Velliquette, J. Welch, M. Crow, R. Devarakonda, S. Heinz and R. Crystal-Ornelas, “ESS-DIVE Reporting Format for Comma-separated Values (CSV) File Structure,” U.S. Department of Energy Office of Scientific and Technical Information, https://doi.org/10.15485/1734841, 2021.
25. A. Nayak, B. Božić and L. Longo , “Data Quality Assessment of Comma Separated Values Using Linked Data Approach,” in Business Information Systems Workshops, Springer International Publishing, 2022, p. 240–250.
26. F. Pezoa, J. L. Reutter, F. Suarez, M. Ugarte and D. Vrgoč, “Foundations of JSON Schema,” in Proceedings of the 25th International Conference on World Wide Web, Montréal, 2016.

Exploring the Efficient Implementation of Knowledge Base Generators for KLM-Style Defeasible Reasoning

Chipo Hamayobe[0009−0008−0987−814X] ⋆ ⋆⋆ ⋆⋆⋆

University of Cape Town
chipo@cs.uct.ac.za

1 Introduction

Knowledge representation and reasoning is a branch of artificial intelligence that structures knowledge through formal logic. This structure allows specific rules to be applied to manipulate and infer new information from the existing knowledge [2].

Knowledge bases serve as structured repositories of information and facts. This research focuses on creating advanced, defeasible knowledge bases essential for handling information that may not be absolute and could have exceptions or conditions. Through defeasible reasoning, these knowledge bases offer a means to model and deliberate over information that is uncertain or incomplete. Utilising an extension to propositional logic introduced by Kraus, Lehmann and Magidor (KLM) [6], the research seeks to represent data within these specialised knowledge bases effectively. The goal is to enhance the understanding and assessment of new entailment relations in the context of defeasible reasoning by developing complex knowledge bases.

This paper unveils a detailed and customisable plan for a non-deterministic generator for defeasible knowledge bases, including a more efficient version. It also introduces a research path to examine how various configurations of defeasible implications affect the time it takes to generate them and compare the performance of both generator versions. These generators will be crafted to produce knowledge bases across a broad spectrum of configurations, facilitating an extensive investigation into diverse structures of knowledge bases. We aim to employ the *Rational Closure* [5] (and *Lexicographic Closure* [3]) `BaseRank algorithm` to ensure the generated knowledge bases are organised to align with the algorithm's ranking process [4].

2 Propositional Logic

Propositional logic is a foundational logical system utilised to understand and represent knowledge about the world. Its language, denoted as $\mathcal{L}$, is constructed

⋆ Supervised by `Professor Tommie Meyer (tmeyer@cs.uct.ac.za)`
⋆⋆ Co-Supervised by `Dr Jonas Haldimann (jonas.haldimann@fernuni-hagen.de)`
⋆⋆⋆ Supported by the `Artificial Intelligence Research Unit (airu.org.za)`

from basic units called *atoms* and a series of *boolean operators*. An atom represents a statement or an assertion, typically signified by lowercase letters from the Latin alphabet, for example, $\mathcal{P} = \{b, c, d, e, ...\}$. Each atom is assigned a truth value, either `true` (T) or `false` (F). By combining these atoms with boolean connectives, such as not ($\neg$), or ($\vee$), and ($\wedge$), implies ($\rightarrow$), and if and only if ($\leftrightarrow$), propositional formulas are formed [1].

The semantics of propositional logic include the notion of validity, where a statement's premises, if `true`, ensure the truth of its conclusion. A statement in propositional logic is considered valid if the conclusion (the *consequent*) necessarily follows from the premises (the *antecedent*). The process of assigning truth values to atoms, known as *interpretations, valuations* or possible worlds, plays a crucial role in this. Mathematically, this assignment process can be represented by a function $u : \mathcal{P} \rightarrow \{T, F\}$, where $\mathcal{P}$ is the set of propositional atoms and $\{T, F\}$ the possible truth values.

When an interpretation assigns the truth value of `true` to a propositional atom, that atom is considered satisfied within that interpretation. This concept of satisfaction is symbolised by the notation $\Vdash$. For instance, given an interpretation u, if for any atom $b \in \mathcal{P}$ it holds that $u(\mathcal{P}) = T$, then it is said that $u \Vdash b$. Conversely, if $u(\mathcal{P}) = F$, it is denoted as $u \nVdash b$.

3 Defeasible Reasoning

Defeasible reasoning, characterised by its *non-monotonic* nature, permits the revision of conclusions drawn from incomplete or uncertain knowledge upon the emergence of conflicting information, aligning more closely with human cognitive processes. Kraus, Lehmann, and Magidor advanced this field through the development of the KLM framework, which augments propositional logic with the preferential consequence relation symbol $\mid\sim$, denoting defeasible implications ($\alpha \mid\sim \beta$, with $\alpha, \beta \in \mathcal{L}$) to signify that the truth of α typically suggests the probable truth of β [8]. Thus, a defeasible knowledge base comprises a collection of such implications.

In contrast to propositional logic, defeasible reasoning lacks a standardised approach for determining defeasible entailment ($\mid\approx$). The KLM framework specifies criteria for a defeasible entailment relation and any method meeting these criteria is deemed LM-Rational.

4 Defeasible Implication Generation

Defeasible implications (DIs), constituting a defeasible knowledge base, comprise two main elements: the antecedent and the consequent, linked by $\mid\sim$ to form a defeasible implication. An aggregation of these implications creates a *defeasible knowledge base*. The `IDefeasibleImplicationBuilder` interface facilitates the generation of DIs for the knowledge base, categorising them into three types: structure, simple, and complex DIs [9].

Within any defeasible knowledge base, atoms serve as the foundational elements of defeasible implications, with their generation overseen by the `IAtomBuilder` interface. To infuse atom creation with a measure of pseudo-randomness, `IAtomBuilder` iteratively selects random characters from a specified character set, dependent on the atom's intended length, thereby guaranteeing the uniqueness of each newly produced knowledge base. It maintains a record of generated atoms to prevent duplications, preserving the knowledge base's intended structure [7].

The *GrowthValidator* function, part of `IAtomBuilder`, will dynamically adjust the atom length based on the volume of atoms produced, ensuring a continuous output of unique atoms from a limited character set. `IAtomBuilder` also allows for selecting diverse character sets—upperlatin, lowerlatin, altlatin, and greek—for atom generation. Adherence to the *Singleton Design Pattern* will guarantee a single instance of `IAtomBuilder` across the knowledge base generation lifecycle, maintaining consistency in the atom list and their lengths amidst concurrent atom generation processes.

5 Knowledge Base Construction

The `IDefeasibleImplicatioAllocation` interface orchestrates the layout of DIs within a knowledge base, manipulating their organisation across varying ranks based on user-defined parameters: the total number of DIs, the number of ranks, and the chosen distribution type, culminating in a sorted set depicting the DIs distribution across ranks. `IDefeasibleImplicatioAllocation` will implement four distinct allocation methods:

- **Uniform Allocation** distributes DIs uniformly across all ranks, with any surplus DIs allocated from the bottom rank upwards.
- **Arbitrary Allocation** disperses DIs across ranks randomly.
- **Consistent Increase Allocation** assigns an increasing number of DIs to successive ranks, adhering to a linear growth model.
- **Consistent Decrease Allocation** distributes DIs in a linearly decreasing fashion across ranks.

The `IKnowledgeBaseGenerator` interface will be tasked with constructing knowledge bases by managing the creation of DIs through the *Generate* function. This method will sequentially construct the knowledge base, starting from rank_0, and requires input arrays detailing DI allocations, the complexities of both antecedent and consequent, connective types, and two boolean values indicating whether only simple DIs should be generated and whether consequents should be reused across ranks. The time complexity for `IKnowledgeBaseGenerator` is expected to be $O(r * n)$, where r represents ranks and n denotes the number of DIs. The process will encompass several phases:

- **Activation**: The knowledge base will be initiated as a set of linked HashSet data structures. Base atoms for rank_0 consequent and antecedent are generated alongside a list of atoms reusable across any rank.
- **Rank Creation Iteration**: This involves iterating through ranks, starting from rank_0, and executing the following steps for each rank:

- Maintain lists for the current rank's generated DIs, reusable atoms for antecedents, and a temporary list for any-rank reusable atoms. Include $\texttt{rank}_0$ antecedent in the current rank's antecedents for reuse within the rank.
- For $\texttt{rank}_0$, construct the baseline DI using the base rank antecedent and consequents.
- For ranks $\texttt{rank}_0$, generate initial structure DIs based on the reuse consequent indicator and the required number of DIs for the rank. Depending on the indicator and DI count, either a new rank consequent is created or a restricted rank creation function is employed.
- For the remaining DIs, if the simple only indicator is `true`, choose among three simple DI strategies. If any rank has no atoms, prioritise recycling atoms over reusing consequents. Both the negation of antecedents and consequent reuse will contribute atoms to any rank atoms reserves for future reuse. If the simple indicator is `false`, utilise the key generation algorithm to determine the connection type and invoke the corresponding complex DI function from the `IDefeasibleImplicationBuilder` interface.
- Negate the $\texttt{rank}_0$ consequents to prepare for subsequent ranks.
- Integrate the generated DI set into the knowledge base, update any rank atom collection with contents from the reserves, and advance the rank counter.

6 System Design and Implementation

The system's implementation and architecture will utilise Java, with Maven for software system management. UML diagrams, other architectural diagrams, and class descriptions will be available in a public repository, as well as in the appendix of the research output document. The software features a command-line interface, allowing direct input or text file specifications for knowledge base generation, including:

- **Command-line Interface**: Enables direct specification entry or text file input for operation.
- **Number of Ranks**: Users specify a non-negative number of ranks for the knowledge base.
- **Allocation**: Users select an allocation type for defeasible implications.
- **Number of DIs**: Users input a number, adhering to a minimum requirement, to ensure structural integrity for the specified ranks and distribution.
- **Simple Only**: Determines whether the knowledge base comprises solely simple DIs or a combination of simple and complex DIs.
- **Reuse Consequent**: Decides whether the knowledge base reuses rankBaseCons across all ranks or generates new ones for each.
- **Antecedent Complexity**: Users specify the permissible number of connectives in complex DI antecedents, choosing from 0, 1, and 2.
- **Consequent Complexity**: Similar to antecedent complexity, users dictate the allowable connectives in complex DI consequents.
- **Connective Types**: Users define permissible connective types in complex DIs, choosing from disjunction (1), conjunction (2), implication (3), bi-implication (4), and mixture (5).
- **Adjustable Connective Symbols**: Users can customise symbols for defeasible implication, disjunction, conjunction, implication, bi-implication, and negation.

- **Adjustable Character Set**: Users select a character set for atom generation: upperlatin, lowerlatin, altlatin, or greek.
- **Generator Type**: Users choose between a standard or optimised generator for the knowledge base creation.
- **Export**: Allows the knowledge base to be displayed and exported to a text file for entailment relation testing.
- **ReGenerate, Settings, and Quit**: Offers options to regenerate the knowledge base, alter specifications, or exit the program.

7 Conclusions

The research objectives aim to culminate in developing a parameterised, non-deterministic defeasible knowledge base generator that aligns with the BaseRank algorithm model. Additionally, an optimised variant of the generator will also be created. A comparative analysis of the envisioned generators will reveal which version offers a considerable speed enhancement compared to the standard generator. It should be noted that a knowledge base comprising complex implications requires more generation time than those with a straightforward impact. These and other practical factors will be thoroughly considered when analysing the material performance of the generators when employed in both *Rational Closure* and *Lexicographic Closure* entailment and justification determinations.

References

1. Ben-Ari, M.: Propositional Logic: Formulas, Models, Tableaux, pp. 1, 7–47. Springer London, London (2012)
2. Brachman, R.J., Levesque, H.J.: Actions. In: Knowledge Representation and Reasoning. pp. 285–303. Elsevier (2004)
3. Casini, G., Meyer, T., Varzinczak, I.: Taking defeasible entailment beyond rational closure. In: Logics in Artificial Intelligence, pp. 182–197. Lecture Notes in Computer Science, Springer International Publishing, Cham (2019)
4. Giordano, L., Gliozzi, V., Olivetti, N., Pozzato, G.: Semantic characterization of rational closure: From propositional logic to description logics. Artificial Intelligence **226**, 1–33 (2015). https://doi.org/https://doi.org/10.1016/j.artint.2015.05.001
5. Kaliski, A.: An Overview of KLM-Style Defeasible Entailment. Master's thesis, Faculty of Science, University of Cape Town, Rondebosch, Cape Town, 7700 (2020)
6. Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logics. Artificial Intelligence **44**(1), 167–207 (1990). https://doi.org/https://doi.org/10.1016/0004-3702(90)90101-5
7. Lang, A.: Extending Defeasible Reasoning Beyond Rational Closure. Honours project, Faculty of Science, University of Cape Town, Rondebosch, Cape Town, 7700 (2023)
8. Lehmann, D., Magidor, M.: What does a conditional knowledge base entail? Artificial Intelligence **55**(1), 1–60 (1992). https://doi.org/https://doi.org/10.1016/0004-3702(92)90041-U
9. Pillay, E.: An Investigation into the Scalability of Rational Closure V2. Honours project, Faculty of Science, University of Cape Town, Rondebosch, Cape Town, 7700 (2022)

Machine Learning Techniques for Improving Real-time APT Detection in Cloud Computing Environments

Mihlali Mbhedeni [1[0000−1111−2222−3333]] and **Dr C. Guragena** [2[1111−2222−3333−4444]]
University of Fort Hare. 1 King William's Town Rd, Alice, 5700. Private Bag X1314, Alice, 5700.

Abstract. This research explores the use of machine learning techniques to enhance the real-time detection of Advanced Persistent Threats (APTs) in cloud computing environments. By leveraging algorithms like Adaboost, Support Vector Machines (SVM), Random Forests, Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs), the study aims to develop a robust model capable of identifying these sophisticated threats. The research involves data collection, preprocessing, model development, and evaluation using real-world dataset simulations in a controlled cloud environment. This study's findings are expected to contribute to the advancement of cloud security practices and provide a foundation for future research in APT detection using machine learning.

Keywords: Advanced Persistent Threats, Machine Learning, Cloud Computing, Real-time Detection, Cybersecurity.

1. Introduction

The rise of Advanced Persistent Threats (APTs) has posed a significant challenge to both private and corporate sectors in recent years. APTs are long-term, targeted cyberattacks where intruders gain unauthorized access to a network and remain undetected for extended periods (Hejase and Kazan, 2020). These threats employ evolving tools and techniques, rendering existing security measures ineffective (Baksi, 2022), and result in severe security threats and financial losses globally (Xiong et al., 2022). APTs are particularly challenging to detect due to their prolonged network presence and potential to cause system crashes through heavy traffic (Joloudari et al., 2020). Despite extensive research on host security, the rapid evolution of these threats has outpaced current defenses (Xiong et al., 2022).

Given the growing sophistication of APTs, traditional detection methods are increasingly inadequate, so addressing these threats requires a comprehensive understanding of APT operations and motivations (Momeni Milajerdi et al., 2019). This study aims to enhance real-time APT detection in cloud computing environments through the application of machine learning techniques, specifically focusing on the Adaboost classifier.

2. Problem Statement

Detecting Advanced Persistent Threats in cloud computing environments is challenging due to their stealthy and persistent nature (Sharma et al., 2023). Existing methods lack real-time capabilities, resulting in delayed detection and response times (Hassan et al., 2020).

3. Research Questions

The main research is, "How can machine-learning techniques be used to improve real-time detection of APTs in cloud computing environments, considering the changing nature of APT attacks and the importance of timely and accurate detection to avoid data breaches and financial losses?" This question will be addressed through the addressing of the following sub questions:

- What are the current advanced machine-learning techniques and algorithms used for APT detection in cloud computing environments, and how do they compare in terms of effectiveness and efficiency?
- What datasets can be efficiently accessed and used, and which ones are suitable for training and testing an APT detection model in cloud environments?
- How can network traffic patterns, system log data, and application behavior be efficiently used to develop and apply machine learning models for real-time APT detection in cloud environments?
- How does the performance of the developed machine learning model, compare with existing techniques when evaluated using real-world dataset simulations for APT detection in cloud environments?

4. Research Aim and Objectives

The aim of this research is to develop and evaluate machine learning-based techniques to improve the real-time detection of APTs in cloud computing environments. The research has the following objectives:

- To Assess existing and most recent machine-learning techniques and algorithms suitable for APT detection in cloud environments.
- To Gather relevant datasets for training and evaluation of the APT detection model.
- To Develop and implement a machine learning model based on network traffic patterns, system log data, and application behaviour that can efficiently identify APTs in real-time.
- To Evaluate the performance of the designed model using real-world dataset simulations and compare it with existing techniques to demonstrate its effectiveness and efficiency in detecting APTs.

5. Related Work

5.1. Evolution and Characteristics of APTs

APTs represent a significant evolution in cyber threats, characterized by their silent, long-term infiltration tactics (Alshamrani et al., 2019). Traditional detection methods, such as signature-based and heuristic-based detection, struggle to identify these sophisticated threats due to their ability to mimic legitimate user behavior and use encrypted communications (Sharma et al., 2023). As a result, the development of advanced detection techniques is crucial.

5.2. Machine Learning for APT Detection

Machine learning has emerged as a promising approach to enhance APT detection. Anomaly detection techniques, such as isolation forests and autoencoders, have been used to identify unusual patterns in network traffic indicative of APTs (Stojanović, 2020). These techniques are valuable for spotting behavioral anomalies, which are essential for detecting advanced threats overlooked by traditional methods (Xiong et al., 2022).

5.3. Supervised and Unsupervised Learning Methods

Supervised learning methods, including Support Vector Machines (SVMs) and Random Forests, have demonstrated effectiveness in categorizing malicious activities by learning from labeled datasets (AL-Aamri et al., 2023). These models excel in managing large datasets and producing precise classifications. Unsupervised learning methods, such as DBSCAN and K-means clustering, identify outliers and group similar data points, aiding in the detection of unusual network traffic without requiring labeled data (Alghamdi and Reger, 2020).

5.4. Deep Learning Techniques

Deep learning techniques, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have been adapted to evaluate sequence data and capture temporal dependencies (Gopinath and Sethuraman, 2023). These models are particularly effective in simulating event sequences over time and identifying patterns indicative of APTs (Stojanović et al., 2020).

5.5. Cloud Security Challenges

The increasing adoption of cloud computing has transformed data storage and access, but it has also introduced new security challenges (Alqahtani and Gull, 2018). APTs exploit vulnerabilities in cloud environments, making robust detection mechanisms essential. Despite the efficiency of cloud computing in data management, industries face cybersecurity risks due to the sensitive data stored on cloud servers (Aljumah and Ahanger, 2020). Machine learning techniques have shown promise in enhancing cloud security by detecting APTs through analysis of network traffic patterns, system log data, and application behavior (Jupalle et al., 2022).

5.6. Current Research and Challenges

Current research highlights the potential of machine learning in improving APT detection. However, challenges remain, including the high rate of false positives and false negatives associated with many detection models (Neuschmied et al., 2022). Advanced detection techniques require significant resources for training and processing, posing challenges for smaller organizations (Alshamrani et al., 2019). Additionally, integrating these models into existing security infrastructures and scaling them to manage large data volumes in dynamic cloud environments is complex and resource-intensive (Niedermaier et al., 2019).

6. Methodology

This study employs a quantitative research design focused on developing and evaluating a machine learning model for APT detection in cloud computing environments. The research process includes the following steps:

6.1. Data Collection: Gathering relevant datasets such as CTU-13 and UNSW-NB15.
6.2. Preprocessing: Public APT datasets will be preprocessed to ensure data quality and consistency.
6.3. Model Development: The primary model will utilize the Adaboost classifier, with baseline models including SVM, Random Forests, CNNs, and RNNs developed for comparison.
6.4. Evaluation: The model will be evaluated using real-world dataset simulations in a controlled cloud environment (Google Cloud) with a focus on real-time detection capabilities.

7. Conclusion

This research aims to advance the detection of Advanced Persistent Threats in cloud computing environments by leveraging machine learning techniques. The study's findings are expected to enhance cloud infrastructure security, safeguard sensitive information from sophisticated cyber threats, and provide a scalable solution for various industries. By improving real-time APT detection, this research contributes to a safer and more resilient digital environment, addressing critical challenges in cybersecurity.

References

1. AL-Aamri, A.S., Abdulghafor, R., Turaev, S., Al-Shaikhli, I., Zeki, A., Talib, S., 2023. Machine Learning for APT Detection. Sustainability 2023, Vol. 15, Page 13820 15, 13820. https://doi.org/10.3390/SU151813820
2. Alghamdi, A.A., Reger, G., 2020. Pattern Extraction for Behaviours of Multi-Stage Threats via Unsupervised Learning. 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment, Cyber SA 2020. https://doi.org/10.1109/CYBERSA49311.2020.9139697
3. Aljumah, A., Ahanger, T.A., 2020. Cyber security threats, challenges and defence mechanisms in cloud computing. IET Communications Research Article. https://doi.org/10.1049/iet-com.2019.0040
4. Alqahtani, A., Gull, H., 2018. Alqahtani: Cloud Computing and Security Issues-A... - Google Scholar [WWW Document]. URL https://scholar.google.com/scholar_lookup?title=Cloud%20computing%20and%20security%20issuesa%20review%20of%20amazon%20web%20services&publication_year=2018&author=A.%20Alqahtani&author=H.%20Gull (accessed 5.9.24).
5. Alshamrani, A., Myneni, S., Chowdhary, A., Huang, D., 2019. A Survey on Advanced Persistent Threats: Techniques, Solutions, Challenges, and Research Opportunities. IEEE Communications Surveys and Tutorials 21, 1851–1877. https://doi.org/10.1109/COMST.2019.2891891
6. Baksi, R.P., 2022. FRAMEWORK AND STRATEGIES TO MITIGATE ADVANCED PERSISTENT THREATS (APT). State University of New York at Buffalo ProQuest Dissertations Publishing.
7. Gopinath, M., Sethuraman, S.C., 2023. A comprehensive survey on deep learning based malware detection techniques. Comput Sci Rev 47, 100529. https://doi.org/10.1016/J.COSREV.2022.100529
8. Hassan, W.U., Bates, A., Marino, D., 2020. Tactical provenance analysis for endpoint detection and response systems. Proc IEEE Symp Secur Priv 2020-May, 1172–1189. https://doi.org/10.1109/SP40000.2020.00096
9. Hejase, H.J., Kazan, H., 2020. ADVANCED PERSISTENT THREATS (APT): AN AWARENESS REVIEW. Journal of Economics and Economic Education Research 21. https://doi.org/10.13140/RG.2.2.31300.65927
10. Joloudari, J.H., Haderbadi, M., Mashmool, A., Ghasemigol, M., Band, S.S., Mosavi, A., 2020. Early detection of the advanced persistent threat attack using performance analysis of deep learning. IEEE Access 8, 186125–186137. https://doi.org/10.1109/ACCESS.2020.3029202
11. Jupalle, H., Kouser, S., Bhatia, A.B., Alam, N., Nadikattu, R.R., Whig, P., 2022. Automation of human behaviors and its prediction using machine learning. Microsystem Technologies 28, 1879–1887. https://doi.org/10.1007/S00542-022-05326-4/TABLES/5
12. Momeni Milajerdi, S., Gjomemo, R., Eshete, B., Sekar, R., Venkatakrishnan, V.N., 2019. HOLMES: Real-time APT detection through correlation of suspicious information flows. Proc IEEE Symp Secur Priv 2019-May, 1137–1152. https://doi.org/10.1109/SP.2019.00026
13. Neuschmied, H., Winter, M., Stojanović, B., Hofer-Schmitz, K., Božić, J., Kleb, U., 2022. APT-Attack Detection Based on Multi-Stage Autoencoders. Applied Sciences 2022, Vol. 12, Page 6816 12, 6816. https://doi.org/10.3390/APP12136816
14. Niedermaier, M., Striegel, M., Sauer, F., Merli, D., Sigl, G., 2019. Analysis and design of security mechanisms in the context of Advanced Persistent Threats against critical infrastructures.
15. Sharma, A., Gupta, B.B., Singh, A.K., Saraswat, V.K., 2023. Advanced Persistent Threats (APT): evolution, anatomy, attribution and countermeasures. J Ambient Intell Humaniz Comput 14, 9355–9381. https://doi.org/10.1007/S12652-023-04603-Y/TABLES/3
16. Stojanović, B., Hofer-Schmitz, K., Kleb, U., 2020. APT datasets and attack modeling for automated detection methods: A review. Comput Secur 92, 101734. https://doi.org/10.1016/J.COSE.2020.101734
17. Xiong, C., Zhu, T., Dong, W., Ruan, L., Yang, R., Cheng, Y., Chen, Y., Cheng, S., Chen, X., 2022. Conan: A Practical Real-Time APT Detection System with High Accuracy and Efficiency. IEEE Trans Dependable Secure Comput 19, 551–565. https://doi.org/10.1109/TDSC.2020.2971484

Comparative Study of Machine Learning Models Using UNSW Datasets

Sunday Adeola AJAGBE[1*][0000-0002-7010-5540], Oluwaseyi Omotayo ALABI [2][0009-0005-0027-5930], Bolaji ADEKUNLE [3][0009-0005-8913-8587], Matthew O. ADIGUN [4][0000-0001-6256-5865]

[1,4]Department of Computer Science, University of Zululand, Kwadlangezwa, South Africa
[2]Department of Mechanical Engineering, Lead City University, Ibadan, Nigeria,
[2] Department of Mechanical Engineering, University of Ibadan, Nigeria.
[3]Department of Data Science, University of Salford, United Kingdom

230015266@stu.unizulu.ac.za*, oalabi4330@stu.ui.edu.ng, B.I.Adekunle@salford.edu.ac.uk, adigunm@unizulu.ac.za.
Corresponding email *

Abstract. The global digitization of most activities of corporate and non-corporate bodies has come to stay. There is a need for machine learning (ML) models to be used as a tool to help prevent unauthorized intrusion into various networks. The strengths and weaknesses of this ML model are concerned with intrusion detection (IDS) tasks. In this research, an experimental approach has been used to compare the strengths and weaknesses of the ML models such as linear SVC, LR, Random Forest (RF), Decision Tree (DT) and XGBoost on the UNSW NB15 datasets to detect intrusion. Data exploration, Feature engineering, selection and a test set of 15%, a validation set of 15%, and a training set of 70% respectively were used for data splitting. Performance evaluation was carried out using accuracy, recall, precision and F1-score. The outcome of the experiment shows a percentage of 92.71% (1, normal) and 7.29% (0, attack) for normal traffic and attack traffic respectively. Performance evaluation results showed that RF and XGBoost outperformed the other ML models. Hence, ML models can effectively be used to detect system attacks.

Keywords: Comparative analysis, Detection Experiment, Internet of Things, Contemporary Machine Learning, Deep learning.

1 Introduction

A branch of Internet of Things (IoT) called industrial IoT (IIoT) is centred on industrial assets and manufacturing process automation. IIoT is closing the gap between information technology and operational technology by integrating control and information systems with physical and business operations [1]. In the IoT, intrusion detection systems (IDSs), are made to identify and guard against malicious activity or unauthorized access to a network, system, or application. They are useful for identifying and stopping any harmful IoT activity. Additionally, they can assist in locating any breaches and attacks, enabling a company to take the necessary precautions before more harm is done. Large volumes of data can be analyzed by machine learning ML model to find trends that might point to attacks [2].

In order to identify the multi-class intrusion attacks in the IoT, the paper introduced an IDS using multiple ML Classifier techniques on the Message Queuing Telemetry Transport-IoT-IDS dataset. A comparison of the results of the result analysis has also been done in this research work to assess the effectiveness of all ML classifier techniques that have been employed. The tested IDS's overall accuracy was 96.76%, 97.38%, 99.98%, 99.98%, and 97.58% when employing the classifiers k-NN, SVM, NB, RF, DT, and Stochastic Gradient Descent, in that order. All the tested ML classifiers' precision, and F1-score were present but the dept of comparison was missing [3].

To identify network- and application-based threats, an intelligent intrusion detection framework (IDF) was established. The three stages of the suggested framework were feature selection, classification, and data pre-processing. For the feature selection stage, the Opposition-based Learning-Rat Inspired Optimizer (OBL-RIO) is intended. Rats' evolutionary tendency selects the important characteristics. The stability of the OBL-RIO characteristics is guaranteed by the fittest value. As the binary class classifier, a 2D-Array-based Convolutional Neural Network (2D-ACNN) was suggested. To operate on the complicated layers, a 2D-array model maintains the input features. It recognizes both typical and unusual traffic. The Netflow-based datasets were used to train and evaluate the suggested framework. 95.20% accuracy, 2.5% false positive rate, and 97.24% detection rate are produced using the suggested framework [4].

Data's security could be jeopardized at any time. The IDS, which are used to detect network intrusion, are used to handle safety. Different categorization methods are utilized to detect different sorts of assaults, hence improving the performance of IDS. Selecting the best one to build an IDS is a difficult undertaking [5]. The most effective approach is to evaluate the various classification algorithms' performance. However, as a measure of classification ability, the confusion matrix has drawn the attention of the majority of studies. As a result, this matrix is frequently used in publications to provide a thorough comparison with the dataset, data preprocessing, feature selection methods, classification, and performance assessment [6]. The goal of this paper is to present a comparison of the application of contemporary ML models used to build and improve IDSs in terms of accuracy, recall, precision and F1-score using the UNSW-NB15 Dataset. Therefore, the following is a summary of the study's primary contributions:

i. Compare the contemporary ML models viz-a-viz linear SVC, LR, Random Forest (RF), Decision Tree (DT) and XGBoost on intrusion detection using the UNSW NB15 dataset.
ii. Evaluating the ML models in terms of detection accuracy, recall, precision, and F1-score

The remaining sections of this study are arranged as follows: The IDSs employing ML-based model comparison and DL-based are covered in Part 2 of the "Review of related works". The "Methodology," which portrays the design and implementation setup for the ML-based IDSs comparative research is presented in Part 3. The “Experimental results and discussion” displays the implementation and results in Part 4. The "Conclusion" that wraps together the study and suggests future work is contained in Part 5.

2 Methodology

The necessary procedures and methodological stages to complete this research have been illustrated in this section. The research methodology is chronologically explained in this section, this includes research design, research procedure how data was acquired, preprocessed, modelling and tested which is similar to [7].

2.1 Acquired Dataset and Description

The raw network packets from the UNSW-NB15 data sets are set up in the cyber range lab at the Austrian Centre for Cyber Security to produce a combination of genuine current regular operations and simulated contemporary attack behavior using the IXIA Perfect-storm approach. One can archive 100 GB of raw traffic (such as Pcap files) with the tcpdump utility. In order to produce a total of 49 features with the class label, twelve algorithms are built and the Argus and Bro-IDS tools are employed. The four CSV files in the original dataset contain 2,540,044 packets [8]. Most researchers have utilized these datasets to assess each component of their created IDSs independently. The UNSW Dataset was downloaded from public repository called Kaggle.

2.2 Data Pre-processing

Mainly in this study, data pre-processing includes data cleaning, data normalization and data splitting as described thus:

2.2.1 Data cleaning: since the original dataset cannot be used to feed the ML models because it has a large number of missing values in several feature columns. During the data cleaning stage, the following features were removed from the dataset.

2.2.2 Data normalization: Data normalization was particularly useful for systems where measurements are frequently reported on widely disparate levels. More reliably constructed neural networks are made possible by min-max normalization. The advantage of this normalization technique is that it precisely preserves all data linkages, meaning it does not lead to bias. While min-max is added, the rising function remains below the proper value range for the classification, but the corresponding distributions of the associated features remain inside the current value range [9].

2.2.3 Data splitting: At this point, the primary dataset was divided into three groups: a test set (15%), a validation set (15%), and a training set (70%).

2.3 Modelling

We suggested comparing five different types of modern ML models for an IDS viz-a-viz linear SVC, LR, DT, RF, and XGB in this study. Therefore, the dataset was subjected to random sampling using an oversampling technique in order to improve its compatibility with the machine learning model.

2.4 Performance Evaluation Measures

This will assess and gauge the effectiveness of ML models for IDS. Additionally, the writers evaluated each model's performance using accuracy and loss criteria. Additionally, because a callback function is used to track validation loss at each training epoch, the training process will end if validation loss has not improved after twenty epochs. After the learning process is complete, the test sets were used to assess the ML model. Evaluating the ML models in terms of detection accuracy, recall, precision, and F1-score [9].

3 Results and Discussion

3.1 Experimental Results

In carrying out the Exploratory Data Analysis (EDA) on the UNSW-NB15 dataset, the outcome of the experiment shows a percentage of 92.71% (1, normal) and 7.29% (0, attack) for normal traffic and attack traffic respectively. Oversampling is the process of creating synthetic data that attempts to randomly generate a sample of the attributes from observations in the minority class, whereas Undersampling is the process of randomly deleting some of the observations from the majority class in order to match the numbers with the minority class. The findings of comparing the four distinct ML models' performances on the extracted and reduced dataset were given, as was covered in section 4.2. Each and every metric has a value between 0 and 1. The model is better the closer the measures' values are near 1.

Table 1: Train Set results of ML models in IDS task

ML models	Accuracy	Recall	Precision
Linear SVC	68.72	68.72	69.21
Logistic Regression	69.92	69.92	70.36
Random Forest	97.87	97.87	97.93
Decision Tree	83.68	83.68	83.96
XGBoost	92.14	92.14	92.30

4.2 Discussion

In addition to the experimental results presented in this paper, as earlier stated in the previous section of this paper, the accuracy of Linear SVC, LR, RF, DT and XGB were to be determined, Table 1 present the results of the train set and the test set respectively. The dataset was a balanced dataset since the difference between the classes (normal and attack) is not much when compared with the UNSW dataset which has 92% to 8%. The performance evaluation shows that RF has high performance followed by the XGBoost model.

4 Conclusion

The general consensus is that network IDS are an efficient and effective way to handle threats to network security. Additionally, ML models in well-established IDS are applied through an effective and efficient detection procedure to meet increasingly complex cyber security threats. An accurate and efficient system that identifies malicious attacks and helps in reducing security risks associated with IoT devices has become mandatory. In this study, it has been established that there are several unauthorized users who are intruding into individual and corporate websites and portals across the globe. Many researchers have applied the use of diverse approaches ranging from simple codes to the use of algorithms and models. This research work has performed comparative experiments using various ML models such as Linear SVC, LR, RF, DT and XGB on the UNSW NB15 datasets to detect intrusion. Performance evaluation results showed that RF and XGBoost outperformed the other ML models. Hence, ML models can effectively be used to detect system attacks. This research maybe expanded in the future and use the paradigm in a real-world dataset with further conclusions and justifications. In addition, deep learning models may be used to gain better insight into the dataset.

References

1. Vitorino, J., Andrade, R., Praça, I., Sousa, O., & Maia, E. (2022). A Comparative Analysis of Machine Learning Techniques for IoT Intrusion Detection. In E. Aïmeur, M. Laurent, R. Yaich, B. Dupont, & J. Garcia-Alfaro (Ed.), Foundations and Practice of Security. FPS 2021. 13291, pp. 191-207. Lecture Notes in Computer Science, Springer, Cham. doi:https://doi.org/10.1007/978-3-031-08147-7_13
2. Ajagbe, S. A., Florez, H., & Awotunde, J. B. (2022). AESRSA: A New Cryptography Key for Electronic Health Record Security. In H. Florez, & H. (. Gomez (Ed.), Communications in Computer and Information Science. 1643, pp. 244-258. Peru: Springer
3. Saran, N., & Kesswani, N. (2023). A comparative study of supervised Machine Learning classifiers for Intrusion Detection in Internet of Things. Procedia Computer Science, 218, 2049-2057
4. Nizamudeen, S. M. (2023). Intelligent intrusion detection framework for multi-clouds – IoT environment using swarm-based deep learning classifier. Journal of Cloud Computing, 12, 134. doi:https://doi.org/10.1186/s13677-023-00509-4
5. Gautam, R. K., & Doegar, E. A. (2018). An Ensemble Approach for Intrusion Detection System Using Machine Learning Algorithms. International Conference on Cloud Computing, Data Science & Engineering (Confluence), (pp. 14-15).
6. Tahri, R., Jarrar, A., Lasbahani, A., & Balouki, Y. (2022). A comparative study of Machine learning Algorithms on the UNSW-NB 15 Dataset. ITM Web Confesrence, 48, p. 03022. doi:https://doi.org/10.1051/itmconf/20224803002
7. Rbah, Y., M., M., Balboul, Y., Chetioui, K., Fattah, M., Mazer, S., . . . Bernoussi, B. (2023). A machine learning based intrusions detection for IoT botnet attacks . AIP Conf. Proc. , 2814, p. 030012. doi:https://doi.org/10.1063/5.0149102
8. Mohy-Eddine, M. G., Benkirane, S., & al., e. (2023). An Ensemble Learning Based Intrusion Detection Model for Industrial IoT Security. Big Data Mining and Analytics, 6(3), 273-287. doi:https://doi.org/10.26599/BDMA.2022.9020032
9. Awotunde, J. B., Chakraborty, C., & Adeniyi, A. E. (2021). Intrusion Detection in Industrial Internet of Things Network-Based on Deep Learning Model with Rule-Based Feature Selection. Wireless Communications and Mobile Computing, 2021, 1-17. doi:https://doi.org/10.1155/2021/7154587

Odometry by means of Stereoscopic Vision and Deep Neural Networks

Anne Elizabeth van der Merwe[1,2][0009−0001−9709−9312], Mathys Cornelius Du Plessis[1,3][0000−0002−9187−3960], and Dieter Vogts[1,4][0000−0002−2554−7518]

[1] Nelson Mandela University, Gqeberha, 6001, South Africa
[2] s219133360@mandela.ac.za
[3] mc.duplessis@mandela.ac.za
[4] dieter.vogts@mandela.ac.za

Abstract. Localisation and autonomous navigation are fundamental tasks in many fields requiring ego-motion estimation techniques to provide efficient and accurate estimations of an agent's translation and rotation within challenging environments. Although several techniques are currently in use, certain challenging scenarios require a more robust approach to solving the problem of ego-motion estimation while maintaining efficiency and accuracy. This study highlights the use of visual odometry and deep learning as a technique for accurate localisation.

Keywords: Deep Neural Networks · Ego-motion · Localisation · Stereoscopic Vision · Visual Odometry.

1 Introduction

1.1 Background

Localisation and motion estimation are important aspects to consider for accurate autonomous navigation in robotics [6]. Several localisation techniques currently exist, each with their advantages and disadvantages for estimating the location and position of a robot [1]. Odometry, defined as the process of estimating the change in position (translation and orientation) over time, is an alternative approach to localisation that is inexpensive and provides better accuracy than traditional techniques. Ego-motion is defined as the translation and orientation of an agent and visual odometry can be defined as the estimation of the ego-motion of an agent through the acquisition of consecutive images from a single or multiple camera setup [1].

Visual odometry has a wide range of applications within many different fields, such as robotics, the automation and agriculture. The navigation ability of visual odometry allows for autonomous navigation of vehicles and robots. This can be used in challenging environments that may not have access to traditional navigation systems such as GPS, or in low-cost scenarios that cannot make use of more advanced and costly methods [1]. A stereo camera setup makes use of a pair of cameras and therefore motion estimation can be done using the depth information calculated from the consecutive pairs of images [8].

Deep learning-based methods have shown great potential for being useful in visual odometry problems. This is due to the powerful capability these methods possess to learn effectively and provide a more robust approach to visual odometry, specifically concerning depth estimation. Learning-based methods currently being used do not meet efficiency requirements due to the unavailability of accurate large-scale data required for learning. In most cases, these methods are specific to each problem making the generalisation ability of these methods incapable of being successfully applied to multiple real-world applications [11].

1.2 Problem Statement

Visual odometry is a multifaceted problem that typically requires solutions that are computationally expensive and complex [5]. Poor lighting conditions and dynamic objects create challenging environments for current visual odometry models, which result in lower estimation accuracy and a decrease in performance [7, 9]. Existing solutions for visual odometry fail to meet the criteria for efficiency and accuracy due to the inability to find a general solution that can apply to many common scenarios in real-world applications [11]. Many visual odometry solutions follow a sensor fusion approach which combines visual odometry with other odometry techniques to improve accuracy, but this approach increases computational cost and complexity [8]. Traditional solutions to visual odometry are challenging to implement because of all the different processes involved in estimating translation and orientation from images [11].

1.3 Research Aim

The main research aim is to provide a method for performing stereo-visual odometry, specifically using a deep neural network. This will be achieved by feeding sequential image pairs into the deep neural network to produce information about translation and orientation. This process can be described by Eq. 1.

$$F(I_{L_{n-1}}, I_{R_{n-1}}, I_{L_n}, I_{R_n}) \to x, y, \phi \tag{1}$$

Where $I_{L_{n-1}}$, $I_{R_{n-1}}$, I_{L_n} and I_{R_n} are the left and right images at time $n-1$ and n, x and y represents the translation and ϕ represents the orientation.

1.4 Significance and Impact

Existing solutions to visual odometry include several steps and many processes which is time-consuming, computationally complex, and expensive. Designing a solution that makes use of deep learning methods to perform odometry can help mitigate the many challenges and limitations of traditional approaches. A convolutional neural network using sequential image pairs obtained from a stereo camera setup as input and providing information regarding translation and orientation as output, allows the visual odometry process to become significantly less complex. A deep stereo visual odometry solution would be beneficial research and crucial in many localisation applications and autonomous navigation problems.

2 Related Work

2.1 Localisation and Autonomous Navigation

The ability to perceive self-motion and the surrounding environment is a skill humans possess through sensory perception, and allows for understanding and navigation through three-dimensional space. Similarly, artificial agents can perform self-localisation and mapping using different sensors and techniques allowing for a certain level of autonomy [3]. Localisation is the ability to obtain the internal system states of robot motion such as locations and orientations. It is crucial for a robot to always maintain knowledge of its position and orientation. Self-localisation is the process in which an agent estimates its position and orientation over time [1].

The traditional localisation technique that is widely used is a Global Positioning System (GPS). GPS is a satellite-based system that provides accurate positioning on the surface of the earth. There are, however, many situations in which GPS cannot provide accurate positioning information [8]. GPS is effective for outdoor navigation and environments in which there is a clear view of the sky. Environments that are indoors, underground, underwater, or in space cannot use GPS accurately due to satellite signal blockage [1]. An alternate approach to using GPS for localisation is ego-motion estimation and odometry [6].

The main goal of ego-motion estimation is to assist in navigation and obstacle detection. In GPS-denied environments, ego-motion estimation techniques provide accurate solutions to tracking, navigation, and surveillance. [6].

2.2 Visual Odometry and Stereoscopic Vision

Visual Odometry is the process in which the ego-motion of an agent is estimated through the use of visual input. This input takes the form of a sequence of consecutive images obtained from a single camera or multiple cameras. The changes induced by the agent's motion can be examined in the sequence of images and ego-motion estimates can be done incrementally [10].

A stereo camera setup consists of two separate lenses, and depth information can be extracted from a single frame obtained from those lenses. Stereo cameras have a baseline distance which is defined as the distance between the two lenses. A larger baseline difference is required to accurately estimate the depth of objects further away [8]. Stereoscopic vision from two camera lenses follows the biological model of how humans estimate depth with their eyes. This is due to the distance between the eyes and the disparity that is produced between the two images obtained from each eye [2]. For a set of camera lenses a rectified image pair is obtained from the two lenses and corresponding pixels from each image are identified. The horizontal difference between the corresponding pixels is the disparity which can be used to calculate the depth of each pixel and a disparity map is produced for each image pair [12].

Visual odometry has more success in indoor environments than outdoor environments due to computational complexity, scale ambiguity, and low lighting conditions. Drifting issues are also present due to the incrementally accumulated errors that are introduced when motion estimation is done after each frame. Sensor fusion has been suggested as a possible solution for addressing some of the challenges mentioned. This includes combining visual odometry with other localisation techniques [8].

2.3 Deep Learning for Localisation

Machine learning is a subset of artificial intelligence which makes use of specifically designed algorithms that can recognise patterns and make predictions on the data they have been provided. Deep learning is a subset of machine learning that uses Artificial Neural Networks (ANN) to recognise patterns in large amounts of data. Unlike machine learning, deep learning models can automatically learn and extract information from data sets. This makes deep learning well suited for tasks such as image classification and recognition, speech recognition, language processing, and autonomous navigation [5].

An ANN is an algorithmic model of a biological neural system which is made up of neurons that are interconnected through a synapse connection. An artificial neuron is modelled after a biological neuron, which receives an input signal and transmits an output signal. The neuron can receive input signals and compute an output signal as a function of the weights of the input signals. [4]. An artificial neural network is a network of artificial neurons and can consist of several layers where each neuron in a layer is connected to other neurons in the next layer. Each neuron will calculate a net input signal as a weighted sum of all the input signals connected to it [4].

A Convolutional Neural Network (CNN) is used to extract features from given data. CNNs can identify patterns from complex data sets and are specifically useful for image recognition, object detection, and image segmentation. CNNs can be used for visual odometry, including depth estimation, feature extraction, feature matching, and pose estimation [11].

3 Experimental Design

The data used for training the network is generated using a simulated environment created in Unity, a widely used game development platform, which consists of various game objects that have different images on them. Two cameras are attached to a game object that follows a predetermined path within the environment capturing left-right image pairs at regular intervals as well as the corresponding ground truth data.

To prepare the data for the neural network, a preprocessing pipeline is proposed for the images and ground truth data. For each data point (x, y and ϕ)), the delta values are calculated by computing the difference between each sequential pair of data points. A data cleaning process is implemented to correct any discrepancies that may arise, normalisation is done and a rotation matrix is applied to transform the data from the global coordinates to the local coordinates.

4 Conclusion

This paper presents a study on visual odometry using stereoscopic images and deep neural networks as a technique for indoor localisation. It highlights the need for efficient and accurate navigation and localisation and mentions the challenges currently faced in achieving accurate indoor localisation.

Acknowledgments. We would like to express our gratitude to the Telkom Center of Excellence for providing funding for this research.

References

1. Aqel, M.O., Marhaban, M.H., Saripan, M.I., Ismail, N.B.: Review of visual odometry: types, approaches, challenges, and applications. SpringerPlus **5**, 1–26 (2016)
2. Bhatti, A.: Current Advancements in Stereo Vision. IntechOpen, Rijeka (Jul 2012). https://doi.org/10.5772/2611
3. Chen, C., Wang, B., Lu, C.X., Trigoni, N., Markham, A.: A survey on deep learning for localization and mapping: Towards the age of spatial machine intelligence (2020)
4. Engelbrecht, A.P.: Computational intelligence: an introduction. John Wiley & Sons (2007)
5. Golroudbari, A.A., Sabour, M.H.: Recent advancements in deep learning applications and methods for autonomous navigation–a comprehensive review. arXiv preprint arXiv:2302.11089 (2023)
6. Khan, N.H., Adnan, A.: Ego-motion estimation concepts, algorithms and challenges: an overview. Multimedia Tools and Applications **76**, 16581–16603 (2017)
7. Kuo, X.Y., Liu, C., Lin, K.C., Lee, C.Y.: Dynamic attention-based visual odometry. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 36–37 (2020)
8. Mohamed, S.A., Haghbayan, M.H., Westerlund, T., Heikkonen, J., Tenhunen, H., Plosila, J.: A survey on odometry for autonomous navigation systems. IEEE access **7**, 97466–97486 (2019)
9. Pillai, S., Leonard, J.J.: Towards visual ego-motion learning in robots. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 5533–5540. IEEE (2017)
10. Scaramuzza, D., Fraundorfer, F.: Visual odometry [tutorial]. IEEE robotics & automation magazine **18**(4), 80–92 (2011)
11. Wang, K., Ma, S., Chen, J., Ren, F., Lu, J.: Approaches, challenges, and applications for deep visual odometry: Toward complicated and emerging areas. IEEE Transactions on Cognitive and Developmental Systems **14**(1), 35–49 (2020)
12. Zhou, K., Meng, X., Cheng, B.: Review of stereo matching algorithms based on deep learning. Computational intelligence and neuroscience **2020** (2020)

Mental Illness Monitoring and Predictive Model using Smart Devices: Machine Learning Approach

Senota Frans Kgoete[1][0009-0009-1704-2630] and Michael Nthabiseng Moeti[2][0009-0009-1704-2630]

[1] Tshwane University of Technology, 109 Market Street, Polokwane Limpopo, South Africa
[2] Tshwane University of Technology, 109 Market Street, Polokwane Limpopo, South Africa
kgoetesf@tut.ac.za,Moetimn@tut.ac.za

Abstract The prevention and treatment of mental illness is a serious social issue. Early prediction and intervention have been serious challenge because of lack of technological devices that can detect the symptoms related to mental illness. The objective of this study is to use real time data acquired from smart devices to build a monitoring and predictive model that can contribute to the prevention of the on set of mental illness. The study intend to apply machine learning methods using passive and active data of an individual person gathered from smart devices. The monitoring and predictive model will show useful accuracy for classifying the risk of mental illness using real time data, suggesting the potential of predictive detection, and preventive intervention measures using smart devices.

Keywords: Mental illness, Mental health, Smartphone Sensors, Smart device, Machine Learning.

1 Introduction

Over the world, millions of individuals suffer from mental diseases. If left untreated, it leads to poor physical health, which increases the risk of unnecessary disability, such as heart attacks, strokes, and fatigue[1]. Furthermore, it interferes with cognitive function through memory processes and attention because it affects life and work. Among the common indications and symptoms of mental illnesses are personality mood swings, excessive anxieties, violent behavior, thinking or talking about suicide or harming oneself, inability to cope with problems and daily activities, strange and grandiose ideas, extreme early mental illness detection and appropriate intervention can significantly reduce the rate of undiagnosed individuals and promote help-seeking of mental health treatment[2].

The most used diagnostic tool for mental illness is called the diagnostic and statistical manual of disorder, which is the most widely used tool for diagnosing a mental illness. It's maintained and published by the American Psychiatric Association; the fifth is the latest (DSM-5). It provides a set of criteria for the classification and diagnosis of numerous disorders, such as daily mood, weight loss, insomnia, and other major depressive disorders. A major depressive disorder diagnosis may be made if a person meets five of the nine criteria within a two-week period (MDD).

DSM offers a common set of guidelines, yet diagnosis of mental illness remains difficult. Significant variations in symptom onset in mental illnesses can make it difficult to fit any standardized profile. For instance, when considering symptom combinations for MDD, there are 1,497 different possibilities [3]. A patient's symptoms may also alter with time. Moreover, severe mental illness tends to have several other comorbidities, such as panic disorder, substance abuse, and depression, which are often comorbid with schizophrenia, which complicate the subjective diagnosis of mental illness [4]. Multiple comorbidities can further complicate the subjective diagnosis of mental disorders. Zimmerman et al.16 reported that a significant portion of psychiatrists and non-psychiatrist physicians frequently do not diagnose patients using the DSM criteria.

The advancement to the Fourth Industrial Revolution (4IR) holds promising solutions for addressing mental illness challenges. Utilizing machine learning algorithms and smart devices, applications can be created to help with this attempt. With the help of these technologies, a multitude of data will be gathered, including social, psychological, and behavioral tracks that will point changes in a patient's overall health. For example, a decreased social contact and mobility may indicate a decline in mental health [5]. As a result, ongoing observation of these signals will offer a special insights into the course of illness. With the usage of these signals, the mental illness monitoring and predictive model will be able to pinpoint unique indicators of the beginning of an illness, which will be utilized to create specific treatment plan. This will greatly enhance therapeutic input and support individuals seeking therapy.

2 Literature Review

Variations in parameters such as movement, sleep duration, heart rate, electrocardiogram, skin temperature, etc., are often associated with psychiatric disorders [6]. Collecting and analyzing data from sensors embedded in the context of daily life has been widely employed for the monitoring of mental health. Aqajari, et al [7] developed a real-time monitoring system that utilize both smart watch and smart phone to collect physiological and contextual to monitor their physiological and activity level and further used to predict stress levels in everyday life using random algorithm. Fazel also used passive-sensing data to monitor mental health status of individuals to track the perceived stress to be used to obtain a better understanding of patient behaviour and improve personalized treatments.
Shukla, et al [8] proposed a personalized mental health tracking and mood prediction system that utilizes patient physiological data collected through personal health devices. Their system leverages a decentralized learning mechanism that combines transfer and federated machine learning concepts using smart contracts, allowing data to remain on users' devices and enabling effective tracking of mental health conditions for psychiatric treatment and management in a privacy-aware and accountable manner.
Zou, et.at [9]proposed a sequenced modeling approach that uses passive sensing data to predict treatment response of patients suffering major depressive disorder. Each was given a smart phone with a self-developed app that record passive data such as their daily phone usage and physical data .The system forecast the treatment success of depression or failure at the early stage to improve patient care .

Due to the inter related nature of mental illness, we propose a comprehensive interpersonal approach which includes behavioral, psychological ,social factors and psychosocial factors to detect the set-in of mental illness and tailor the interventions to individual’s needs.

3 Proposed Research Aim

The main aim of this research is to develop mental illness monitoring and predictive model using smart devices:

. The specific objectives of the research are to:

1. To explore the current technologies and determine the challenges of current mental health care support systems in the literature.
2. To explore the technologies that can be considered in developing a mental health support system.
3. To design and prototyped mental illness monitoring and predictive for mental illness monitoring support.
4. To implement the proposed prototyped of mental illness monitoring and predictive model for mental illness monitoring
5. To evaluate the performance of the proposed mental illness monitoring and predictive model for mental illness monitoring.

3.1 Methodology

To achieve the desired technical objectives, Design science research (DSR) will be adopted in conjunction with a case study approach in this study.

1. An in-depth investigation will be conducted to discover the existing challenges, determine the technologies that can be used, determine the mechanism to collect and extract data and identify suitable machine learning techniques with high predictive performance
2. Data Collection and Analysis - An ethical clearance will be requested from the institution's ethical clearance committee, and a consent letter will be issued to the participants stating the aim and also to make them aware that their personal information will not be disclosed. The primary data collection method that will be deployed is a survey. A questionnaire survey will be administered to university students to gather psychosocial experiences and environmental data. A psychologist will be identified, together with patients who are willing to take part in the study then (i). their psychosocial data such as trauma, childhood abuse, stigma, social relating, mental thinking, spiritual well-being, emotional feeling) will be gathered from a patient, (ii) physiological data (physical health metrics such as heart rate, steps taken, sleep patterns, and calories burned), gender,

nutrition information) will be recorded, and (iii) environmental data factors (climate, experience of racism, place of work/live) will be collected via a mobile phone.

3. Modeling: The gathered data about behavioral, physiological, psychosocial, and social data will be utilized in Machine learning algorithms such as Naive Bayes, random forest, decision tree, K-Nearest Neighbor, artificial neural networks/deep learning, and compared based on their performance and accuracy, to predict the set-in of mental illness.

4. Experimentation by simulation and usability testing: A smartphone application for Android will be created to gather information, transfer it to the cloud, analyze it, and look for trends or connections that might point to a patient's mental health condition.

5. The user-cantered evaluation usability testing method will be adopted. It will be used to assess the usability and user experience of the application, interface, and the application itself. The Usability testing will be unmoderated and remote.

4 Conclusion

Our goal in this ongoing research is to create a model that uses sensor data from smartphones to identify underlying behaviors. Using a combination of environmental stressors, social, psychological, behavioral ,traumatic experiences, and genetic predisposition, our model will uses machine learning techniques to monitor and predict when mental illness will be apparent. This innovative method will reduces the requirement for in-person sessions with mental health specialists by enabling people to remotely access customized treatment programs and proactively monitor their mental health status. The suggested approach shows potential in resolving a number of issues, including financial effects and stigma associated with the illness.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Organization, W.H., *World mental health report: Transforming mental health for all.* 2022.
2. Fowler, D.G., et al., *Early detection and psychosocial intervention for young people who are at risk of developing long term socially disabling severe mental illness: should we give equal priority to functional recovery and complex emotional dysfunction as to psychotic symptoms?* Clinical Neuropsychiatry, 2010. **7**(2): p. 63-72.
3. Fava, M. and K.S. Kendler, *Major depressive disorder.* Neuron, 2000. **28**(2): p. 335-341.

4. Launders, N., et al., *The impact of comorbid severe mental illness and common chronic physical health conditions on hospitalisation: A systematic review and meta-analysis.* PloS one, 2022. **17**(8): p. e0272498.
5. Fazeli, S., et al. *Passive monitoring of physiological precursors of stress leveraging smartwatch data*. in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2022. IEEE.
6. Mohr, D.C., K. Shilton, and M. Hotopf, *Digital phenotyping, behavioral sensing, or personal sensing: names and transparency in the digital age.* NPJ digital medicine, 2020. **3**(1): p. 45.
7. Aqajari, S.A.H., et al., *Context-aware stress monitoring using wearable and mobile technologies in everyday settings.* medRxiv, 2023: p. 2023.04. 20.23288181.
8. Shukla, M. and O. Seneviratne. *MentalHealthAI: Utilizing Personal Health Device Data to Optimize Psychiatry Treatment*. in *AMIA Annual Symposium Proceedings*. 2023. American Medical Informatics Association.
9. Zou, B., et al., *Sequence modeling of passive sensing data for treatment response predictpredictprediion in major depressive disorder.* IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023. **31**: p. 1786-1795.

Development of Security Threats Identification Model for Internet of Things (IoT) Smart Office Devices

Mdzobo Malwande [0000-001-9877-1644] and Jere Nobert [0000-0001-8966-2753]
Alice Campus, 1 King Williams Town Road Private Bag X1314 Alice 5700. Tel: +27 (0)40 602 2441

Abstract: The proliferation of Internet of Things (IoT) devices in smart office environments has revolutionized workplace efficiency and connectivity. However, this increased interconnectivity also brings significant security challenges, as IoT devices often become targets for cyber-attacks due to their diverse and often insufficiently secured nature. This paper presents the development of a comprehensive Security Threats Identification Model specifically tailored for IoT smart office devices. The proposed model leverages a multi-layered approach, incorporating machine learning algorithms, real-time monitoring, and anomaly detection techniques to identify and mitigate potential security threats. By analyzing data patterns and device behavior's, the model can detect unauthorized access, data breaches, and other malicious activities in real-time. The effectiveness of the model is evaluated through a series of simulations and real-world testing scenarios, demonstrating its capability to enhance the security posture of smart office environments. This work aims to provide a robust framework for securing IoT devices, thereby safeguarding sensitive corporate data and ensuring the integrity and reliability of smart office systems.

Keywords: IoT, smart office, security threats, machine learning, anomaly detection, real-time monitoring.

1 INTRODUCTION

The integration of IoT devices in smart office environments has become increasingly prevalent, driven by the need for enhanced efficiency, productivity, and connectivity Ahanger & Aljumah, (2019). These devices, ranging from smart thermostats to security cameras and voice assistants, offer unprecedented convenience and operational benefits Lonzetta et al., (2018). However, the very features that make IoT devices attractive also introduce significant security vulnerabilities. The heterogeneity of devices, coupled with often inadequate security measures, creates an expanded attack surface for cybercriminals Samaila et al., (2018). The security landscape of IoT devices is further complicated by the diverse range of manufacturers and varying levels of firmware security, leaving many devices susceptible to unauthorized access and exploitation Sontan & Samuel, (2024).

2. The primary challenge lies in the lack of comprehensive security frameworks tailored to the unique characteristics of IoT devices in smart offices Lonzetta et al., (2018). Traditional security solutions often fall short in addressing the dynamic and interconnected nature of IoT ecosystems. This research aims to fill this gap by developing a Security Threats Identification Model that leverages advanced machine learning algorithms and real-time monitoring to proactively detect and mitigate security threats Fraga-Lamas et al., (2016).

2. Aim and Objectives

The aim of this research is to develop an AI-driven system capable of detecting and mitigating security threats in university smart office environments. To achieve this aim, the following objectives will be pursued:

Objectives

- To conduct a comprehensive review of existing literature on IoT device security and threat identification models.
- To identify the unique security challenges posed by IoT devices in smart office environments.
- To propose a set of security threat identification models tailored to IoT smart office devices.
- To evaluate the effectiveness and feasibility of the proposed models through empirical testing and validation.

Research Questions

What are the key security threats facing IoT devices in smart office environments?

Sub-Questions

- How do existing security threat identification models address the specific vulnerabilities of IoT smart office devices?
- What are the essential components of effective security threat identification models tailored to IoT smart office devices?
- How can the proposed models be validated and implemented in real-world smart office settings?
- How can the effectiveness and feasibility be evaluated of the proposed models through empirical testing and validation.

3.3 Related Literature

The literature review will serve as a foundational component of the proposed master's thesis, focusing on IoT device security and threat identification models within smart office environments Ali & Awad, (2018). The review will explore existing research, academic papers, and industry reports related to IoT security, with a specific emphasis on its application in smart office settings. Key areas of interest will include:

IoT Security Challenges and Vulnerabilities.

This section will explore how these principles are applied to ensure the integrity, confidentiality, and availability of data transmitted and processed by IoT devices in smart office environments. Authentication ensures that only authorized users or devices can access the network or specific resources Maxsud Tulqin, (2023). The security of IoT devices has been a major concern, with numerous studies highlighting the vulnerabilities inherent in these devices Chen & Esmaeilzadeh, (2024). The study by van Daalen, (2023) addresses common issues including weak authentication mechanisms, lack of encryption, and poor device management practices. The finding was that encryption protects the confidentiality of data transmitted and processed by IoT devices, preventing unauthorized access. Research by Wijesundara et al., (2024). emphasizes the need for stronger security protocols and the implementation of device-specific security measures.

Existing Security Models

Several security models have been proposed to address IoT vulnerabilities. These include network segmentation, which isolates IoT devices from critical network resources, and the use of encryption protocols to protect data in transit Chen & Esmaeilzadeh, (2024). However, these models often lack the capability to adapt to the evolving threat landscape, necessitating more dynamic and intelligent solutions Fadhel et al., (2024).

Machine Learning and Anomaly Detection

Machine learning and anomaly detection techniques have shown promise in enhancing IoT security. Studies by Shipena, E.F.N. and Gamundani, (2024) demonstrate the effectiveness of machine learning algorithms in identifying abnormal device behaviors and potential security threats. These approaches rely on analyzing large datasets to establish baseline behaviors and detect deviations indicative of malicious activity.

Real-Time Monitoring and Response

Real-time monitoring is critical for timely threat detection and mitigation. Research by M. M. Hossain et al., (2015) highlights the importance of continuous monitoring and rapid response mechanisms in maintaining the security of IoT environments. Implementing such measures can significantly reduce the window of opportunity for attackers and minimize the impact of security breaches.

4. Methodology

The proposed Security Threats Identification Model will be developed using a multi-layered approach that integrates machine learning algorithms, real-time monitoring, and anomaly detection Herzalla et al., (2023). The methodology consists of the following steps:

Data Collection: Gathering comprehensive datasets on network traffic patterns, device behaviors, and historical security incidents from smart office environments Hadyaoui & Cheniti-Belcadhi, (2023).

Model Development: Utilizing machine learning techniques to analyze the collected data and identify patterns indicative of security threats. This includes training algorithms on labelled datasets to distinguish between normal and anomalous behaviors Raja Sindiramutty, (2023).

Real-Time Monitoring: Implementing real-time monitoring capabilities to continuously track device activities and network traffic Telo, (2023)

5. Conclusion

The development of a Security Threats Identification Model for IoT smart office devices represents a significant advancement in addressing the security challenges posed by these environments. By leveraging machine learning and real-time monitoring, the proposed model can detect and mitigate security threats in real-time, thereby enhancing the overall security posture of smart offices. The findings from this research will provide valuable insights into the development of robust security frameworks for IoT devices, contributing to the safeguarding of sensitive corporate data and ensuring the integrity and reliability of smart office systems

6. References

1. Abuzainab, N., Isler, V., Yener, A., Erpek, T., Davaslioglu, K., Sagduyu, Y. E., Shi, Y., MacKey, S. J., Patel, M., Panettieri, F., & Qureshi, M. A. (2019). QoS and Jamming-Aware Wireless Networking Using Deep Reinforcement Learning. Proceedings - IEEE Military Communications Conference MILCOM, 2019-Novem, 610–615. https://doi.org/10.1109/MILCOM47813.2019.9020985
2. Ahanger, T. A., & Aljumah, A. (2019). Internet of things: A comprehensive study of security issues and defense mechanisms. IEEE Access, 7, 11020–11028. https://doi.org/10.1109/ACCESS.2018.2876939
3. Ali, B., & Awad, A. I. (2018). Cyber and physical security vulnerability assessment for IoT-based smart homes. Sensors (Switzerland), 18(3), 1–17. https://doi.org/10.3390/s18030817
4. Chen, Y., & Esmaeilzadeh, P. (2024). Generative AI in Medical Practice: In-Depth Exploration of Privacy and Security Challenges. Journal of Medical Internet Research, 26(1). https://doi.org/10.2196/53008
5. Fadhel, M. A., Duhaim, A. M., Saihood, A., Sewify, A., Al-Hamadani, M. N. A., Albahri, A. S., Alzubaidi, L., Gupta, A., Mirjalili, S., & Gu, Y. (2024). Comprehensive systematic review of information fusion methods in smart cities and urban environments. Information Fusion, 107(February), 102317. https://doi.org/10.1016/j.inffus.2024.102317

6. Fraga-Lamas, P., Fernández-Caramés, T. M., Suárez-Albela, M., Castedo, L., & González-López, M. (2016). A Review on Internet of Things for Defense and Public Safety. Sensors (Basel, Switzerland), 16(10), 1–44. https://doi.org/10.3390/s16101644
7. Hadyaoui, A., & Cheniti-Belcadhi, L. (2023). Ontology-based group assessment analytics framework for performances prediction in project-based collaborative learning. Smart Learning Environments, 10(1). https://doi.org/10.1186/s40561-023-00262-w
8. Herzalla, D., Lunardi, W. T., & Andreoni, M. (2023). TII-SSRC-23 Dataset: Typological Exploration of Diverse Traffic Patterns for Intrusion Detection. IEEE Access, 11(September), 118577–118594. https://doi.org/10.1109/ACCESS.2023.3319213
9. Hossain, M. M., Fotouhi, M., & Hasan, R. (2015). Towards an Analysis of Security Issues, Challenges, and Open Problems on the Internet of Things. 2015 IEEE World Congress on Services, 21–28. https://doi.org/10.1109/SERVICES.2015.12
10. Lonzetta, A. M., Cope, P., Campbell, J., Mohd, B. J., & Hayajneh, T. (2018). Security vulnerabilities in Bluetooth technology as used in IoT. Journal of Sensor and Actuator Networks, 7(3), 1–26. https://doi.org/10.3390/jsan7030028
11. Maxsud Tulqin, U. (2023). Identification and Autentification.
12. Raja Sindiramutty, S. (2023). Autonomous Threat Hunting: A Future Paradigm for AI-Driven Threat Intelligence.
13. Samaila, M. G., Neto, M., Fernandes, D. A. B., Freire, M. M., & Inácio, P. R. M. (2018). Challenges of securing Internet of Things devices: A survey. Security and Privacy, 1(2), 1–32. https://doi.org/10.1002/spy2.20
14. Shipena, E.F.N. and Gamundani, A. M. (2024). A Review of Internet of Things (IoT) Smart Office Data Security Threats (pp. 375–381). https://doi.org/https://doi.org/10.1108/
15. Sontan, A. D., & Samuel, S. V. (2024). The intersection of Artificial Intelligence and cybersecurity : Challenges and opportunities The intersection of Artificial Intelligence and cybersecurity: Challenges and opportunities. February. https://doi.org/10.30574/wjarr.2024.21.2.0607
16. Telo, J. (2023). Smart City Security Threats and Countermeasures in the Context of Emerging Technologies. International Journal of Intelligent Automation and Computing, 6, 31–45.
17. van Daalen, O. L. (2023). The right to encryption: Privacy as preventing unlawful access. Computer Law and Security Review, 49, 105804. https://doi.org/10.1016/j.clsr.2023.105804
18. Wijesundara, W. M. A. B., Lee, J. S., Tith, D., Aloupogianni, E., Suzuki, H., & Obi, T. (2024). Security-enhanced firmware management scheme for smart home IoT devices using distributed ledger technologies. International Journal of Information Security. https://doi.org/10.1007/s10207-024-00827-x

An Enterprise Architecture (EA) implementation framework to enhance business-IT alignment in Public Health Organisations in South Africa.

Naleli Ntšonyana[1], Monelo Nxozi[2], Marlien Herselman[3]

Rhodes University, Makhanda, 6139, South Africa
nalelintsonyana@gmail.com, m.nxozi@ru.ac.za, mherselman@csir.co.za

Abstract. The public health sector is transforming by using digital and health technology to increase the effectiveness of the organisations' capabilities to deliver public services. To guide this transformation and improve alignment between business and Information Technology (IT) strategies, enterprise architecture (EA), is increasingly becoming a significant consideration. This study, therefore, aims to develop an EA implementation framework that a) addresses the GWEA implementation challenges that have hindered the progression of proper implementation and application of the framework, b) provides an alternative EA framework that insists on using both business and IT personnel to align the organisation strategies. Finally, c) underscores that business and IT alignment should be continuous across the organisations' strategic, tactical, and operational levels. The development of the framework will also contribute to public health organisations rendering improved services. As such, in creating the framework, the study seeks to answer the main research question: *What constitutes the components of an enterprise architecture implementation framework for aligning business and IT strategies in public health organisations?* The study adopts a social science research methodology framework, which is the research onion that follows a sequential research process. This helps justify each element in the research design and bases the justifications on the research questions and objectives. A scoping review is applied to identify gaps in the evidence base where little to no research has been carried out on the use of EA in aligning business and IT strategies in public health organisations in South Africa; hence, it is used to provide a broad overview of this topic. It follows an interpretive qualitative approach, and its ultimate deliverable is an enterprise architecture implementation framework that enhances business-IT alignment in public health organisations in South Africa.

Keywords: Enterprise Architecture, Strategic alignment models, Government Wide Enterprise Architecture, Business IT alignment.

1 Introduction

EA is used to align business and IT strategies and to use IT to support business objectives throughout an organisation [1]. It is an organisational approach that aims to coordinate, integrate, and align multidisciplinary business and IT initiatives within organisations [2], [3]. It also aligns the strategies by organising the structural context for aligning technology and business planning with strategic planning as its primary driver [4]. It is, therefore, considered a practical approach to addressing business IT alignment issues due to its multiple viewpoints and artefacts [1].

In the healthcare sector, organisations have also increasingly embraced EA to promote digital transformation and alignment between business and IT strategies [4], [5], [6]. It is important to note that in developing this implementation framework, the strategic alignment model (SAM) is also used as a seminal model to align business and IT strategies in public health organisations in South Africa. The model, usually cited for its inherent argument in highlighting the significance of alignment, illustrates the degree to which IT activities and capabilities support an organisation's strategy and the shared responsibility between business executives across all levels of the organisation [7].

The adoption of EA frameworks and models has noticeably increased in healthcare because the absence of frameworks is believed to signify the lack of formal procedures and, consequently, any official structures in organisations [8] . Within the healthcare sector, EA is primarily used as a reference during the implementation of IT, guiding the execution process and serving as a concluding evaluation tool that highlights procedures and focuses on operational procedures and activities [9]. It also provides resources in models and roadmaps and introduces specific directions and constraints to undertakings in organisations [2].

In the South African public sector, the GWEA framework was instituted and designed to establish minimum standards for using an EA approach to develop and construct national and departmental plans and blueprints for the government of South Africa [10] . However, the implementation of the framework has been unsuccessful due to a lack of proper implementation, resulting in an ineffective architecture implementation process [11]. The advantage of the framework is that it is generic and applicable to all public and private entities that engage in EA planning programs across all levels of government [11]. Despite this, there is a lack of sufficient guidelines provided by the framework for government implementation [10].

1.1 Background

Due to the size and complexity of enterprise information systems, organisations need help achieving alignment between business and IT [12]. In public health organisations, a gap exists in implementing frameworks that facilitate IT alignment to improve service delivery and realise good governance [13], [14]. It has been noted that social aspects such as organisational structure, lack of effective communication and planning, scarce

EA skills, user-limited knowledge and resistance to innovation persisted in the public sector and adversely impacted EA implementation [11], [15].

In South Africa, as in many developing countries, the implementation of EA has been slower than initially anticipated [16]. Despite the potential benefits that EA can offer to public organisations, these benefits have yet to be fully realised due to the formative nature of the field of EA, which requires further research [10], [16]. The lack of an EA framework in public organisations can lead to negative consequences such as poor system implementation and raise questions about the viability of these organisations to provide services [10]. Therefore, the implementation of GWEA is deemed low despite the substantial benefits to public sector organisations, as is evident by the scarce literature that holistically integrates business and IT alignment in the public health sector through enterprise architecture [17], [18].

Implementing EA in healthcare services necessitates a proactive approach, and the TOGAF framework offers a structured process for implementing holistic and coordinated improvements to service delivery [3], [19]. It is preferred because it provides a detailed process and focuses on operational activities [9], [20], [21].

It is essential to recognise and articulate the assumed relationships among the key factors or constructs, and the rationale for these assumptions may be derived from established theoretical or empirical work found in the literature [22]. The underlying presumptions that inform the development of the conceptual framework are grounded in institutional theory and are designed to situate organisations within a social context that comprises norms, culture, and values [23].

2 Proposed EA Implementation Framework

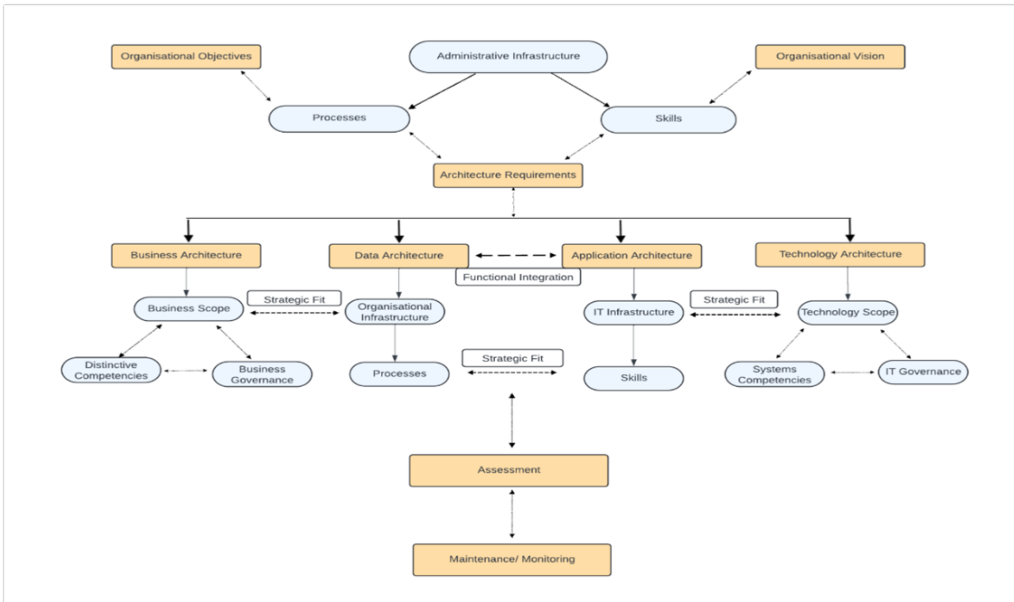

Fig. 1. The Proposed EA Implementation Framework Adapted from [24], [25], [26]

2.1 EA implementation framework constructs

The TOGAF is the chosen EA framework to implement the proposed conceptual framework. It can support the strategic planning process and optimise the use of IT to achieve business objectives while enhancing the quality of services in organisations and optimising the use of existing IT platforms [20]. It uses the same mechanistic step-by-step logic as all architecture planning methodologies [27]. The implementation of TOGAF follows a five-step process with comprehensive and necessary procedures to establish the methodology and guide the development of a proficient and focused EA framework [19], illustrated in Fig 1 above.

Stage 1, which is the initial phase/problem identification stage, is where organisational objectives and vision are set, and this is done by business and IT personnel and their varied operational objectives and goals from their departments. The planning stage is stage 2; here, the data gathering done in stage 1 is validated, where the established objectives and vision are mapped out to create architecture requirements. These are the guiding principles of implementation. Stage 3 is the execution stage and is the iterative primary process of the framework where the strategies formulated are put into practice. There is a strategic fit between business and data architectures and between the application and technology architectures.

The strategic fit shows interdependencies between the interrelationships that exist between internal and external domains [25]. The functional integration between the data and application layers indicates that the integration between business and IT domains is a functional domain of the organisation [25], [28] . Stage 4, the assessment stage, evaluates the impact of EA implementation by comparing the architectural requirements with the identified governing IT and business principles executed in Stage 3. The final stage, the monitoring and maintenance stage, reinforces the idea that business-IT alignment is a continuous process and measures the efficiency and status of EA implementation.

The relationships between all the stages are bidirectional, enabling organisations to make the necessary changes at any stage of the implementation process. This further elaborates that business-IT alignment is a continuous process that allows business and IT elements to evolve in the same direction [29] .

2.2 EA implementation across organisational levels

EA at a strategic level helps senior managers make informed decisions, improves governance, and ultimately enhances organisational integration and coherency among lower levels of organisations [16]. At a tactical level, EA emphasises managerial autonomy and control, which signifies that managerial control aims to establish and oversee specific objectives in collaboration with operational managers [30]. Organisations' operational function ensures adaptability within technical and business interoperability [30]. At this level, EA alignment entails integrating various layers through underlying model concepts, primarily focused on EA's business processes and application layers [31].

Disclosure of Interests. The authors have no competing interests to declare relevant to this article's content.

References

1. M. Zhang, H. Chen, and A. Luo, “A Systematic Review of Business-IT Alignment Research with Enterprise Architecture,” *IEEE Access*, vol. 6, no. April, pp. 18933–18944, 2018.

2. A. K. Ajera, E. Hustada, and P. Vassilakopouloua, “Enterprise architecture in hospitals: Resolving incongruence issues,” in *In Health and Wellbeing e-Networks for All-Proceedings of the 17th World Congress on Medical and Health Informatics*, IOS Press, Aug. 2019, pp. 659–663.

3. A. K. S. Ajer, E. Hustad, and P. Vassilakopoulou, “Enterprise architecture operationalization and institutional pluralism: The case of the Norwegian Hospital sector,” *Information Systems Journal*, vol. 31, no. 4, pp. 610–645, Jul. 2021.

4. Y. Masuda, A. Zimmermann, M. Viswanathan, M. Bass, O. Nakamura, and S. Yamamoto, “Adaptive enterprise architecture for the digital healthcare industry: A digital platform for drug development,” *Information (Switzerland)*, vol. 12, no. 2, pp. 1–26, 2021.

5. A. K. Ajer, E. Hustad, P. Vassilaskopoulou, and D. H. Olsen, “Strengthening the use of Enterprise Architecture: An institutional work perspective,” 2021. [Online]. Available: https://aisel.aisnet.org/ecis2021_rp/130

6. J. Jonnagaddala, G. N. Guo, S. Batongbacal, A. Marcelo, and S. T. Liaw, “Adoption of enterprise architecture for healthcare in AeHIN member countries,” *BMJ Health Care Inform*, vol. 27, no. 1, pp. 1–8, 2020.

7. B. Fritscher and Y. Pigneur, “A Visual Approach to Business IT Alignment between Business Model and Enterprise Architecture,” *In Sustainable Business: Concepts, Methodologies, Tools, and Applications*, pp. 543–566, 2020.

8. A. Haghighathoseini, H. Bobarshad, F. Saghafi, M. S. Rezaei, and N. Bagherzadeh, “Hospital enterprise Architecture Framework (Study of Iranian University Hospital Organization),” *Int J Med Inform*, vol. 114, pp. 88–100, Jun. 2018.

9. M. Rozario, A. A. Zainuddin, and P. Manokaran, “Analysis of a survey on the role of enterprise architecture in the field of healthcare technology by knowledge-based research,” *International Journal of Computing and Digital Systems*, vol. 13, no. 1, pp. 547–557, 2023.

10. R. Vidmar, E. Kotzé, D. Merwe, and J. Mentz, “An Approach to Design and Evaluation of e-Government Theory for South Africa,” in *Proceedings Annual Workshop of the AIS Special Interest Group for ICT in Global Development*, Dec. 2019.

11. M. Makovhololo and N. Ruxwana, “Use of DOI to establish the effects of GWEA implementation: A case of South African Government,” *International Journal of*

Computer Science and Information Security (IJCSIS), vol. 15, no. 9, pp. 86–92, 2017.

12. D. Dumitriu and M. A. M. Popescu, "Enterprise architecture framework design in IT management," *Procedia Manuf*, vol. 46, pp. 932–940, 2020.

13. M. Makovhololo, P. Makovhololo, and T. Sekgwelelo, "The significance of Enterprise Architecture in driving Digital Transformation on Public sectors," *International Journal of Applied Mathematics Electronics and Computers*, vol. 9, no. 3, pp. 35–42, 2021.

14. Y. P. Putra and A. Hadiana, "Designing Enterprise Architecture for Public Health Center Based on TOGAF Architecture Development Method," *IOP Conf Ser Mater Sci Eng*, 879, no. 1, 2020.

15. D. Duong Dang, S. Pekkola, and D. Duong, "Root Causes of Enterprise Architecture Problems in the Public Sector," in *The Pacific-Asia Conference on Information Systems*, Chiayi, Jun. 2016.

16. S. Patel, J.-P. Van Belle, and M. Turnip, "Factors Influencing the Adoption of Enterprise Architecture in the South African Public Sector," *Information Technology and Systems: ICITS 2021*, vol. 1, pp. 590–602, 2021.

17. N. A. Ahmad, S. M. Drus, and H. Kasim, "Factors That Influence the Adoption of Enterprise Architecture by Public Sector Organizations: An Empirical Study," *IEEE Access*, vol. 8, pp. 98847–98873, 2020.

18. P. Bhattacharya, "Aligning enterprise systems capabilities with business strategy: An extension of the Strategic Alignment Model (SAM) using Enterprise Architecture," *Procedia Comput Sci*, vol. 138, pp. 655–662, 2018.

19. M. N. Alwi, D. Hindarto, A. Marina, and D. Yudhakusuma, "Efficiency and Effectiveness: Enterprise Architecture Strategies for Healthcare Service," *International Journal Software Engineering and Computer Science (IJSECS)*, vol. 3, no. 3, pp. 386–397, Dec. 2023.

20. A. S. Girsang and A. Abimanyu, "Development of an enterprise architecture for healthcare using togaf adm," *Emerging Science Journal*, vol. 5, no. 3, pp. 305–321, 2021.

21. D. Tarenskeen, R. Bakker, and S. Joosten, "Applying the v Model and Axiomatic Design in the Domain of IT Architecture Practice," in *Procedia CIRP*, Elsevier B.V., 2015, pp. 263–268.

22. G. van der Waldt, "Constructing conceptual frameworks in social science research," *The Journal for Transdisciplinary Research in Southern Africa*, vol. 16, no. 1, pp. 1–9, Feb. 2020.

23. M. Sony and P. S. Aithal, "A Resource-Based View and Institutional Theory-based analysis of Industry 4.0 Implementation in the Indian Engineering Industry," *International Journal of Management, Technology, and Social Sciences (IJMTS)*, vol. 5, no. 2, pp. 2581–6012, 2020.

24. F. Gonzalez-Lopez and G. Bustos, "Integration of Business Process Architectures within Enterprise Architecture Approaches: A Literature Review," *EMJ - Engineering Management Journal*, vol. 31, no. 2, pp. 127–140, Apr. 2019.

25. J. Henderson and N. Venkatraman, "Strategic Alignment: A model for organizational Transformation via Information Technology," Cambridge, MA, 1990.

26. J. C. Henderson and N. Venkatraman, "Strategic alignment; Leveraging Information Technology for transforming organizations," *IBM Systems Journal*, vol. 38, no. 2.3, pp. 472–484, 1999.

27. S. Kotusev, "A Comparison of the Top Four Enterprise Architecture Frameworks," in *British Computer Society (BCS)*, 2021, pp. 1–10.

28. D. Hindarto, "Application Of Customer Service Enterprise Architecture In The Transportation Industry," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 5, no. 2, pp. 682–692, Aug. 2023.

29. M. Zhang, H. Chen, Y. Mao, and A. Luo, "An approach to measuring business-IT alignment maturity via DoDAF2.0," *Journal of Systems Engineering and Electronics*, vol. 31, no. 1, pp. 95–108, Feb. 2020.

30. R. Bemthuis, M. E. Iacob, and P. Havinga, "A design of the resilient enterprise: A reference architecture for emergent behaviors control," *A reference architecture for emergent behaviors control.*, vol. 20, no. 22, pp. 1–39, Nov. 2020.

31. P. Andre, D. Tamzalit, A. Benjilany, H. Bruneliere, and P. André, "A Review of Core Operational Business-IT Alignment A Review of Core Operational Business-IT Alignment A Review of Core Operational Business-IT Alignment," *In 31st International Conference on Information Systems Development (ISD 2023)*, 2023.

Financial Transactions Monitoring Using an AI Predictive Model in the Malawi Banking Sector

Mwayi Mwalwimba[1,2], Prof Dieter Vogts[2,3] and Prof Janet Wesson[3,4]
[1]Nelson Mandela University, PO Box 77000, Port Elizabeth, South Africa
[2] s228371961@mandela
[3] dieter.vogts@mandela.ac.za
[4] Janet.wesson@mandela.ac.za

Abstract. Financial Transaction monitoring is a virtual process within financial institutions aimed at scrutinising vast amounts of transaction data to detect fraudulent activities. The process involves analysis of transaction data to identify suspicious patterns and trends that may indicate financial crimes, fraud, money laundering or other illicit activities for further investigation to ensure the financial system's integrity. Financial institutions, including banks, are using traditional rule-based transaction monitoring processes. These processes face challenges of the volume of transactions, the complexity of financial transactions, and continuously evolving methods fraudsters use to conduct fraudulent activities through financial institutions. These challenges result in many false positives, where the process cannot distinguish between suspicious and legitimate transactions. This research project aims to use Artificial Intelligence to design and develop an AI Predictive Model to address challenges faced by traditional transaction monitoring processes.

Keywords: Artificial intelligence, Machine learning, transaction monitoring, suspicious financial transactions, financial fraud detection, fraudulent transactions, financial crimes, fraud prediction

1 Introduction

Financial crimes such as illegal banking transactions negatively affect the integrity of financial institutions. Criminals use financial institutions such as the banking sector as instruments for laundering illicit funds [1,2]. This act presents the financial institutions with supervisory, financial security and reputation challenges. Hence, it is necessary to implement measures that would help to preserve a good reputation and avoid any risks resulting from these acts [3]. As a result of the magnitude of illicit financial flows and crimes, the Financial Action Task Force (FATF) [1] formulated recommendations that countries should adhere to help prevent these crimes. The recommendations include transaction monitoring; the worrying situation is that financial institutions are using traditional rule-based processes, which result in the inability to identify suspected fraudulent activities within financial transactions.

Traditional financial transaction monitoring processes use predefined thresholds as rules to detect fraud. The processes are non-agile and hence limited in adapting to the changing nature of fraudulent activities and the methods used [16]. Furthermore, the processes are ineffective, time-consuming, and impractical, and financial crime detection must evolve accordingly [6]. With advancements in technology, the FATF developed and promoted the use of digital solutions for AML/CFT based on AI and its different subsets, such as ML and natural language processing, to help with better transaction monitoring through facilitating data collection, processing and analysis and help banks identify and manage financial crimes such as money laundering and terrorist financing (ML/TF) [4]. AI offers a promising solution to the transaction monitoring process but has some challenges regarding its adoption. This research aims to use AI/ML algorithms to design and develop an AI predictive model for the transaction monitoring process within the banking sector.

1.1 Research Problem

Traditional transaction monitoring systems detect financial crimes and fraud through predefined rules concentrating on transactions equal to or beyond threshold values [7,8]. Hence, the methods are limited in fraud detection accuracy in a context where fraudulent activities are rare, split into small amounts and hidden among legitimate transactions [9,10]. The continuous increase in financial transaction data [8] and the changing nature of fraudulent activities results in a high risk of undetected fraudulent transactions, false positives, and false negatives [7,11]. Furthermore, heavy reliance on human intervention [12] makes traditional monitoring systems prone to error, hampering the timely identification of unusual transaction patterns, emerging fraud trends or infrequent patterns of suspicious activities [13].

The current status of the transaction monitoring systems gives room for improvement through AI-based ML algorithms. Even though AI-based ML models are proving effective in fraud detection, there are gaps and recommendations for further studies regarding use of AI algorithms for fraud detection and recommended further research. Subsequently, a study by [17] recommended using a combination of evolutionary algorithms and ML-based financial transaction feature engineering methods to enhance anomaly and fraud detection. Whereas studies by [14,15] focused on the gap in the existing literature about the interaction between AI and money laundering from an African perspective and recommended that more research studies are required from an African perspective to capture contextual factors that can influence the effectiveness of AI in combating money laundering in the African context. Traditional systems are ineffective in predicting and detecting fraudulent activities; for that reason, I plan to use AI to enhance the process.

1.2 Problem Statement

Financial institutions, including the banking sector, face challenges monitoring vast amounts of transaction data for financial crime detection. With continuous increases in financial transaction data and the constantly changing nature of fraudulent activities and methods used by criminals within the banking sector, transactional rule-based transaction monitoring processes find it difficult to detect trends and patterns of fraudulent activities accurately. Due to the drawbacks of current processes, the research project plans to leverage AI-based ML algorithms to design a robust, resilient and adaptive model that could enhance the transaction monitoring process.

1.3 Proposed Solution

The project will leverage AI based ML algorithms to design and develop an AI predictive model for transaction monitoring. It will also try to find ways to convert financial transactions to images for Machine Learning to train and identify patterns and trends of suspicious transactions. AI model will be evaluated for effectiveness and accuracy in identifying fraudulent activities by distinguishing between legitimate and suspicious transactions, reducing the rate of false positives and adopting the continuously changing fraudulent activities methods.

2 Research Aim and Questions

The research project aims to leverage AI-based ML algorithms to design and develop a predictive model for enhancing transaction monitoring processes for financial crime detection within financial institutions, specifically focusing on improving accuracy and efficiency. A prototype will be implemented to demonstrate the effectiveness of the model. The project's scope will focus on designing and implementing a prototype to analyse behaviour and networks of transactions, as well as using classification and clustering to convert transactions into images that ML can understand to detect patterns and trends of fraudulent activities. The research question will guide the project:

Main Research Question:

How can developing and implementing an AI-based predictive model enhance the detection of financial crimes in Malawi's banking sector?

2.1 Significance of the study

The study will make a meaningful contribution towards enhancing the transaction monitoring process in the banking sector while focusing on achieving accurate and effective financial crime and fraud detection. The study falls within the applied research; it will contribute practical knowledge and applicable procedures on how financial institutions can apply AI in developing robust, resilient and adaptive predictive models, which are crucial when dealing with ever-changing patterns of fraudulent activities.

3 Research Design and Methodology

The project will apply Design Science Research Methodology.

The researcher plans to use real-world/actual financial transaction datasets from the Malawi banking sector.

3.1 Research Design

The Design Science Research methodology will be used to design,develop and evaluate the proposed AI predictive model.

3.2 Experiment Design

The project will use real world financial transaction dataset from Malawi financial institutions especially the banking sector. The dataset will be used for training AI algorithms.

4 Conclusion

In this paper, we leverage AI to design and develop AI predictive model for transaction monitoring to detect patterns and trends of financial crimes from real world transaction dataset. Different classification and clustering ML techniques will be applied and trained with transaction data. Different data analysis methods will be applied and the model will be evaluated for accuracy and effectiveness.

References

1. FATF, “International Standards on Combating Money Laundering and the Financing of Terrorism & Proliferation,” *FATF*, Paris, France, 2012-2023, www.fatf-gafi.org/en/publications/Fatfrecommendations/Fatf-recommendations.html,
2. M.F.Tolipovich, “Improving the system for fighting against laundering the proceeds obtained from criminal activities in banks.” World Economics and Finance Bulletin, Vol. 32, p. 56-59,2024
3. B.Basaran-Brooks, “Money laundering and financial stability: does adverse publicity matter?.” Journal of Financial Regulation and Compliance, Vol. 30.2, pp. 196-21, 2022.
4. FATF, “Opportunities and Challenges of New Technologies for AML/CFT,” *FATF*, Paris, France,2021, https://www.fatf-gafi.org/publications/fatfrecommendations/documents/opportunities-challenges-newtechnologies-aml-cft.html
5. W.Hilal, S.A. Gadsden, and J. Yawney, “Financial fraud: a review of anomaly detection techniques and recent advances.” Expert Systems with Applications, 193,116429,2022.
6. J.West & M.Bhattacharya, “Intelligent financial fraud detection: A comprehensive review. Computers & Security”, Vol. 57, pp. 47-66, 2016, https://doi.org/10.1016/j.cose.2015.09.005.
7. D.Chau and M.van Dijck Nemcsik, “Anti-money laundering transaction monitoring systems implementation: Finding anomalies,” John Wiley & Sons,2020.
8. Z.Huang *et al.*, “Application of machine learning-based k-means clustering for financial fraud detection,” *Academic Journal of Science and Technology*, Vol. 10(1), pp. 33-39,2024.
9. X.Zhu *et al*., “Intelligent financial fraud detection practices in post-pandemic era,” The Innovation, Vol. 2(4), 2021.
10. A.Thommandru & B.Chakka, “Recalibrating the banking sector with blockchain technology for effective anti-money laundering compliances by banks,” Sustainable Futures, Vol. 5, 100107, 2023.
11. A.Gautam, “Evaluating the impact of artificial intelligence on risk management and fraud detection in the banking sector,” AI, IoT and the Fourth Industrial Revolution Review, Vol. 13(11), pp. 9-18, 2023.
12. B. Dhamodharan, “Beyond traditional methods: a novel approach to anomaly detection and classification using AI techniques,” Transactions on Latest Trends in Artificial Intelligence, Vol. 3(3), 2022.
13. I. Benchaji, S. Douzi, and B. El Ouahidi, “Credit card fraud detection model based on LSTM recurrent neural networks,” Journal of Advances in Information Technology, Vol. 12(2), 2021.
14. M.Dzingirai, “Effects of artificial intelligence on money laundering in Southern Africa.” In ICABR Conference, Cham: Springer Nature Switzerland, pp. 483-500,2023.
15. Tagbo, Selorm Kofi, and Adebayo Felix Adekoya. "A systematic literature review of machine learning techniques in financial fraud prevention and detection." International Journal of Society Systems Science, 14(4) (2023): 303-348.

16. Kotagiri, Anudeep, and Abhinay Yada. "Improving Fraud Detection in Banking Systems: RPA and Advanced Analytics Strategies." International Journal of Machine Learning for Sustainable Development 6.1 (2024): 1-20.

17. I.D. Mienye and Y.Sun, "A machine learning method with hybrid feature selection for improved credit card fraud detection. Applied Sciences, Vol. 13(12), p.7254,2023.

Big Data Analytics for Enhancing Student Retention in South African Universities: An Examination of Critical Success Factors

Ganizani Mainoti[1]

[1] University of Fort Hare, East London, SouthAfrica
gmainoti@ufh.ac.za

1 Introduction

In recent years, higher education institutions worldwide have faced the challenge of improving student retention rates, ensuring that students progress successfully through their academic journeys and ultimately graduate [1]. This issue is particularly pressing in South African universities, given the diverse student population and unique socioeconomic factors influencing student success [2]. To address this challenge, an emerging approach has gained attention: utilizing big data analytics [3]–[5]. Big data analytics offers the potential to extract valuable insights from vast amounts of student data, enabling institutions to identify early warning signs of student attrition and implement targeted interventions [6]. This research paper examines the critical success factors of applying big data analytics to enhance student retention in South African universities. By investigating the factors contributing to this approach's effectiveness, this study seeks to provide valuable insights and recommendations for institutions aiming to improve student retention rates and support student success in the South African higher education context.

2 Background and Research Problem

Student retention has been a long-standing challenge in South African higher education [2], [7]. The issue of low student retention rates in South African universities can be attributed to a range of factors, including socioeconomic challenges, academic preparation, access to resources and support, and institutional factors [8]. South Africa has a high unemployment rate, with many students coming from low-income families, often resulting in a lack of financial resources to support their studies. This financial strain can lead to students dropping out of their studies to find employment to support their families. Additionally, many students come from under-resourced schools, which can lead to a lack of academic preparedness, making it difficult for them to cope with the academic demands of higher education. Moreover, South African universities face institutional factors that impact student retention [9]. The theoretical and social support provided by universities can be inadequate, leading to feelings of isolation and disengagement among students [10]. Additionally, administrative and bureaucratic challenges such as registration delays, poor communication, and inadequate academic advising can exacerbate the problem [10].

In response to this challenge, universities and policymakers have implemented a range of initiatives to improve student retention rates, such as providing financial aid, improving academic support structures, and enhancing student engagement through digital platforms and learning analytics [11]–[13]. However, the problem of low student retention rates in South African universities remains a significant challenge that requires ongoing attention and innovation [14]. The potential of big data analytics to address this problem has been noted in the literature [2], [15], [16]. Still, there is limited research on how it can be effectively implemented in the South African context. This study aims to contribute to the existing literature by investigating and exploring the factors influencing successful big data analytics initiatives in higher education.

3 Methodology

This study employed a literature review methodology to examine the critical success factors of big data analytics for enhancing student retention in South African universities. A literature review is a systematic and comprehensive approach to identifying, evaluating, and synthesizing existing scholarly works relevant to the research topic [17]. The primary objective of this study was to identify and analyze the critical success factors associated with big data analytics in improving student retention in South African universities. The study aimed to provide an in-depth understanding of the current state of research and the key findings in this area. An extensive search of academic databases, including but not limited to Google Scholar, JSTOR, and IEEE Xplore, was conducted to conduct the literature review. The search involved keywords and combinations relevant to the research topic, such as "big data analytics," "student retention," "South African universities," and "critical success factors." The inclusion criteria for selecting relevant studies were based on the following factors: studies published in peer-reviewed journals, conference proceedings, and reputable academic sources; and studies focused on big data analytics and its applications in student retention. The selected studies were thoroughly reviewed and analyzed to extract relevant information. Key data points and findings were recorded, including the critical success factors identified, methodologies employed, and outcomes achieved. The data were organized and synthesized to identify common themes, patterns, and relationships among the studies.

4 Literature Review

Based on the research conducted by [18], the comprehensive retention rate within South African universities is approximately 75%, although certain institutions have reported rates as low as 50%. The study further highlights notable variations in retention rates among different universities, with historically disadvantaged institutions generally exhibiting lower retention rates than their more established counterparts. Factors that have been identified as contributing to low retention rates in South African universities include inadequate academic preparation, financial constraints, lack of support services, and the need for students to work while studying [19]. Language barriers, inadequate academic infrastructure, and lack of student engagement have also been identified as contributing factors [20].

In response to these challenges, South African universities have implemented a range of strategies to improve student retention rates. These include academic and social support services such as tutoring, mentorship, and counseling. Universities have also implemented academic interventions, such as bridge programs, learning communities, and orientation programs, to improve student preparedness for university study [21].

4.1 Big Data and Its Potential for Managing Student Retention

Big data is a term used to describe large volumes of structured and unstructured data that can be analyzed to reveal patterns, trends, and associations. In recent years, big data has become increasingly relevant in higher education, where it is used to inform decision-making and improve student outcomes [2]. In the context of student retention, big data has the potential to provide insights into student behavior, academic performance, and engagement that can inform interventions aimed at improving retention rates [16]. Table 1 below summarizes areas of big data analytics application in higher education.

Table 1: Application of big data analytics in higher education

Area of Higher Education	Examples
Student retention	By analyzing student data, institutions can identify factors contributing to student retention, such as academic performance, engagement, and socioeconomic status [24]. Institutions can then develop targeted interventions to improve retention rates.
Academic performance	Predictive analytics can be used to identify at-risk students and provide targeted support and resources to help them succeed [25]. Learning analytics can be used to identify which course materials are most effective and which students need additional support and feedback [26].
Resource allocation	By analyzing data on resource utilization, institutions can make informed decisions about staffing, space utilization, and procurement [22]. This can help institutions allocate resources more efficiently and effectively.
Teaching and Learning	Learning analytics can be used to analyze student interactions with course content, identify areas for improvement and innovation, and personalize learning experiences [27].
Research	Big Data Analytics can analyze research outputs, identify trends, and predict future research directions [28]
Institutional Operations	Big Data Analytics can be used to optimize enrollment management, financial aid, and facilities management by identifying trends and patterns that inform strategic decision-making and resource allocation [29], [28].
Governance and Policy	Big Data Analytics can be used to analyze institutional performance, compliance, and risk management [30].

5 Critical Success Factors for Big Data Analytics in Higher Education

To optimize the utilization of big data analytics in the context of student retention within higher education, diligent attention must be paid to the critical success factors inherent in this endeavor. These factors encompass a range of components that contribute to the effective implementation and utilization of data-driven methodologies within higher education. By acknowledging and addressing these factors, educational institutions can fully leverage the advantages of big data analytics while upholding ethical standards, engaging stakeholders, and safeguarding data privacy. Table 2, presented below, offers a succinct overview of the principal critical success factors associated with big data analytics in higher education. A comprehensive understanding and integration of these factors will empower institutions to make well-informed decisions and enact interventions that optimize their efforts toward student retention.

Table 2: Critical success factors for big data analytics in higher education

Critical Success Factors	Explanation
1. Data Governance and Infrastructure	Establishing robust data governance frameworks and infrastructure to ensure standardized data collection processes, accuracy, quality, and security [39], [40].
2. Stakeholder Engagement and Collaboration	Engaging administrators, faculty, staff, and students to foster a shared understanding, buy-in, and participation in data-driven initiatives [37].
3. Skilled Workforce and Capacity Building	Developing a skilled workforce through training and professional development programs enhances analytical capabilities within the institution [41].
4. Ethical Data Use and Privacy Protection	Establishing guidelines for ethical data practices, including privacy protection, data anonymization, and compliance with regulations and policies [42], [43].
5. Actionable Insights and Intervention Strategies	Translating data insights into actionable strategies and interventions to positively impact student retention efforts and improve educational outcomes [44].
6. Scalability and Flexibility	Ensuring that the big data analytics infrastructure can handle increasing data volumes, adapt to emerging technologies, and accommodate evolving needs.
7. Continuous Evaluation and Improvement	Regularly evaluating and monitoring data-driven initiatives, measuring their effectiveness, and refining strategies based on ongoing data analysis [45].

6 Conclusion

Exploring the challenges and identifying critical success factors provides valuable insights for higher education institutions aiming to leverage big data analytics in their student retention efforts. The critical success factors include data governance and infrastructure, stakeholder engagement and collaboration, workforce development, ethical data use and privacy protection, actionable insights and intervention strategies, scalability and flexibility, and continuous evaluation and improvement. Ongoing research is necessary to keep pace with the evolving technological landscape and explore emerging analytical techniques and tools to enhance student retention efforts. As big data analytics advances, assessing new approaches and algorithms' effectiveness and ethical implications is crucial.

References

1. T. Thies, "International students in higher education: the effect of student employment on academic performance and study progress," High. Educ., Nov. 2022, doi: 10.1007/s10734-022-00950-5.
2. N. Cele, "Big data-driven early alert systems as means of enhancing university student retention and success," South African J. High. Educ., vol. 35, no. 2, pp. 56–72, 2021, doi: 10.20853/35-2-3899.
3. D. West, D. Heath, and H. Huijser, "Let's talk learning analytics: A framework for implementation in relation to student retention," J. Asynchronous Learn. Netw., vol. 20, no. 2, pp. 30–50, 2016, doi: 10.24059/olj.v20i2.792.
4. A. Merceron, P. Blikstein, and G. Siemens, "Learning Analytics: From Big Data to Meaningful Data," J. Learn. Anal., vol. 2, no. 3, pp. 4–8, Feb. 2016, doi: 10.18608/jla.2015.23.2.
5. C. Umezuruike and H. N. Ngugi, "Imminent Challenges of Adoption of Big Data in Educational Systems in Sub-Saharan Africa Nations," Int. J. Recent Technol. Eng., vol. 8, no. 5, pp. 4544–4550, 2020, doi: 10.35940/ijrte.e6885.018520.
6. C. Mi, "Student Performance Early Warning based on Data Mining," Int. J. Performability Eng., 2019, doi: 10.23940/ijpe.19.03.p11.822833.
7. H. Visser and D. van Zyl, "Assessment of academic readiness to achieve student success and retention," South African J. High. Educ., vol. 27, no. 2, pp. 330–352, 2016, doi: 10.20853/27-2-249.
8. T. A. Thomas and D. Maree, "Student factors affecting academic success among undergraduate students at two South African higher education institutions," South African J. Psychol., vol. 52, no. 1, pp. 99–111, Feb. 2022, doi: 10.1177/0081246320986287.
9. B. V. Mtshweni, "COVID-19: Exposing unmatched historical disparities in the South African institutions of higher learning," South African J. High. Educ., vol. 36, no. 1, pp. 234–250, 2022, doi: 10.20853/36-1-4507.
10. M. Sadoughi and F. Hesampour, "Relationship between Social Support, Loneliness and Academic Adjustment among University Students," Int. J. Acad. Res. Psychol., vol. 3, no. 1, pp. 48–55, 2016, doi: 10.46886/ijarp/v3-i1/2455.
11. D. Potter, D. Jayne, and S. Britt, "Financial Anxiety Among College Students: The Role of Generational Status," J. Financ. Couns. Plan., vol. 31, no. 2, pp. 284–295, 2020, doi: 10.1891/JFCP-17-00033.

12. M. Krawitz, J. Law, and S. Litman, “How higher-education institutions can transform themselves using advanced analytics,” Public Sect. Pract., no. August, p. 6, 2018, [Online]. Available: https://www.mckinsey.com/~/media/McKinsey/Industries/Social Sector/Our Insights/How higher education institutions can transform themselves using advanced analytics/How-higher-education-institutions-can-transform-themselves-using-advanced-analytics.pdf
13. M. Teräs, J. Suoranta, H. Teräs, and M. Curcher, “Post-Covid-19 Education and Education Technology ‘Solutionism’: a Seller’s Market,” Postdigital Sci. Educ., vol. 2, no. 3, pp. 863–878, 2020, doi: 10.1007/s42438-020-00164-x.
14. R. Chetty and S. Pather, “Challenges in Higher Education in South Africa,” Tell. Stories Differ. Engag. 21st Century Students Through Digit. Storytell., no. July, pp. 1–6, 2016, [Online]. Available: https://www.researchgate.net/publication/305213223_Challenges_in_Higher_Education_in_South_Africa
15. K. Sin and L. Muthu, “"Application of Big Data in Education Data Mining and Learning Analytics – a Literature Review ",” ICTACT J. Soft Comput., vol. 05, no. 04, pp. 1035–1049, 2015, doi: 10.21917/ijsc.2015.0145.
16. K. Adam, N. A. A. Bakar, M. A. I. Fakhreldin, and M. A. Majid, “Big Data and Learning Analytics: A Big Potential to Improve e-Learning,” Adv. Sci. Lett., vol. 24, no. 10, pp. 7838–7843, 2018, doi: 10.1166/asl.2018.13028.
17. H. Snyder, “Literature review as a research methodology: An overview and guidelines,” J. Bus. Res., vol. 104, no. August, pp. 333–339, 2019, doi: 10.1016/j.jbusres.2019.07.039.
18. P. Moodley and R. Singh, “Addressing student dropout rates at South African universities,” vol. 17, no. 17, pp. 91–115, 2015.
19. P. Machika and B. Johnson, “Postgraduate students experience of poverty and academic success at a university of technology in South Africa,” South African J. High. Educ., vol. 29, no. 6, 2016, doi: 10.20853/29-6-542.
20. T. D. Sikhwari, C. Maphosa, L. Masehela, and C. Ndebele, “Exploring Students’ Views on Factors Affecting Academic Performance in a South African University,” Int. J. Educ. Sci., vol. 10, no. 3, pp. 442–450, 2015, doi: 10.1080/09751122.2015.11890367.
21. K. Linden, “Improving Student Retention by Providing Targeted Support to University Students who do not Submit an Early Assessment Item. A Practice Report,” Student Success, vol. 12, no. 3, pp. 67–73, 2022, doi: 10.5204/SSJ.2152.
22. S. Steiner, “Harnessing Data to Make Better-informed Decisions,” Scientia, 2022, doi: 10.33548/SCIENTIA768.
24. M. Ngoma, P. D. Ntale, and E. Abaho, “Social-Economic Factors, Student Factors, Student Academic Goals and Performance of Students in Institutions of Higher Learning in Uganda,” Africa Educ. Rev., vol. 14, no. 2, pp. 106–121, Apr. 2017, doi: 10.1080/18146627.2017.1286941.
25. S. Ratnapalan and A. Jarvis, “How to Identify Medical Students at Risk of Academic Failure and Help Them Succeed? An Interview with a Medical Educator,” Med. Sci. Educ., vol. 30, no. 2, pp. 989–994, Jun. 2020, doi: 10.1007/s40670-020-00940-1.
26. G. Siemens and P. Long, “Penetrating the fog: Analytics in learning and education,” Educ. Rev., vol. 46, no. 5, 2020.
27. C. Herodotou, B. Rienties, B. Verdin, and A. Boroowa, “Predictive Learning Analytics ‘At Scale’: Guidelines to Successful Implementation in Higher Education,” J. Learn. Anal., vol. 6, no. 1, Apr. 2019, doi: 10.18608/jla.2019.61.5.
28. D. Lesjak, S. Natek, and F. Kohun, “Big data analytics in higher education,” Issues Inf. Syst., vol. 22, no. 4, pp. 320–333, 2021, doi: 10.48009/4_iis_2021_346-359.

29. M. J. Dennis, "How big data can help recruitment and retention," Enroll. Manag. Rep., vol. 21, no. 5, p. 3, Jul. 2017, doi: 10.1002/emt.30320.
30. M. E. Khatib, H. A. Shehhi, and M. A. Nuaimi, "How {Big} {Data} and {Big} {Data} {Analytics} {Mediate} {Organizational} {Risk} {Management}," J. Financ. Risk Manag., vol. 12, no. 01, pp. 1–14, 2023.
37. S. R. Mathur and K. M. Corley, "Bringing ethics into the classroom: Making a case for frameworks, multiple perspectives and narrative sharing," Int. Educ. Stud., vol. 7, no. 9, pp. 136–147, 2014, doi: 10.5539/ies.v7n9p136.
39. S. S. Chaurasia, D. Kodwani, H. Lachhwani, and M. A. Ketkar, "Big data academic and learning analytics: Connecting the dots for academic excellence in higher education," Int. J. Educ. Manag., vol. 32, no. 6, pp. 1099–1117, 2018, doi: 10.1108/IJEM-08-2017-0199.
40. M. Attaran, J. Stark, and D. Stotler, "Opportunities and challenges for big data analytics in US higher education: A conceptual model for implementation," Ind. High. Educ., vol. 32, no. 3, pp. 169–182, 2018, doi: 10.1177/0950422218770937.
41. M. Barbero et al., "Big data analytics for policy making," European Commission & Deloitte. p. 122 pages, 2016. [Online]. Available: https://joinup.ec.europa.eu/sites/default/files/document/2016-07/dg_digit_study_big_data_analytics_for_policy_making.pdf
42. R. Ferguson, "Ethical challenges for learning analytics," J. Learn. Anal., vol. 6, no. 3, pp. 25–30, 2019, doi: 10.18608/jla.2019.63.5.
43. D. Murchan and F. Siddiq, "A call to action: a systematic review of ethical and regulatory issues in using process data in educational assessment," Large-Scale Assessments Educ., vol. 9, no. 1, 2021, doi: 10.1186/s40536-021-00115-3.
44. M. Easter et al., "A Framework for Student Success Analytics," 2022, [Online]. Available: https://library.educause.edu/resources/2022/5/a-framework-for-student-success-analytics
45. G. Gray, G. Cooke, P. Murnion, P. Rooney, and K. C. O'Rourke, "Stakeholders' insights on learning analytics: Perspectives of students and staff," Comput. Educ., vol. 187, no. May, p. 104550, 2022, doi: 10.1016/j.compedu.2022.104550.

Evaluation of YOLO Algorithms for Real-Time Fish Detection in Underwater Environments

Daanyaal Salie[1][0000−0001−5642−6906] and Dane Brown[2,3][0000−0001−7395−7370]

[1] Rhodes University, Makhanda, Eastern Cape, South Africa
g19s6325@campus.ru.ac.za
[2] Rhodes University, Makhanda, Eastern Cape, South Africa
d.brown@ru.ac.za

Abstract. This research investigated the generalization capabilities of AI models for underwater fish detection. YOLOv5, YOLOv8, and YOLOv9 were evaluated using two datasets: DeepFish from tropical Australia and Seychelles from the Aldabra Islands. YOLOv9 consistently outperformed the other models, achieving the highest validation and test mAPs on both datasets. This showcased its superior feature learning and generalization within homogenous data splits. Notably, YOLOv9 achieved an mAP of 0.953 and 0.955 on DeepFish and Seychelles, respectively.
The Seychelles-trained models were further evaluated on a different Seychelles Reference Library dataset to assess model generalization on unseen heterogeneous data collected using a different camera perspective. YOLOv8 only marginally outperformed YOLOv5 by 0.001 mAP. Regardless, these models could remove the burden of manual annotation towards acquiring larger labelled training data. These findings highlight the potential for precise fish detection and emphasize the need for diverse datasets to enhance model generalization. In the future, further experimentation may be prudent on diverse datasets to ascertain the strengths and weaknesses of these models in different environments.

Keywords: Object Detection, YOLO, Deep Learning, Computer Vision, Marine Research

1 Introduction

Studying fish populations and their behaviours is essential for understanding and protecting marine ecosystems. Traditional research methods, like visual surveys by divers or net capture, are labour-intensive, costly, and potentially disruptive to marine life [1]. Underwater camera systems, such as Baited Remote Underwater Video Stations (BRUVs) and Autonomous Underwater Vehicles (AUVs), offer a non-invasive alternative, revolutionizing marine research by enabling observation and recording of fish behaviour in their natural habitats [2]. However, the vast amount of video data generated requires automated analysis tools.

Manual annotation of underwater video data, a prerequisite for training object detection models, is time-consuming and expensive. Researchers have turned to the YOLO (You Only Look Once) family of object detectors to automate this process [3]. While earlier versions like YOLOv3 [4] and YOLOv4 [5] showed promise, they struggled with accurately identifying small fish in complex underwater backgrounds [2]. Subsequent iterations, YOLOv5 [10] through YOLOv8 [9], have demonstrated improved performance, with YOLOv5 being particularly effective in certain underwater scenarios [7]. The latest version, YOLOv9 [8], introduces innovative features like Programmable Gradient Information (PGI) and the Generalized Efficient Layer Aggregation Network (GELAN) that offer the potential further to enhance object detection performance in challenging underwater environments.

This research investigates the potential of YOLOv5, YOLOv8, and YOLOv9 architectures to address these challenges by leveraging their unique advancements. Specifically, we aim to assess their ability to detect small fish in complex backgrounds accurately, a critical step towards semi-automated annotation and streamlined data analysis in marine research.

2 Methodology

This research comprehensively evaluates the object detection performance of YOLOv5, YOLOv8, and YOLOv9 in underwater environments. These models were selected for their unique architectural advancements: YOLOv5 features an anchor-free split head for greater flexibility and adaptive detection, YOLOv8 employs advanced backbone and neck architectures to enhance feature extraction and detection performance, while YOLOv9 incorporates PGI and GELAN to improve gradient generation, model convergence, and parameter utilization. A two-pronged approach is employed to assess real-world performance and generalization capabilities:

- **Benchmark Analysis:** Utilizes the DeepFish dataset [6], a standard benchmark comprising high-resolution video footage from 20 marine habitats in tropical Australia, with 3405 training images, 973 validation images, and 486 test images. This analysis establishes baseline performance metrics under controlled conditions.
- **Generalization Analysis:** To assess the models' ability to adapt to novel underwater environments, a diverse dataset of 14,272 training images and 1,996 validation images was curated from Seychelles, encompassing 364 fish species inhabiting the Aldabra Islands was used for initial evaluation. The Seychelles Reference Images dataset, a separate collection of high-quality fish images from the same region but a different camera (diver) perspective, was reserved as an unseen test set. Both datasets were sourced from the South African Institute for Aquatic Biodiversity (SAIAB). This approach evaluated the models' performance on data distinct from their training set, thus gauging their real-world applicability in diverse marine ecosystems.

All models were trained using transfer learning from the COCO dataset with consistent parameters (300 epochs, image size 640 x 640, learning rate 0.01, weight decay 0.0005, batch size 16) on a machine equipped with an AMD Ryzen 7 3800X CPU, RTX 4090 GPU, and 64GB RAM. By combining these analyses, this research aims to offer insights into the strengths and weaknesses of each model, guiding the development of more robust and accurate underwater object detection models for marine research applications.

3 Results and Discussion

This section presents the results of the benchmark and generalization analyses, comparing the performance of YOLOv5, YOLOv8, and YOLOv9 in underwater fish detection tasks.

3.1 Benchmark Analysis (DeepFish Dataset)

YOLOv9 consistently outperformed YOLOv5 and YOLOv8 across all evaluated metrics on the DeepFish dataset (Table 1). YOLOv9 achieved the highest mAP of 0.953, indicating its superior ability to locate fish accurately. Furthermore, YOLOv9 demonstrated a slight edge in the F1 score of 0.930 compared to 0.927 and 0.926 achieved by YOLOv5 and YOLOv8, respectively, suggesting a better balance between precision and recall in fish detection. YOLOv9 also exhibited the highest precision of 0.963, signifying fewer false positive detections while maintaining a recall comparable to the other models.

Table 1. DeepFish YOLO Detection Results

Model	mAP@50	F1	Precision	Recall
YOLOv5	0.944	0.927	0.954	0.903
YOLOv8	0.950	0.926	0.951	0.903
YOLOv9	0.953	0.930	0.963	0.901

3.2 Generalization Analysis (Seychelles and Reference Library)

The generalization analysis further affirmed YOLOv9's superior performance. When trained on the Seychelles dataset, YOLOv9 consistently outperformed both YOLOv5 and YOLOv8 across all metrics, achieving a mAP of 0.955, an F1 score of 0.895, a precision of 0.893, and a recall of 0.898 (Table 2). This demonstrates YOLOv9's exceptional accuracy and balanced performance in fish detection on the training dataset.

YOLOv9 maintained its lead when tested on the unseen Seychelles Reference Library, achieving an mAP of 0.674 and an F1 score of 0.648 (Table 3). This indicates a strong ability to generalize to novel fish species and underwater environments not encountered during training.

While YOLOv8 underperformed YOLOv9 on the Seychelles dataset, with an mAP of 0.949 compared to YOLOv9's 0.955, it marginally surpassed YOLOv5 on the unseen Seychelles Reference Library, achieving an mAP of 0.659 compared to YOLOv5's 0.658. This suggests that YOLOv8 may possess slightly better generalization capabilities when faced with novel data, albeit not statistically significant in this study. The marginal difference in performance between YOLOv8 and YOLOv5 on the unseen dataset highlights the need for further investigation with larger and more diverse datasets to draw definitive conclusions about their relative generalization abilities.

Table 2. Seychelles YOLO Detection Results

Model	mAP@50	F1	Precision	Recall
YOLOv5	0.951	0.888	0.889	0.888
YOLOv8	0.949	0.887	0.887	0.888
YOLOv9	0.955	0.895	0.893	0.898

Table 3. Seychelles Reference Library YOLO Detection Results

Model	mAP@50	F1	Precision	Recall
YOLOv5	0.658	0.641	0.731	0.571
YOLOv8	0.659	0.634	0.768	0.540
YOLOv9	0.674	0.648	0.740	0.577

3.3 Discussion

The results presented in Tables 1, 2, and 3 validate YOLOv9's superior performance in underwater fish detection, demonstrating its effectiveness in various underwater scenarios. Its consistent superiority across evaluation metrics on both benchmark and training datasets underscores the effectiveness of its architectural enhancements, particularly the PGI module.

However, the generalization analysis on the unseen Seychelles Reference Library dataset revealed nuanced findings. While YOLOv9 maintained its lead, its performance notably decreased compared to the training dataset, suggesting potential challenges in handling novel fish species or environments. Conversely, YOLOv8, despite trailing slightly on the training set, exhibited more consistent performance across both datasets, suggesting a subtle advantage in generalization to new scenarios.

These findings underscore the complexity of underwater object detection, emphasizing the need for models balancing high performance and adaptability. The marginal difference in generalization between YOLOv8 and YOLOv5 highlights the importance of diverse dataset testing for accurate assessment. Additionally, current dataset limitations underscore the need for richer, more varied data to enhance feature learning and model generalization.

4 Conclusion

This research demonstrates YOLOv9's exceptional performance in underwater fish detection, highlighting the importance of diverse and representative datasets for training robust object detection models. These findings contribute to developing automated tools that promise to enhance our understanding and conservation of underwater ecosystems.

References

1. Xu, S., Zhang, M., Song, W., Mei, H., He, Q. & Liotta, A. A Systematic Review and Analysis of Deep Learning-Based Underwater Object Detection. *Neurocomputing.* **527** pp. 204-232 (2023), https://doi.org/10.1016/j.neucom.2023.01.056
2. Muksit, A., Hasan, F., Emon, M., Haque, M., Anwary, A. & Shatabda, S. YOLO-Fish: A Robust Fish Detection Model to Detect Fish in Realistic Underwater Environment. *Ecol. Informatics.* **72** pp. 101847 (2022), https://doi.org/10.1016/j.ecoinf.2022.101847
3. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference On Computer Vision And Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016.* pp. 779-788 (2016), https://doi.org/10.1109/CVPR.2016.91
4. Redmon, J. & Farhadi, A. YOLOv3: An Incremental Improvement. *CoRR.* **abs/1804.02767** (2018), http://arxiv.org/abs/1804.02767
5. Bochkovskiy, A., Wang, C. & Liao, H. YOLOv4: Optimal Speed and Accuracy of Object Detection. *CoRR.* **abs/2004.10934** (2020), https://arxiv.org/abs/2004.10934
6. Saleh, A., Laradji, I., Konovalov, D., Bradley, M., Vázquez, D. & Sheaves, M. A Realistic Fish-Habitat Dataset to Evaluate Algorithms for Underwater Visual Analysis. *CoRR.* **abs/2008.12603** (2020), https://arxiv.org/abs/2008.12603
7. Gašparović, B., Mauša, G., Rukavina, J. & Lerga, J. Evaluating YOLOV5, YOLOV6, YOLOV7, and YOLOV8 in Underwater Environment: Is There Real Improvement?. *2023 8th International Conference On Smart And Sustainable Technologies (SpliTech).* pp. 1-4 (2023)
8. Wang, C., Yeh, I. & Liao, H. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *CoRR.* **abs/2402.13616** (2024), https://doi.org/10.48550/arXiv.2402.13616
9. Jocher, G., Chaurasia, A. & Qiu, J. Ultralytics YOLOv8. (2023), https://github.com/ultralytics/ultralytics/
10. Jocher, G. Ultralytics YOLOv5. (2020), https://github.com/ultralytics/yolov5

Evolving Robotic Systems with Improved General Capabilities Using Neural Modules that Support Perception, Higher Cognition, and Memory-Based Adaptation

Sithuliswa Mthembu[1][0000−0002−0234−7388] and Mathys C. du Plessis[2][0000−0002−9187−3960]

[1] Nelson Mandela University, Gqeberha, 6001, South Africa
s219046379@mandela.ac.za
[2] Nelson Mandela University, Gqeberha, 6001, South Africa
mc.duplessis@mandela.ac.za

Abstract. In biological organisms, innate mechanisms often underpin complex behaviours. Animals possess specialised sensory organs and neural circuits pre-tuned to process environmental information such as visual and auditory cues. The absence of these modules limits the behavioural range and impairs vital tasks like food gathering due to incomplete environmental representation. Inspired by these biological principles, this research explores whether integrating analogous artificial neural network (ANN) modules specialised for distinct sensor modalities into a central network controller can enhance intelligent behaviours in evolutionary robotics. The central network combines information from specialised modules to dictate the robot's behaviour. The first research objective addresses the issue of incomplete information. In ER, robotic agents often lack comprehensive data about their internal state (e.g., orientation, position) and external state (e.g., obstacles in their environment). The hypothesis is that specialised perceptual modules will improve task performance by better representing internal and external environments. The second objective explores decoupling sensory and control capabilities. Traditional ER approaches use a monolithic ANN with numerous input neurons for all sensor modalities, leading to inefficient training and execution. The proposed approach decouples these functions into separate ANN modules, allowing for more targeted training and potentially better performance. Experiments will compare modular, decoupled controllers with traditional monolithic ANNs in tasks of increasing complexity. The goal is to determine if this approach can improve task performance and unlock new or unexpected behaviours and optimisation pathways, mirroring the advantages seen in biological systems.

Keywords: Artificial Neural Network, Evolutionary Robotics, Sensory Decoupling, Monolithic Controllers, Modular Controllers, Specialised Sensory Organs, Environmental Representation, Sensor Modalities.

1 Introduction

This paper reports the work in progress of an investigation conducted to determine the extent to which incomplete information affect the performance of a robot controller in Evolutionary Robotics (ER). As an initial step, the current work explores the performance of a basic robotic controller architecture in the context of a checkpoint navigation task. In this setup, a Khepera III mobile robot is tasked with visiting a series of randomly ordered checkpoints within a bounded rectangular arena. The report begins by evaluating a base controller design that accepts instructional inputs specifying the current target checkpoint. This establishes a baseline understanding of the limitations and performance characteristics of a relatively simple control approach. This work-in-progress paper thus lays the foundation for an ongoing exploration of techniques to enhance robotic performance through the use of specialised neural modules.

2 Task Definition

The research experiments took place on a rectangular platform where each corner served as a potential goal destination (refer to Fig 1). The central task involved the robot moving from one corner to another in a random order. If a controller reached its instructed corner, it would be instructed to reach a different, randomly chosen, corner. The robot was always started in the same position and orientation, but the order of the instructed corners would vary. This process continued until the controller had been instructed to reach all four corners

Fig. 1. The rectangular environment.

3 Controller Architecture

The controller used was an LSTM with 4 inputs. These were the hot encoded instructional inputs which tells the controller which corner to navigate to in the platform. For instance, goal 1 in Fig 2 had the encoding 1000, goal 2 had the encoding 0100, etc. Preliminary experimentation determined that a hidden layer size of 27 neurons provided better controller performance.

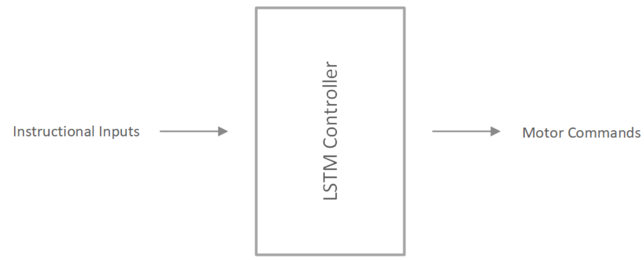

Fig. 2. The LSTM controller with instructional inputs.

4 Results

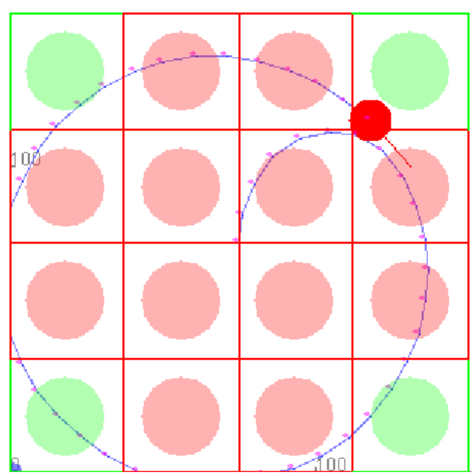

Fig. 3. Typical example of the controller paths.

The controller path shown in Fig 3 indicates that the evolved controllers predominantly exhibited a strategy of moving in a circular motion along the walls of the platform. In fact, across all the evaluated controllers, the same wall-following behavior in a clockwise direction was observed. Despite the varying order of checkpoint visitation, the controllers did not demonstrate any significant variation in their overall navigation strategy. They simply followed the perimeter of the platform, completing the task by visiting each checkpoint in succession as they encountered them during the circular motion. This consistent wall-following behavior suggests that the evolved controllers were unable to develop more sophisticated navigation strategies to efficiently visit the checkpoints in the specified order. The results highlight the challenge of evolving controllers that can adapt their navigation strategies to the unknown conditions, especially its internal state relative to its environment.

5 Discussion and Conclusion

The results reveal that the evolved controllers exhibited a consistent strategy of simply following the walls of the platform in a clockwise direction, regardless of the randomised order of checkpoint visitation. This behavior suggests that the controllers were unable to develop more sophisticated navigation strategies to efficiently visit the checkpoints. One key factor that may have contributed to this limited performance is the incomplete information available to the controllers. The controllers only had access to the instructional inputs, which indicated the target checkpoint to reach, but lacked any explicit representation of the robot's internal state, such as its current position and orientation within the environment. To address this limitation of incomplete information, a potential solution could be to equip the controllers with additional sensory inputs that provide a frame of reference for the robot's location and orientation. Animals use environmental cues, such as the sun, for orientation and navigation [7, 2, 5]. Inspired by this, robot controllers could be augmented with auditory information as a similar frame of reference. Fig 4 illustrates the proposed approach, where the robot is equipped with two microphones, one on each side, similar to the human ears. These microphones could capture the auditory cues in the environment, and the controller could learn to map the perceived sound information to the robot's internal state and relative position within the environment.

Fig. 4. Robot is equipped with two microphones.

The sound file (shown in Fig 5) used in this experiment was a periodic snap sound, played every half a second. To analyse the auditory cues, each audio signal was divided into equal segments, and the Interaural Level Difference (ILD) of each segment was calculated and plotted in the graphs shown in Fig 6.

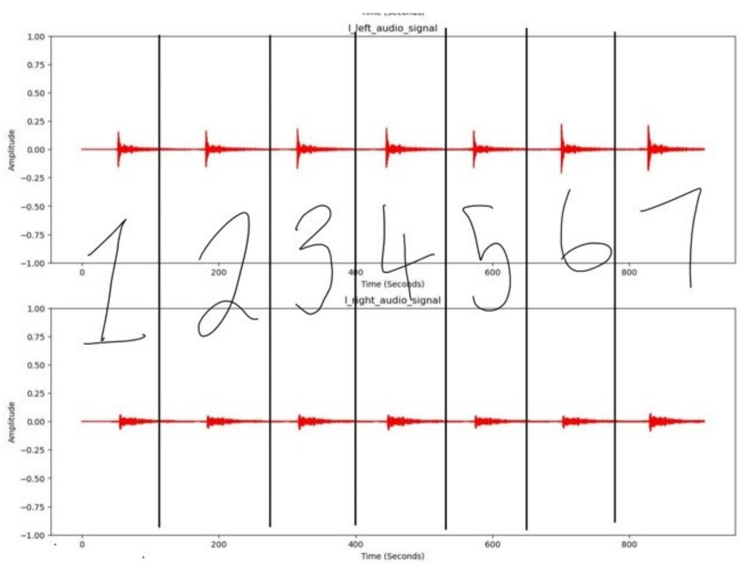

Fig. 5. The periodic snap sound.

As shown in Fig 6, the difference in sound intensity between the two microphones, known as the Interaural Level Difference (ILD), can provide valuable information about the direction of the sound source [4, 6, 1, 3]. For instance, a higher amplitude in the right microphone would indicate that the sound is coming from the right side of the robot.

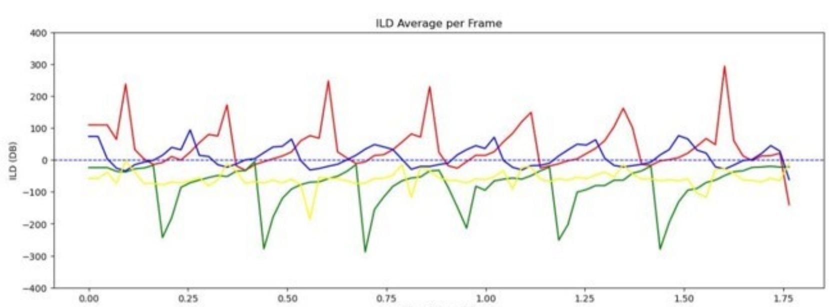

Fig. 6. The difference in sound intensity between the two microphones.

The graphs in Fig 6 illustrate the relationship between the ILD and the direction of the sound source. The red line, which is the highest above the zero line, represents the ILD when the sound source is located to the left of the microphones. Conversely, the green line, which is the lowest below the zero line, corresponds to the ILD when the sound source is positioned to the right of the microphones. The blue line, slightly above the zero line, indicates the ILD when the sound source is in front of the microphones. Finally, the yellow line, slightly below the zero line, represents the ILD when the sound source is behind the microphones. The distinct patterns observed in these graphs demonstrate that the ILD can effectively convey information about the direction of the sound source. By training the controller to interpret these auditory cues, it can potentially develop a better understanding of its own position and orientation within the environment, which can then be used to enhance the robot's navigation capabilities. The controller can then be able to map the perceived sound to the corresponding information about its internal state, allowing it to localise the sound source and use this as a frame of reference for its own movements.

References

1. Clarkson, J: Human capability and product design, pp. 165–198. Elsevier (2008). https://doi.org/10.1016/B978-008045089-6.50009-5
2. Dacke, M, Baird, E, Byrne, M, Scholtz, C, Warrant, E: Dung beetles use the milky way for orientation. Current Biology **23**, 298–300 (2 2013). https://doi.org/10.1016/j.cub.2012.12.034
3. Desai, D, Mehendale, N: Robotic ear: Audio signal processing for detecting direction of sound. SSRN Electronic Journal (2021). https://doi.org/10.2139/ssrn.3891347
4. Grothe, B, Pecka, M, McAlpine, D: Mechanisms of sound localization in mammals. Physiological Reviews **90**, 983–1012 (7 2010). https://doi.org/10.1152/physrev.00026.2009
5. el Jundi, B, Warrant, E. J, Byrne, M. J, Khaldy, L, Baird, E, Smolka, J, Dacke, M: Neural coding underlying the cue preference for celestial orientation. In: Proceedings of the National Academy of Sciences **112**, 11395–11400 (9 2015). https://doi.org/10.1073/pnas.1501272112
6. Tollin, D, Yin, T: Sound Localization: Neural Mechanisms, pp. 137–144. Elsevier (2009). https://doi.org/10.1016/B978-008045046-9.00267-9
7. Warrant, E, Dacke, M: Vision and visual navigation in nocturnal insects. Annual Review of Entomology **56**, 239–254 (1 2011). https://doi.org/10.1146/annurev-ento-120709-144852

Part II

Author Index

Author Index