Journal of Telecommunications and the Digital Economy

Volume 11, Number 3
September 2023

Published by
Telecommunications Association Inc.

ISSN 2203-1693

The *Journal of Telecommunications and the Digital Economy* is published by TelSoc four times a year, in March, June, September and December.

Journal of Telecommunications and the Digital Economy

Volume 11, Number 3
September 2023

Table of Contents

Editorial Team

The *Journal* is published by the Telecommunications Association (TelSoc), a not-for-profit society registered as an incorporated association. It is the Australian telecommunication industry's oldest learned society. The *Journal* has been published (with various titles) since 1935.

Editorial

Technological Methods Against Disinformation

Leith H. Campbell
Managing Editor

Abstract: This editorial introduces the September issue and highlights four papers that are concerned with automatically detecting fake news and "phishing" attempts. The discussion is in the context of a recent proposal by the Australian Government to restrict the spread of misinformation and disinformation. It is noted that all the proposed methods use Artificial Intelligence (AI), suggesting that bias, which can be introduced by AI training processes, would be worthy of further research. The other papers are briefly described.

This issue also includes an obituary for Harry Wragge, a former head of Telecom/Telstra Research Laboratories. We also note the passing of John Burke, an influential member of this *Journal*'s Editorial Advisory Board.

Keywords: Editorial, Fake news, AI methods

Approaches to Disinformation

Every week, it seems, we have more news about "phishing" scams, privacy breaches and the malicious spreading of misinformation and disinformation through social media. Fear and uncertainty, in some demographics at least, is holding back the further spread of the digital economy. Application developers and platform providers will find it increasingly more difficult and complex to assure users that their privacy and security are not at risk.

The "regulation" of disinformation and criminality on the Internet has, for decades, been largely in the hands of the big technology companies, particularly Google, Microsoft and Apple, together with the specialized anti-virus providers. Their methods, while widely known, at least in outline, in the tech. community, have not been subject to more general scrutiny.

Given the widespread concern about privacy and security, governments have started to introduce more formal regulation. The European Union's General Data Protection Regulation,

http://doi.org/10.18080/jtde.v11n3.809

introduced from 2016, with amendments in 2018, has had a fundamental impact worldwide on the processing and storage of customer data and the underlying processes of websites. Now, other governments are considering more detailed regulation of content.

In June 2023, the Australian Government issued an exposure draft of proposed new legislation, the *Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023*. It would give the Australian Communications and Media Authority (ACMA) powers to act "if industry efforts in regard to misinformation and disinformation are inadequate". It would enable the ACMA to request that the industry "develop a code of practice covering measures to combat misinformation and disinformation on digital platforms", with the threat that the ACMA could "create and enforce an industry standard (a stronger form of regulation), should a code of practice be deemed ineffective" (Australian Government, 2023a). This approach is similar to the one taken for the issue of news dissemination on digital platforms (Wilding, 2021).

The Government was careful to say that "private" communications, including closed chat groups, would not be covered by the legislation, and that the ACMA would not have the power to demand public content be removed. However, the proposed legislation has excited debate on whether or not it goes too far in restricting freedom of speech. The submissions to the public consultation have been available since 22 September 2023 (Australian Government, 2023b). The concerns of the Australian Human Rights Commission (Finlay, 2023) are typical: the "broad definitions [of misinformation and disinformation] used here risk enabling unpopular or controversial opinions or beliefs to be subjectively labelled as misinformation or disinformation, and censored as a result". Other submissions suggest that trust in government, generally, is low.

The Minister, in her public statements, has put some emphasis on better record-keeping and more effective systems and processes adopted by digital platforms (Minister for Communications, 2023). It is worthwhile to ask how well automated systems will perform in identifying disinformation and fake news. While the systems used by the large digital platforms are proprietary, there is now further academic interest in researching and extending the underlying technology.

In this issue, we publish four papers on the general topic of automatically identifying disinformation or phishing attempts: *Building a Fortress Against Fake News* (Fahad *et al.*, 2023); *Language-Independent Models for COVID-19 Fake News Detection* (Wong *et al.*, 2023); *Phishing Message Detection Based on Keyword Matching* (Tham, Ng & Haw, 2023); and *Improving Phishing Email Detection Using the Hybrid Machine Learning Approach* (Palanichamy & Murti, 2023). These papers offer some insight into suitable methods for

http://doi.org/10.18080/jtde.v11n3.809

detecting disinformation and how well they could be expected to perform. The performance results should be taken as indicative only, given that proprietary systems, developed over many years through several cycles of development, are likely to perform better.

It should be noted that these systems to a greater or lesser extent all depend on Artificial Intelligence (AI) methods. While some uses of AI may turn out to be problematic, this is the benign end of the subject: a collection of data analysis procedures, some predating the "AI" tag, that help to identify features of interest in large datasets. They are now just part of the arsenal of techniques available to researchers. Like all AI, they usually come with no guarantees of providing the best possible solutions; nor should one dismiss the possibility of inherent bias, identified in other applications (Lima, Pisker & Corrêa, 2023), introduced in the "training" of the system. Bias in such systems, given that social and political prejudices may play a part in judgments about what is disinformation, would be a worthy and important topic of further research.

John Burke

It is with sadness that we note the passing of John Burke, who was a member of the Editorial Advisory Board from the time of the *Journal*'s reformation in 2013. His initiative to found and lead what became TelSoc's Broadband Futures Group, based on discussions in the Advisory Board, was of prime importance in supporting and enhancing the *Journal*. A total of 17 related papers have been published since 2019, providing a firm fact base for TelSoc's discussions and advocacy on the need for universal broadband access.

A full obituary for John Burke will be published in a later issue.

Elsewhere in This Issue

In addition to the four papers mentioned above, this issue contains seven other papers on a wide variety of topics.

In the Digital Economy section, we publish three papers. *Proposal of a Measurement Scale and Test of the Impacts on Purchase and Revisit Intention* looks at how to measure the effectiveness of marketing on websites. *ICT-driven Transparency: Empirical Evidence from Selected Asian Countries* provides some good news on how the use of ICT can improved transparency of governments. *Blockchain Technology for Tourism Post COVID-19* proposes some ways of improving the systems for social restrictions and travel in light of experience from the COVID-19 pandemic.

In the Telecommunications section, there are seven papers in total, the first four of which are concerned with fake news and phishing detection. Of the others, the first two, by the same

http://doi.org/10.18080/jtde.v11n3.809

authors, *Big Data Analytics in Tracking COVID-19 Spread Utilizing Google Location Data* and *Utilizing Mobility Tracking to identify Hotspots for Contagious Disease Spread*, describe how to use Google location data to assist in tracking disease spread. The final paper, *Customer Churn Prediction through Attribute Selection Analysis and Support Vector Machine*, explores improved ways for identifying customers in danger of churning away, an important issue for telecommunications and other companies.

In the Biography section, we publish an obituary of Harry Wragge, a former head of Telecom/Telstra Research Laboratories and an influential engineer in shaping the standards and technologies underpinning Australia's telecommunications networks.

As always, we encourage you to consider submitting articles to the *Journal* and we welcome comments and suggestions on which topics or special issues would be of interest. Feedback on the current issue would be welcome.

References

Australian Government. (2023a). Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023—Fact sheet. Department of Infrastructure, Transport, Regional Development, Communications and the Arts, June 2023. Available at https://www.infrastructure.gov.au/sites/default/files/documents/communications-legislation-amendment-combatting-misinformation-and-disinformation-bill-2023-factsheet-june2023.pdf

Australian Government. (2023b). New ACMA powers to combat misinformation and disinformation. Department of Infrastructure, Transport, Regional Development, Communications and the Arts, September 2023. Available at https://www.infrastructure.gov.au/have-your-say/new-acma-powers-combat-misinformation-and-disinformation

Fahad, N., Goh, K. O. M., Hossen, M. I., Tee, C., & Ali, M. A. (2023). Building a Fortress Against Fake News: Harnessing the Power of Subfields in Artificial Intelligence. *Journal of Telecommunications and the Digital Economy*, *11*(3), 68–83. https://doi.org/10.18080/jtde.v11n3.765

Finlay, L. (2023). Why Misinformation Bill risks Freedoms it Aims to Protect. Australian Human Rights Commission. Available at https://humanrights.gov.au/about/news/opinions/why-misinformation-bill-risks-freedoms-it-aims-protect

Lima, R. M. de, Pisker, B., & Corrêa, V. S. (2023). Gender Bias in Artificial Intelligence: A Systematic Review of the Literature. *Journal of Telecommunications and the Digital Economy*, *11*(2), 8–30. https://doi.org/10.18080/jtde.v11n2.690

Minister for Communications. (2023). Interview with Andy Park, ABC RN Drive. Transcript available at https://minister.infrastructure.gov.au/rowland/interview/interview-andy-park-abc-rn-drive

Palanichamy, N., & Murti, Y. S. (2023). Improving Phishing Email Detection Using the Hybrid Machine Learning Approach. *Journal of Telecommunications and the Digital Economy*, *11*(3), 120–142. https://doi.org/10.18080/jtde.v11n3.778

Tham, K.-T., Ng, K.-W., & Haw, S.-C. (2023). Phishing Message Detection Based on Keyword Matching. *Journal of Telecommunications and the Digital Economy*, *11*(3), 105–119. https://doi.org/10.18080/jtde.v11n3.776

Wilding, D. (2021). Regulating News and Disinformation on Digital Platforms: Self-Regulation or Prevarication? *Journal of Telecommunications and the Digital Economy*, *9*(2), 11–46. https://doi.org/10.18080/jtde.v9n2.415

Wong, W. K., Juwono, F. H., Chew, I. M., & Lease, B. A. (2023). Language Independent Models for COVID-19 Fake News Detection: Black Box versus White Box Models. *Journal of Telecommunications and the Digital Economy*, *11*(3), 84–104. https://doi.org/10.18080/jtde.v11n3.789

Proposal of a Measurement Scale and Test of the Impacts on Purchase and Revisit Intention

Salma Ayari
E.S.C., La Manouba University, Tunisia

Imène Ben Yahia
ESC, La Manouba University, Tunisia

Abstract: Online immersion is considered as a determining factor of web surfers' reactions. Its importance may be greater in a 3D-enriched environment. However, little research has explored it in marketing and less has investigated its impact on consumer behaviour in an enriched commercial website. In addition, when it comes to its operationalization, many weaknesses are noticed in the existing literature. Accordingly, the objective of this study is two-fold: in order to test the impact of immersion on purchase and revisit intentions to a 3D-enriched commercial website, a scale measurement of immersion tailored to this specific context is proposed. Following Churchill's framework and the recommendations of Rossiter, a number of methodological instruments, including two focus groups (the first with 4 experts; the second with 18 consumers) and three surveys (first: 140 students; second: 350 Internet users; third: 200 Internet users), are used. The confirmatory factor analysis resulted in an 8-item scale which seems to exhibit evidence of reliability and validity. The predictive validity was confirmed since the impacts of immersion on the intentions to buy and revisit the website are significant. The proposed scale measure may help academics conduct better and more reliable studies on consumer behaviour online.

Keywords: Scale measure, online immersion, merchant website, Internet users, Digital Marketing

Introduction

Much research has been conducted to examine consumers' behaviour while visiting commercial websites. Indeed, this type of website presents the products and services of companies and may lead to a purchase. Attracting visitors and retaining them is therefore the big challenge of the e-commerce industry. Thanks to virtual reality technologies, managers

http://doi.org/10.18080/jtde.v11n3.657

enrich their commercial websites to offer unique and long-lasting experience to visitors. Previous studies, mainly in psychology, have drawn attention to the phenomenon of immersion online. In fact, the Internet stimulates senses (Zhang, Phang & Zhang , 2022; Chen & Lin, 2022; Volle, 2000) and increases immersion, which may lead to a real and unique experience (Carù & Cova, 2006; Holbrook & Hirschman, 1982; Mathwick & Rigdon, 2001) producing specific consumer reactions (Ayari, Ben Yahia & Debabi, 2022).

When it comes to marketing, online immersion is considered as a determining factor of web surfers' reactions (Demangeot & Broderick, 2007; Pentina & Taylor ,2010; Charfi & Volle, 2011; Charfi, 2012). Its importance may be greater in an enriched 3D environment with virtual technologies (Ayari & Ben Yahia, 2023). In fact, the development of immersive virtual reality technologies allows Internet users to live virtual experiences (Schnack, Wright & Elms 2021; Banfi, 2021) by interacting with an artificial environment (real or imaginary) (Coban, Bolat & Goksu, 2022) and exercising cognitive activities (Smith, 2019) impacting their performance (Leung, Hazan & Chan, 2022).

Despite the growing importance of immersion, little research has explored it in marketing (Wang *et al.*, 2021; Volle & Charfi, 2011) and less has investigated its impact on consumer behaviour in a 3D-enriched commercial website (Kowalczuk , Siepmann & Adler, 2021; Banfi, 2021). In fact, directly manipulating a virtual object leads to an easier online immersion of Internet users (Schlosser, 2003; Coban, Bolat & Goksu , 2022; Leung, Hazan & Chan, 2022). Besides, when it comes to its operationalization, many weaknesses are noticed in the existing literature. Thus, its conceptualization, measurement and impacts on consumers' behaviour is not explored enough (Daassi & Debbabi, 2021; Schnack, Wright & Elms , 2021), For instance, some authors consider Immersion as a process and as a state. Others consider it as an experience. Also, so far, authors either use scales of immersion developed in other offline contexts (Rosza *et al.*, 2022) or adapt scales that had not been rigorously developed. For instance, Charfi & Volle (2011) used the scale measure of Fornerino, Helme-Guizon & Gotteland (2008), which was designed within the context of a cinematographic experience, video games and leisure, and does not take into account the specificity of the web. Yet, the context of a website may stimulate reactions and behaviours different from another context; especially since a commercial website enriched by virtual reality technologies is more interactive than other traditional web sites (Antunes & Correia, 2022).

Accordingly, the objective of this study is two-fold: in order to test the impact of immersion on purchase and revisit intentions to the commercial website, a scale measurement of immersion tailored to the context of commercial web sites will be proposed. In fact, as

http://doi.org/10.18080/jtde.v11n3.657

highlighted by previous researchers on the conditions to create a new measurement scale (Kalafatis, Sarpong & Sharif, 1995; Frikha, 2019), two reasons justify our proposal: the existing instruments of immersion are not applicable to the context of 3D-enriched commercial websites and do not take into account certain specificities of the web.

The proposed scale measure may help academicians perform better and more reliable studies on consumer behaviour online. At a managerial level, it may help managers improve traffic on their websites by offering Internet users a unique, rich and effective immersive experience at all levels.

This paper is then structured as follows. First, the concept of immersion is defined and confusion about its conceptualization clarified. Second, the impacts in 3D-enriched websites will be developed. Third, the existing measurements of online immersion will be presented and their weaknesses highlighted. The methodology will later detail the different stages of the development of the measurement scale following Churchill's paradigm (Churchill, 1979) enriched by the recommendations of Rossiter (2002). Finally, results will be exposed and discussed, before concluding with the contributions and limitations of the study.

Defining Online Immersion

The literature on immersion distinguishes between three conceptualizations of immersion: as a process and as a state and as telepresence. In fact, some authors consider immersion as telepresence of Internet users (Grinberg *et al.*, 2014). However, several researchers have stipulated that presence or telepresence is the antecedent of immersion (Hoffman & Novak, 2009; Griffith & Chen, 2004). In addition, according to Carù & Cova (2006), immersion is "the process of accessing the optimal experience called the state of flux". It represents the steps that allow the consumer to reach the experience when interacting with the experiential context. The authors added that it arises from the interaction between a consumer and an enclosed, secure and themed experiential context (Carù & Cova, 2003) during which the individual connects to this context and disconnects from the real world. In other words, it is a "dip" or a gradual entry into the experiential context. In the same line, it also presents as "a strong moment experienced by the consumer and resulting from a partial or complete process of appropriation on his part" (Carù & Cova, 2006, p. 60).

For instance, Fornerino, Helme-Guizon & Gotteland (2008) defined immersion as "the state of intense activity in which the consumer finds himself when he fully accesses the experience". Indeed, it represents the set of reactions manifested by the individual during the experience. These reactions can be cognitive, sensory, affective, social or even physical. In this context, Tamás *et al.* (2022) have studied immersion while using entertainment and digital communication applications. A consumer who strongly immerses in an experiential

http://doi.org/10.18080/jtde.v11n3.657

environment is involved, absorbed and fully engaged (Lombard & Ditton, 1997). He or she forgets the external reality, losing consciousness of what he or she is in the real world in favour of a new self in the experiential context.

Immersion in 3D-enriched Websites

Hyper-real environments have an important role in the process of accessing the experience, making it easier for consumers to immerse themselves in the context and improving the quality of the visit. In effect, virtual reality technologies can provoke emotional reactions in Internet users similar to those caused by physical environments, often leading to immersion.

In the same framework of analysis, virtual environments aim to create positive effects on Internet users, both cognitively and emotionally (Hoffman & Novak, 2009). Immersion in experiential environments therefore leads to behavioural changes (Vézina, 1999). In other words, the intention to buy is a redundant concept in the literature on merchant sites (Poddar *et al.*, 2009). This is a concept at the heart of the concerns of managers. In addition, to understand the conative reactions of Internet users within a merchant website, researchers have often studied the intention to revisit the site (Hausman & Siekpe, 2009).

The escape experienced at the time of the visit prompts the user to return to the site and visit similar sites. Therefore, we formulate the following hypothesis:

H1: Immersion positively influences (a) purchase intention and (b) intention to revisit the site.

Pointing Out the Weaknesses of Existing Scale Measures

To operationalize the concept of online immersion, previous research has opted for an existential phenomenological approach. Data was collected through written accounts or interviews, during which the researcher asks the subject to contextualize a specific experience and to relate it to the first person, step by step, as it was lived (Thompson, Locander & Pollio, 1989). Despite the importance of the concept however, many weaknesses are witnessed in the existing papers operationalizing immersion. For instance, Mathwick & Rigdon (2004, p. 330) proposed a measurement scale of a three-item-dimension, which was later translated by Simon (2007). This scale is not adequate for immersion, because it only takes into account the "escape" factor identified by Mathwick & Rigdon (2004). Also, which is worthy of note, the authors did not proceed to scale purification and empirical validation. As highlighted by Fornerino, Helme-Guizon & Gotteland (2008), Mathwick & Rigdon' scale (2004, p. 330) is not developed properly in line with Churchill's procedures (1979). Besides, the theoretical online immersion items established by Mathwick & Rigdon (2004) are not

http://doi.org/10.18080/jtde.v11n3.657

empirically validated. Furthermore, other items are observed and the scale failed to include important items.

Also, in their research, Grinberg *et al.* (2014) used the Barfield *et al.* (1995) scale to measure immersion. However, this scale was created to measure the presence of Internet users and not to measure their immersion. In addition, it is measured by only one item which has not been used in previous literature. Hudson *et al.* (2019) adapted the scale from Jennett *et al.* (2008) to measure the immersion of Internet users during a virtual reality (VR) underwater seascape exploration. This scale was created in a game context. In addition, it seems that the process of its creation is not rigorous.

Noticing the shortcomings of previous measurement scales of immersion, Fornerino, Helme-Guizon & Gotteland (2008) proposed a new 15-item scale to measure immersion in a cinematographic context. The authors have finally confirmed a unidimensional scale consisting of six items reflecting the cinematographic context. This scale was created to measure immersion in a context different from the context of the web. Yet, Charfi & Volle (2011) used it to study consumer behaviour online. Table 1 exposes the measurement scales and highlights their limitations when applied online.

Table 1. Existing measurement scales of online immersion

Authors	Objective	Scale measuring immersion	Limitations
Mathwick & Rigdon, (2004)	Measure consumer online immersion.	Three items divided into a single dimension called "escape". It was translated by Simon (2007).	This measure reflects the distortion of time that manifests itself in the Internet user, as well as insensitivity to any attentional solicitation outside of the visit experience. They did not proceed to scale purification and empirical validation.
Jennett *et al.* (2008)	Measure consumer online immersion.	Scale is designed within the context of games. Four items measured on a 7-point Likert scale. This scale is adapted by Hudson *et al.* (2019).	A context different from the context of the web. It does not take into account the specificity of the web.
Fornerino, Helme-Guizon & Gotteland (2008)	Measure consumer online immersion.	Scale is designed within the context of a cinematographic experience, video games and leisure. This scale was subsequently used and adapted by Charfi & Volle (2011).	A context different from the context of the web. It does not take into account the specificity of the web.

Research Method

In the following sections, the different steps leading to the development of a scale that may help researchers operationalize immersion while studying web surfers' behaviour will be presented. To this effect, it seems convenient to proceed with a triangulation approach, including both qualitative and quantitative methods. Based on Churchill's recommendations

http://doi.org/10.18080/jtde.v11n3.657

(1979), the procedure is threefold: (a) generating an initial pool of items; (b) data collection and purification of measures; and (c) estimating the scale's validity. The following table presents the steps of Churchill (1979).

Table 2. Application of Churchill's paradigm (1979)

Steps	Studies
1st step: Specify the domain of the construct	• Definition of the concept of online immersion • Qualitative study: 18 consumers questioned about their browsing behaviour on merchant websites (semi-structured individual interviews)
2nd step: Generate a sample of statements	• Drafting of 12 items • Submission to 4 experts • Test content validity This leads to the deletion of two items
3rd step: First data collection	• Data collection: 140 consumers asked about their browsing behaviour in merchant websites • Selection of a 5-point Likert format • Exploratory factor analysis (analysis principal component factorial) This leads to the deletion of two items
4th step: Second data collection	• Data collection: 350 consumers asked about their browsing behaviour • Selection of a 5-point Likert format • Exploratory factor analysis (analysis principal component factorial). The number of items kept is 8
5th step: Purification phase	• Confirmatory factor analysis: PLS 3 • Evaluation of convergent and discriminating validity based on responses from 350 consumers • Testing predictive validity: the effect of online immersion on the intentions of Internet users (200 consumers)

Specifying the domain of the construct and generating statements

The simple recourse to literature seems insufficient to us to define such a controversial and variously treated topic. Indeed, in spite of the plethora of works on this topic, it remains poorly known. Admittedly, this is not a new concept, since the literature dealing with immersion in a commercial website is abundant. Understanding our research more precisely then requires a confrontation in the field of the different interpretations. In addition to the literature review, several steps are in order to generate an initial pool of items. Specifically, two focus groups are held. The first one is held with four experts who are teachers-researchers specialized in online immersion and merchant websites.

Following this experts meeting, at first, we sought to check whether the study of online immersion within a web enriched by virtual reality techniques needs a specific measurement tool. On this latter point, the experts have insisted on the web specificity. One of them has emphasized that 'when in a 3D context, it becomes different'. Another specified that 'a site is a number of links … there are people who talk about sites as if they were people or locations'.

http://doi.org/10.18080/jtde.v11n3.657

Another expert then highlighted that 'we will not find the same concepts, the same items …'. The experts have finally agreed that an online immersion measurement instrument specific to a web context needs to be developed.

The meeting's objective was then to discuss the items of online immersion and the items that could represent each dimension. Another participant added that 'if you log on to a web site, you don't know where you are already, you cannot situate yourself. You don't know what … well … how you can reach such or such objectives'. The experts noted: 'Can we convert, transpose this concept?' The experts have finally agreed that an online immersion measurement instrument specific to a web context needs to be developed.

Then, semi-structured interviews were conducted with 18 Internet users, which constitutes a collection method allowing specific themes to be addressed (Evrard, Pras & Roux, 2009). Internet users have been asked to browse the matterport.com website. Eighteen people agreed to participate in this study. To find out more about our sample, we asked for their gender, age category, profession and income. We thus met 10 men (55.6%) and 8 women (44.4%). The people are between 25 and 44 years old; the majority are in the age category between 35 and 44, in this case 10 people, and 8 people are in the age category between 25 and 34 years old. Theoretical saturation is reached after the 18th interview. After welcoming and thanking the participants for their collaborations, they were exposed to the site in question and called upon to explore it. Then, the topics of the interview guide were discussed. The site selected for the study mobilizes virtual reality devices (virtual agents and 3D environments).

Respondents described their concentration during navigation and their implications in the offer: *"I was concerned and focused all the time*"; "*Avatars speak directly to us about what makes us more involved"; "It's like I'm at the agency; I think it's a real success … "; "The conversation between the two virtual agents allows us to listen to the message and stay focused …"; "I was focused with what the avatars were saying, I didn't see the time pass"*. In some cases, when the user integrates the experiential components of the site, he or she is engaging in the experience. This immersive experience leads to a disconnection from the real world, similar to the feeling of being present in the virtual environment. This is justified by the following verbatims:

> *"As absorbed by the site…. I was very comfortable"; "I was at the heart of the site, I participated fully, I gave my choices …"; "I forgot the people around me … the real world"; "I would like to buy a house"; "I was curious, intrigued, I wanted to find out what was going on"; I was waiting for the rest, I was listening to the presenters"*.

http://doi.org/10.18080/jtde.v11n3.657

Thus, the results identify the main levels of immersion. According to the responses, there are people who are drawn to the experience. These individuals have expressed their focus and involvement. Others are more committed. They are so out of touch with the real world that they lose their spatio-temporal landmarks. We can therefore conclude that the immersion is explained by the implication, the concentration and the commitment. The majority of the respondents admitted their curiosity, concentration and disconnection. According to the results, we note that the qualitative study allowed us to identify two new items of immersion (I was curious… I wanted to find out what they wanted to present to me, I focused, disconnected, curious involved).

Following the recommendations of Rossiter (2002), several items were formulated. Afterwards, a qualitative pretest was conducted by submitting the statements to four experts. The experts gave their opinion on the clarity of the proposals. Two proposals were deleted, as they were deemed too redundant or not applicable to our construct. Some proposals have also been reformulated. Indeed, the qualitative study allowed us to specify the domains of the constructs and to produce an initial list of items. Thereby, the review of the literature and the qualitative study allowed us to identify 10 items of online immersion.

Exploratory analysis

Table 3. Factor analysis and reliability test results

Item	Quality of representation	Factor loading
The site created a new world which suddenly disappeared at the end of the visit	0.863	0.923
At times, I lost consciousness of my surroundings	0.876	0.932
During the visit, my body was in front of the screen, but my mind was in the world created by the site	0.858	0.915
The site made me forget the realities of the outside world	0.887	0.939
While viewing the site, what happened before the visit or what would happen afterwards no longer mattered	0.857	0.912
Visiting the site made me forget my immediate surroundings	0.874	0.928
I was curious… I wanted to find out what they wanted to present to me	0.814	0.898
I focused, disconnected, curious involved	0.756	0.869
I felt detached from the outside world	0.301	0.411
I felt completely immersed	0.402	0.422
Eigenvalue % of variance explained Cronbach's alpha (standardized) Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy	6.866 78.735 0.902 0.898	

In the previous section, a set of items representing different items of online immersion adequate for a web surfing context are outlined. To ensure the reliability of these items, a website (https://www.darellamma.com/darellamma), which is likely to generate the

http://doi.org/10.18080/jtde.v11n3.657

immersion of Internet users, is chosen. In fact, directly manipulating a virtual object leads to an easier online immersion of Internet users (Schlosser, 2003; Coban, Bolat & Goksu, 2022; Leung, Hazan & Chan, 2022). Within laboratory conditions, 140 Tunisian business students are invited to surf this web site, which offers stays in a guest house, during a 15-min period. Once they finish, the participants are administered a questionnaire consisting of the previously generated items and are asked to rate the items on a 5-point Likert scale ranging from 1 ('strongly disagree') to 5 ('strongly agree'). The collected data are processed by a principal components factor analysis and a reliability test.

These preliminary results seem to be interesting in that the generated items are often represented in the literature. Besides, six items have been debated by several authors (Charfi & Volle, 2011; Fornerino, Helme-Guizon & Gotteland, 2008) who supported the idea that these items represent a unique one dimension rather than separate dimensions. However, the last two items of the qualitative study have been deleted. Therefore, eight items are retained.

Test and confirmatory analysis of the scale

A new data collection phase is undertaken to attest to the validity of the scale. To this effect, the scale is tested again using the previously mentioned website. In fact, the previous purification stage was conducted for Internet users within laboratory conditions. In this phase, the objective is to check for the psychometric quality of items. For the choice of this website, eight different site links have been created, which the surfers have no knowledge of and are likely to immerse in the context. The choice of unknown websites is motivated by the concern to neutralize any familiarity effects with the website. Then, a sample of individuals is asked to navigate the websites and rate them on their potential to immerse. Then, these web sites are classified and the one which seems to mostly generate immersions is chosen (https://pyntopyn.com/DarEllamma/?i=3). The scale has been tested over a heterogeneous population of web surfers and in off-laboratory conditions. Consistent with previous research (Fornerino, Helme-Guizon & Gotteland, 2008), participants are asked to surf on the website for 15 minutes and to respond to a questionnaire. A pool of 360 questionnaires is collected. Nevertheless, some of them are eliminated because the participants knew the website (a question was included for the purpose). Therefore, 350 questionnaires are retained; 52 % of the respondents are female and 48% are men. A majority of these respondents have university education and belong to different socioeconomic categories. Following Gerbing & Anderson's (1988) recommendations, an exploratory factor analysis (EFA) is performed before conducting a confirmatory factor analysis (CFA). To carry out an EFA, we must verify the factorization conditions of the measurement scale relating to immersion. It is worth

http://doi.org/10.18080/jtde.v11n3.657

noting that the obtained results during this second phase corroborate those collected during phase 1 in two major aspects. First, the KMO index (0.962) shows a value greater than 0.5 and Bartlett's test is significant. Furthermore, we performed an EFA on the eight items of the measurement scale. This solution explains 85.875% of the variance. Thus, we have found that all the factor contributions of the immersion variable vary between 0.872 and 0.947, so all the items of this measurement scale have been kept. The results of the EFA are presented in the table.

Table 4. Purification test of the Immersion variable

Item	Quality of representation	Factor loading
The site created a new world which suddenly disappeared at the end of the visit	0.872	0.934
At times, I lost consciousness of my surroundings	0.887	0.942
During the visit, my body was in front of the screen, but my mind was in the world created by the site	0.873	0.934
The site made me forget the realities of the outside world	0.897	0.947
While viewing the site, what happened before the visit or what would happen afterwards no longer mattered.	0.861	0.928
Visiting the site made me forget my immediate surroundings	0.897	0.947
I was curious... I wanted to find out what they wanted to present to me	0.824	0.908
I focused, disconnected, curious involved	0.760	0.872
Eigenvalue % of variance explained Cronbach's alpha (standardized) Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy	6.870 85.875 0.976 0.962	

Second, after having verified the unidimensionality, we move on to the internal consistency of the selected items. Cronbach's alpha for each dimension must be greater than 0.6. For the dimension chosen for immersion, the value of Cronbach's alpha, 0.976, is therefore close to 1, which reflects good internal consistency of the items. Principal component factor analysis of the eight retained items shows the one-dimensional character of the scale. These factors together explain 85.875% of the total variance. The contributions of the items to immersion are all significant, since their values are greater than 0.5. The EFA made it possible to retain only one dimension concerning this measurement scale.

From these results we can conclude that Bartlett's sphericity test (p = 0.000) makes it possible to safely reject the hypothesis of nullity of the correlation coefficients. The KMO index = 0.962 explains a level of appreciation judged to be meritorious by Kaiser, Meyer and Olkin.

Following this second purification phase, a CFA is applied using Smart PLS 3. The indicators for the evaluation of the quality of the measurement can be grouped according to Roussel *et*

http://doi.org/10.18080/jtde.v11n3.657

al. (2002) into three different categories. According to Roussel *et al.* (2002) and given the large number of indices, it is advisable to retain 4 indices. Table 5 summarizes the thresholds of the various indices according to Roussel *et al.* (2002).

Table 5. Key values of the adjustment indices used

CR: Composite Reliability	> 0.7	Nunnally & Bernstein (1994)
AVE: Average Variance Extracted	> 0.5	Fornell & Larcker (1981)
T-value	> 1.96	Hensler *et al.* (2009)
Cronbach's Alpha	> 0.7	Nunnally & Bernstein (1994)

By calculating the PLS Algorithm on our entire sample (350 respondents), we check the convergent validity for each of the constructs. Thus, convergent validity is examined by calculating the Composite reliability index (CR), Cronbach's alpha index, and the Average Variance Extracted, AVE. The acceptability thresholds required for the measurement criteria are shown in Table 6.

Table 6. Convergent validity criteria

	Cronbach's Alpha	**Composite Reliability (CR)**	**Average Variance Extracted (AVE)**
Online immersion	0.976	0.980	0.859

From Table 6, the composite reliability (CR) exceeds the required threshold of 0.7 (Chin, Peterson & Brown, 2008), and the AVE (shared mean variance) exceeds the required threshold of 0.5 (Fornell & Larcker, 1981). Consequently, the convergent validity of our model is thus assured, especially since discriminant validity is assessed by examining the factorial contributions (loadings) of the items to their respective constructs.

We have checked, in particular, if, for each construct, the factor contributions are greater than the cross-factor contributions between each item and the other constructs. Thus, the discriminant validity is assured, because our construct has factorial contributions which are greater than the cross-factorial contributions. The discriminant validity is also evaluated according to Fornell & Larker (1981) by checking that the square root of the AVE for each construct exceeds the inter-construct correlations concerning it.

Predictive validity of the scale: mediating role of immersion between virtual technologies and loyalty

As a follow-up to these results, we proceeded to estimate the structural relationships between immersion and users' intent to revisit and purchase. To this end, purchase intention was measured using the Yoo & Donthu (2001) scale, while the intention to revisit the site was measured using the Demangeot & Broderick (2007) scale. These scales have been shown to have excellent psychometric qualities (Yoo &Donthu, 2001; Demangeot & Broderick, 2007).

http://doi.org/10.18080/jtde.v11n3.657

For the scale measuring purchase intention, it turns out that only one factor has an eigenvalue greater than one. The items therefore all seem to be strongly linked to a factor. They are very well represented (quality of representation> 0.8). Bartlett's test is significant (p = 0.000) and the KMO index reaches a value of 0.875. The data are therefore factorizable. Likewise, the scale exhibits excellent internal consistency (Cronbach's alpha = 0.970).

Factor analysis performed on the intention-to-revisit scale revealed that it is a one-dimensional scale. We are led to retain a single factor explaining 89.296% of the variance. In addition, the items are well represented (> 0.8) and linked to the factor (> 0.8). Bartlett's test records a significant value (p = 0.000) and the KMO is 0.917. This proves that the factor analysis is of good quality. This scale is also reliable (Cronbach's alpha = 0.970).

We can currently estimate the structural relationships between constructs. The correlation relationships between the constructs are estimated by examining the standardized correlation coefficients (path-coefficients) and the statistical T-values (obtained following the Bootstrapping analysis), which express the degree of significance of the correlations. A correlation relationship is significant if the statistical T value (or student's *t*) is greater than the threshold of 1.96 (if p <0.05). The positive correlation coefficients that are close to 1 assume a strong correlation link between the constructs.

Table 7. Search model result

Correlation relationship	Correlation coefficients (standardized)	T-statistic	P value
Immersion -> purchase intent	-0.402	4.858	0.000
Immersion -> intention to revisit the site	0.430	5.687	0.000

Table 8. Regression results on PLS

	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistic (|O/STDEV|)	P Value
Immersion -> purchase intent	-0.402	-0.405	0.083	4.858	0.000
Immersion -> intention to revisit the site	0.430	0.426	0.076	5.687	0.000
Online immersion		**R Squared**		**R Squared Adjusted**	
		0.914		0.912	

Two hundred Internet users were selected (part of the 350 consumers questioned during the second data collection). The results of the analysis show that there is a significant link between immersion and intention (intention to buy: t = 4.858; p = 0.000; intention to revisit

http://doi.org/10.18080/jtde.v11n3.657

the site: t = 5.687; p = 0.000). Examining the relationship between the different variables, we found that web-driven online immersion, measured by our scale, has a significant impact on some surfers' responses. The obtained results confirm the predictive validity of the scale and confirm **H1**.

Discussion and Conclusion

As a consequence of the shortcomings of the literature review regarding a valid scale measure of online immersion, this research develops, first, a scale measure of it following the Churchill paradigm enriched with Rossiter's (2002) recommendations, before testing the impacts on the revisit and purchase intentions on a 3D-enriched commercial website. The study results in a unidimensional scale represented by 8 items. Unlike previous research, this scale is adjusted to the context of commercial websites, especially enriched ones, and is developed rigorously following the Churchill paradigm as a threefold process: (a) generating an initial pool of items; (b) data collection and purification of measures; and (c) estimating the scale's validity. Similar to previous research (Charfi & Volle, 2011; Fornerino Helme-Guizon & Gotteland 2008), our results confirm the uni-dimensionality of the scale. In addition, they confirm that immersion fosters online consumer behaviour. It is determining in enriched commercial websites.

At a methodological level, several phases were followed to develop a valid and reliable measurement scale to accurately operationalize the concept of online immersion. In addition, the development of this measurement scale and the assessment of its validity used a triangulation of several data collection methods and a review of the literature to cross-check the relevance of items generated with respect to previous research on the subject.

At a theoretical level, the developed scale measure may support the digital marketing field by helping researchers conduct valid studies on online consumer behaviour. In past researches, immersion in video games and movies has been examined but less immersion in commercial websites. The proposed scale may help academics to conduct better and more reliable studies on online consumer behaviour.

At a managerial level, companies need to evaluate the performance of their websites. They may use the current scale measure to assess the capacity of their website to stimulation immersion, since it is a determinant of net-surfers' retention. It may also help administrators improve their website traffic by providing a unique, rich and effective experience at all levels.

This study has some limitations as well. First, our empirical tests are exclusively conducted on a sample of Tunisian web surfers who are not necessarily representative of the world's Internet users. Future research may test the scale in other contexts and on other categories

http://doi.org/10.18080/jtde.v11n3.657

of social presence online, like social media. Second, developing a valid and a reliable measurement scale is a long and an ongoing process. Accordingly, we recommend the use of this measurement scale for future online immersion-based research so as to further test its validity. Third, this scale is tailored to a 3D-enriched commercial website. However, visual social media platforms, like Instagram, may stimulate immersion and may lead to purchase (Ben Yahia *et al.*, 2018). Future research may investigate immersion on Instagram considering the characteristics of the platform, the number and quality of the interactions, and also the social presence.

References

Antunes, R. F., & Correia, L. (2022). Virtual simulations of ancient sites inhabited by autonomous characters: Lessons from the development of easy-population. *Digital Applications in Archaeology and Cultural Heritage*, *26*, e00237. https://doi.org/10.1016/j.daach.2022.e00237

Ayari, S., & Ben Yahia, I. (2023). Impacts of immersion on loyalty to guesthouse websites: the simultaneous effect of 3D decor and avatars in a hyper-real environment. *Journal of Marketing Communications*, 1–16. https://doi.org/10.1080/13527266.2023.2193824

Ayari, S., Ben Yahia, I., & Debabi, M. (2022). Measuring e-browsing behaviour and testing its impact on online immersion. *Journal of Telecommunications and the Digital Economy*, *10*(2), 111–125. http://doi.org/10.18080/jtde.v10n2.546

Banfi, F. (2021). The evolution of interactivity, immersion and interoperability in HBIM: Digital model uses, VR and AR for built cultural heritage. *ISPRS International Journal of Geo-Information*, *10*(10), 685. https://doi.org/10.3390/ijgi10100685

Barfield, W., Zeltzer, D., Sheridan, T., & Slater, M. (1995). Presence and performance within virtual environments. In Barfield, W., & Furness, T. A. (eds), *Virtual environments and advanced interface design*, New York: Oxford University Press, pp. 473–513.

Carù, A., & Cova, B. (2003). Approche empirique de l'immersion dans l'expérience de consommation: les opérations d'appropriation. *Recherche et Applications en Marketing (French Edition)*, *18*(2), 47–65. https://doi.org/10.1177/076737010301800203

Carù, A., & Cova, B. (2006). Expériences de marque: comment favoriser l'immersion du consommateur? *Décisions Marketing*, *41*, 43–52.

Charfi, A. A., & Volle, P. (2011). L'expérience d'immersion en ligne: un nouvel outil pour les sites marchands. *Revue Française du Marketing*, *234/235*, 49–65.

Charfi, A. A. (2012). L'immersion en ligne: l'effet conjoint de l'agent virtuel et des environnements en 3D. *Actes de la 11ème Journée du e marketing, Paris* la Sorbonne, 7 Septembre

Chen, Y., & Lin, C. A. (2022). Consumer Decision-Making in an Augmented Reality Environment: Exploring the Effects of Flow via Augmented Realism and Technology

Fluidity. *Telematics and Informatics*, *71*, 101833. https://doi.org/10.1016/j.tele.2022.101833

Chin, W. W., Peterson, R. A., & Brown, S. P. (2008). Structural equation modeling in marketing: Some practical reminders. *Journal of marketing theory and practice*, *16*(4), 287–298. https://doi.org/10.2753/MTP1069-6679160402

Churchill Jr, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of marketing research*, *16*(1), 64–73. https://doi.org/10.1177/002224377901600110

Coban, M., Bolat, Y. I., & Goksu, I. (2022). The potential of immersive virtual reality to enhance learning: A meta-analysis. *Educational Research Review*, *36*, 100452. https://doi.org/10.1016/j.edurev.2022.100452

Daassi, M., & Debbabi, S. (2021). Intention to reuse AR-based apps: The combined role of the sense of immersion, product presence and perceived realism. *Information & Management*, *58*(4), 103453. https://doi.org/10.1016/j.im.2021.103453

Demangeot, C., & Broderick, A. J. (2007). Conceptualising consumer behaviour in online shopping environments. *International journal of retail & distribution management*. *35*(11), 878–894. https://doi.org/10.1108/09590550710828218

Evrard, Y., Pras, B., & Roux, E., (2009). *Market – Etudes et recherches en marketing*, 4ième edition. Dunod, Paris.

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research*, *18*(3), 382–388. https://doi.org/10.1177/002224378101800313

Fornerino, M., Helme-Guizon, A., & Gotteland, D. (2008). Expériences cinématographiques en état d'immersion: effets sur la satisfaction. *Recherche et Applications en Marketing (French Edition)*, *23*(3), 95–113. https://doi.org/10.1177/076737010802300304

Frikha, A. (2019). *La mesure en marketing: Opérationnalisation des construits latents*. ISTE Group, London, UK.

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of marketing research*, *25*(2), 186–192. https://doi.org/10.1177/002224378802500207

Griffith, D. A., & Chen, Q. (2004). The influence of virtual direct experience (VDE) on on-line ad message effectiveness. *Journal of Advertising*, *33*(1), 55. https://doi.org/10.1080/00913367.2004.10639153

Grinberg, A. M., Careaga, J. S., Mehl, M. R., & O'Connor, M. F. (2014). Social engagement and user immersion in a socially based virtual world. *Computers in Human Behavior*, *36*, 479–486. https://doi.org/10.1016/j.chb.2014.04.008

Hausman, A. V., & Siekpe, J. S. (2009). The effect of web interface features on consumer online purchase intentions. *Journal of business research*, *62*(1), 5–13. https://doi.org/10.1016/j.jbusres.2008.01.018

Holbrook, M. B., & Hirschman, E. C. (1982). The experiential aspects of consumption: Consumer fantasies, feelings, and fun. *Journal of consumer research*, *9*(2), 132–140.https://doi.org/10.1086/208906

Hoffman, D. L., & Novak, T. P. (2009). Flow online: lessons learned and future prospects. *Journal of interactive marketing*, *23*(1), 23–34. https://doi.org/10.1016/j.intmar.2008.10.003

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. In *New challenges to international marketing*. Emerald Group Publishing Limited. https://doi.org/10.1108/S1474-7979(2009)0000020014

Hudson, S., Matson-Barkat, S., Pallamin, N., & Jegou, G. (2019). With or without you? Interaction and immersion in a virtual reality experience. *Journal of Business Research*, *100*, 459–468. https://doi.org/10.1016/j.jbusres.2018.10.062

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International journal of human-computer studies*, *66*(9), 641–661. https://doi.org/10.1016/j.ijhcs.2008.04.004

Kalafatis, S. P., Sarpong Jr, S., & Sharif, K. J. (2005). An Examination of the Stability of Operationalisations of Multi-Item Marketing Scales: A look at the evidence for the usefulness, reliability and validity of projective techniques in market research. *International Journal of Market Research*, *47*(3), 255–266. https://doi.org/10.1177/147078530504700301

Kowalczuk, P., Siepmann, C., & Adler, J. (2021). Cognitive, affective, and behavioral consumer responses to augmented reality in e-commerce: A comparative study. *Journal of Business Research*, *124*, 357–373. https://doi.org/10.1016/j.jbusres.2020.10.050

Leung, G., Hazan, H., & Chan, C. S. (2022). Exposure to nature in immersive virtual reality increases connectedness to nature among people with low nature affinity. *Journal of Environmental Psychology*, *83*, 101863. https://doi.org/10.1016/j.jenvp.2022.101863

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of computer-mediated communication*, *3*(2), JCMC321. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Mathwick, C., Malhotra, N., & Rigdon, E. (2001). Experiential value: conceptualization, measurement and application in the catalog and Internet shopping environment. *Journal of retailing*, *77*(1), 39–56. https://doi.org/10.1016/S0022-4359(00)00045-2

Mathwick, C., & Rigdon, E. (2004). Play, flow, and the online search experience. *Journal of consumer research*, *31*(2), 324–332. https://doi.org/10.1086/422111

Nunnally, J., & Bernstein, I., (1994) *Psychometric Theory*. New York: McGraw Hill, 3éme edition.

Pentina, I., & Taylor, D. G. (2010). Exploring source effects for online sales outcomes: the role of avatar-buyer similarity. *Journal of Customer Behaviour*, *9*(2), 135–150. https://doi.org/10.1362/147539210X511344

Poddar, A., Donthu, N., & Wei, Y. (2009). Web site customer orientations, Web site quality, and purchase intentions: The role of Web site personality. *Journal of Business research*, *62*(4), 441–450. https://doi.org/10.1016/j.jbusres.2008.01.036

Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International journal of research in marketing*, *19*(4), 305–335. https://doi.org/10.1016/S0167-8116(02)00097-6

Rózsa, S., Hargitai, R., Láng, A., Osváth, A., Hupuczi, E., Tamás, I., & Kállai, J. (2022). Measuring Immersion, Involvement, and Attention Focusing Tendencies in the Mediated Environment: The Applicability of the Immersive Tendencies Questionnaire. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.931955

Roussel, P., Durrieu, F., Campoy, E. & El Akremi, A. (2002). *Méthodes d'équations structurelles: recherche et applications en gestion*. Paris, Economica.

Schlosser, A. E. (2003). Experiencing products in the virtual world: The role of goal and imagery in influencing attitudes versus purchase intentions. *Journal of consumer research*, *30*(2), 184–198. https://doi.org/10.1086/376807

Schnack, A., Wright, M. J., & Elms, J. (2021). Investigating the impact of shopper personality on behaviour in immersive Virtual Reality store environments. *Journal of Retailing and Consumer Services*, *61*, 102581. https://doi.org/10.1016/j.jretconser.2021.102581

Simon, F. (2007). Les composantes de l'expérience virtuelle de recherche d'information: imagination, plaisir et immersion. *Actes de la 12ieme Journées de la Recherche en Marketing de Bourgogne*, 1–20.

Smith, S. A. (2019). Virtual reality in episodic memory research: A review. *Psychonomic bulletin & review*, *26*(4), 1213–1237. https://doi.org/10.3758/s13423-019-01605-w

Tamás, I., Rózsa, S., Hargitai, R., Hartung, I., Osváth, A., & Kállai, J. (2022). Factors influencing schizotypal personality trait-dependent immersion and digital media usage: Adaptation and maladaptation. *Acta Psychologica*, *230*, 103735. https://doi.org/10.1016/j.actpsy.2022.103735

Thompson, C. J., Locander, W. B., & Pollio, H. R. (1989). Putting consumer experience back into consumer research: The philosophy and method of existential-phenomenology. *Journal of consumer research*, *16*(2), 133–146. https://doi.org/10.1086/209203

Vézina, R. (1999). Pour comprendre et analyser l'expérience du consommateur. *GESTION-MONTREAL*, *24*(2), 59–65.

Volle, P., & Charfi, A. A. (2011). Valeur perçue et comportements en ligne en état d'immersion: le rôle modérateur de l'implication et de l'expertise. In *27ème Congrès International de l'Association Française du Marketing, Bruxelles, Belgique* (pp. 1–25).

Volle, P. (2000). Du marketing des points de vente à celui des sites marchands: spécificités, opportunités et questions de recherche. *Revue Française du Marketing*, *177/178*, 83–100.

Wang, Q. J., Escobar, F. B., Da Mota, P. A., & Velasco, C. (2021). Getting started with virtual reality for sensory and consumer science: Current practices and future perspectives. *Food Research International*, *145*, 110410. https://doi.org/10.1016/j.foodres.2021.110410

Yahia, I. B., Al-Neama, N., & Kerbache, L. (2018). Investigating the drivers for social commerce in social media platforms: Importance of trust, social support and the platform perceived usage. *Journal of Retailing and Consumer Services*, *41*, 11–19. https://doi.org/10.1016/j.jretconser.2017.10.021

Yoo, B., & Donthu, N. (2001). Developing and validating a multidimensional consumer-based brand equity scale. *Journal of business research*, *52*(1), 1–14. https://doi.org/10.1016/S0148-2963(99)00098-3

Zhang, Q., Phang, C. W., & Zhang, C. (2022). Does the internet help governments contain the COVID-19 pandemic? Multi-country evidence from online human behaviour. *Government information quarterly*, *39*(4) 101749. https://doi.org/10.1016/j.giq.2022.101749

ICT-driven Transparency: Empirical Evidence from Selected Asian Countries

Ajmal Hussain
University of the Punjab (PU)

Abstract: We are living in the digital era and ICTs have become necessities in this contemporary world. The aim of this study is to investigate transparency in Asia through ICT's diffusion by using Driscoll-Kraay standard error technique. We used panel data for 17 Asian countries from 2010 to 2019 and we use control of corruption as a proxy for transparency checking. The results show that ICTs leave a positive effect on the control of corruption. Other determinants of transparency in this paper, such as political stability and effective governance, have a positive effect on control of corruption. ICT policy can play an important role in curbing corruption. So, there is a strong need for ICT diffusion, suggesting that effective governance helped to reduce corruption in the Asian region and establish a surveillance-based system in public institutions.

Keywords: ICTs, Transparency, Control of corruption, Governance, Political stability

Introduction

Transparency is an indispensable cornerstone of accountable governance. A diminishment of corruption within state or local spheres augments the imperative for enacting the tenets of efficacy and lucidity in the redistribution of social funds. However, corruption, a pervasive and endemic malaise, continues to afflict numerous Asian countries, impeding their socio-economic development and undermining public trust in governance (Adam & Fazekas, 2021). This entrenched venality, often manifested through bribery, extortion, collusion, fraud, embezzlement, misappropriation, trading influence, illegal enrichment, obstruction of justice, abuse of position, and money laundering, engenders a vicious cycle of unequal resource allocation, hindering equitable development. Therefore, corruption is a threat to economic progress in developing countries (Fisman & Svensson, 2007; Welsch, 2008). That is why it is very harmful to society, because it may generate poverty, decrease the availability of money, destroy the trust in government, and reduce economic growth.

http://doi.org/10.18080/jtde.v1n3.658

Nowadays, digital technologies are frequently perceived as a means to enhance trust, transparency, and accountability through the dissemination of information to the public. The dissemination of information can be done easily in the present world by Information and Communication Technology (ICT), such as computers, smart-phone, smart gadgets, fast Internet and networking connections (Maiti *et al.*, 2020). It brings individuals' and countries' interactions closer (Castellacci, 2006). Thus, this world is presently known as the world of ICT: it has turned the whole world into a global village. Innovations speed up automation through Artificial Intelligence (AI), reduce the level of difficulty in work, cut down the information gap among agents, decrease distance barriers, save time, spread knowledge easily, help in the governance system, improve transparency, empower the individual capabilities, and so on (Castellacci & Tveito, 2016). Moreover, ICT has a strong connection with all domains of life and can be used as a tool to solve problems. It is playing a vital role in the overall development of the world. The impact of ICT is very complicated from the past, both in depth and spread, due to modern innovations (Crafts, 2004; Sala-I-Martín *et al.*, 2012).

A common theme of different research strands shows the relationship among economic indicators but isolates other aspects of life. ICT diffusion does not only affect economic indicators, but also influences governance, accountability, transparency, and institutional policies. ICT tools that facilitate the gathering, storage, and processing of data can significantly contribute to the detection, prevention, and prosecution of corrupt practices. Some Asian countries, such as India, Bangladesh, and Pakistan, have shown a high trend of corruption with huge ICT diffusion in the last two decades, as the diffusion of ICT is not equally applicable to all countries for showing the impact on corruption. According to the report of Transparency International (2021), Bangladesh, China, India, and Pakistan are ranked low in transparency. The latest Global Corruption Barometer (GCB) survey (2020) shows that 74% of Asian citizens believe that corruption is a serious issue in their region and one out of five bribes public servants. Unfortunately, recent studies have shown that ICT has not been utilised effectively against corruption in Asian countries. This leads to a gradual increase in corruption in these economies. Corruption is a problem that has negative effects not only on Asian economies, but on other nations as well. Additionally, Transparency International Report (2021) shows that the Asian region stands second in terms of corruption. The system of these countries will be damaged if this situation continues for a long period.

Meanwhile, recent studies suggest that ICT is an important tool to curb corruption in Asia (Liu *et al.*, 2021; Suardi, 2021) and developed and developing economies (Bhattacherjee & Shrivastava, 2018; Gouvea *et al.*, 2022; Mouna *et al.*, 2020), such as ASEAN (Darusalam *et al.*, 2021; Hartani *et al.*, 2020), European Union (Androniceanu *et al.*, 2021), and Africa (Kouladoum, 2022). ICT is a tool that can help in the investigation against corruption through

http://doi.org/10.18080/jtde.v11n3.658

four distinct channels: generating electronic records; cyber investigation; whistleblowers; and institutional collaboration. Dirienzo *et al.* (2007) investigated whether the increase in access to ICT and the exchange of information make it easy to provide checks and balances for public officials and to check the effectiveness of government. Corruption has become risky under the shadow of ICT. This advances higher transparency and effective governance and decreases the menace of corruption (Adam, 2020). According to the International Telecommunication Union (ITU) Report (2021), the mobile markets in Asia are growing tremendously, which reveals a rising trend in Internet use. South Asian countries—India (84.3), and Pakistan (76.4)—have subscription rates per 100 people that are higher than those of the other Asian nations. Furthermore, the World Bank Report (2020) also argues that the move towards a digital government and revolutionary advancements in technology present opportunities and risks for the fight against corruption. Developing nations might develop and spread the use of novel technological innovations to solve problems in the public sector. The cost of corruption is lower in economies that have switched from a natural resource-based to a digital and innovation-driven economy. These initiatives have been effective in different countries. Some examples of digital platforms and tools raising people's concerns, gathering data, having an impact, and aiding in the fight against corruption are the OPEN systems in South Korea, the JAGA app in Indonesia, IPaidABribe in India, WhatDoTheyKnow in the UK, and K-Monitor in Hungary. Therefore, there is a dire need to examine the linkage between ICT and control of corruption using a robust technique.

This research contributes to the existing knowledge in multiple aspects. The empirical study of the impacts of ICT on corruption is in its infancy due to inadequate measurement (Žuffová, 2020). This topic is still debatable due to its inconclusive results in Asian regions — such as no relationship (Mouna *et al.*, 2020), negative (Hartani *et al.*, 2020), positive (Sassi & Ben Ali, 2017; Suardi, 2021), and U-shaped (Darusalam *et al.*, 2021) — whereas fewer studies used a composite index of ICT (Darusalam *et al.*, 2021; Kouladoum, 2022). In order to fill this gap in the literature, this study examines the impact of a composite index of ICT (fixed broadband, fixed phones, mobile phones, and Internet users) on corruption by analysing panel data of 17 Asian countries from 2010 to 2019 by employing Driscoll-Kraay standard error technique.

The rest of the paper is devised in the following way. The subsequent section describes the literature review on ICT and control of corruption. In the third section, this paper explains theoretical considerations, and the following section deals with methodology. The final section includes concluding remarks.

http://doi.org/10.18080/jtde.v11n3.658

Literature Review

The literature has attempted to furnish the effect of ICT diffusion to make the world transparent, but the results are mixed and inconclusive in some studies. Empirical analysis of ICT development on corruption is very rare, due to a lack of measurements (Žuffová, 2020). Transparency is the key and crucial aspect of ICT adoption. These opportunities are now possible as a result of the rapid adoption of ICT across all economic relations. Corruption is reducing in developed countries with ICT diffusion, while the same trend is doubtful in the case of developing countries (Mahmood, 2004), and supported by the findings of Heeks (1998) and Wescott (2001). Heeks (1998) observed that, although ICT can occasionally deter corruption, it does not significantly reduce it. Investments in ICT infrastructure are frequently ineffectual at reducing corruption. This is due to the fact that ICT diffusion can give rise to "upskilling" of corruption and decrease competition for upskilled corrupt public officials and servants. Some of these workers possibly had access to private data, which they could use for their own benefit (Wescott, 2001). In the same vein, Sturges (2004) conducted a study in order to ascertain the impact of ICT on the widespread corruption of politicians, governments, higher administration, and the private business sector. He has also shown in his mixed and ambiguous findings that it is challenging to employ ICT for the benefit of the poor, making the needy and the poor the victims of corruption. Moreover, Vasudevan (2006) investigated whether or not ICT for development is helpful to reduce corruption and found mixed results.

Additionally, recent studies have also raised doubts regarding ICT's actual effectiveness in curbing corruption (Charoensukmongkol & Moqbel, 2014; Garcia-Murillo, 2013) due to the implementation of intra- and inter-institutional flows that ensure that only individuals with permission can access those data and information. Charoensukmongkol & Moqbel (2014) observed that ICT investment does, to some extent, minimise corruption, but excessive ICT expenditure may open up new opportunities for misconduct and corruption. Garcia-Murillo (2013) questioned the effectiveness of ICT in reducing corruption by claiming that such systems are frequently used to give the electorate a favourable impression of government operations in order to win re-election. He also claimed that these systems frequently have no impact on corruption outcomes because they are not accompanied by any significant process or role changes in the corrupt system that would not benefit the people in power.

Another strand of literature supporting ICT as a tool for reducing corruption also came into existence. The reason is that the ICT environment has shifted from being dominated by specialised systems to one that now comprises widely adopted and compatible solutions. Moreover, ICT infrastructure is undergoing rapid and pronounced development and it is helpful to make government effective and reduce corruption (Poliak *et al.*, 2020; Russell,

http://doi.org/10.18080/jtde.v11n3.658

2020). Thus, the use of ICT makes public officials more efficient and capable. It also improves monitoring mechanisms and increases transparency and human empowerment through the spread of information. Mostly, it is considered an anti-corruption instrument, but it also has some negative effects when ICT tools are used instead of anti-corruption. With the advancement in ICT infrastructure, E-signatures and time-stamping services create ease for management solutions.

Therefore, ICT is a source of reducing corruption through the theory of network society (Soper, 2007) and causes an increase in more information to the public (Castells, 2000). So, ICT tools are used to curb or determine corruption due to the advancement in ICT; their diffusion reduces the discretionary power of public administrators and consequently causes a decrease in corruption (Jha, 2020; Jha & Sarangi, 2014; Longe *et al.*, 2020). According to De Sousa (2018), corruption can be reduced in five major ways: 1) raising awareness through ICTs; 2) online monitoring; 3) reducing direct contact through mobile phones, the Internet, and telephones; 4) effective control of financial transactions; and 5) initiating anti-corruption campaigns. Thus, ICTs can be utilized by the government to reduce corruption. Moreover, Shim & Eom (2008) argued that corruption decreases with ICT due to a reduction in physical interaction; and fast Internet adoption has a positive relationship with the control of corruption (Lio *et al.*, 2011). In a similar line, Andersen (2009) discloses that the implementation of e-government is successful in reducing corruption for a selected panel of 100 countries by taking a timespan of ten years. A study in 2017 provided a massive argument about controlling corruption by ICT (Sassi & Ben Ali, 2017). Lidman (2011) and Sassi & Ben Ali (2017) argue that public officials can be traced through mobile phones and the Internet, and by recording their conversations to ask for a bribe; such fear also reduces corruption. These studies suggested that only policy-based technologies reduce corruption. This is why it is of the utmost importance to spread awareness of these technologies among the country's citizens so that they can use them to combat corruption.

However, that corruption can be reduced through ICT is doubtful (Kim *et al.*, 2009) but recent studies propose some distinct arguments that it curbs corruption. Lincényi & Čársky (2021) and Remeikienė *et al.* (2020) argued that the use of ICT improves governance and accountability, and reduces corruption because information diffusion in society enables government officials to have a higher chance of being caught and prosecuted (Cho & Choi, 2004). Ben Ali & Gasmi (2017) examined the relationships between the adoption of ICT and corruption utilising a panel of 175 countries for the years 1996 to 2014, and they reached the same conclusion: digital inclusion is a powerful instrument for fighting against corruption. Moreover, Androniceanu *et al.* (2021) investigated the influence of ICT integration on the control of corruption in the administrations of the EU by using panel data from 2010 to 2019,

http://doi.org/10.18080/jtde.v11n3.658

and showed that it has a significant effect on reducing corruption. According to Afzal *et al.* (2021), greater adoption of the Internet and mobile phones among selected Pacific-Asian economies fosters transparency and good governance.

In addition, Bhattacherjee & Shrivastava (2018) employed the hypothetico-deductive technique in a study using general deterrence theory (GDT). ICT use affects corruption by enhancing the certainty and swiftness of punishments associated with it, and ICT investments may have a limited impact on corruption without ICT laws. Other evidence showed that technology adoption slows the rate of corruption in low-income and high-income countries, while it is insignificant in the case of middle-income countries due to the digital divide (Mouna *et al.*, 2020). Kouladoum (2022) investigated the effects of ICT on corruption in Africa. As estimation methods, the fixed- and random-effects models are used. The two-stage least square (2SLS) and Lewbel techniques are chosen to deal with the issue of probable endogeneity due to the flaws in the fixed- and random-effects models. The results show that Internet use, mobile phone use, and the composite ICT indicator all favourably influence the improvement of corruption control in Africa. Furthermore, Hartani *et al.* (2020) examined the ICT-corruption nexus by using cross-sectional data on associated ASEAN country-related variables. For the purpose of estimating different relationships among variables, the studies include the IPS unit root test, Pedroni cointegration, and FMOLS estimate. The researchers verified that the use of ICT and e-government could lessen corruption in ASEAN nations and proved ICT as a tool to reduce corruption. Suardi (2021) examined the effect of ICT diffusion on the corruption perception index (CPI) in Asia and concluded that it reduces corruption. Its implementation has raised public perceptions of corruption, with telecommunications infrastructure having the most profound impact. According to Gouvea *et al.* (2022), countries that have made a transition from a resource-based economy to a system that is innovation-driven and digital have lower levels of corruption. This is based on panel data from 147 countries during a seven-year period from 2013 to 2019. Corruption is inversely related to ICT indicators like Internet usage and e-government.

The current strand of literature illustrates the limitations of technology as a way to guarantee transparency in government interactions. De Sousa (2018) argued that ICT cannot end corruption on its own. He contends that the proper institutional framework should be used to train public officials. His findings show that ICT dissemination works well with education and training. Darusalam *et al.* (2021) examined the impact of ICT on corruption control in Asian countries over a 33-year period from 1984 to 2016. ICT and the control of corruption have a non-linear, inverted U-shaped relationship, which suggests that ICT in these countries does not lower the rate of corruption. The findings of Darusalam *et al.* (2021) also show that

http://doi.org/10.18080/jtde.v11n3.658

government efficiency and education must be added to ICT in order to effectively combat corruption. Their findings of a non-linear effect suggest that ICT may make corruption easier.

As shown by the studies mentioned above, research on the relationship between ICT and corruption is not theoretically supported, somewhat inconclusive, and unable to clearly explain when ICT reduces corruption and when it does not. The previous literature on the impact of technology on the control of corruption has overlooked fixed telephone and fixed broadband indicators in favour of looking only at the Internet and mobile phone penetration (Darusalam *et al.*, 2021; Kouladoum, 2022) and mobile phone penetration (Sassi & Ben Ali, 2017), whereas Suardi (2021) examined the effect of ICT diffusion on the corruption perception index (CPI) in Asia and used the ICT infrastructure index as a proxy. Darusalam *et al.* (2021) used the panel ARDL model and Suardi (2021) used the fixed effect and random effect model to demonstrate their results. In order to fill this gap in the literature, this study examines the impact of a composite index of ICT (fixed broadband, fixed phones, mobile phones, and Internet users) on corruption, using panel data of 17 Asian countries from 2010 to 2019 by employing Driscoll-Kraay standard error technique.

Theoretical Considerations

Conceptual framework

The advancement in ICT enables human beings to perform efficiently in every field of life and it also changes the behaviour of individuals. Let us understand the functions of ICT which cause a change in the attitude and behaviour of people in the contemporary world.

Many influential research studies have shown the influence of ICT on human life (Jorgenson & Stiroh, 2000; Oliner & Sichel, 2000; van Ark *et al.*, 2008). Many pieces of evidence show all domains of life are affected by it, while domains of life such as working, private, and environmental life, human capabilities, psychological functioning, cultural values, and beliefs create a heterogeneity problem (Castellacci & Tveito, 2016); whereas there are many advantages of ICT development, it also has reverse effects on human well-being, such as cyberbullies, privacy risk, leakage of information, and increasing corruption. If it cannot be effectively used, it will leave negative impacts on governance, economic development, education, and corruption. On the other hand, leakage of information can also create a problem for individuals as well as the country. False information and rumours could be spread through it. ICT can be very advantageous and hence can be used efficiently. The following conceptual framework borrowed from Castellacci & Tveito (2016) and slightly changed is incorporated for this research agenda.

http://doi.org/10.18080/jtde.v11n3.658

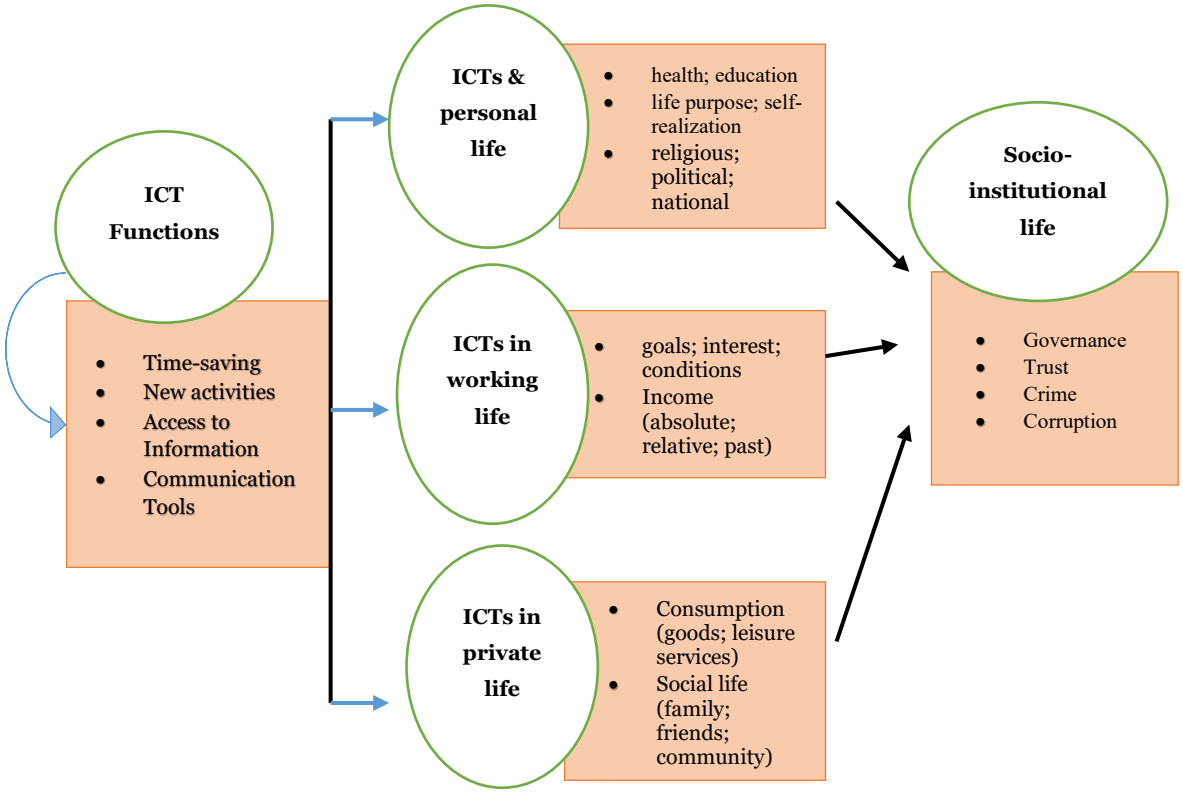


Figure 1. Conceptual Framework (Castellaci & Tveito, 2016)

In the realm of personal life, ICTs wield the power to disseminate information with unprecedented celerity. This unfettered flow of data can act as both a conduit for enlightenment and reduction in nefarious practices because corruption often thrives in the interstices of opacity. Citizens armed with instant access to government proceedings, budget allocations, and policy implementations are more likely to hold public officials accountable for their actions. In the domain of working life, the integration of ICTs has ushered in an era of heightened efficiency and interconnectedness. The expedited transfer of funds, procurement processes, and contract adjudications, can also bring down grounds for corruption to fester. Turning to the private sphere, the omnipresence of ICTs has indelibly altered the dynamics of personal relationships and communication. Automated record-keeping systems and digital audit trails engender an environment where transactions are subject to meticulous scrutiny, minimizing the interstices in which corrupt acts can flourish. Simultaneously, ICTs permeate the private lives of individuals, augmenting their ability to monitor and report irregularities. Crowdsourcing platforms and mobile applications enable citizens to report instances of bribery, extortion, or other corrupt practices with immediacy and anonymity.

Crime opportunity theory

This study is backed by Crime opportunity theory (COT). COT states that an offender does official matters that are very simple to perform but nonetheless offer great incentives and favourable returns (Faisal *et al.*, 2016). According to this theory, two factors have significance

http://doi.org/10.18080/jtde.v11n3.658

for the execution of a crime. The first element is the presence of a criminal or offender and the second is the state of a location where a criminal is physically present to do a certain crime. For committing a crime, these two elements must be present together (Grasmick *et al.*, 1993). In our study, crime refers to corruption, and the offender or criminals can be conceived of as government officials who are in charge of offering a range of services to the public. The belief that crime can be effectively averted by altering the circumstances in which it will occur is another crucial part of this theory. This can be accomplished by introducing various activities in the system. ICT interrupts the aforementioned two factors and reduces corruption. According to our study, the impact of ICT diffusion is being examined in relation to a decline in the rate of corruption (Hartani *et al.*, 2020; Jeffery & Zahm, 1993).

Data and Methodology

Data and data sources for the study

In this study, we investigate panel data for 10 years ranging from 2010 to 2019. Based on data availability, 17 countries from the Asian continent are taken for this study. The countries are Bangladesh, Cambodia, China, Cyprus, Indonesia, India, Iran Islamic Republic, Israel, Japan, Kuwait, Malaysia, Pakistan, Saudi Arabia, Singapore, Thailand, Türkiye, and United Arab Emirates. The main goal of this study is to check the relationship between ICT diffusion on control of corruption in Asian countries. The following question is addressed in this research:

1. Is ICT diffusion affecting the control of corruption?

Variables are borrowed from different research works. Control of corruption (CoC) is a dependent variable in this study which has been used by Darusalam *et al.* (2019). Four independent variables are taken as determinants of corruption, namely, political stability (PS), GDP per capita, effective governance (EG), and ICT. Bhattacherjee & Shrivastava (2018) used political stability and effective governance as determinants of corruption. GDP per capita is also used as a determinant and many studies used the ICT development index to analyze corruption, such as ICT exposure index, and different proxies (Internet users per 100 inhabitants, mobile cellular subscriptions per 100 people) (Androniceanu *et al.*, 2021; Darusalam *et al.*, 2021; Fisman & Svensson, 2007; Gouvea *et al.*, 2022; Hartani *et al.*, 2020). In this study, we used an ICT exposure index that has been constructed using the following four variables:

1. Mobile Cellular subscriptions (per 100 people)
2. Internet users (per 100 people)
3. Number of Secure Internet servers

http://doi.org/10.18080/jtde.v11n3.658

4. Fixed telephone subscriptions (per 100 people).

To construct the ICT exposure index, Principal Component Analysis has been used. See Hanafizadeh *et al.* (2009) for ICT index construction. Definitions and data sources are described in Table 1.

Table 1. Measures and Indicators

Variable	Definition	Indicators/Scale/Source
CoC	Control of corruption	Reflects perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests.
		Scale: -2.5 – 2.5 (High Corrupt – Low Corrupt)
		Source: The Worldwide Governance Indicators (WGI)
ICT	Information communication and technology exposure index	Mobile Cellular subscriptions (per 100 people) Internet users (per 100 people) Secure Internet servers Fixed telephone subscriptions (per 100 people)
		Source: World Development Indicators (WDI)
PS	Political stability	Orderly transfers, Armed conflict, Violent demonstrations, Social Unrest, International tensions, Cost of Terrorism, Frequency of political killings, Frequency of disappearances, Frequency of tortures, Political terror scale, Security Risk Rating, The intensity of Internal conflicts (Ethnic, religious, or regional), Intensity of violent activities (Political), Intensity of social conflicts (except Land)
		Scale: -2.5 – 2.5 (weak – strong) stability
		Source: The Worldwide Governance Indicators (WGI)
EG	Effective Governance	Reflects perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.
		Scale: -2.5 – 2.5 (weak – strong) governance
		Source: The Worldwide Governance Indicators (WGI)
GDP	Gross Domestic Product per capita	Source: World Bank Data (WB)

Estimable model

To analyze the relationship between control of corruption (CoC) and its explanatory variables, the function given below is constructed:

$$\mathbf{CoC} = \boldsymbol{f}(\mathbf{PS}, \mathbf{GDP}, \mathbf{EG}, \mathbf{ICT}) \quad (1)$$

Here, transparency is measured by the control of corruption (CoC), which is a function of the ICT Index, gross domestic product (GDP), effective governance (EG), and political stability (PS), which are expected to be linked to control of corruption.

Econometric specification of this function is as follows:

$$\boldsymbol{CoC_{it}} = \boldsymbol{u_i} + \boldsymbol{\delta_t} + \boldsymbol{\beta_1}(\boldsymbol{PS_{i,t}}) + \boldsymbol{\beta_2}(\boldsymbol{GDP_{i,t}}) + \boldsymbol{\beta_3}(\boldsymbol{EG_{i,t}}) + \boldsymbol{\beta_4}(\boldsymbol{ICT_{i,t}}) + \boldsymbol{\varepsilon_{i,t}} \quad (2)$$

http://doi.org/10.18080/jtde.v11n3.658

The subscript i is for countries and t is for time. All methods have been elaborated earlier in the literature, where u_i, δ_t take into account the unobserved country-specific effects and $\varepsilon_{i,t}$ indicates the error terms, which are assumed to be i.i.d. with null mean and variance σ^2. $\beta_1, \beta_2, \beta_3,$ and β_4 are the coefficients of political stability (PS), gross domestic product (GDP), effective governance (EG), and ICT Index, respectively.

There are several pooling models, but panel effects such as fixed effect and random effect are mostly used for panel data analysis. This study consists of 17 Asian countries and entails balanced panel data ranging from 2010 to 2019 for 10 years. There are 170 (17x10=170) observations. See Mehmood *et al.* (2013) on how to estimate a panel effects model. Panel data estimation is discussed in the following section.

Panel Data Estimation and Discussion

The subsections below follow a sequence to estimate the panel dataset. To analyse the data, both descriptive and inferential techniques are used. The results of both techniques are described below with pertinent analysis.

Descriptive analysis

The dependent variable control of corruption (CoC) is used as a proxy for transparency in these analyses. Independent variables are PS, GDP, EG, and ICT. A set of descriptive results is shown in Table 2.

The mean value of CoC is .072 with an estimated standard deviation of .929 for all 17 Asian countries over 10 years from 2010 to 2019. The minimum value of CoC for the whole data set is -1.31 for Cambodia in 2018, which has the highest corruption, and the maximum value is 2.18 for Singapore in 2010, which has the lowest corruption. The ICT exposure index has a mean value of 0.009, which is very small, and the highest standard deviation of one.

Table 2. Descriptive Analysis

Variable	Mean	Standard Deviation	Minimum	Maximum
CoC	.072	.929	-1.31	2.18
PS	-.413	.991	-2.81	1.62
GDP	3.938	.567	2.95	4.787
EG	.399	.874	-.94	2.24
ICT	.009	1	-2.015	2.404

The minimum value of the ICT exposure index for the whole data set is -2.015 in 2010 for Bangladesh and the maximum value of ICT is 2.404 for UAE in the year 2016. The mean value of the whole data set of Political stability is -.413, which means that, on average, all countries are passing through political instability. The minimum value of the political stability index is

http://doi.org/10.18080/jtde.v11n3.658

-2.81 in the year 2011 for Pakistan, which means the highest political instability was in Pakistan in 2011. The maximum value of the Political stability index is 1.62 for Singapore in 2017. Cambodia has the minimum values of GDP and effective governance in the year 2010 and Singapore has the maximum values for both in 2010. Singapore has a maximum value for four variables in our data set. Although the ICT exposure index has low values for Singapore, it has very strong control over corruption due to the highest values of the other four variables. In Cambodia, the CoC index has the highest negative values because of its bad performance in the three major indicators.

Inferential analysis

In inferential analysis, statistical estimation is done on the pre-defined econometric model. In this study, we use three estimation techniques for better results and choose which is the best. The primary goal of this research is to analyse the relationship between ICT and control of corruption (CoC) and how other control variables, political stability, effective governance, and economic development, affect corruption. Thus, CoC is used as a dependent variable. PS, GDP, EG, and, ICT exposure index were regressed on CoC, and econometric results are appended below.

Test for multi-collinearity

A primary goal of economic analysis is to determine whether there is multi-collinearity among the independent variables. As a general rule, variables have severe multi-collinearity if their variance inflation factors (VIFs) are greater than 10, which often occurs when R^2 reaches 0.90. According to Damodar Gujarati (2022), if the value of VIF is less than 5, a multi-collinearity problem does not exist. Table 3 shows the VIF results for investigating multi-collinearity in panel (a). The set of selected variables shows that there is no multi-collinearity.

Ordinary Least Squares (OLS) will be a partial specification if there is country-level heterogeneity (differing social and cultural norms, for example) and a fixed/random effects model should be estimated. The subsequent two tests are essential for developing a good estimation method for panel data analysis.

Breusch and Pagan Lagrangian multiplier test for random effects or OLS

The post-estimation test assists in selecting between an OLS and a random effects regression. The null hypothesis of the Lagrangian Multiplier (LM) test is that there is no significant difference across countries, or that variances across countries are equal to zero. Table 3 shows the results of LM for panel effects. Breusch and Pagan Lagrangian multiplier (1980) in panel (b) justifies that there is a panel effect because χ^2 (01) = 377.62 and Prob.> χ^2=0.0000. Variance across entities is zero and the probability value also affirms the panel effect.

http://doi.org/10.18080/jtde.v11n3.658

Hausman test

The Hausman test is performed to select the best option between fixed effects model (FEM) and random effects model (REM) and the results are appended in Table 3, panel (c). FEM and REM are compared having the null hypothesis for the final decision. The result shows that FEM is preferred, because the p-value<0.01 and the null hypothesis is accepted. In this case, if we perform REM, it generates biased estimators. Therefore, we prefer the fixed effects model.

Table 3. Diagnostics Tests

<table>
<tr><td colspan="3">Panel (a): Investigating multicollinearity</td><td colspan="2" rowspan="2">Panel (b): Breusch and Pagan Lagrangian multiplier (LM) for exploring panel effects</td></tr>
<tr><td>Variable</td><td>VIF</td><td>1/VIF</td></tr>
<tr><td>PS</td><td>4.56</td><td>0.219348</td><td>Ho: No panel effect</td><td>$\chi^2(01) = 377.62$</td></tr>
<tr><td>GDP</td><td>4.47</td><td>0.2238</td><td colspan="2">P-value >= 0.00000</td></tr>
<tr><td>EG</td><td>2.67</td><td>0.37502</td><td colspan="2" rowspan="3">This test investigates the existence of the panel effects. The results shown in the left column justify the existence of panel effects, because the null hypothesis is not rejected.</td></tr>
<tr><td>ICT</td><td>2.21</td><td>0.45291</td></tr>
<tr><td>Mean VIF</td><td colspan="2">3.48</td></tr>
<tr><td colspan="5">Panel (c): Hausman Test to choose between Fixed and Random effects</td></tr>
<tr><td colspan="2">Ho: Fixed effect
H1: Random effect</td><td colspan="2">$\chi^2(4) = 14.62$</td><td>P-value > χ^2= 0.0056</td></tr>
<tr><td colspan="5">Panel (d): Modified Wald Test for Group Wise Heteroskedasticity</td></tr>
<tr><td colspan="2">Ho: $\sigma(i)^2 = \sigma^2$ for all i</td><td colspan="2">$\chi^2(17) = 594.27$</td><td>P-value > χ^2 = 0.0000</td></tr>
<tr><td colspan="5">Panel (f): Wooldridge Test for Serial Correlation</td></tr>
<tr><td colspan="2">Ho: no first order autocorrelation</td><td colspan="2">F(1, ,16) = 30.030</td><td>P-value > F = 0.0001</td></tr>
</table>

Comparisons between POLS, RE, and FE models

Pooled OLS: Pooled OLS model's F-statistic value is 700.15 and at a 1% level of significance. R^2 explains the variation in the dependent variable, CoC, due to independent variables, PS, GDP, EG, and ICT. According to the value of R^2, independent variables (PS, GDP, EG, and ICT) explain 94.44% of variation in the dependent variable (CoC), though, with a high R^2, there may be a problem of autocorrelation. The value of adjusted R^2 shows that there is approximately no difference and its value is 0.9430. According to Pooled OLS estimation, slope parameters for all variables are $(\beta_{PS}^{POLS})_{1\%}$=.1125, $(\beta_{GDP}^{POLS})_{1\%}$=.436, $(\beta_{EG}^{POLS})_{1\%}$=.754, $(\beta_{ICT}^{POLS})_{1\%}$=.1099, respectively, and show their potential toward CoC.

Fixed Effects Estimates: A second technique of panel data analysis, regression estimation through the fixed effects model, also shows there is a positive relationship between ICT and CoC. Its results show that, except for GDP, all other variables are significant at a 1% level. By using FEM, GDP is insignificant, which suggests excluding this variable. The incline coefficients corresponding to each individual variable are observed to be as follows: $(\beta_{PS}^{FEM})_{1\%}$=.1678, $(\beta_{GDP}^{FEM})_{insig}$=-.0661, $(\beta_{EG}^{FEM})_{1\%}$=.4966, $(\beta_{ICT}^{FEM})_{1\%}$=.0599. Using FEM technique, ICT coefficient reduces to $(\beta_{ICT}^{FEM})_{1\%}$=.0599 from $(\beta_{ICT}^{POLS})_{1\%}$=.1099, and it also

http://doi.org/10.18080/jtde.v11n3.658

decreases the coefficient of EG but increases PS compared with POLS. The value of R^2 also decreased and independent variables explain the variation of 91.43% in CoC. The value of adjusted R^2 is 0.8990, which is slightly lower than R^2. FEM's F-statistics value is 22.94 and at a 1% significant level.

Table 4. Pooled OLS, Panel Effects & Driscoll-Kraay Method – A Comparison

Dependent Variable (CoC)			**Constant**	**Independent variables**			
				PS	**GDP**	**EG**	**ICT**
Coefficients	**Pooled OLS**		-1.899*** (.241)	.1125*** (.028)	.436*** (.0642)	.754*** (.0413)	.1099*** (.0253)
	Panel Effects	**FEM**	.203 (.954)	.1678*** .0514	-.0661 (.243)	.4966*** (.0682)	.0599*** (.0214)
		REM	-2.0443*** (.491)	.1478*** (.0448)	.495*** (.1259)	.568*** (.0618)	.0886*** (.0199)
	Driscoll-Kraay Method (DKM)		-.20317*** (1.738)	.1678*** (.0587)	-.0661 (.445)	.4966*** (.0999)	.0599*** (.024)
Techniques	Pooled OLS		Panel Effects		Driscoll-Kraay Method		
			Fixed Effects	Random Effects			
R^2	0.9444		0.9143	0.9529	0.8632		
Adjusted R^2	0.9430		0.8990	0.9398	0.8241		
Model Significance	F(4, 165) = 700.15 Prob. > F = 0.0000		F(4, 149) = 22.94 Prob. > F = 0.0000	Wald $\chi^2(4)$ = 336.00 Prob. >χ^2 = 0.0000	F(4, 9) = 59.64 Prob. > F = 0.0000		
*** Significant at 1%							

Random Effects Estimation: REM results are also given in Table 4, which shows that there is a positive relation between ICT and CoC. Pursuant to REM estimation, the slope coefficients for each variable stand at $(\beta_{PS}^{REM})_{1\%}$=.1478, $(\beta_{GDP}^{REM})_{1\%}$=.495, $(\beta_{EG}^{REM})_{1\%}$=.568, $(\beta_{ICT}^{REM})_{1\%}$=.0886, respectively. The value of R^2 is greater than the POLS model and FEM which explains 95.29% variation in the dependent variable (CoC) due to independent variables. For REM, the model significance is shown through the Wald chi-squared test and its value is 336. In this model, the value of coefficients for all variables is increased. For the best choice between FEM and REM, the Hausman test confirms that FEM is the best. Therefore, REM estimation to analyze the panel effect and its results may be spurious.

Test for serial correlation

Serial correlation is typically not anticipated when the time span is less than 20 years. Serial correlation lowers standard errors of coefficients and raises R^2. The micro panel data used in this study (t = 10 < 20) reduces the likelihood of a serial correlation. But this test is used for the sake of precision. Table 3's statistics are interesting in that they indicate that the null hypothesis is rejected because p-value is less than 1%, and that there is a serial correlation among the residuals. OLS coefficients are hence probably biased, inconsistent, and ineffective.

http://doi.org/10.18080/jtde.v11n3.658

Hence, we employ the Driscoll-Kraay standard error methodology to render OLS estimations robust and efficacious in addressing the aforementioned issue.

Test for heteroskedasticity

The error term ε can be heteroskedastic if the variance of the conditional distribution of ε_i given $X_i[\mathrm{var}(\varepsilon_i|X_i)]$ is non-constant for i= 1, 2,…,n, and specifically does not depend on X; else, ε is homoscedastic. Heteroskedasticity can lead to inaccurate estimations of standard error coefficients and, consequently, of their t-values. OLS estimates may not be biased in this situation but generate wrong standard errors. The Modified Wald Test for Group Wise Heteroskedasticity in Table 3 suggested that there is a problem with heteroskedasticity.

Fixed effects estimates with Driscoll and Kraay (DK) standard errors

The Driscoll and Kraay (DK) standard error technique is regarded as one of the best approaches if there is a possibility of heteroskedasticity, spatial dependence, and serial dependency in the data. The fixed effects regression with Driscoll and Kraay standard errors (SE) is required by the results of the Wooldridge test for serial correlation and the Modified Wald test for group-wise heteroskedasticity. Therefore, the Driscoll & Kraay (1998) standard error technique is used to examine the effect of ICT on the control of corruption for a panel of Asian countries, because the DK methodology is a flexible, non-parametric method. Additionally, the DK covariance estimator performs with both balanced and unbalanced panel data and is capable of handling missing values. DK estimations can deal with cross-sectional and temporal dependency patterns. The results demonstrate no unexpected shift in the statistical significance of the fixed effects estimates. In two steps, the individual fixed-effects estimator is used. First, using xtreg command in Stata for OLS, all variables $z_{it} \in \{y_{it}, x_{it}\}$ are transformed.

$$\tilde{\boldsymbol{z}}_{\boldsymbol{it}} = \mathbf{z}_{\boldsymbol{it}} - \bar{\boldsymbol{z}}_{\boldsymbol{it}} + \bar{\bar{\mathbf{z}}} \tag{3}$$

Equation (3) is about transforming a variable z_{it}, which could be either y_{it}(a dependent variable) or x_{it} (an independent variable). $\tilde{\boldsymbol{z}}_{\boldsymbol{it}}$ is a newly transformed variable. Equation (4) calculates the average value, $\bar{\boldsymbol{z}}_{\boldsymbol{it}}$, of the transformed variable, $\tilde{\boldsymbol{z}}_{\boldsymbol{it}}$, across all time periods $\mathbf{T}_{\boldsymbol{i}}$ for a given entity i.

$$\bar{\boldsymbol{z}}_{\boldsymbol{it}} = \mathbf{T}_{\boldsymbol{i}}^{-1} \sum_{\boldsymbol{t}=1}^{\mathbf{T_i}} \mathbf{z}_{\boldsymbol{it}} \tag{4}$$

$$\bar{\bar{\mathbf{z}}} = (\sum \mathbf{T}_{\boldsymbol{i}})^{-1} \sum \boldsymbol{i} \sum \boldsymbol{t}\, \mathbf{z}_{\boldsymbol{it}} \tag{5}$$

Here, $\bar{\bar{\mathbf{z}}}$ is the constant term, which is calculated from equation (5).

http://doi.org/10.18080/jtde.v11n3.658

Secondly, as a result, this work takes into account a linear model and uses Fixed Effects Estimates with Driscoll and Kraay Standard Errors for estimation of a linear model which can be expressed as follows:

$$\tilde{\boldsymbol{y}}_{i,t} = (\widetilde{\boldsymbol{X}'}_{i,t})\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}_{i,t} \quad (6)$$

where $\tilde{\boldsymbol{y}}_{i,t}$ is the dependent variable, and $\widetilde{\boldsymbol{X}'}_{i,t}$ denotes independent variables. Also, i is the index of countries, i = 1, 2, 3 ...,17, and t is a study period, t = 2010, 2012,.....,2019.

To solve the above diagnostic problem, we use a Driscoll-Kraay standard error model and it shows more significant results than the previous two models. The Driscoll-Kraay model also shows that ICT reduces corruption in the case of Asian countries. Slope parameters for the Driscoll-Kraay technique are $(\beta_{PS}^{DKM})_{1\%}$=.1678, $(\beta_{GDP}^{DKM})_{insig}$=-.0661, $(\beta_{EG}^{DKM})_{1\%}$=.4966, $(\beta_{ICT}^{DKM})_{1\%}$=.4966, respectively. The coefficients of PS, GDP, EG, and ICT decreased as compared with other techniques. This model's significance is shown through F-test and its value is 59.64, which is very high compared with the previous two estimation techniques, and shows a high level of significance at 1% level. All previous variables are confirmed through the Driscoll-Kraay model.

The relationship of political stability with control of corruption is evident through their slope parameters in all techniques DKM, FEM, REM, and POLS, i.e. $(\beta_{PS}^{DKM})_{1\%}$=0.1678, $(\beta_{PS}^{FEM})_{1\%}$=0.1678, $(\beta_{PS}^{REM})_{1\%}$=0.1478, $(\beta_{PS}^{POLS})_{1\%}$=0.1125, respectively. Political stability leads to consistency in policies and regulations, contributes to a stronger rule of law and judicial independence, and raises the political will. When governments are constantly changing due to instability, there is a higher likelihood of policy flip-flops and ad-hoc decision-making. Such uncertainty can create opportunities for corruption as individuals may exploit regulatory loopholes or manipulate changing policies for personal gain. In stable political environments, bureaucratic processes and procedures are more likely to be streamlined and efficient, and can reduce opportunities for bribery and extortion.

The slope parameters for government effectiveness and control of corruption are positively related in all addressed techniques, i.e., $(\beta_{EG}^{DKM})_{1\%}$=0.4966, $(\beta_{EG}^{FEM})_{1\%}$=0.4966, $(\beta_{EG}^{REM})_{1\%}$=0.568, $(\beta_{EG}^{POLS})_{1\%}$=0.754, respectively. The results are consistent with the prior studies (Bhatnagar, 2000; Poliak *et al.*, 2020; Russell, 2020). The findings additionally suggest that the mastery over corruption escalates proportionally alongside the augmentation of governmental efficacy within the chosen Asian nations. The government's effectiveness further enriches institutional calibre, amplifying the enforcement of legal frameworks and the command over corruption within the corresponding assemblage of nations.

http://doi.org/10.18080/jtde.v11n3.658

The purpose of this study is to check that the relationship between ICT and CoC is confirmed in all addressed techniques, i.e., $(\beta_{ICT}^{DKM})_{1\%}$=.0599, $(\beta_{ICT}^{FEM})_{1\%}$=.0599, $(\beta_{ICT}^{REM})_{1\%}$=.0886, $(\beta_{ICT}^{POLS})_{1\%}$=.1099, respectively. The ICT exposure index leaves a very small impact in this case, but the ICT exposure index again achieved its position at a 1% level of significance. Furthermore, ICT shows a positive relationship with control of corruption; this finding is corroborated by different regional studies. The magnitude of the ICT coefficient for reducing corruption is 0.87 in Europe (Androniceanu *et al.*, 2021), 8.254 for selected Asian countries (Suardi, 2021), and 0.0256 in Africa (Kouladoum, 2022). The finding is also corroborated by other studies (Darusalam *et al.*, 2021; Gouvea *et al.*, 2022; Hartani *et al.*, 2020). The first reason is that Internet technology improves the implementation of the law and regulations by limiting the public administration's discretion. Secondly, the ability of society to report corruption-related acts and the accelerated dissemination of information made possible by mobile phones make them effective instruments for detecting corruption. Third, online communication is faster than traditional processes, and, by implementing such technology, government agencies can take prompt preventative action in the event of suspicious or malicious activity.

Conclusion

This study delved into the impact of Information and Communication Technology (ICT) on control of corruption within a cohort of 17 Asian nations in a time span from 2010 to 2019. Employing the Driscoll-Kraay methodology for standard error computation, the research pursued the establishment of foundational outcomes, thereby confronting the challenges posed by autocorrelation and heterogeneity. The study adopted four measures of ICT that are the number of individuals using the Internet, fixed broadband subscriptions, mobile cellular subscriptions, and secure Internet servers.

Corruption hampers the growth track to pursue inclusive, equitable, and sustainable economic growth and development. Additionally, it affects the distribution of resources within and between regions. This study safely concluded, with a fixed effects Driscoll-Kraay OLS estimation, that ICT is a determinant of corruption and that ICT reduces corruption, contrary to other studies (Charoensukmongkol & Moqbel, 2014; Heeks, 1998), which show that there is no significant relation between ICT and CoC; whereas the results are substantiated by additional scholarly investigations (Darusalam *et al.*, 2021; Gouvea *et al.*, 2022; Hartani *et al.*, 2020). The amelioration of corruption within a nation through the adept application of ICT is poised to culminate in the augmentation of transparency and accountability. Consequently, this synergistic effect is anticipated to improve a nation's economic prosperity as well. It also shows that there is a positive relationship between effective governance and control of

http://doi.org/10.18080/jtde.v11n3.658

corruption, as in previous studies (Bhatnagar, 2000; Poliak *et al.*, 2020; Russell, 2020). ICT is significant at 1% in the final regression to support the objective. Hence, ICT is a potential variable to explain CoC.

ICT policy can play an important role in curbing corruption, so there is a strong need for ICT diffusion, as it has a strong positive relation with CoC. The results suggest that effective governance helped to reduce corruption in the Asian region and establish a surveillance-based system in public institutions. Access to ICT tools and platforms should be made easy for all parties to ensure the dissemination of information.

References

Adam, I. O. (2020). Examining E-Government development effects on corruption in Africa: The mediating effects of ICT development and institutional quality. *Technology in Society*, *61*. https://doi.org/10.1016/j.techsoc.2020.101245

Adam, I., & Fazekas, M. (2021). Are emerging technologies helping win the fight against corruption? A review of the state of evidence. *Information Economics and Policy*, *57*, 100950. https://doi.org/10.1016/j.infoecopol.2021.100950

Afzal, M. S., Khan, A., Qureshi, U. U. R., Saleem, S., Saqib, M. A. N., Shabbir, R. M. K., Naveed, M., Jabbar, M., Zahoor, S., & Ahmed, H. (2021). Community-Based Assessment of Knowledge, Attitude, Practices and Risk Factors Regarding COVID-19 Among Pakistanis Residents During a Recent Outbreak: A Cross-Sectional Survey. *Journal of Community Health*, *46*(3), 476–486. https://doi.org/10.1007/s10900-020-00875-z

Andersen, T. B. (2009). E-Government as an anti-corruption strategy. *Information Economics and Policy*, *21*(3), 201–210. https://doi.org/10.1016/j.infoecopol.2008.11.003

Androniceanu, A., Nica, E., Georgescu, I., & Sabie, O. M. (2021). The influence of the ICT on the control of corruption in public administrations of the EU member states: A comparative analysis based on panel data. *Revista Administratie Si Management Public*, *37*, 41–59. https://doi.org/10.24818/amp/2021.37-03

Ben Ali, M. S., & Gasmi, A. (2017). Does ICT diffusion matter for corruption? An Economic Development Perspective. *Telematics and Informatics*, *34*(8), 1445–1453. https://doi.org/10.1016/j.tele.2017.06.008

Bhatnagar, S. (2000). Social Implications of Information and Communication Technology in Developing Countries: Lessons from Asian Success Stories. *The Electronic Journal of Information Systems in Developing Countries*, *1*(1), 1–9. https://doi.org/10.1002/j.1681-4835.2000.tb00004.x

Bhattacherjee, A., & Shrivastava, U. (2018). The effects of ICT use and ICT Laws on corruption: A general deterrence theory perspective. *Government Information Quarterly*, *35*(4), 703–712. https://doi.org/10.1016/j.giq.2018.07.006

Breusch, T. S., & Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *The review of economic studies*, *47*(1), 239-253. https://doi.org/10.2307/2297111

Castellacci, F. (2006). Innovation, diffusion and catching up in the fifth long wave. *Futures*, *38*(7). https://doi.org/10.1016/j.futures.2005.12.007

Castellacci, F., & Tveito, V. (2016). The Effects of ICTs on Well-being: A Survey and a Theoretical Framework. Working Papers on Innovation Studies 20161004, Centre for Technology, Innovation and Culture, University of Oslo. https://ideas.repec.org/p/tik/inowpp/20161004.html

Castells, M. (2000). Materials for an exploratory theory of the network society. *British Journal of Sociology*, *51*(1), 5–24. https://doi.org/10.1111/j.1468-4446.2000.00005.x

Charoensukmongkol, P., & Moqbel, M. (2014). Does Investment in ICT Curb or Create More Corruption? A Cross-Country Analysis. *Public Organization Review*, *14*(1), 51–63. https://doi.org/10.1007/s11115-012-0205-8

Cho, Y. H., & Choi, B. D. (2004). E-government to combat corruption: The case of Seoul metropolitan government. *International Journal of Public Administration*, *27*(10), 719–735. https://doi.org/10.1081/PAD-200029114

Crafts, N. (2004). Steam as a General Purpose Technology: A Growth Accounting Perspective. *The Economic Journal*, *114*(495), 338–351. https://doi.org/10.1111/j.1468-0297.2003.00200.x

Darusalam, D., Janssen, M., Sohag, K., Omar, N., & Said, J. (2021). The Influence of ICT on the Control of Corruption: A Study Using Panel Data From ASEAN Countries. *International Journal of Public Administration in the Digital Age*, *8*(1). https://doi.org/10.4018/IJPADA.20210101.oa2

Darusalam, D., Said, J., Omar, N., Janssen, M., & Sohag, K. (2019). The Diffusion of ICT for Corruption Detection in Open Government Data. *Knowledge Engineering and Data Science*, *2*(1), 10. https://doi.org/10.17977/um018v2i12019p10-18

De Sousa, L. (2018). Open government and the use of ICT to reduce corruption risks. In Delicado, A., Domingos, N., & de Sousa, L. (eds), *Changing societies: legacies and challenges, Vol. III. The diverse worlds of sustainability* (pp. 179–202). Lisboa: Imprensa de Ciências Sociais. https://doi.org/10.31447/ics9789726715054.07

Dirienzo, C. E., Das, J., Cort, K. T., & Burbridge, J. (2007). Corruption and the Role of Information. *Journal of International Business Studies*, *38*(2), 320–332. https://doi.org/10.1057/palgrave.jibs.8400262

Driscoll, J. C., & Kraay, A. C. (1998). Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics*, *80*(4), 549–560. https://doi.org/10.1162/003465398557825

Faisal, M., Shabbir, M. S., Javed, S., & Shabbir, M. F. (2016). Measuring service quality and customer satisfaction in Pakistan: Evidence based on Carter Model. *International Business Management*, *10*(20), 5011–5016. https://doi.org/10.36478/ibm.2016.5011.5016

Fisman, R., & Svensson, J. (2007). Are corruption and taxation really harmful to growth? Firm level evidence. *Journal of Development Economics*, *83*(1), 63–75. https://doi.org/10.1016/j.jdeveco.2005.09.009

Garcia-Murillo, M. (2013). Does a government web presence reduce perceptions of corruption? *Information Technology for Development*, *19*(2), 151–175. https://doi.org/10.1080/02681102.2012.751574

Global Corruption Barometer (2020). Transparency International — Global corruption barometer 2020. Transparency.org. Retrieved from https://www.transparency.org/en/gcb/asia/asia-2020

Gouvea, R., Li, S., & Montoya, M. (2022). Does transitioning to a digital economy imply lower levels of corruption? *Thunderbird International Business Review*, *64*(3), 221–233. https://doi.org/10.1002/tie.22265

Grasmick, H. G., Tittle, C. R., Bursik, R. J., & Arneklev, B. J. (1993). Testing the Core Empirical Implications of Gottfredson and Hirschi's General Theory of Crime. *Journal of Research in Crime and Delinquency*, *30*(1), 5–29. https://doi.org/10.1177/0022427893030001002

Gujarati, D. N. (2022). *Basic econometrics*. Prentice Hall.

Hanafizadeh, M. R., Saghaei, A., & Hanafizadeh, P. (2009). An index for cross-country analysis of ICT infrastructure and access. *Telecommunications Policy*, *33*(7), 385–405. https://doi.org/10.1016/j.telpol.2009.03.008

Hartani, N. H., Cao, V. Q., & Nguyen, A. Q. (2020). Reducing corruption through e-government adoption, information and communication technology in Asean countries. *Journal of Security and Sustainability Issues*, *9*(May), 202–213. https://doi.org/10.9770/JSSI.2020.9.M(16)

Heeks, R. (1998). Information Technology and Public Sector Corruption. Information Systems for Public Sector Management Working Paper no. 4. http://dx.doi.org/10.2139/ssrn.3540078

Information technology union (2021), Digital trends in Asia and Pacific 2021 information and communication trends and development in the Asia-Pacific region, 2017-2020. http://handle.itu.int/11.1002/pub/81803ba9-en

Jeffery, C. R., & Zahm, D. L. (1993). Crime prevention through environmental design, opportunity theory, and rational choice models. In Clarke, R. V., & Felson, M. (eds), *Routine Activity and Rational Choice*, (pp. 323–350). https://doi.org/10.4324/9781315128788

Jha, C. K. (2020). Information Control, Transparency, and Social Media. In Information Resources Management Association (Ed.), *Media Controversy: Breakthroughs in Research and Practice* (pp. 399–417). IGI Global. https://doi.org/10.4018/978-1-5225-9869-5.ch023

Jha, C. K., & Sarangi, S. (2014). Social Media, Internet and Corruption. Departmental Working Papers 2014-03, Department of Economics, Louisiana State University.

Jorgenson, D. W., & Stiroh, K. J. (2000). U.S. Economic Growth at the Industry Level. *American Economic Review*, *90*(2), 161–167. https://doi.org/10.1257/aer.90.2.161

Kim, S., Kim, H. J., & Lee, H. (2009). An institutional analysis of an e-government system for anti-corruption: The case of OPEN. *Government Information Quarterly*, *26*(1), 42–50. https://doi.org/10.1016/j.giq.2008.09.002

Kouladoum, J. C. (2022). Technology and control of corruption in Africa. *Journal of International Development*. https://doi.org/10.1002/jid.3723

Lidman, R. (2011). Is the Internet Mightier than the Sword. In Piaggesi, D., Sund, K., & Castelnovo, W. (Eds.), *Global Strategy and Practice of E-Governance: Examples from Around the World* (pp. 338–354). IGI Global. https://doi.org/10.4018/978-1-60960-489-9.ch019

Lincényi, M., & Čársky, J. (2021). Research of citizens' behavior in a political campaign in searching for and monitoring political advertising in The Slovak Republic. *Insights into Regional Development*, *3*(1), 29–40. https://doi.org/10.9770/IRD.2021.3.1(2)

Lio, M.-C., Liu, M.-C., & Ou, Y.-P. (2011). Can the internet reduce corruption? A cross-country study based on dynamic panel data models. *Government Information Quarterly*, *28*(1), 47–53. https://doi.org/10.1016/j.giq.2010.01.005

Liu, X., Latif, Z., Danish, Latif, S., & Mahmood, N. (2021). The corruption-emissions nexus: Do information and communication technologies make a difference? *Utilities Policy*, *72*. https://doi.org/10.1016/j.jup.2021.101244

Longe, O. B., Bolaji, A. A., & Boateng, R. (2020). ICT for Development in Nigeria: Towards an Alignment with ICT4D 2.0 Goals. In Information Resources Management Association (Ed.), *Wealth Creation and Poverty Reduction: Breakthroughs in Research and Practice* (pp. 213–223). IGI Global. https://doi.org/10.4018/978-1-7998-1207-4.ch012

Maiti, D., Castellacci, F., & Melchior, A. (2020). Digitalisation and Development: Issues for India and Beyond. In Maiti, D., Castellacci, F., & Melchior, A. (eds), *Digitalisation and Development* (pp. 3–29). Singapore: Springer. https://doi.org/10.1007/978-981-13-9996-1_1

Mehmood, B., Shahid, A., & Ahsan, S. B. (2013). Covariates of international tourism in Asia: A fixed effects-Driscoll and Kraay approach. *Academicia: An International Multidisciplinary Research Journal*, *3*(9), 51–61.

Mouna, A., Nedra, B., & Khaireddine, M. (2020). International comparative evidence of e-government success and economic growth: technology adoption as an anti-corruption tool. *Transforming Government: People, Process and Policy*, *14*(5), 713–736. https://doi.org/10.1108/TG-03-2020-0040

Oliner, S. D., & Sichel, D. E. (2000). The Resurgence of Growth in the Late 1990s: Is Information Technology the Story? *Journal of Economic Perspectives*, *14*(4), 3–22. https://doi.org/10.1257/jep.14.4.3

Poliak, M., Baker, A., Konecny, V., & Nica, E. (2020). Regulatory and Governance Mechanisms for Self-Driving Cars: Social Equity Benefits and Machine Learning-based Ethical Judgments. *Contemporary Readings in Law and Social Justice*, *12*(1), 58. https://doi.org/10.22381/CRLSJ12120208

Remeikienė, R., Gasparėnienė, L., Chadyšas, V., & Raistenskis, E. (2020). Links between corruption and quality of life in European Union. *Entrepreneurship and Sustainability Issues*, *7*(4), 2664–2675. https://doi.org/10.9770/jesi.2020.7.4(7)

Russell, H. (2020). Sustainable Urban Governance Networks: Data-driven Planning Technologies and Smart City Software Systems. *Geopolitics, History, and International Relations*, *12*(2), 9. https://doi.org/10.22381/GHIR12220201

Sala-I-Martín, X., Bilbao-Osorio, B., Blanke, J., Crotti, R., Drzeniek Hanouz, M., Geiger, T., & Ko, C. (2012). The Global Competitiveness Index 2012-2013: Strengthening Recovery by Raising Productivity. In Schwab, K. (Ed.), *The Global Competitiveness Report 2012–2013 Full Data Edition*, Geneva: World Economic Forum. Available at https://www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2012-13.pdf

Sassi, S., & Ben Ali, M. S. (2017). Corruption in Africa: What role does ICT diffusion play. *Telecommunications Policy*, *41*(7–8), 662–669. https://doi.org/10.1016/j.telpol.2017.05.002

Shim, D. C., & Eom, T. H. (2008). E-Government and Anti-Corruption: Empirical Analysis of International Data. *International Journal of Public Administration*, *31*(3), 298–316. https://doi.org/10.1080/01900690701590553

Soper, D. (2007). ICT Investment Impacts on Future Levels of Democracy, Corruption, and E-Government Acceptance in Emerging Countries. *AMCIS 2007 Proceedings*. 227. https://aisel.aisnet.org/amcis2007/227

Sturges, P. (2004). Corruption, Transparency and a Role for ICT? *IJIE: International Journal of Information Ethics*, *2*(11), 1–9.

Suardi, I. (2021). E-Government, Governance and Corruption in Asian countries. *Emerging Markets: Business and Management Studies Journal*, *8*(2), 137–150. https://doi.org/10.33555/embm.v8i2.180

Transparency International. (2021). The global coalition against corruption: https://www.transparency.org/en/publications/annual-report-2021

van Ark, B., O'Mahony, M., & Timmer, M. P. (2008). The Productivity Gap between Europe and the United States: Trends and Causes. *Journal of Economic Perspectives*, *22*(1), 25–44. https://doi.org/10.1257/jep.22.1.25

Vasudevan, R. (2006). Changed governance or computerized governance? Computerized property transfer processes in Tamil Nadu (India). *2006 International Conference on Information and Communication Technologies and Development*, 101–109. https://doi.org/10.1109/ICTD.2006.301846

Welsch, H. (2008). The welfare costs of corruption. *Applied Economics*, *40*(14), 1839–1849. https://doi.org/10.1080/00036840600905225

Wescott, C. G. (2001). E-Government in the Asia-pacific region. *Asian Journal of Political Science*, *9*(2), 1–24. https://doi.org/10.1080/02185370108434189

World Bank Data. Retrieved from https://data.worldbank.org/

World Bank. (2020). Enhancing Government Effectiveness and Transparency, The Fight Against Corruption. https://doi.org/10.1596/34533

World Development Indicators. Retrieved from https://databank.worldbank.org/source/world-development-indicators

World Development Report. (2020). World Development Report 2020: trading for development in the age of global value chains. https://digitallibrary.un.org/record/3850531?ln=en

Worldwide Governance Indicator. Retrieved from https://databank.worldbank.org/source/worldwide-governance-indicators#

Žuffová, M. (2020). Do FOI laws and open government data deliver as anti-corruption policies? Evidence from a cross-country study. *Government Information Quarterly*, *37*(3), 101480. https://doi.org/10.1016/j.giq.2020.101480

Blockchain Technology for Tourism Post COVID-19

Mohd Norman Bin Bakri
Faculty of Information Science and Technology, Multimedia University, Malaysia

Han-Foon Neo
Faculty of Information Science and Technology, Multimedia University, Malaysia

Chuan-Chin Teo
Faculty of Information Science and Technology, Multimedia University, Malaysia

Abstract: During the pandemic, the tourism industry was one of the most severely impacted sectors. As vaccines are now widely available, each government is working to develop a system that can generate a digital vaccine certificate and PCR lab test result to verify that a person has been fully vaccinated or has a negative PCR test result, in order to allow them to enter business premises, travel overseas or cross state borders. However, the use of centralised systems in the development of the digital COVID-19 pass system results in a number of challenges, including the system's high susceptibility to failures, sluggish and inefficient information transmission, and vulnerability. The goal of this research is to offer a new digital COVID-19 pass based on the proposed "SmartHealthCard" blockchain technology. SmartHealthCard is a decentralised application (dApp) encrypting and hashing user data and safely storing it in a distributed database. Privacy preservation, GDPR compliance, self-sovereignty, KYC compliance and data integrity are featured. This initiative has the potential to benefit the public, healthcare professionals, service providers and the government. SmartHealthCard enables quick verification of tamper-proof COVID-19 tests/vaccines, aiding in COVID-19 transmission control while respecting the user's right to privacy.

Keywords: blockchain, tourism, COVID-19, decentralised, privacy.

Introduction

The worldwide pandemic of Coronavirus Disease 2019 (COVID-19) has caused tremendous negative effects on the lives of billions of people, with substantial health, societal, and economic implications. Despite the deployment of effective restrictive public health measures, such as tight travel restrictions, the spreading of COVID-19 persists (Pavli & Maltezou, 2021).

http://doi.org/10.18080/jtde.v11n3.764

Tourism is one of the most important economic sectors in the world, accounting for 7% of global commerce in 2019. Overall, tourism is the world economy's third largest export industry. This amounts to more than 20% of the Gross Domestic Product (GDP) in some countries. It is the most heavily impacted industry hit by the COVID-19 pandemic, which caused a global effect on economies, lives, public services and opportunities. International tourism has dropped by an astounding 73 percent in 2020 due to the COVID-19 pandemic, and demand for international travel remained low at the start of 2021 (Pavli & Maltezou, 2021).

The impact of the COVID-19 pandemic on the travel tourism business has been greatly underestimated since the sprouting of COVID-19 in China. Up until this day, tourism agencies and policymakers still do not have a comprehensive understanding of the crisis's consequences and potential, which can have a serious influence on the industry (Škare *et al.*, 2021). According to the World Travel and Tourism Council (WTTC), COVID-19 might put up to 75 million workers at risk of losing their jobs. The GDP loss from travel tourism might be as high as US$ 2.1 trillion in 2020. WTTC also expects a daily loss of one million jobs in the travel tourism sector (Škare *et al.*, 2021).

Even though the COVID-19 pandemic threat is receding, surges in the number of cases and new variants are on-going. Governments all around the globe are still dealing with the threats that have wreaked havoc on people's lives and economies in over 190 countries, resulting in over 82 million illnesses and 1.8 million deaths (Abid *et al.*, 2021).

Apart from mitigation strategies for COVID-19, the economic reopening plan has risen to the top of the priority list for all governments, businesses, and individuals. One of the problems facing public officials and governments is to efficiently govern their respective economies, open workplaces, permit travel, and avoid new outbreaks of disease.

Different technical alternatives, such as movement papers and tracking applications, are being investigated. However, all are vulnerable to deception and fabrication, and can have an impact on essential freedoms or be socially undesirable. More specifically, because of the nature of trackable applications, public concerns about privacy have been a roadblock to existing solutions. Because of Google/Apple contact tracing capabilities, privacy and secrecy of personal information are jeopardised. Furthermore, apps for Bluetooth-based traceability require the user's gadget to stay in an active broadcasting mode, which drains the battery. In the meantime, Bluetooth technology includes security flaws, such as a weak wireless interface and the identification and disclosure of physical hardware (Abid *et al.*, 2021). Furthermore, there is a considerable possibility of replay attacks on the trackable network that can generate widespread panic.

http://doi.org/10.18080/jtde.v11n3.764

Using a supposed "risk-free certificate"/"immunity passport"/"health certificate" or other secure health document is a potential option. The main concept is that a proof of vaccination may be used to create a certificate that exempts a person from the most stringent government requirements. Its goal is to provide a credential in a digital yet printable format that is tamper-proof and globally provable to anybody who has been vaccinated or has received an authorised PCR/antibody test result. This health credential enables public authorities to control access to sensitive or critical facilities, such as airports, schools, hospitals, workplaces, and other public places, while taking into account the remaining uncertainties about the virus, changing health policies, and the validity period of the test result. As compared to traceability applications, the health credential protects user privacy and will only be checked at frontiers (such as schools, hospitals and airports), saving the battery life of devices because it works offline and consumes no energy.

France, Italy, United States, United Kingdom, China, Estonia, Chile and Germany have all stated that they want to test such credentials. Unfortunately, many government-tested and deployed solutions give little technical specifics, making them difficult to fully comprehend or evaluate (Abid *et al.*, 2021). It is generally known, however, that some of the systems are centralised or dependent on third parties, posing security and privacy concerns.

In light of certain governments' interest and the emergence of a number of commercial alternatives, a scholarly examination of COVID-19 health credentials is required. It is critical, in particular, to give precise technological solutions and to identify existing limits in order for healthcare authorities to be properly informed. Furthermore, a large percentage of developing nations lack the technological and economic capabilities for such developments.

To aid in the fight against this global health crisis, blockchain technology has the advantage of playing a critical role in COVID-19 prevention and assisting in the implementation of government rules and standards while maintaining confidence among all parties. Indeed, due to its characteristics of resilience, integrity and transparency, the emerging Blockchain solution, which is an immutable, distributed, and tamper-proof record database with global computational groundwork (i.e., smart contracts), tends to provide effective COVID-19 solutions based on a great amount of trust and accuracy (Abid *et al.*, 2020).

Blockchain Technology

Blockchain is a system of recording information in a way that makes it difficult or impossible to change, hack or cheat the system. A blockchain is essentially a digital ledger of transactions that is duplicated and distributed across the entire network of computer systems on the blockchain. Each block in the chain contains a number of transactions, and every time a new transaction occurs on the blockchain, a record of that transaction is added to every

http://doi.org/10.18080/jtde.v11n3.764

participant's ledger. The decentralised database managed by multiple participants is known as Distributed Ledger Technology (DLT). Blockchain is a type of DLT in which transactions are recorded with an immutable cryptographic signature called a hash. This means, if one block in one chain was changed, it would be immediately apparent it had been tampered with. If hackers tried to corrupt a blockchain, they would have to change every block in the chain, across all of the distributed versions of the chain. Blockchains such as Bitcoin and Ethereum are constantly and continually growing as blocks are added to the chain, which significantly adds to the security of the ledger.

Bitcoin's blockchain is used in a decentralized way. However, private, centralized blockchains, where the computers that make up its network are owned and operated by a single entity, do exist. In a blockchain, each node has a full record of the data that has been stored on the blockchain since its inception. For Bitcoin, the data is the entire history of all Bitcoin transactions. If one node has an error in its data, it can use the thousands of other nodes as a reference point to correct itself. This way, no one node within the network can alter information held within it. Because of this, the history of transactions in each block that make up Bitcoin's blockchain is irreversible. If one user tampers with Bitcoin's record of transactions, all other nodes would cross-reference each other and easily pinpoint the node with the incorrect information. This helps to establish an exact and transparent order of events.

Because of the decentralized nature of Bitcoin's blockchain, all transactions can be transparently viewed by either having a personal node or by using blockchain explorers that allow anyone to see transactions occurring live. Each node has its own copy of the chain that gets updated as fresh blocks are confirmed and added. Blocks on Bitcoin's blockchain store data about monetary transactions. It turns out that blockchain is actually a reliable way of storing data about other types of transactions too. Some companies that have already incorporated blockchain technology include Walmart, Pfizer, AIG, Siemens and Unilever.

Blockchain technology has the potential to be utilized in various industries. By integrating blockchain into banks, consumers can see their transactions processed in as little as 10 minutes, basically the time it takes to add a block to the blockchain, regardless of holidays or the time of day or week. With blockchain, banks also have the opportunity to exchange funds between institutions more quickly and securely. In the stock trading business, for example, the settlement and clearing process can take up to three days (or longer, if trading internationally), meaning that the money and shares are frozen for that period of time. Health-care providers can leverage blockchain to securely store their patients' medical records. When a medical record is generated and signed, it can be written into the blockchain, which provides patients with the proof and confidence that the record cannot be changed. These personal

http://doi.org/10.18080/jtde.v11n3.764

health records could be encoded and stored on the blockchain with a private key, so that they are only accessible by certain individuals, thereby ensuring privacy.

Blockchain-related applications

This section describes four blockchain-related mobile apps to create a digital health/vaccine/immunity passport for users, healthcare providers and service providers.

AOKpass

AOKpass (ICC United Kingdom, 2021) is a blockchain-based platform and mobile app that allows a user to securely validate his or her health status with any third parties while maintaining the privacy of his/her underlying personal health data. Users retain complete control over their personal health data, such as health credentials and test results, which are solely saved on the user's smartphone that does not involve any external database or centralised system. This app is built on the Ethereum permissionless blockchain and employs distributed ledger technology. To secure user data and preserve system security, the AOKpass platform also uses hashing and encryption techniques.

AOKpass is based on the Ethereum public Blockchain network ("Ethereum"), which is a worldwide, open-source decentralised computing platform. Unlike centralised computing networks, which are supported and secured by a single or small number of private databases and/or servers (or "nodes"), Ethereum is supported and secured by a global decentralised public network of discrete nodes (the total number of nodes at any given time).

Ethereum is a platform for the deployment of decentralised applications (dApps), which are programs that operate and access the appropriate functional advantages of Blockchain technology through decentralised Blockchain networks.

A cryptographic 'hash' is a sophisticated digital signature, often known as a cryptographic 'proof', that is a crucial part of a Blockchain. The key characteristics of a hash are that it is obtained precisely from the certificate information; nevertheless, the certificate information cannot be derived backwards from the hash. The hash acts as a secure signature that can be confirmed by anybody who has been given an AOKpass. AOKpass uses the SHA 256 hashing algorithm, which is generally acknowledged as an industry standard.

AOKpass uses Amazon Web Services (AWS) as a third-party service provider to securely store a restricted set of data, which does not include AOKpass users' personal health information. The AOKpass security paradigm is built on two principles, simplicity and the storing and management of as little data as feasible. The backend infrastructure of the AOKpass system is made up of serverless AWS functionality and other provisioned AWS cloud services, effectively offloading any infrastructure-level security threat to AWS operations. The only possible threat

http://doi.org/10.18080/jtde.v11n3.764

is the compromise of the AWS credentials, or the credentials of a few other infrastructure services (AWS, Google Play, Apple Developer Account, Pulumi, Cloudflare). These credentials are kept on secure management platforms and only shared with those who have a need to know.

For the sensitive information that AOKpass does keep (such as attestor emails and names), AOKpass encrypts it utilising a cold encryption key that is not stored anywhere on the network. Without access to a cold private key, which is required to sign possible attestation sources, approval of attestors is likewise impossible. AOKpass's design, in addition to the security of its infrastructure, permits "third party" attestation providers to host their own attestation infrastructure if extra regulatory needs require it.

CommonPass

CommonPass (HandyVisas, 2020) is a digital health software application that allows users to show standardised, provable documentation where they have been vaccinated against COVID-19 or have tested negative for the virus. CommonPass creates a digital health pass that a user produces for border crossings and travel, or to indicate compliance with the venue or destination's health admission regulations, using secure and private processes.

The CommonPass app is free to download from Google Play Store or Apple App Store. It may be used by anybody to keep track of their test results and immunization status. If travelling with CommonPass, the destination may require a unique invitation code to certify that the user fulfils all of the relevant travel regulations.

For pass generation, CommonPass only uses the most current laboratory results from a particular supplier. If a CommonPass was produced during a previous test, it will remain available until the user deletes it. Other than the most recent test results, other test results may be retrievable and used in future editions of the program.

Only the mobile device has access to the CommonPass. Outside healthcare professionals, partners, contractors, or vendors are not allowed to share or store anyone's CommonPass. For troubleshooting purposes, only a de-identified (HIPAA compliant) version of the information is temporarily retained on the systems (two weeks for test results, and 30 days for vaccination records).

Because Personally Identifiable Information (PII) is just stored on the user's own device, the user has control over it. When a particular airline or venue needs to verify the user's status, CommonPass servers save it using a cryptographic algorithm that allows the system to react on the user's behalf only to people who know the user, not to strangers. After two weeks, the status is automatically removed from the servers.

http://doi.org/10.18080/jtde.v11n3.764

IBM Digital Health Pass

The use of the IBM Digital Health Pass (2021) helps to manage and regulate the sharing of COVID-19 health credentials. It only allows for the verification of COVID-19 immunisation records issued digitally by Digital Health Pass participants, including pharmacies, laboratories and clinicians. A user may keep track of COVID-19 health certificates stored in the wallet, which is safely encrypted on a smartphone, or print them out as secure QR codes. It is helpful for those who do not have access to a smartphone.

In business, a staff member may use the Digital Health Pass Verifier app to ensure the COVID-19 health certificate is valid when a user visits a participating company. They may also want a picture ID with the user's name and birth date to ensure the pass belongs to the owner. On the same device, adults can keep COVID-19 health certificates for youngsters or seniors under their dependency.

When a participating company scans the encrypted QR code, it should only be able to determine whether or not the pass is still valid, as well as any personal data that the user has given them permission to access for a COVID-19 test result or vaccine verification. It contains data such as the user's name and birth date that is required to validate their identity.

Digital Health Pass simply uses a unique combination of numbers and characters to represent personal health information, as a QR code. Users have complete control over their data, deciding what to share, with whom, and for what reason. Therefore, the encrypted wallet is inaccessible to IBM and other verifying companies.

Digital Health Pass is not a contact tracker or a location tracking app. It is a secure modern replacement to paper COVID-19 vaccination cards or test results that allows the users to manage and share that they have been tested negative or vaccinated for COVID-19 whenever they want.

CovidPass

During the COVID-19 pandemic, CovidPass (2020) is "health passport" software that promises to resurrect worldwide travel, major events, and gatherings without jeopardising health and safety or privacy. The goal is to allow healthy persons to travel or attend events while avoiding unaffordable total lockdowns. It is safe software that does not expose personal information by relying on insecure Bluetooth technologies. Instead, it uses a closed loop system with end-to-end encryption, making it hard to attack.

CovidPass provides a secure, safe, and long-term solution for re-allowing travel, worldwide tourism and large-scale events. It can assist in addressing the problems that these businesses have faced since the beginning of the pandemic. Indeed, the essential steps implemented thus

http://doi.org/10.18080/jtde.v11n3.764

far to combat the spread of COVID-19 have affected everyone, regardless of whether they are virus-free or infected and have thus had a significant impact on companies and economies throughout the world. This also includes many large-scale events, such as the Tokyo Summer Olympics and the UEFA Euro 2020, that have been postponed as well.

CovidPass provides a smooth and safe encryption solution for results of COVID-19 screening tests from accredited medical laboratories. Individuals who desire to travel or attend large events are provided with these test results, allowing them to offer the same results to the authorities or organisations that have sought access. Those who pass the serological or PCR test receive a secure QR code on their smartphone, which they may show at airline check-in counters, border crossings or event gates to certify their safe status.

Several unique characteristics of CovidPass include:

1) Technology: CovidPass stores encrypted data from COVID-19 screening tests using Blockchain Technology, providing a consistent, inalterable display of PCR or serological test findings.
2) Privacy: CovidPass is not a contact tracing program, but it solves users' privacy concerns. Its goal is to not only aid and promote the economy, but to also resolving people's worries about utilising contact tracing applications.
3) Expertise: CovidPass draws on the knowledge of specialists in the domains of medicine, consumer apps, government relations and tourism to build a complete, flexible solution.

CovidPass is a response to the current problems that the tourist and events industries are facing. It aims to restore trust in safe travel and social interactions by addressing the concerns of governments, corporations, and individuals. Millions of people will be able to return to flights, hotels, stadiums, conferences, and other venues as a result of this.

Research Method

The hardware requirement for this research is a Windows Desktop PC/Laptop, Apple Desktop, or MacBook to build and maintain the SmartHealthCard dApp system. The software requirements needed are draw.io, Node.js, Cloud MongoDB database and uPort Smart Contract.

As for user, the hardware and software needed for users is an Android OS mobile smartphone to install and register the uPort Mobile app. The user also needs to be registered in the SmartHealthCard system by a healthcare provider via the SmartHealthCard dApp.

Equivalent to the user, an issuer or a verifier needs to install and register with the uPort Mobile app. Additionally, they need to install and register with the SmartHealthCard dApp, which works on Windows Desktop PC/Laptop, Apple Desktop, and MacBook. It should be noted that

http://doi.org/10.18080/jtde.v11n3.764

both issuer and verifier must be registered and verified by Local Authorities first in order to install and register with the SmartHealthCard dApp.

Conceptual framework

This research proposes a new Blockchain-based privacy-preserving digital COVID-19 credential platform, SmartHealthCard, for issuing and confirming COVID-19 vaccine and PCR test certificates. SmartHealthCard seeks to stop COVID-19 from spreading while adhering to privacy regulations. For instance, it is compliant with the General Data Protection Regulation (GDPR) and Know Your Customer (KYC), as well as preserving user autonomy. The suggested method will be used not only for COVID-19 testing, but also for COVID-19 vaccinations, which are now accessible in several countries. The characteristics of the proposed conceptual framework are:

1) Privacy preservation: To reserve encrypted user data, including COVID-19 findings, SmartHealthCard uses an off-chain IPFS (Benet, 2014) storage (InterPlanetary File System) Infura (2022). Only the IPFS hash is saved on the Blockchain, ensuring that sensitive data is never exposed to those scanning the Blockchain.
2) General Data Protection Regulation compliance: SmartHealthCard is GDPR-compliant because it uses well-known data-protection standards, including JSON Web Tokens (JWT) (2013), ERC1056 Lightweight Ethereum Identity (Thorstensson, 2018), and W3C verifiable credentials (VC) (2022), ensuring that users remain in charge over their personal data.
3) Self-sovereignty: The user is the owner of his/her identities in SmartHealthCard and has full autonomy over his/her personal information. It enables the selective disclosure idea, which allows the user to exchange specific bits of data with specified trustworthy partners.
4) KYC-compliance: Because it checks the identification of various users before onboarding them, SmartHealthCard is KYC-compliant. This enables more reliable communication and collection of genuine data in real time. As a result, the suggested strategy would act to be the foundation for real-time supervision of the community health condition, as well as the progress of deconfinement and pandemic management.
5) Integrity: This research can confirm the genuineness of the digital COVID-19 credentials by comparing the hash value of the information supplied by the users and the one which is already recorded in the Blockchain ledger, because the hash value of the information is recorded immutably in the Blockchain.

http://doi.org/10.18080/jtde.v11n3.764

Research design

Application's installation, DIDs' generation, and Issuer and Verifier registration

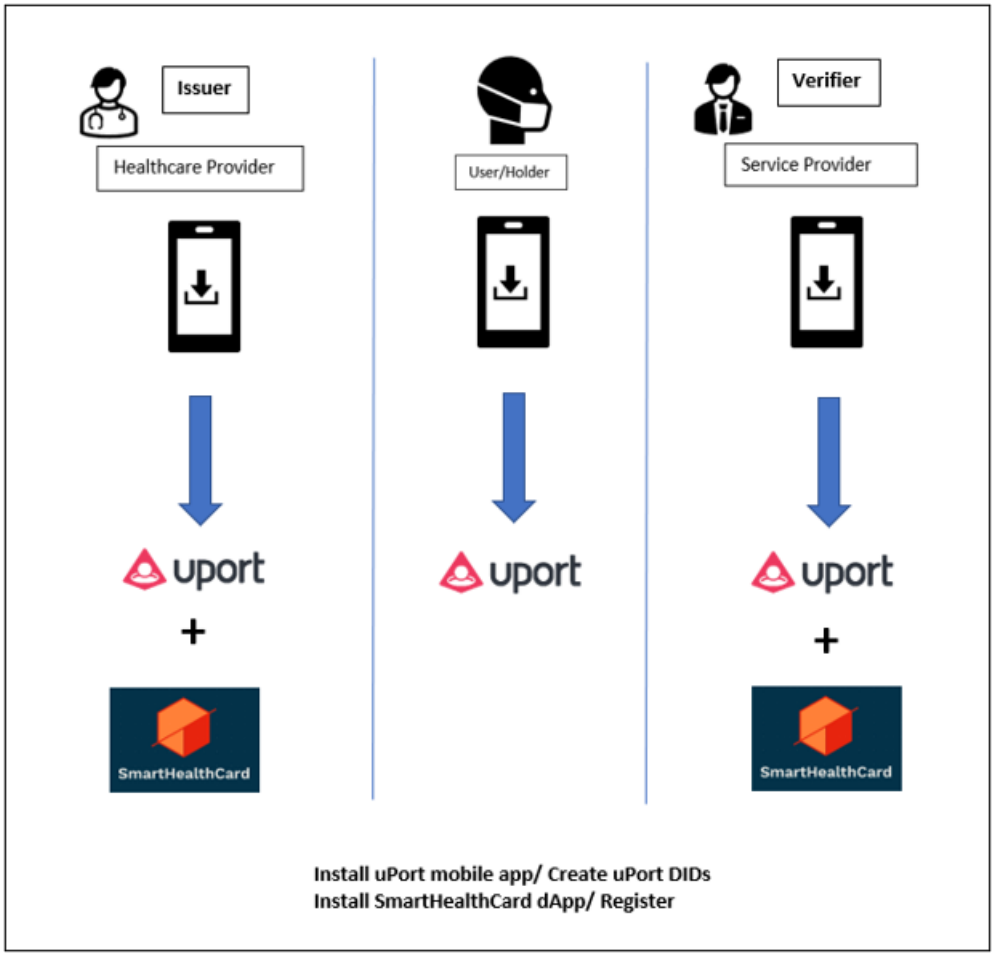


Figure 1. App configurations

Figure 1 depicts several SmartHealthCard configurations, including the generation of Decentralized IDs (DIDs) and registration of the verifier and issuer. This research assumes that all main actors have the SmartHealthCard and uPort mobile (Braendgaard, 2018) applications loaded before the COVID-19 vaccine/test and COVID-19 credential issuing steps:

- Both the verifier and issuer install and download the uPort Credential Mobile Wallet app, the self-sovereign identity Wallet, and give the authority with the personal information and uPort DID. After that, the authority verifies the issuer's or verifier's eligibility and registers the service/medical ID, uPort DID, and other personal information on the Ethereum Blockchain.
- Installing the SmartHealthCard dApp and logging in with his/her uPort DID enables the healthcare provider to obtain the "Issuer" role. The same steps are applied for the service provider to get authorisation for "Verifier" role.

http://doi.org/10.18080/jtde.v11n3.764

- The holder additionally downloads and registers for the uPort app from the Apple App Store or Google Play Store. Creating a uPort Identity is as simple as generating a standard Ethereum key pair account, where there are no gas charges and all Ethereum accounts are legitimate identities. Furthermore, uPort enables identities to be denoted as an object capable of doing tasks like validating messages from other DIDs, signing communications, and updating their DID-document. It enables the holders to regain access to their identity in the event of a broken or lost phone.

The holder retains full autonomy over his or her identity and all related information and will not lose access due to the loss of the private key. It is worth noting that, because the ERC-1056 standard is utilised, a holder just has to construct an Ethereum key pair and not a smart contract for key management or a transaction. As a result, the identity generation procedure is very quick and easy, whereby millions of identities may be generated in one day, guaranteeing strong alignment with a government-sponsored identity initiative.

Holder's access to SmartHealthCard

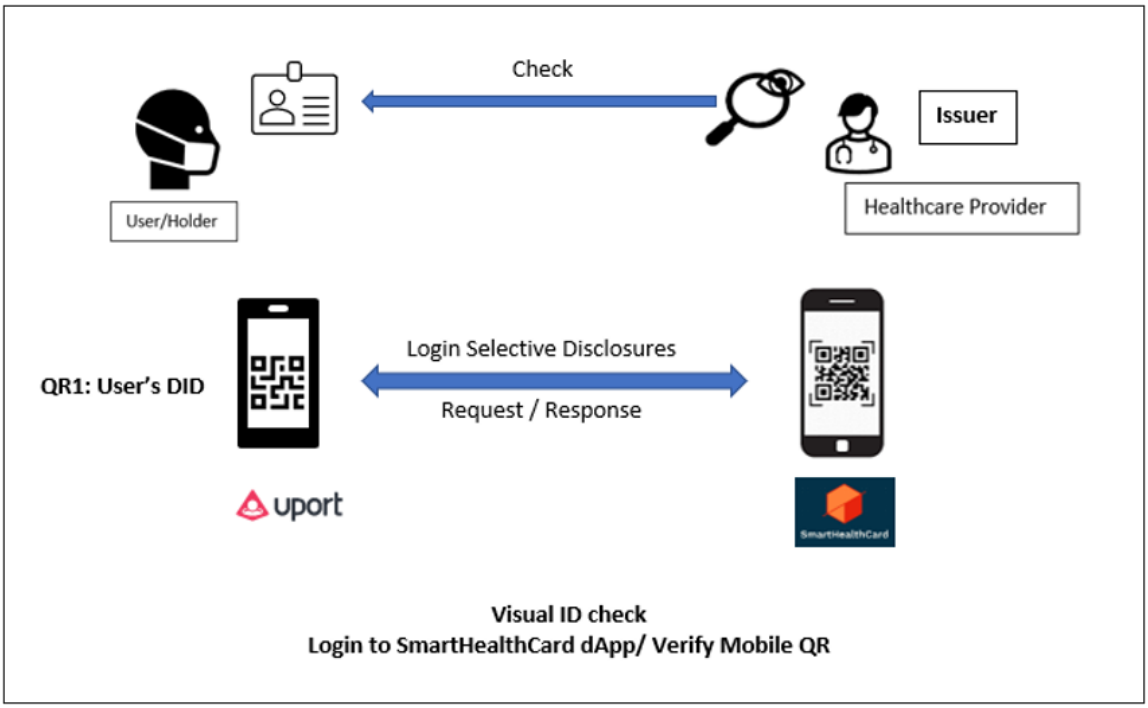


Figure 2. Interaction between holder and the apps

Figure 2 shows the holder's initial interaction with the SmartHealthCard dApp, which takes the form of a login selective disclosure response/request. A holder has now generated his or her uPort DID and is ready to go to the healthcare provider for the COVID-19 vaccine/test and get registered on the SmartHealthCard platform.

In order to do so, the issuer examines the holder's official physical ID (identity card or passport) and scans the uPort DID's QR code, "QR1", to engage with the holder's identification

http://doi.org/10.18080/jtde.v11n3.764

for the first time. Alternatively, the issuer examines to see if the holder already has a valid credential, because each holder may only have one valid certificate at a time. The holder must then log in to SmartHealthCard. The holder's uPort DID and personal information (for example, passport or identity card number) are requested, and the holder has the ability to accept or reject the request using his/her uPort mobile application. Thus, a selective disclosure response/request is triggered. It is the primary way of validating a holder's credentials, and therefore provides total authority over the personal data. After successfully logging in, the healthcare professional can run PCR or antibody tests, as well as administer a COVID-19 vaccination.

Issuing COVID-19 certificates

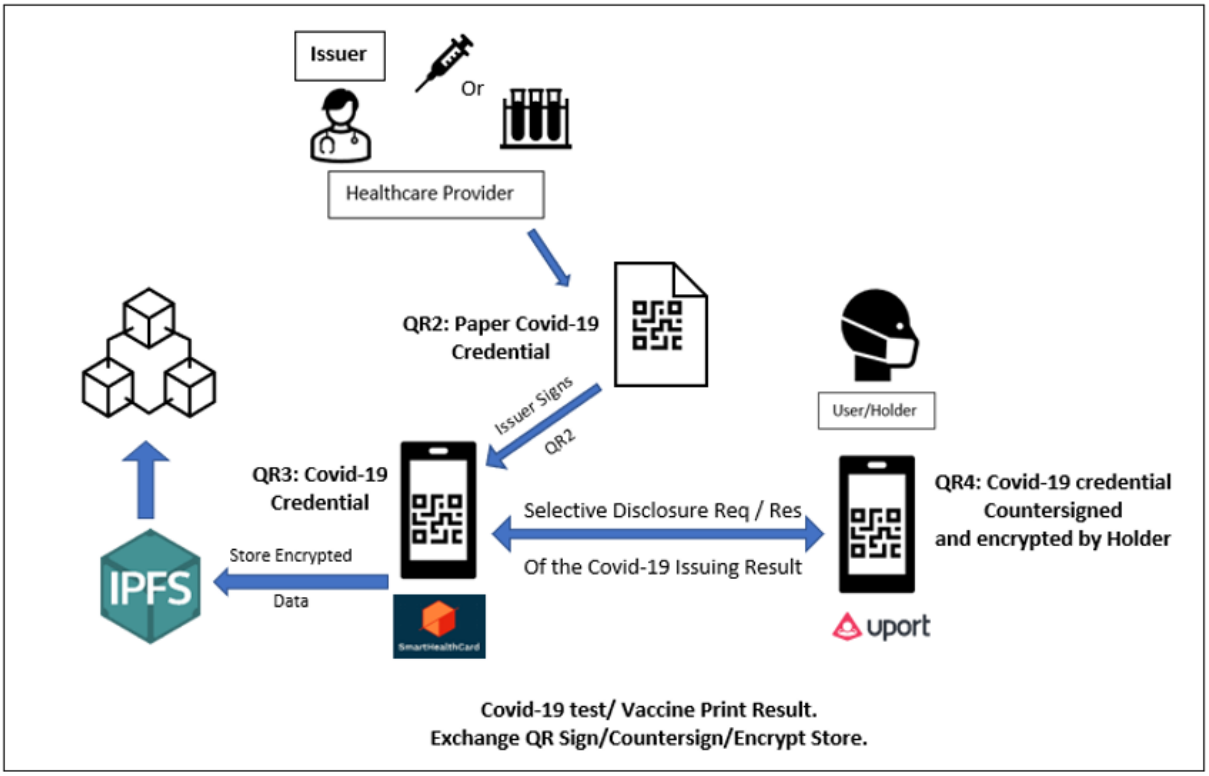


Figure 3. Interaction between holder and the apps

The process of issuing COVID-19 certificates is illustrated in Figure 3, where the issuer performs the COVID-19 vaccination or test. The result of a COVID-19 test is obtainable in about 48 hours for a PCR test and 15 minutes for an antibody test. In this case, a positive result indicates the absence of the virus or the existence of antibodies over a certain threshold. The issuer issues a printed paper version of the COVID-19 certificate once the result is ready. The issuer then uses the SmartHealthCard dApp to scan the printed QR code, "QR 2", to create a digitally signed vaccine/test result as a new QR code, "QR 3". Next, the "QR 3" is sent to the holder, who scans it using the uPort mobile app and digitally countersigns it as a recipient acceptance, resulting in the holder generating and owning a new QR code, "QR 4". To be more

http://doi.org/10.18080/jtde.v11n3.764

specific, this information is transmitted via a selective disclosure response/request, in which the user can confirm or deny.

Meanwhile, the COVID-19 credentials and signatures are encrypted by the holder. The holder also signs the request with his/her device's private key and transmits the result. When the issuer receives the selective disclosure answer, it preserves encrypted personal information, together with COVID-19 digital certificates, in a secure off-chain IPFS storage (Benet, 2014). Only the IPFS hash (SHA-256 hash) is saved on-chain as a data pointer, ensuring that sensitive information is never disclosed to anyone scanning the Blockchain. Lastly, the hash of the encrypted information reflects the QR code "QR 4", which belongs to the holder.

It is worth noting that the "QR 2" code and the printed QR code, which is not digitally signed, may act as a fallback version, backup or rescue version in the event of a lost or stolen mobile phone or a specific desire of the verifier or holder, particularly during initial familiarisation with the digital COVID-19 credential.

Verification of COVID-19 certificates

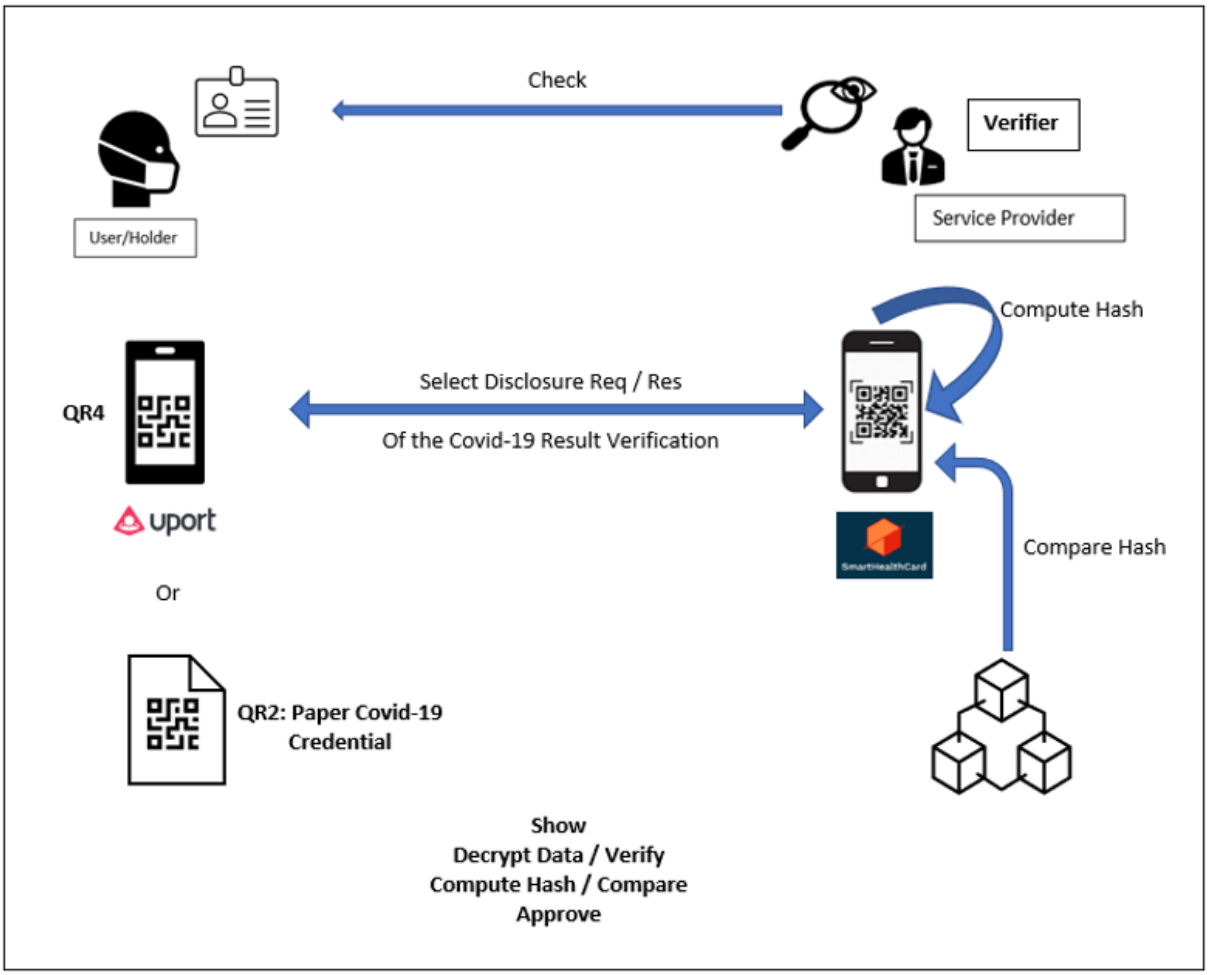


Figure 4. Certificate verification

http://doi.org/10.18080/jtde.v11n3.764

The process of verification of COVID-19 certificates is shown in Figure 4. The holder now possesses the counter-signed and signed COVID-19 vaccine/test certificates (QR 4) as well as a backup/rescue certificate (QR 2), that may offer the verifier with a provably legitimate or valid COVID-19 credential. To stop someone impersonating him or her, the holder must show not only the COVID-19 certificate, but also the evidence of identification. As a result, the holder/user must present the same valid physical ID that were used during the registration stage at the time of verification. The verifier must decode the holder data in order to validate COVID-19 credentials. He or she can then check the COVID-19 result, the physical ID number, the uPort DID and both signatures (holder and issuer).

The uPort process utilises the Box Public Key Authenticated Encryption Algorithm (Ecrypt, 2019) to decrypt holder/user data and offers an ERC-1098 cross-client method (Alabi, 2018) for requesting decryption/encryption, allowing the latest generation of decentralised apps to securely store users' private information in databases. In this technique, Ethereum key pairs must never be utilised directly for encryption; instead, the user must create a random ephemeral key pair for encryption and acquire an encryption key pair from the account's private key for decryption. To decrypt data, the verifier has to acknowledge the holder's secret key and acquire user approval.

Continuing the verification of COVID-19 certificates, the verifier computes the content hash ("QR4"), then matches it up to the hash recorded in the Blockchain, assuring permanence and data integrity. Lastly, the verifier can authenticate COVID-19 credential acceptance and broadcast it in a secure manner. In the end, the verifier/service provider may certify that COVID-19 credentials have been accepted and safely proclaim the user's admittance.

It is worth noting that different verifier institutions, such as hospitals, testing facilities, authorities and airline agents, can do the same operation. This would allow them to not just restrict access to public areas, but also to gain access to accurate data and create anonymised statistics. It would make it easier for the government to keep track of the population's health in real time.

http://doi.org/10.18080/jtde.v11n3.764

Architecture design

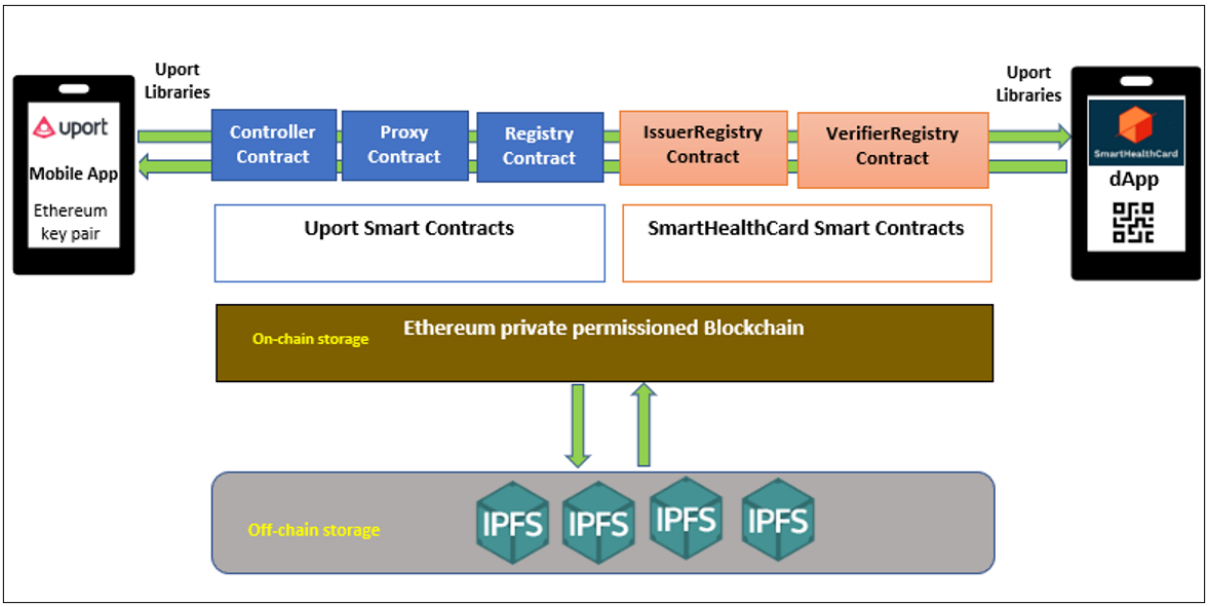


Figure 5. SmartHealthCard's architecture

Figure 5 shows the proposed SmartHealthCard platform architecture, which includes a uPort mobile app, uPort smart contracts, SmartHealthCard dApp and SmartHealthCard smart contracts. The SmartHealthCard system, in particular, uses uPort libraries and tools to build and control identities, as well as exchange and request certified information between them. The Decentralized Identifier (DID) specification is followed by uPort identities. Furthermore, a holder's personal information is encrypted and reserved off-chain in IPFS, with just the hash of the encrypted information being saved on-chain in the Ethereum private permissioned Blockchain, that will be used for confirmation process.

Verifiable credentials

The World Wide Web Consortium (W3C) developed the Verifiable Credentials (VC) standard to sort out digital certificates, authentications and claims in a safe and privacy-preserving way (W3C, 2022). The primary notions are built on the concept of Public Key Infrastructure (PKI). It is intended to standardise document format standards that make them machine-readable and communicative, as well as to generalise PKI, which is often expensive and centralised. The generalisation goes to a distributed or decentralised registry for cryptographic keys, often (not always) stored on a Blockchain since this enables each public key to have its own distinctive address, namely a Decentralized Identifier (DID).

http://doi.org/10.18080/jtde.v11n3.764

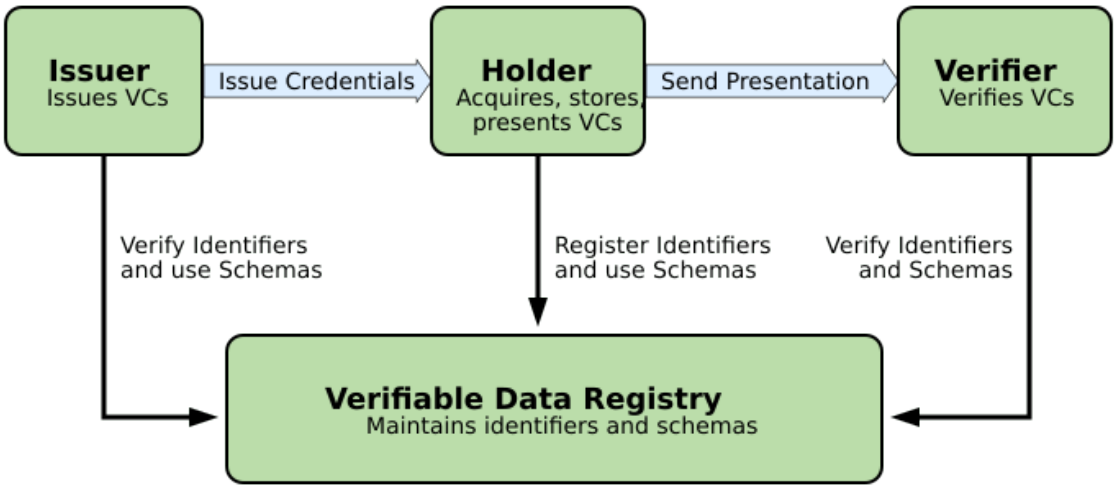


Figure 6. Information flow and responsibilities in Verifiable Credentials Model (adapted from W3C, 2022)

Figure 6 depicts the information flow and many responsibilities in the verifiable credential model (W3C, 2022). Through a verified data registry, the subject should establish globally unique IDs. The holder requests that the issuer create a VC by associating properties with identifiers. The topic of the VC they are preserving is generally, but not always, the holder. A parent, for example, may keep track of their children's VCs. The issuer checks the holder's identities and attributes, as well as its legal authority to hold the subject's VC, before issuing it. The issued VC must be kept by the holder. Finally, the holder may present the verifier with a provable appearance of his or her credentials. The issuer does not know the identities of the verifier in this model, which is a significant change from present identity management systems.

uPort

uPort is an Ethereum Blockchain-based user-centric information and self-sovereign identity platform. The uPort infrastructure consists of a self-sovereign wallet in a mobile app, a modern web application/decentralized application authentication mechanism and associated developer libraries. Figure 7 depicts the overall design and operation of the uPort identity handling system. For identity-related information, any app or user in uPort can communicate with an "Application Contract". This process has an impact on two primary contracts: 1) "Proxy Contract", which serves as an immutable and universal user identifier; 2) "Controller Contract", which manages identity access control logic. The app communicates with the "Proxy Contract" via the "Controller Contract", which transmits a request to the appropriate app. The "Proxy Contract" communicates with all application contracts on the Blockchain as a permanent identification and establishes a layer between application contracts and the user's private key (in the digital wallet).

http://doi.org/10.18080/jtde.v11n3.764

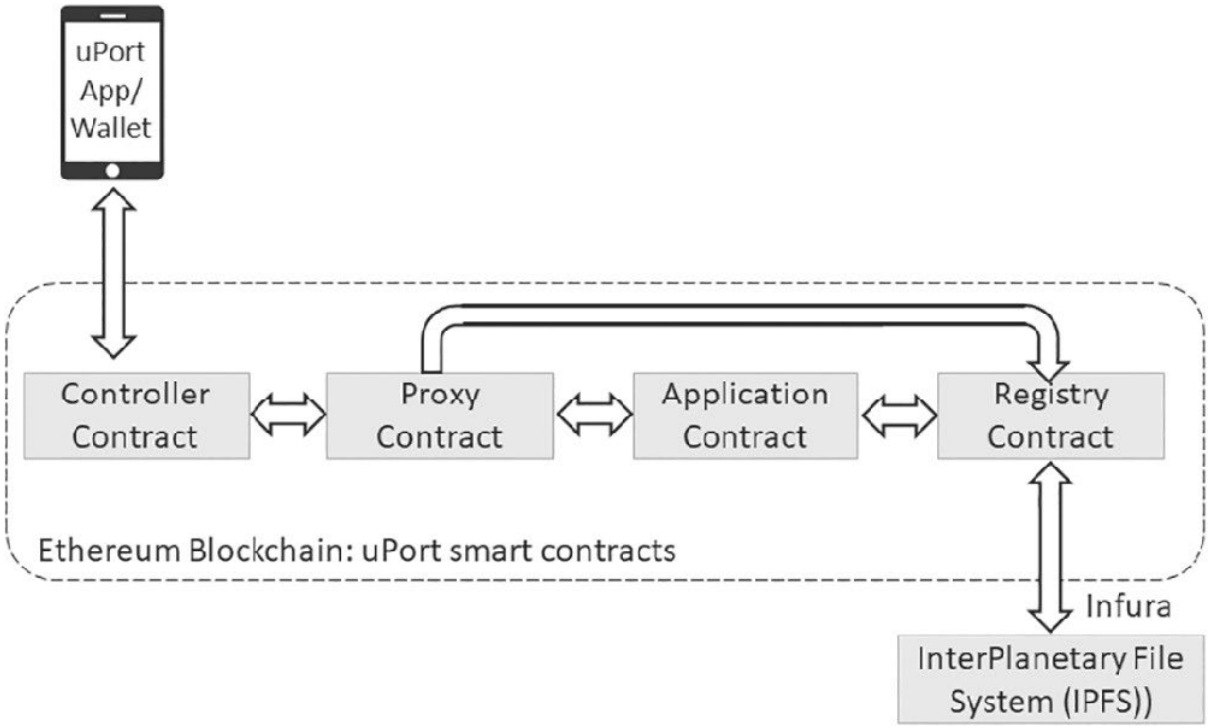


Figure 7. uPort Identity Handling System (adapted from Naik & Jenkins, 2020)

In order to do so, uPort makes use of Infura's standard RPC interface (2022), which provides an infrastructure for communicating with the Ethereum network. Furthermore, by submitting a transaction to the uPort Sensui server, which subsequently supplies appropriate Ether to pay the transaction fee, users can make a transaction without having any Ether in their wallet. Finally, data relating to uPort identification will be encrypted and reserved off-chain (which is on IPFS). This is accomplished by utilising a "Registry Contract" to create a cryptographic connection to an external data structure that can only be changed by the "Proxy Contract". To interact with the off-chain network, uPort requires the Infura interface.

It is worth noting that uPort is based on the W3C VC standard and includes additional methods to assist users to protect their personal information, for instance the notion of Selective Disclosure. Thus, the holder may choose whatever aspects of his or her VC to disclose with a verifier using this approach, while keeping the rest hidden. This is unquestionably a significant step forward in protecting users' rights to personal data privacy.

Implementation

SmartHealthCard dApp system components are shown in Figure 5. There are two forms of smart contracts, the smart contract for SmartHealthCard dApp and uPort Credential Mobile Wallet.

The registration procedure of the Issuer and Verifier is carried out on-chain to ensure transparency. However, only the Holder's data is encrypted and saved off-chain. Two smart contracts are created to accomplish this, the "IssuerRegistry" and the "VerifierRegistry"

http://doi.org/10.18080/jtde.v11n3.764

contract. The fundamental features of both contracts are built on events, which are used to alert authority listeners of what is happening. This also reduces on-chain costs and makes use of the Blockchain's immutable logs.

Furthermore, this research has created a modifier which guarantees that only the authorised Ethereum addresses are permitted to conduct the activities. The modifier is, in fact, a Solidity component that is applied to ensure that specific criteria are satisfied before performing a function. As a result, if access is refused, the function will not be triggered, and the Blockchain transaction that dealt with the call will be revoked.

As for uPort smart contracts, they contain a proxy, controller and registry contract. Proxy contract is the holders' permanent identity that is linked to their private key. This enables the holders to replace their private key without affecting their long-term identification. Controller contract is a component which restricts access to the proxy contract and enables the holders to restore their identity in the event that the holders' private key or mobile devices are lost. Registry contract is to establish a cryptographic connection between the holders' off-chain personal data and uPort identity.

SmartHealthCard dApp

We utilise uPort Credentials in SmartHealthCard to enable the generation and validation of identity data. This is an uPort library that facilitates secure communication between parties by enabling the activity of identity generation within the SmartHealthCard dApp and allowing data to be signed and verified. These bits of information, known as credentials, are presented as signed JSON Web Tokens (JWTs) (autho, 2013) and will be shown as QR codes. The uPort Transports library also helps to transport COVID-19 certificates between the SmartHealthCard dApp and the Holder through the uPort Credential Mobile Wallet app.

Obtaining SmartHealthCard dApp identity

To construct the SmartHealthCard server side using uPort-credentials, the first step is to create an application identity. As the identity is on the Ethereum blockchain, it complies with ERC-1056 protocol (Thorstensson, 2018). Hence, it is applicable to use it for signing requests.

It is important to remember that the private key should remain secret. It is just presented here for reference. This research uses sample application identities (e.g., private keys) to issue and verify credentials on a server. A sample of identity creation is shown in Figure 8.

http://doi.org/10.18080/jtde.v11n3.764

```
webpack 5.26.0 compiled with 8 warnings in 10429 ms
PS F:\Degree-3year-sem1\FYP\Report\Source Code\SmartHealthCard_Version2\dApp> node
Welcome to Node.js v16.13.2.
Type ".help" for more information.
> const { Credentials } = require('uport-credentials')
undefined
> Credentials.createIdentity()
{
  did: 'did:ethr:0x6cc3b48d3ac4d4bf04f8d52e69e9e0e8cc2c4de2',
  privateKey: '038d70d97f451cf33111fc36c319a575304c2f42a139b34aa17615f010efbfd8'
}
>
```

Figure 8. Obtaining SmartHealthCard dApp identity

Access to SmartHealthCard dApp

To access the app, the identification data is sought, and the sharing of the required information is approved by a uPort client, which acts for the Ethereum identity. This is known as a selective disclosure request. After the provided data has satisfied the SmartHealthCard server-side business logic, the holder will be regarded as authorised to use the validated certificates that he or she has agreed to share.

The SmartHealthCard server-side login service, which employs uPort for verification, consists of the following components:

- The production of a disclosure request message in the form of a JWT, which will be ingested by the mobile app and shown as a QR code;
- To return selective disclosure responses, which is named a callback server.

Generating and issuing COVID-19 certificates

An authorised healthcare practitioner should complete the generation and issuance of COVID-19 certificates. Holders will be able to construct their digital identities and offer actual values to the SmartHealthCard dApp by attesting facts about them. Furthermore, they may have a frictionless “evidence of being a person” validation across the decentralised web. In providing a certificate to a holder, the SmartHealthCard dApp will cryptographically sign a claim for that holder, so attesting to the veracity of a piece of information about the holder. Anybody with access to the DID of the SmartHealthCard application may then verify that a given identification certificate came from the SmartHealthCard dApp. For example, during onboarding, the SmartHealthCard dApp asks for and confirms a holder’s complete name, country, phone number or physical ID number, after which the user can acquire a COVID-19 outcome certificate.

Issuing a certificate at a high level entails, on behalf of the SmartHealthCard application, cryptographically signing user data; and holders can get their COVID-19 certificate as a JWT by scanning an issuing a QR code or receiving a push message.

http://doi.org/10.18080/jtde.v11n3.764

Requesting COVID-19 certificates

The uPort Credentials process is used to request COVID-19 certificates from the SmartHealthCard dApp. An authorised service provider should complete this activity. Requesting COVID-19 certificates follows the exact steps as submitting a disclosure request. Requesting a verification entails, at a high level, on behalf of the SmartHealthCard dApp, cryptographically signing a request to expose the holder's information and send a JWT request to the holder using a verification QR code or a push notification.

COVID-19 certificate request encryption and decryption

The uPort process employs the ERC 1098 encryption mechanism (Alabi, 2018) that employs an ephemeral transmitting key and box method tweet-nacl (Ecrypt, 2019) approach. It enables the Verifier to decrypt the message without first resolving the holder's public key.

The Holder should use this encryption technique:

1) Make the signed JWT payload as usual;
2) JWT is padded to the nearest multiple of 64 bytes using \os;
3) Using nacl.box.keyPair(), make an ephemeral keypair;
4) Using nacl.randomBytes(nacl.box.nonceLength), generate a random nonce of 24 bytes;
5) Use nacl.box(ephemeralKeyPair.secretKey, recipient publicKey, nonce, message) to encrypt the resultant JWT;

In a JSON payload, combine the base64 encoded versions of the ciphertext values, ephemPublicKey and nonce as well as the version of x25519-xsalsa20-poly1305; the Verifier needs to recognise the Holder's secretKey and needs to apply the following technique to decrypt the request:

1) Verify if the version field contains the string x25519-xsalsa20-poly1305 to proceed;
2) Decode the base64 encoded ciphertext attributes, ephemPublicKey and nonce;
3) use nacl.box.open(receiverEncryptionPrivateKey, ephemPublicKey, nonce, ciphertext) to decrypt the message;
4) Remove any trailing \os in the payload;
5) Decode JWT in the usual way.

Screen Interfaces

Screen interfaces of the issuing and verification process done by the issuer and verifier through the SmartHealthCard dApp, which supports the self-sovereign uPort Credential Mobile Wallet app on user's smartphone, are shown in this section (Figures 9–14). Subsequently, the COVID-19 credential verification process by the healthcare provider/issuer at the hospital is shown in

http://doi.org/10.18080/jtde.v11n3.764

Figures 15–18. Figure 19 shows the COVID-19 credential verification process by the service provider/verifier at the airport when the result is invalid (expired) due to exceeding the valid time (which is only four minutes).

Figure 9. Homepage

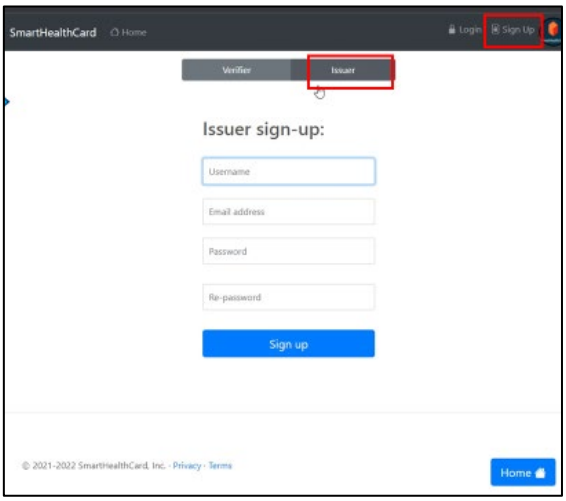


Figure 10. Issuer and verifier sign-up page

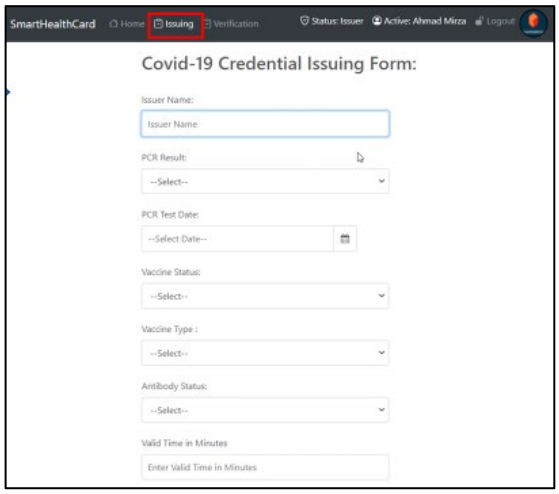


Figure 11. Credential creation

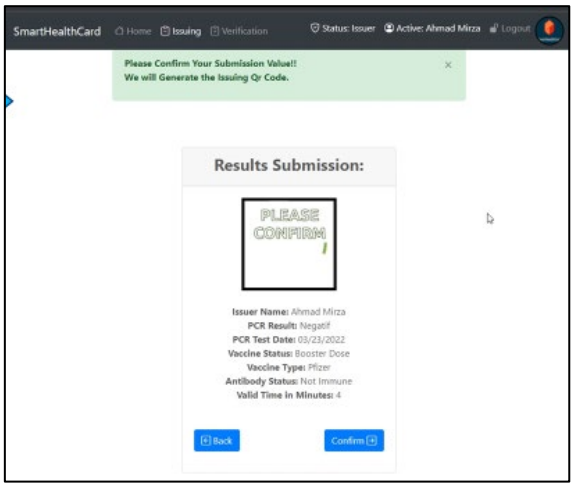


Figure 12. Credential confirmation

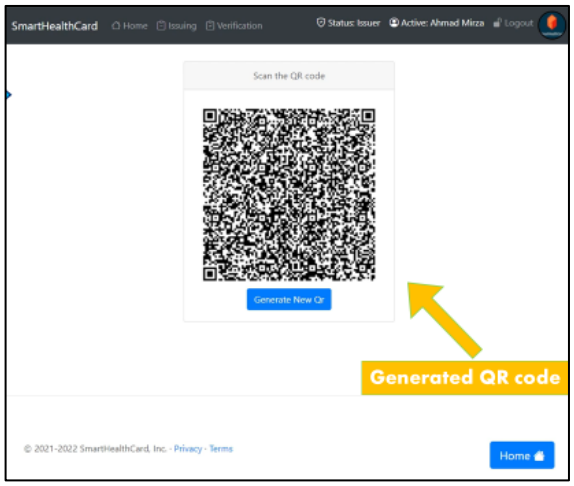


Figure 13. User scans the issuing QR code

http://doi.org/10.18080/jtde.v11n3.764

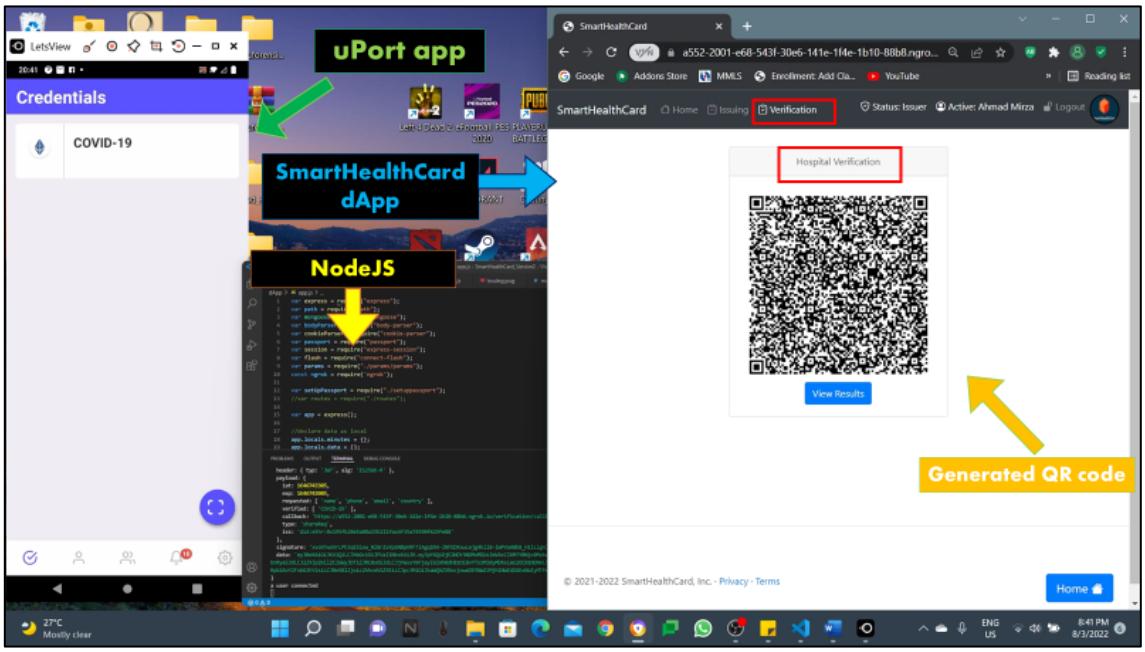


Figure 14. Credential is stored in uPort mobile wallet

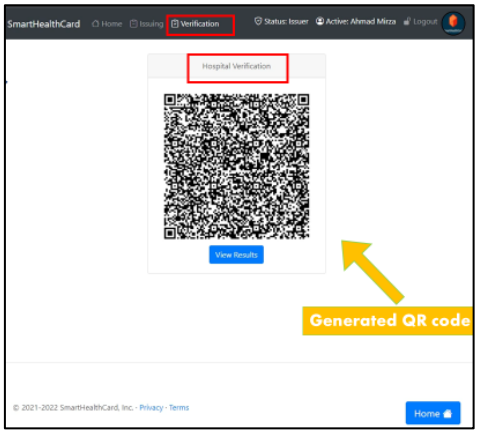


Figure 15. Hospital verification QR code page generated by the issuer

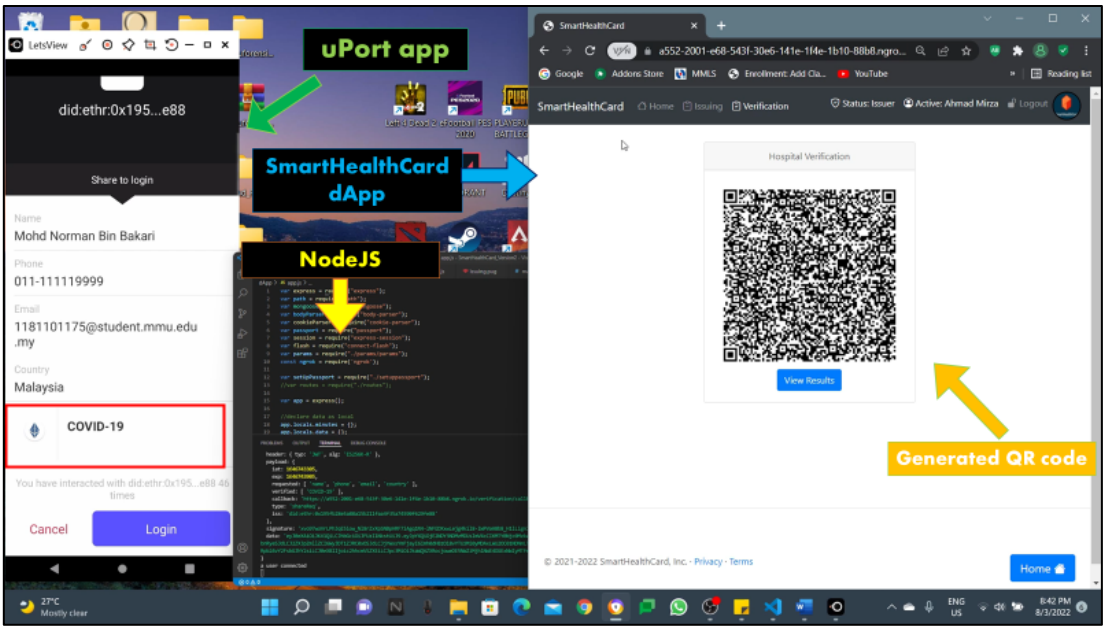


Figure 16. User scans the hospital verification QR code

http://doi.org/10.18080/jtde.v11n3.764

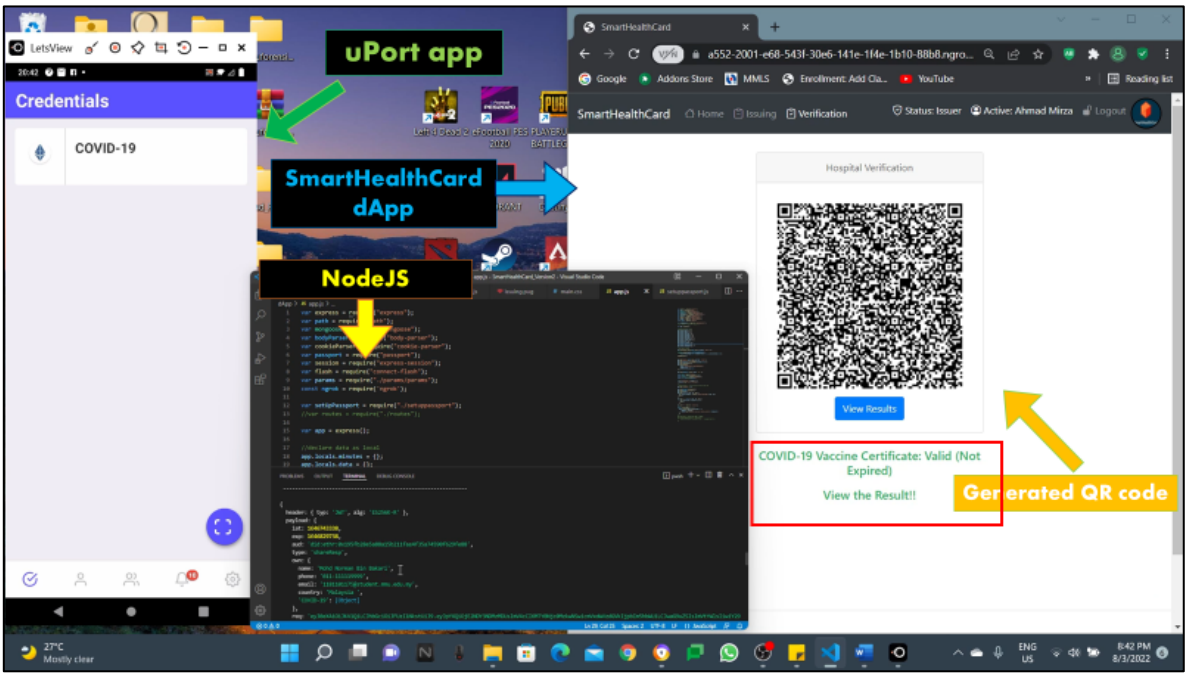


Figure 17. The credential is verified and valid

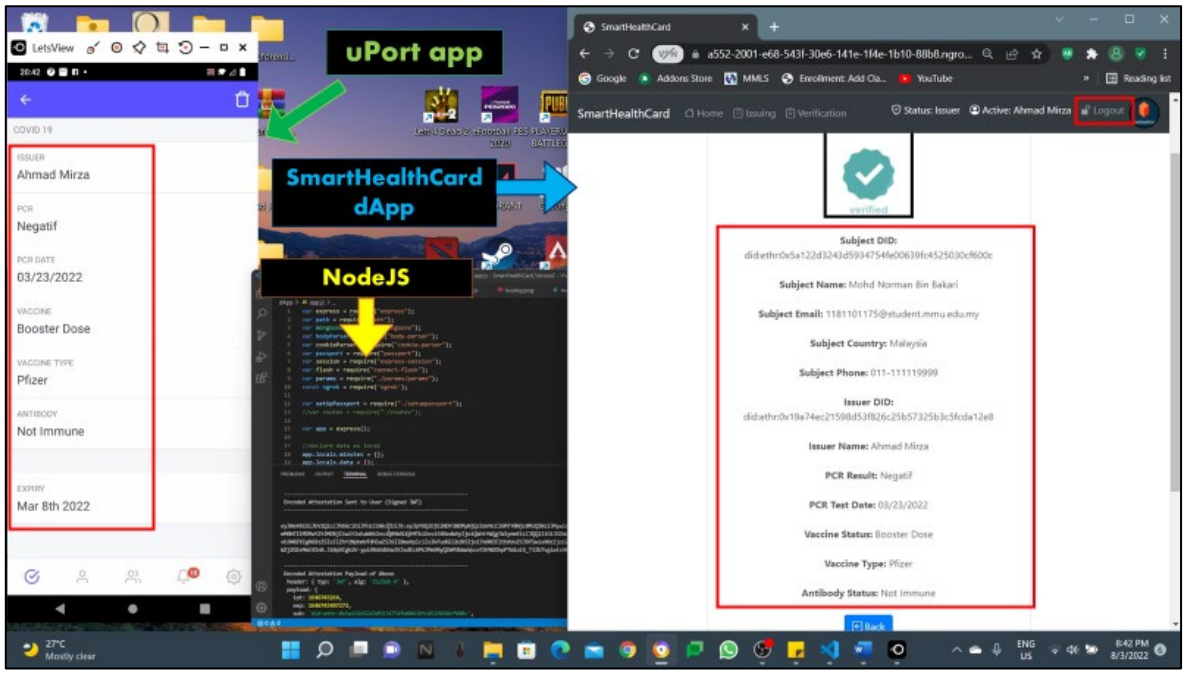


Figure 18. Issuer views the verified credential

Conclusion

When the COVID-19 pandemic illness spread at an unprecedented rate throughout the world in 2021, both major and small economic sectors experienced the effects of government-imposed limitations and regulations, such as social distancing and movement control orders. The tourism industry was one of the most affected economic sectors. As vaccines become more widely available, each government has been working to develop a system that can generate a digital vaccine certificate and PCR lab test result to verify that a person is fully vaccinated or

http://doi.org/10.18080/jtde.v11n3.764

has a negative PCR test result, in order to allow them to enter business premises, travel, cross state borders, and a variety of other activities. Each country will be able to reclaim its business activities, which have been harmed for several years. However, the use of centralised systems in the development of the digital COVID-19 pass system results in a number of issues and limitations, including the system's high sensitivity to failures, slow and inefficient information exchange, and vulnerability in data security and privacy protection for users.

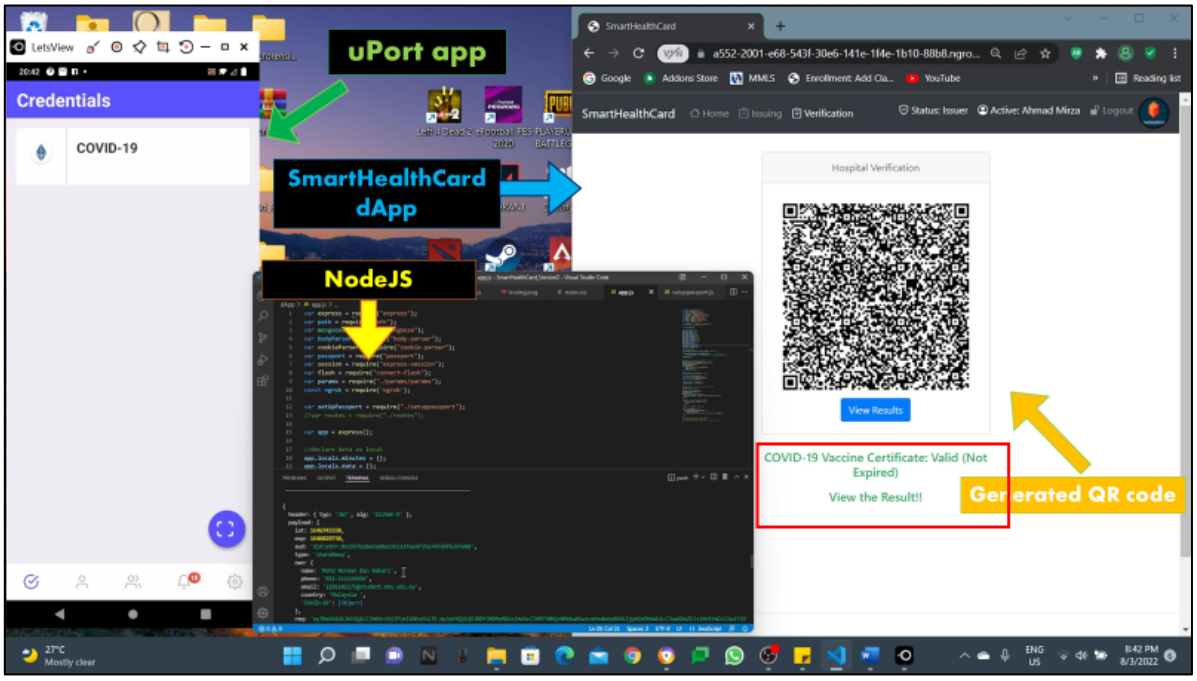


Figure 19. The credential is verified by the verifier and invalid (expired)

As a result, the goal of this research is to offer a new digital COVID-19 pass that uses the "SmartHealthCard" blockchain-based system solution. SmartHealthCard is a decentralised application (dApp) that replaces the old, centralised approach by encrypting and hashing user data and safely storing it in a distributed database. Privacy preservation, GDPR compliance, self-sovereignty, KYC compliance, and data integrity are additional characteristics of SmartHealthCard. This initiative has the potential to benefit the user, healthcare professional, service provider, and the government. The suggested platform enables quick validation of tamper-proof COVID-19 tests/vaccinations, aiding in COVID-19 transmission control while respecting the user's right to privacy.

In principle, a secure COVID-19 credential would serve as evidence that someone has been vaccinated against COVID-19, recovered from COVID-19 or tested negative in a COVID-19 PCR test. Thus, this facilitates safe, unrestricted travel while also removing a person from most government controls. Lastly, this secure COVID-19 certificate may aid public authorities in limiting access to vital or sensitive institutions, such as airports, schools, hospitals, and other public places.

http://doi.org/10.18080/jtde.v11n3.764

Acknowledgement

A version of this paper was presented at the third International Conference on Computer, Information Technology and Intelligent Computing, CITIC 2023, held in Malaysia, 26–28 July 2023.

References

Abid, A., Cheikhrouhou, S., Kallel, S., & Jmaiel, M. (2020). How blockchain helps to combat trust crisis in COVID-19 pandemic? *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, 764–765. https://doi.org/10.1145/3384419.3430605

Abid, A., Cheikhrouhou, S., Kallel, S., & Jmaiel, M. (2021). NovidChain: Blockchain-based privacy-preserving platform for COVID-19 test/vaccine certificates. *Journal of Software: Practice and Experience*, *54*(4), 841–867. https://doi.org/10.1002/spe.2983

Alabi, T. (2018). Add web3.eth.encrypt and web3.eth.decrypt functions to JSON-RPC. *Ethereum/EIPs #1098*. Retrieved from https://github.com/ethereum/EIPs/pull/1098

auth0. (2013). JSON Web Tokens. Retrieved from https://jwt.io/

Benet, J. (2014). IPFS - Content Addressed, Versioned, P2P File System. *Computer Science: Networking and Internet Architecture*, ArXiv. https://doi.org/10.48550/arXiv.1407.3561

Braendgaard, P. (2018). Next Generation uPort Identity App released. Retrieved from https://medium.com/uport/next-generation-uport-identity-app-released-59bbc32a83a0

CovidPass. (2020). Balancing Public Safety & Re-Opening Borders to Travellers. Retrieved from https://www.covid-pass.tech/

Ecrypt. (2019). Public-key authenticated encryption: crypto_*box*. *NaCl: Networking and Cryptography library*. Retrieved from http://nacl.cr.yp.to/box.html

HandyVisas. (2020). CommonPass Health App to Facilitate Travel in 2021. Retrieved from https://www.handyvisas.com/news/commonpass-travel-health-app/

IBM Digital Health Pass. (2021). IBM Watson Health is now Merative. Retrieved from https://www.ibm.com/my-en/products/digital-health-pass/individuals

ICC United Kingdom. (2021). ICC AOKpass — General Overview. Retrieved from https://iccwbo.uk/products/icc-aokpass-general-overview

Infura (2022). "Every Blockchain Journey Begins with a Single Step". Retrieved from https://infura.io/

Naik, N., & Jenkins, P. (2020). uPort Open-Source Identity Management System: An Assessment of Self-Sovereign Identity and User-Centric Data Platform Built on Blockchain. *2020 IEEE International Symposium on Systems Engineering (ISSE)*, 1–7. https://doi.org/10.1109/isse49799.2020.9272223

Pavli, A., & Maltezou, H. C. (2021). COVID-19 Vaccine Passport for Safe Resumption of Travel. *Journal of Travel Medicine*, *28*(4). https://doi.org/10.1093/jtm/taab079

Škare, M., Soriano, D. R., & Porada-Rochoń, M. (2021). Impact of COVID-19 on the Travel and Tourism Industry. *Technological Forecasting and Social Change*, *163*, 120469. https://doi.org/10.1016/j.techfore.2020.120469

Thorstensson, J. (2018). ERC: Lightweight Identity. *Ethereum/EIPs #1056*. Retrieved from https://github.com/ethereum/EIPs/issues/1056

W3C. (2022). Verifiable Credentials Data Model 1.0. Retrieved from https://www.w3.org/TR/vc-data-model/

Building a Fortress Against Fake News

Harnessing the Power of Subfields in Artificial Intelligence

Nafiz Fahad
Faculty of Science & Technology, American International University-Bangladesh, Dhaka, Bangladesh

Kah Ong Michael Goh
Faculty of Information Science & Technology (FIST), Multimedia University, Melaka, Malaysia

Md Ismail Hossen
Faculty of Science & Technology, American International University-Bangladesh, Dhaka, Bangladesh

Connie Tee
Faculty of Information Science & Technology (FIST), Multimedia University, Melaka, Malaysia

Md Asraf Ali
Faculty of Science & Technology, American International University-Bangladesh, Dhaka, Bangladesh

Abstract: Given the prevalence of fake news in today's tech-driven era, an urgent need exists for an automated mechanism to effectively curb its dissemination. This research aims to demonstrate the impacts of fake news through a literature review and establish a reliable system for identifying it using machine (ML) learning classifiers. By combining CNN, RNN, and ANN models, a novel model is proposed to detect fake news with 94.5% accuracy. Prior studies have successfully employed ML algorithms to identify false information by analysing textual and visual features in standard datasets. The comprehensive literature review emphasises the consequences of fake news on individuals, economies, societies, politics, and free expression. The proposed hybrid model, trained on extensive data and evaluated using accuracy, precision and recall measures, outperforms existing models. This study underscores the importance of developing automated systems to counter the spread of fake news and calls for further research in this domain.

Keywords: machine learning (ML), hybrid model, automated system, accuracy, fake news.

http://doi.org/10.18080/jtde.v11n3.765

Introduction

In today's digital age, the identification of fake news has emerged as a pressing concern. The proliferation of social media platforms as primary sources of news consumption and sharing has led to the rapid dissemination of both genuine and fake stories, posing serious consequences for society. Differentiating between various forms of false news on platforms like Twitter remains a major obstacle in the effective detection of fake news (Altheneyan & Alhadlaq, 2023). This is why the problem of fake news is a significant issue that puts the credibility of social networks at risk (Rawat *et al.*, 2023). Detecting fake news manually is not possible, which is why an automatic system is required. This system is created using the subfields of artificial intelligence (AI). Similarly, demonstrating the effects of fake news automatically is not feasible; therefore, a systematic literature review (SLR) is necessary to illustrate the impacts of fake news.

Objectives

Our work aims to prevent fake news by using the power of AI's subfields. In essence, our target is to make an efficient automated system to prevent fake news and our objectives are:

1. To create a dependable and precise system that uses machine learning (ML) classifiers to effectively identify fake news.
2. To make a hybrid model with the help of CNNs (convolution neural networks), RNNs (recurrent neural networks), and ANNs (artificial neural networks). No one has done this before.
3. To find out the impacts of fake news with achieve higher accuracy, precision, recall and F_1 scores, which would ensure that the system can detect most fake news accurately.

Literature Review

In the realm of literature, numerous publications exhibit a keen fascination with the identification of fake news.

Altheneyan & Alhadlaq (2023) said that the proliferation of social media platforms as primary sources of news consumption and sharing has led to the rapid dissemination of both genuine and fake stories. The prevalence of misinformation on these platforms has serious consequences for society. Detecting and differentiating between various forms of false news on platforms like Twitter pose a major obstacle to effective fake news detection, which is why the researchers used distributed learning to detect fake news based on ML. In their study, they used a distributed Spark cluster to construct a stacked ensemble model which achieved 93.40% accuracy; the highest among all the approaches (Altheneyan & Alhadlaq, 2023). In

http://doi.org/10.18080/jtde.v11n3.765

contrast, Rawat *et al.* (2023) revealed that using supervised ML algorithms and suitable tools facilitates the differentiation of false information from authentic news by categorising them accordingly (Rawat *et al.*, 2023). They also recognised the potential of using ML techniques as a possible solution for detecting fake news. In contrast, Sharma *et al.* (2023) used Hybrid Ensemble Model with Fuzzy Logic, which achieved 86.8% accuracy for the two datasets in their proposed model.

Singhal *et al.* (2019) developed a framework called SpotFake to identify fake news, which integrates language models with a pre-trained VGG-19 model on ImageNet to incorporate contextual information. The concatenation technique combined text and visual features to create a multimodal fusion module. The investigation revealed accuracy rates of 77.77% and 89.23% on the publicly available Twitter and Weibo datasets, respectively. Building upon the ideas presented in Singhal et al. (2019), the authors introduced SpotFake+ (Singhal *et al.*, 2020), an enhanced version of SpotFake that utilised transfer learning to extract semantic and contextual information from lengthy news articles and images.

Aslam *et al.* (2021) focused on developing an ensemble model based on deep learning techniques, specifically designed for identifying and detecting fake news. The proposed research aims to stop the spread of rumours and fake news by automatically categorising news articles, enabling people to determine if a news source is reliable or not. They developed an ensemble-based deep learning model to categorise news articles as either fake or real. The dataset underwent pre-processing techniques, and natural language processing (NLP) techniques were specifically applied to the statement attribute. Two different deep learning models were used: a deep learning dense model for nine attributes excluding the statement, and a Bi-LSTM-GRU-dense deep learning model for the statement attribute. The results of the study were highly significant, achieving an accuracy of 0.898 when using the statement feature. Although their proposed study has yielded noteworthy outcomes, there is still room for improvement. Further investigations are necessary to test the model using additional datasets of fake news.

Ahmad *et al.* (2022) worked on developing an improved deep-learning model to create an efficient mechanism for detecting fake news. Their goal was to extensively investigate the challenges associated with automatically detecting rumors on social media. To achieve this, we employ a novel combination of content-based and social-based features specifically designed for identifying rumours. Additionally, we employ a bidirectional LSTM-RNN classifier, a deep learning model, to analyse text data in order to enhance the accuracy of our rumour detection system. Amad *et al*'s bidirectional LSTM-RNN classifiers detect fake news or true news properly. A limitation of their approach is that more comprehensive testing is required to gain a better understanding of how deep learning can effectively detect rumours.

http://doi.org/10.18080/jtde.v11n3.765

Additionally, the presence of a large amount of unlabelled data on social media poses a challenge, and developing models that can work without relying on labelled data becomes necessary.

Although the previously mentioned studies focused on extensive findings regarding the identification of fake news, there is still a need for further updates and expansion due to the significant increase in the volume of topical publications. It is also evident that no one can ensure that the system can detect most fake news accurately and help humans protect themselves from fake news, whereas our study achieves these issues.

SLR to reveal impacts of fake news

Using a methodical examination of published works, an SLR proves to be a highly efficient strategy for uncovering the impacts of fake news. The following table demonstrates the data we have gathered from diverse research papers to demonstrate the impacts of fake news.

Table 1. Impacts of fake news

Serial Number	Citation	Extracted information
1	Ahinkorah *et al.*, 2020	People argue that different types of purposeful damage, and different rewards like money, social recognition and political advantages, often motivate the spread of false information.
2	De Oliveira & Albuquerque, 2021	False information can directly impact a person's ability to stay alive.
3	Leeder, 2019	Students are led astray by false information.
4	Shu *et al.*, 2019	Fake news was created to confuse people and cause doubt, making it harder for them to tell what is true.
5	Bakir & McStay, 2018	The use of emotions to grab people's attention and make money for advertisers is a key factor in the issue of fake news. It also highlights the pressures to create automated fake news that caters to the emotions and behaviours of online social groups.
6	Zafarani *et al.*, 2019	False information impacts financial markets and also leads to significant trade disruptions in economies.
7	Sullivan, 2019	People cannot tell if information is trustworthy, so they believe fake news. As a result, people who want to take advantage of the situation spread wrong information online, which can be dangerous and have a frightening impact on real people.
8	Bago *et al.*, 2020	Bigger changes in society and politics, along with the right to express oneself, have encountered difficulties due to the spread of false information.
9	Almenar *et al.*, 2021	The types of wrong information that people receive also differ based on gender. We know that men and women have different behaviours on social media and when consuming news, but these differences are not as noticeable when it comes to false information. The main idea is that, as in other aspects of life, women tend to worry more than men because of the spread of false information.
10	Butler *et al.*, 2018	Fake news is a big issue that makes it hard for people to trust real news sources. It is a sizeable problem because it also makes it harder for the government to be trusted. Fake news hurts real news sources by rendering them seem less reliable, which means

http://doi.org/10.18080/jtde.v11n3.765

Serial Number	Citation	Extracted information
		people are not equipped with the right information to be involved in a democracy.
11	Stewart, 2021	Fake news can cause harm by spreading violent threats and misleading information that can hurt people mentally or in other ways. It can also be dangerous when it misleads the public about important matters like elections or health.
12	Naeem *et al*, 2021	Reading false information can make people mentally unwell, and it can even put their health in danger.
13	Lakshmanan *et al.*, 2019	False information occasionally discourages individuals.
14	Ghosh & Shah, 2018	Sometimes, individuals read false information and it makes them feel sad and unable to stop doing it repeatedly.
15	Bhatt *et al.*, 2018	False information can damage the democratic system.
16	Pearson, 2017	Fake news destroys public safety.
17	Creech, 2020	False information spread on platforms like Facebook and Twitter also undermines trust.
18	Ho *et al.*, 2022	Fake information and false news on social media usually have negative effects and sometimes annoy everyday people, authorities and/or the government.
19	Khan *et al.*, 2022	Fake news is made to provoke extreme feelings, influence political activities, or create conflict and confusion in society.

Requirement of a hybrid model

After considering the literature review aforementioned in the above table found that the existing classifiers' results were not adequate compared to the proposed models', that the existing classifiers' accuracies were also less than the proposed models', and also found that, due to inaccuracy, no one can properly detect fake or true news and there is still room for improvement, we will first use the existing model followed by making a hybrid model to find an efficient way to detect fake or real news.

Methodology

Our objectives will be fulfilled with the help of ML and deep learning. First, we need to collect a dataset of news so that we can use that data to predict fake news or real news. Basically, we use a dataset consisting of several thousands of news articles and label them as either real news or fake news. Once we have labelled the dataset, we will pre-process the data. A lot of work is involved in this pre-processing step when compared to numerical data because computers and systems don't understand the text or characters. We therefore needs to find a suitable way to convert this text present in the news into meaningful numbers that the machine can understand. Once we pre-process the data or convert the text into meaningful numbers, we then need to split the dataset into training and test data because we need to train our ML model with the training dataset. So, we feed this training data which is pre-processed to our supervised ML classifiers and deep learning.

http://doi.org/10.18080/jtde.v11n3.765

Proposed method

Figure 1 presents a visual representation of the suggested strategies and sequential actions involved in the proposed method. This diagram illustrates the sequence of these steps in the proposed approach.

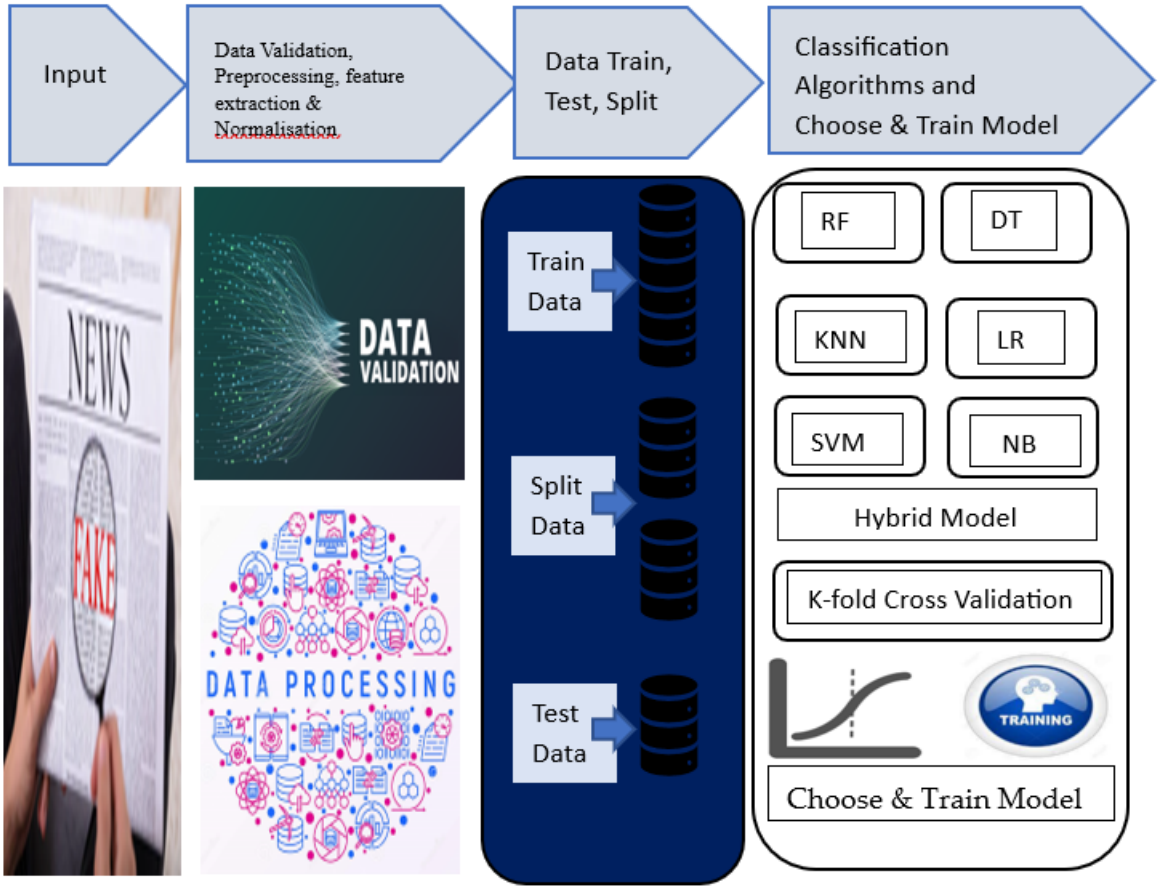


Figure 1. Block diagram

Data collection

The study employs a widely used dataset accessible on Kaggle, comprising 7796 entries and four columns. The initial column acts as a unique identifier for the news, while the second and third columns hold the news article's title and text correspondingly. The fourth column contains labels denoting whether the news is categorised as REAL or FAKE (Mahmoud, 2022).

Data validation

To verify the dataset's validity, we employ the langdetect library to determine if the text in each row of the CSV file is written in English. This process involves iterating through the DataFrame, extracting the text from each row, and using the is_english() function to assess its language. If the text is identified as English, a corresponding message is printed. Conversely, if the text is found to be non-English, an appropriate message is printed. Notably, the dataset

http://doi.org/10.18080/jtde.v11n3.765

in question is used by Mahmoud and Kokiantonis for their respective experiments (Mahmoud, 2022; Kokiantonis, 2022).

Data pre-processing

We pre-process the dataset by performing label encoding to convert categorical labels into numerical values. We also remove rows with missing values and eliminate duplicate rows from the DataFrame.

Feature extraction

For the feature extraction process, we use the TF-IDF (Term Frequency – Inverse Document Frequency) technique to convert the text data into numerical features. This is done using the 'TfidfVectorizer' class from the 'sklearn.feature_extraction.text' module. The text data is first combined from the 'title' and 'text' columns, and then the vectorizer is fit on this combined text data. The 'fit' operation calculates the term frequencies and inverse document frequencies. Next, the 'transform' operation converts the text data into TF-IDF feature vectors.

Normalisation

We also use the normalisation process. The feature vectors are normalised using the 'normalise' function from the 'sklearn.preprocessing' module, resulting in normalised TF-IDF vectors that can be used as input for an ML model.

Classification algorithms and hybrid model

Taking into account the characteristics of our dataset, we employ Logistic Regression (LR), Decision Tree (DT), K-Nearest Neighbours (KNN), Random Forest (RF), Support Vector Machine (SVM) and Naïve Bayes (NB) algorithms to investigate the behaviour of the data when subjected to different classifiers. Furthermore, a hybrid model is constructed by combining CNNs, RNNs and ANNs.

LR

Given that we are categorising text or content using a wide range of features and aiming for a binary output (such as true/false or authentic/fake article), the LR model is employed. The LR model offers a straightforward cost-function equation, allowing for classification between binary or multiple classes. In order to achieve optimal results for a specific dataset, we fine-tune the hyperparameters. Various parameters are assessed before obtaining the LR model, which serve as a benchmark for accuracy. Additionally, logistic regression utilises a sigmoid function to transform the output into a probability value. The objective is to attain an optimal probability that minimises the cost function (Pérez-Rosas *et al.*, 2017).

http://doi.org/10.18080/jtde.v11n3.765

DT

DT, an indispensable tool, exhibits a flow-chart-like structure primarily designed for addressing classification problems. The DT relies on conditions or 'tests' applied to attribute results at internal nodes to determine its branches. As attributes are evaluated, the leaf nodes are assigned class labels. The path from the root to the leaf represents a classification rule, making it versatile for both categorical and dependent variables. Notably, DTs excel in identifying crucial factors and illustrating their relationships, contributing significantly to the development of new variables and insightful data exploration. These tree-based learning algorithms, also known as CART, play a crucial role in constructing accurate predictive models through supervised learning techniques. Their strength lies in effectively capturing non-linear relationships and offering solutions for classification or regression problems (Meel & Vishwakarma, 2021).

KNN

By omitting the need for a dependent variable, KNN demonstrates its ability to make predictions for specific data outcomes. Sufficient training data is provided to enable KNN to accurately determine the exact cluster or category to which a given data point belongs. The value of K determines the number of neighbouring data points considered, and the KNN model calculates the distance between a new data point and its nearest neighbours. If K is set to 1, the new data point is assigned to the class with the closest distance.

RF

RF is a methodology that utilises the amalgamation of numerous DTs or algorithms with similar characteristics in a collection of trees. The RF technique is applicable to both classification and regression tasks.

SVM

SVM is a type of supervised ML model employed by specific classification algorithms. This model excels in solving problems where data is divided into two distinct groups. By training the SVM model with a set of data, it can effectively classify future instances. In scenarios with limited samples, SVM demonstrates superior speed and performance compared to other models. The SVM classifier can be visualised as a straightforward two-dimensional line. It takes data points as input and generates a hyperplane that separates different categories. This line serves as the decision boundary, with one side representing the 'blue' category and the other side representing the 'red' category. The proximity of a data point to the hyperplane determines its assigned tag, with the nearest point having the largest influence and vice versa.

http://doi.org/10.18080/jtde.v11n3.765

Hybrid model

The hybrid model which we name the NFCRA (Nafiz Fahad CNNs, RNNs, and ANNs) model combines CNNs, RNNs, and ANNs to classify news articles as fake or real. It processes the text data by organising it into sequences, then creates separate parts for each architecture. The outputs from each part are combined and passed through a dense layer, using a sigmoid function to classify them into two categories. The model is trained using the training data, and evaluated with the test data.

CNN

CNN is a sophisticated ML framework initially developed for interpreting visual information, yet it can also be effectively employed to handle textual information in natural language-processing endeavours. It uses filters or kernels to scan through input text, extracting local patterns or features. The input text is represented as numerical vectors using word or character embeddings. Convolutional layers perform convolutions on the input text, generating feature maps. Pooling layers downsample the feature maps to capture important features and reduce noise. Finally, fully connected layers make predictions or classifications. CNNs can be effectively trained on labelled text datasets by employing optimisation techniques such as stochastic gradient descent. Additionally, they can be further customised to excel in specific NLP tasks through fine-tuning.

RNN

RNN is a type of deep learning model for processing sequential data, including text. It uses feedback loops to capture information from previous time steps and maintain context. RNNs can have different architectures such as simple RNNs, LSTMs and GRUs, with varying memory and capability for handling long-term dependencies. They are widely used in NLP tasks such as text classification and sequence generation. RNNs can be trained using labelled text datasets and optimisation techniques, but have some limitations like vanishing gradients.

ANN

ANN is a computational model used in ML to analyse textual information by employing a network of interconnected nodes. The input text is represented as numerical vectors, which are processed through hidden layers with activation functions. The output layer produces predictions or classifications. ANN learns from labelled text datasets through parameter updates during training. ANN is versatile for various NLP tasks but may have limitations in capturing sequential dependencies and require large amounts of data for training.

Data analysis techniques

We evaluate our data using accuracy, precision, recall and F_1 scores as our analysis metrics. The analysis techniques are depicted below.

http://doi.org/10.18080/jtde.v11n3.765

Accuracy

Accuracy is the predominant metric used to assess the proportion of correctly predicted outcomes, whether they are true or false. The following equation can be employed to calculate the accuracy of a model.

$$Accuracy = \frac{TP+FP}{(TP+FP+TN+FN)} \quad (1)$$

Precision

Precision is a measure that evaluates how accurately a model predicts positive outcomes. In our study, we determine precision by dividing the count of correctly predicted positive results by the overall count of positive predictions.

$$Precision = \frac{TP}{(TP+FP)} \quad (2)$$

Recall

Recall, which represents the total number of correctly classified instances excluding the true class, is a key measure in our experiment. It specifically refers to the percentage of articles among all accurately predicted articles that were correctly anticipated.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F_1

The F_1 score is a single number that combines precision and recall, giving an overall measure of how well a model performs in tasks where it has to classify areas into two categories.

$$F1 = \frac{2\times(Precision \times Recall)}{Precsion+Recall} \quad (4)$$

Result

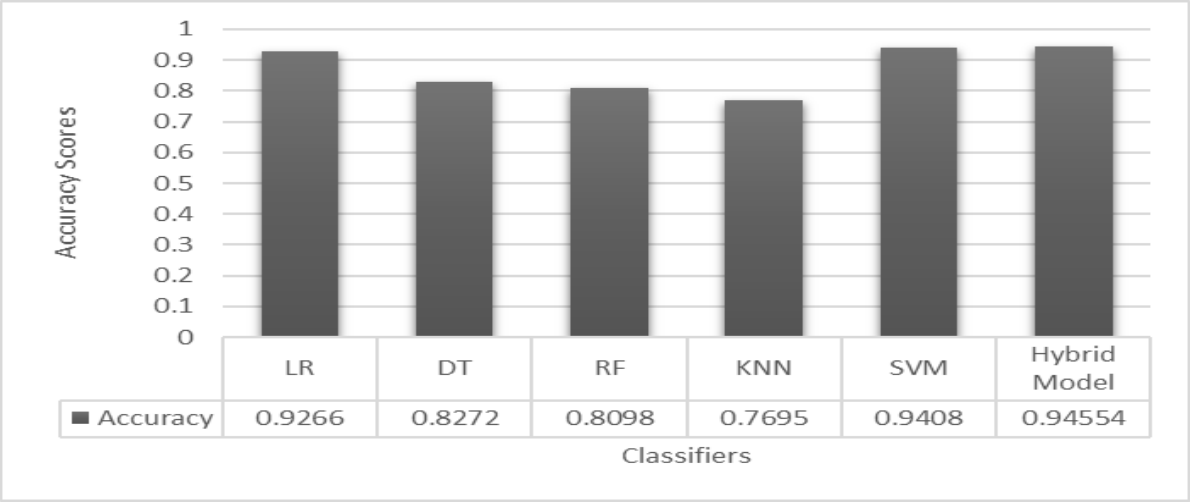


Figure 2. Accuracy vs classifiers

http://doi.org/10.18080/jtde.v11n3.765

This section provides an overview of the outcomes obtained from the experimental analysis conducted in the research study. The study evaluated various classifiers, including LR, DT, RF, KNN, SVM and a hybrid model, with the aim of assessing the variables of recall, accuracy and precision. The findings regarding accuracy, precision and recall are presented here. Additionally, the performances of the proposed systems for each classifier and the hybrid model are reported. On the test data, the LR, DT, RF, KNN, SVM and hybrid models achieved accuracy scores of 92.66%, 82.72%, 80.98%, 76.95%, 94.08% and 94.5% respectively. Notably, the hybrid model exhibited the highest accuracy score, reaching 94.5%.

Figure 3. Precision vs classifiers

Additionally, precision score experiments were carried out to evaluate the effectiveness of the proposed approach. The precision score measures the proportion of accurate positive predictions out of all positive observations, indicating how often the positive forecasts turn out to be correct. A higher precision score is desirable in this context. The precision scores obtained for LR, DT, RF, KNN, SVM and hybrid model were 90.5%, 81.2%, 76.8%, 91.6% and 94.4% respectively. Figure 3 presents the results of all classifier precision experiments, clearly demonstrating that the hybrid model achieves the most favourable outcomes.

The calculation of the recall score involves determining the number of actual positive predictions in relation to all the actual label classes. Recall, also referred to as sensitivity, represents the percentage of true positive findings. A higher recall score indicates better performance. According to the obtained recall scores, LR, DT, RF, KNN, SVM and hybrid model achieved 94.56%, 83.74%, 85.92%, 95%, 96.35% and 94.56% respectively. Once again, SVM achieved the highest rating in terms of performance (96.35%). The experimental recall scores are presented in Figure 4.

http://doi.org/10.18080/jtde.v11n3.765

Figure 4. Recall vs classifiers

Providing a comprehensive summary of all the classifiers, Figure 5 displays a chart presenting accuracy, precision, recall and F_1 scores. The chart proves that the hybrid model outperforms other models because accuracy, precision, and F_1 scores are higher than the existing classifiers we used.

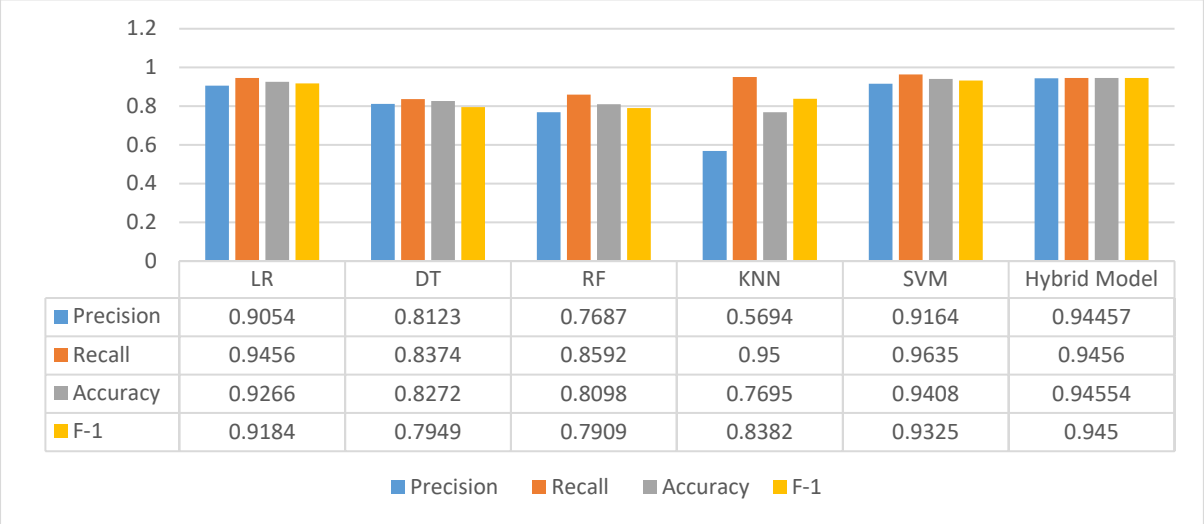


Figure 5. Summary of classifiers and hybrid model with accuracy, precision, recall and F1 scores

Comparison

Table 2 demonstrates that our research paper has been compared to several recently published papers, and that our model's accuracy is superior to the models in those papers.

Table 2. Comparison of models

Best **a**pproach	Accuracy of best approach
Ensemble model (Altheneyan & Alhadlaq, 2023)	93.40%
Hybrid Ensemble Model with Fuzzy Logic (Sharma ***et al.***, 2023)	86.8%
Our hybrid model	94.5%

http://doi.org/10.18080/jtde.v11n3.765

Discussion

It is clear that fake news is a significant problem in today's digital world. As people spend more time on the Internet, they are more likely to come across fake news, which can have serious consequences. Many researchers have tried to develop systems to prevent fake news, and various ML classifiers have been used for this purpose. The literature review also shows that fake news has various effects, including the potential to cause harm to individuals, misguide students, affect financial markets, erode the legitimacy and credibility of traditional news outlets, and hinder people's ability to distinguish between truth and lies.

The objectives of building a fortress against fake news are also clearly stated in the objective section. One of the primary objectives is to develop a reliable and accurate system for detecting fake news based on accuracy, precision and recall. Achieving a higher level of accuracy, precision and recall would ensure that the system can detect most fake news accurately. Another goal is to construct a hybrid model by using CNNs, RNNs and ANNs through the implementation of deep learning techniques. This approach has not been previously attempted and could lead to more effective fake news detection.

Overall, this research highlights the importance of preventing fake news and the need for automated systems to achieve this goal. It also underscores the potential harm that fake news can cause and the various effects it can have. By developing a reliable and accurate system for detecting fake news, we can reduce the impacts of fake news and help people differentiate truth and false news. The use of ML classifiers and deep learning techniques along with a hybrid model, could also lead to more effective and efficient fake news detection. In essence, this study's accuracy is best and this study's accuracy is superior to the recently published papers.

Conclusion

The proliferation of fake news is a major issue with grave repercussions for individuals, communities and the global community. With the increasing use of the Internet and social media, fake news is spreading at an alarming rate, making it difficult for people to differentiate between true and false news. However, based on a comprehensive analysis of published research, it is evident that fake news exerts numerous detrimental effects on individuals. However, a glimmer of optimism arises from the emergence of automated systems, which employ advanced ML and deep learning algorithms to effectively identify and uncover fake news. By evaluating different ML classifiers and making a hybrid model using deep learning, a reliable and accurate system can be developed that would ensure that most fake news can be detected with high levels of accuracy, which is 94.5%. It is imperative that we continue to

http://doi.org/10.18080/jtde.v11n3.765

research and develop systems to prevent the spread of fake news and protect ourselves from the harmful effects it can have on our lives and society.

Acknowledgements

A version of this paper was presented at the third International Conference on Computer, Information Technology and Intelligent Computing, CITIC 2023, held in Malaysia on 26–28 July 2023.

This work was supported by the TM R&D Fund (Project no. RDTC/221045 and SAP ID: MMUE/220017).

References

Ahinkorah, B. O., Ameyaw, E. K., Hagan Jr, J. E., Seidu, A. A., & Schack, T. (2020). Rising above misinformation or fake news in Africa: Another strategy to control COVID-19 spread. *Frontiers in Communication, 5*, 45. https://doi.org/10.3389/fcomm.2020.00045

Ahmad, T., Faisal, M. S., Rizwan, A., Alkanhel, R., Khan, P. W., & Muthanna, A. (2022). Efficient fake news detection mechanism using enhanced deep learning model. *Applied Sciences*, *12*(3), 1743. https://doi.org/10.3390/app12031743

Almenar, E., Aran-Ramspott, S., Suau, J., & Masip, P. (2021). Gender differences in tackling fake news: Different degrees of concern, but same problems. *Media and Communication*, *9*(1), 229–238. https://doi.org/10.17645/mac.v9i1.3523

Altheneyan, A., & Alhadlaq, A. (2023). Big data ML-based fake news detection using distributed learning. *IEEE Access*, *11*, 29447–29463. https://doi.org/10.1109/ACCESS.2023.3260763

Aslam, N., Ullah Khan, I., Alotaibi, F. S., Aldaej, L. A., & Aldubaikil, A. K. (2021). Fake detect: A deep learning ensemble model for fake news detection. *Complexity*, *2021,* 1–8. https://doi.org/10.1155/2021/5557784

Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, *149*(8), 1608. https://psycnet.apa.org/doi/10.1037/xge0000729

Bakir, V., & McStay, A. (2018). Fake news and the economy of emotions: Problems, causes, solutions. *Digital Journalism*, *6*(2), 154–175. https://doi.org/10.1080/21670811.2017.1345645

Bhatt, G., Sharma, A., Sharma, S., Nagpal, A., Raman, B., & Mittal, A. (2018). Combining Neural, Statistical and External Features for Fake News Stance Identification. WWW '18: Companion Proceedings of The Web Conference 2018. https://doi.org/10.1145/3184558.3191577

Butler, A. (2018). Protecting the Democratic Role of the Press: A Legal Solution to Fake News. *Washington University Law Review*, *96*(2), 419–440.

Creech, B. (2020). Fake news and the discursive construction of technology companies' social power. *Media, Culture & Society*, *42*(6), 952–968. https://doi.org/10.1177/0163443719899801

De Oliveira, D. V. B., & Albuquerque, U. P. (2021). Cultural evolution and digital media: Diffusion of fake news about COVID-19 on Twitter. *SN Computer Science*, *2*(6), 1–12. https://doi.org/10.1007/s42979-021-00836-w

Ghosh, S., & Shah, C. (2018). Towards automatic fake news classification. *Proceedings of the Association for Information Science and Technology*, *55*(1), 805–807. https://doi.org/10.1002/pra2.2018.14505501125

Ho, K. K., Chan, J. Y., & Chiu, D. K. (2022). Fake news and misinformation during the pandemic: What we know and what we do not know. *IT Professional*, *24*(2), 19–24. https://doi.org/10.1109/MITP.2022.3142814

Khan, A., Brohman, K., & Addas, S. (2022). The anatomy of 'fake news': Studying false messages as digital objects. *Journal of Information Technology*, *37*(2), 122–143. https://doi.org/10.1177/02683962211037693

Kokiantonis, A. (2022) *News.csv*. https://www.kaggle.com/datasets/antonioskokiantonis/newscsv/code. Accessed on 26 April 2023.

Lakshmanan, L. V., Simpson, M., & Thirumuruganathan, S. (2019). Combating fake news: a data management and mining perspective. *Proceedings of the VLDB Endowment*, *12*(12), 1990–1993. https://doi.org/10.14778/3352063.3352117

Leeder, C. (2019). How college students evaluate and share "fake news" stories. *Library & Information Science Research*, *41*(3), 100967. https://doi.org/10.1016/j.lisr.2019.100967

Mahmoud, A. (2022) *News.csv*. https://www.kaggle.com/code/ahmedxmahmoud/fake-news-detection/input. Accessed on 2 April 2023.

Meel, P., & Vishwakarma, D. K. (2021). A temporal ensembling based semi-supervised ConvNet for the detection of fake news articles. *Expert Systems with Applications*, *177*, 115002. https://doi.org/10.1016/j.eswa.2021.115002

Naeem, S. B., Bhatti, R., & Khan, A. (2020). An exploration of how fake news is taking over social media and putting public health at risk. *Health Information & Libraries Journal*, *38*(2), 143–149. https://doi.org/10.1111/hir.12320

Pearson, M. (2017). Teaching Media Law in a Post-truth Context: Strategies for Enhancing Learning about the Legal Risks of Fake News and Alternative Facts. *Asia Pacific Media Educator*, *27*(1), 17–26. https://doi.org/10.1177/1326365x17704289

Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2017). Automatic detection of fake news. https://doi.org/10.48550/arXiv.1708.07104

Rawat, G., Pandey, T., Singh, T., Yadav, S., & Aggarwal, P. K. (2023, January). Fake news detection using machine learning. 2023 International Conference on Artificial Intelligence and Smart Communication (AISC), pp. 759–762, IEEE. https://doi.org/10.1109/AISC56616.2023.10085488

Sharma, D. K., Singh, B., Agarwal, S., Pachauri, N., Alhussan, A. A., & Abdallah, H. A. (2023). Sarcasm detection over social media platforms using hybrid ensemble model with fuzzy logic. *Electronics*, *12*(4), 937. https://doi.org/10.3390/electronics12040937

Shu, K., Wang, S., & Liu, H. (2019). Beyond news contents: The role of social context for fake news detection. *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pp. 312–320. https://doi.org/10.1145/3289600.3290994

Singhal, S., Shah, R. R., Chakraborty, T., Kumaraguru, P., & Satoh, S. (2019). SpotFake: A Multi-modal Framework for Fake News Detection. IEEE International Conference on Multimedia Big Data (BigMM), pp. 39–47. https://doi.org/10.1109/BigMM.2019.00-44

Singhal, S., Kabra, A., Sharma, M., Shah, R. R., Chakraborty, T., & Kumaraguru, P. (2020). SpotFake+: A Multimodal Framework for Fake News Detection via Transfer Learning (Student Abstract). *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(10), 13915–13916. https://doi.org/10.1609/aaai.v34i10.7230

Stewart, E. (2021). Detecting Fake News: Two Problems for Content Moderation. *Philosophy & Technology*, *34*, 923–940. https://doi.org/10.1007/s13347-021-00442-x

Sullivan, M. C. (2019). Libraries and fake news: What's the problem? What's the plan?. *Communications in Information Literacy*, *13*(1), 7. https://doi.org/10.15760/comminfolit.2019.13.1.7

Zafarani, R., Zhou, X., Shu, K., & Liu, H. (2019). Fake News Research. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 3207–3208. https://doi.org/10.1145/3292500.3332287

Language Independent Models for COVID-19 Fake News Detection

Black Box versus White Box Models

W. K. Wong
Sensor Research Group, Curtin University Malaysia, Miri, Malaysia

Filbert H. Juwono
Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, China

Ing Ming Chew
Department of Electrical and Computer Engineering, Curtin University Malaysia, Miri, Malaysia

Basil Andy Lease
Curtin Malaysia Research Institute, Curtin University Malaysia, Miri, Malaysia

Abstract: In an era where massive information can be spread easily through social media, fake news detention is increasingly used to prevent widespread misinformation, especially fake news regarding COVID-19. Databases have been built and machine-learning algorithms have been used to identify patterns in news content and filter the false information. A brief overview, ranging from public domain datasets through the deployment of several machine learning models, as well as feature extraction methods, is provided in this paper. As a case study, a mixed language dataset is presented. The dataset consists of tweets of COVID-19 which have been labelled as fake or real news. To perform the detection task, a classification model is implemented using language-independent features. In particular, the features offer numerical inputs that are invariant to the language type; thus, they are suitable for investigation, as many regions in the world have similar linguistic structures. Furthermore, the classification task can be performed by using black box or white box models, each having its own advantages and disadvantages. In this paper, we compare the performance of the two approaches. Simulation results show that the performance difference between black box models and white box models is not significant.

Keywords: Fake news, black box model, white box model, machine learning, COVID-19

http://doi.org/10.18080/jtde.v11n3.789

Introduction

Fake news is a term often used to describe fabricated or distorted news or stories to mislead others. The spread of fake news can have a number of negative impacts on individuals and society as a whole. Fake news can spread false information, leading to confusion and misunderstanding about important issues, such as COVID-19 (Rocha *et al.*, 2021). On the polarization of society, fake news can be used to fuel political or social division by presenting one-sided or biased information (Gupta *et al.*, 2023). Another result of fake news propagation is damage to reputation of individuals, companies, or organizations (Domenico *et al.*, 2021). Often, interference in elections is prone to be organized by unethical entities. In politics, fake news can be used to influence the outcome of elections by spreading false information about candidates or issues (Grossman & Helpman, 2023). Moreover, the spread of fake news can undermine democratic processes by spreading disinformation and sowing confusion among citizens. Looking at the above, the spread of fake news can erode trust in traditional news sources and journalism, making it harder for people to separate fact from fiction. A more devastating effect of fake news spread is the spreading of false information about health and medical issues (Waszak *et al.*, 2018). The effect of fake news spread becomes worse in the COVID-19 pandemic time, when it leads to harmful or even deadly consequences (Ferreira Caceres *et al.*, 2022).

The detection of fake news can be a challenging task, as it often involves identifying and evaluating the veracity of information that is presented as true. Some techniques for detecting fake news include fact-checking, using multiple sources to verify information, and looking for patterns of misinformation (Lin *et al.*, 2019; Zhou & Zafarani, 2019). Additionally, there are various tools and software, such as browser extensions and apps, that can help users identify fake news (Nordberga *et al.*, 2020). The most common technique used in the tools for detecting fake news is Artificial Intelligence (AI). In general, there are various AI techniques for identifying fake news, such as Natural Language Processing (NLP), machine learning, and deep learning.

Various NLP techniques can be used to automatically identify patterns in language and structure that are commonly associated with fake news (Probierz *et al.*, 2021). Machine learning techniques have been used to detect fake news by recognizing patterns and features of real and fake news and using these models to classify new, unseen news articles (Imbwaga *et al.*, 2022). Besides machine learning, deep learning techniques have also been popularly used for detecting fake news (Hu *et al.*, 2022).

The objective of this paper is to compare the performance of two model approaches in detecting fake news. We focus on COVID-19 fake news detection from Twitter news feeds

http://doi.org/10.18080/jtde.v11n3.789

(tweets) as the case study. In particular, model development, feature extraction, challenges, and public domain dataset are discussed in this paper. Moreover, we introduce a curated COVID-19 dataset of mixed Malay-English tweets, which is available on GitHub. A Support Vector Machine (SVM) is implemented as an evaluation to the separability of the classes. The rest of the paper is structured as follows. First, we discuss four fake news detection techniques. Then, challenges and remarks on the existing approaches are presented. Next, the discussion of black-box models and white-box models is presented, followed by a case study including a mixed Malay-English Twitter dataset, simulation results, and analysis. Finally, a conclusion is drawn.

Fake News Detection Techniques

In line with the objectives, this section provides some description of fake news detection strategies. According to Shu *et al.* (2017), fake news detection research works can be divided into four main categories: Data-oriented, feature-oriented, model-oriented, and application-oriented. In accordance with this division, we present the discussion in four parts. Furthermore, several of the significant problems are explored.

Data-oriented approach

The data-oriented approach in fake news detection focuses on the collection and annotation of data sets for training and testing the fake news detection models. This approach includes the process of identifying and collecting large volumes of news articles, as well as manually annotating them as real or fake. This approach aims to improve the quality and diversity of the data sets used to train and test fake news detection models. In this approach, various characteristics of the dataset, such as temporal and psychological aspects, are studied. Comprehensive datasets have been created to serve as benchmark datasets. For example, the CHECKED dataset created by Yang *et al.* (2021) includes textual, visual, temporal, and network information related to Chinese COVID-19 fake news. Melo & Figueiredo (2020) created the first Brazilian fake news dataset containing information about hashtags, media, and retweets related to COVID-19 news. Hayawi *et al.* (2022) created a dataset specifically for COVID-19 vaccine news from Twitter. Memon & Carley (2020) created a novel dataset that categorizes COVID-19 online communities into users who post misinformation and users who spread true information. Patwa *et al.* (2021) created a dataset of fake news related to COVID-19 from social media posts and articles for an online competition. Cui & Lee (2020) released the CoAID dataset, which includes fake news from social media and websites, as well as user engagement with the news. Shahi & Nandini (2020) created a fake news dataset in 40 languages related to COVID-19 and categorized the tweets into 11 categories based on the

http://doi.org/10.18080/jtde.v11n3.789

topic. Alam *et al.* (2020) released a large fake news dataset of 16,000 COVID-19 tweets in Arabic, Bulgarian, Dutch, and English languages.

In terms of temporal perspective, the spread of fake news on social media shows distinct patterns that differ from those of real news. Murayama *et al.* (2021) used two datasets of fake news items that spread on Twitter to describe the propagation of fake news as a two-stage process. Kim *et al.* (2018) tracked the time events when a story was posted to determine which story and when it would be sent for verification to fact-checkers. From a psychological standpoint, the echo-chamber effect plays an important role in capturing the intentions aspect of fake news spreading in social media. Törnberg (2018) created a simulation model to investigate the interactions between echo chambers contributing to the viral spread of misinformation on the network. Abonizio *et al.* (2020) included the sentiment polarity (negativity or positivity of a text) in their assessment, measuring the negativity or positivity of a text as part of the input features into the fake-news detection model.

Feature-oriented approach

The feature-oriented approach in fake news detection focuses on identifying and extracting relevant features from news articles that can be used to train and test fake news detection models. This includes identifying patterns in language, analysing the sentiment or tone of a text, and analysing the structure of a news article. This approach aims to improve the feature representation of news articles, which can lead to better performance of fake news detection models. According to Shu *et al.* (2017), there are two main data sources: news content and social context. For the news content data source, linguistic-based and visual-based techniques can be used to extract features from text information. Linguistic-based techniques involve extracting word features from text, which can be in either a static or dynamic form. Word embeddings represent words using vectors and are a common practice in the NLP approach. Wang *et al.* (2020) conducted a study where static representations of words, such as one-hot encoding, Bag-of-words (BoW), and Term Frequency-Inverse Document Frequency (TF-IDF), were used in the early stages of NLP. These embeddings, however, suffer from high dimensional vectors that are often as large as the vocabulary size, making them hard to use. For example, in one-hot encoding, words are represented with a one-zero vector, where all values are zero except the single value, which is one, corresponding to the word column. BoW has been used in Rusli *et al.* (2020) while Term Frequency (TF), which is similar to BoW, has been used in Jiang *et al.* (2021), and TF-IDF has been used by several authors (Hayawi *et al.*, 2022; Rusli *et al.*, 2020; Jiang *et al.*, 2021; Abdelminaam *et al.*, 2021).

Advanced static word embeddings, such as Word2Vec, have been utilized in studies by Oliveira *et al.* (2020), Ivancová *et al.* (2021) and Verma *et al.* (2021). On the other hand, Global Vectors

http://doi.org/10.18080/jtde.v11n3.789

(GloVe) embedding has been implemented in studies by Hayawi *et al.* (2022), Jiang *et al.* (2021) and Abdelminaam *et al.* (2021). Additionally, there are fake news detection models that utilize dynamic word embeddings, such as Bidirectional Encoder Representations from Transformers (BERT), as seen in studies by Hayawi *et al.* (2022), Kar *et al.* (2020) and Hande *et al.* (2021). Other dynamic word embeddings like XLNet, Efficiently Learning an Encoder that Classifies Token Replacement Accurately (ELECTRA), and Robustly Optimized BERT Pretraining Approach (RoBERTa) have been implemented in research by Hande *et al.* (2021) and Das *et al.* (2021).

In this paper, static embedding or linguistic-based features are investigated. Static features with Language-Independent (Lang-IND) characteristics have been employed as multiple languages are investigated for fake news detection. The Lang-IND features focus on capturing high-level structures rather than specific terms from a language. In particular, linguistic-based features are extracted from text content to capture the organization of documents at different levels, including characters, words, sentences, and documents. To capture different writing styles, common lexical features are examined at the character and word level, such as total words, characters per word, frequency of large words, and unique words. Common syntactic features focus on sentence-level features, such as frequency of function words, phrases, punctuation, and part-of-speech tagging.

Sutter *et al.* (2017) used 25 language-independent features (basic frequencies of part-of-speech tags) and five language-dependent features to measure the quality of translation work from English to French and French to English done by students and compare it to the work of professionals. Abonizio *et al.* (2020) extracted language-independent features, such as complexity, stylometric, and psychological features, from textual data to detect fake news in English, Portuguese and Spanish. They found that using purely stylometric features, such as Part-of-Speech tag (POS-tag) diversity (POS-tag is a label given to each word to denote its part of speech), the ratio of named entities to text size, the ratio of quotation marks to text size, and the frequency of unrecognized words, in combination with Random Forest (RF), XG Boost, and SVM classifiers, led to an increase in model accuracy. Faustini & Covões (2020) explored features that can be used regardless of the source platform and extracted a mix of 13 features (complexity and stylometric) from news content. In addition to these features, they also extracted Word2Vec features from text and used the sum of all 100 values in the vector as the 14th feature. In the MM-COVID fake news dataset paper published in Li *et al.* (2020), features extracted from news content and social engagement patterns were described as language-invariant features for six languages (English, Spanish, Portuguese, Hindi, French and Italian). To create Lang-IND features, Veselý *et al.* (2012) trained all languages simultaneously using a Multilingual Artificial Neural Network (MANN) by modelling each language using a separate

http://doi.org/10.18080/jtde.v11n3.789

output layer. Vogel & Meghana (2020) focused mainly on features that were not tied to a specific language in order to determine fake news spreaders and attempt to block fake news from spreading at the earliest stage. They captured high-level textual features of various stylistic and psychological features, such as emojis, hashtags, upper phrases, user mentions, neutral and negative polarity.

Faustini & Covões (2019) extracted eight numerical features directly from raw texts, such as the proportion of uppercase characters, exclamation marks, question marks, text that contains exclamation marks, the number of unique words, sentences, characters and words per sentence. In addition to the eight features, the authors used POS tagging to extract the proportion of adjectives, adverbs, and nouns and three other features, including the sentiment of the message, the proportion of swear words, and the proportion of spelling errors. However, these additional features were not considered Lang-IND as they relied on tools or libraries that had been trained with a specific language to extract the features.

Some studies have combined both static and dynamic embeddings. For example, Kar *et al.* (2020) proposed a method for detecting fake news by using BERT embeddings and combined them with three stylometric features from tweet text, two user engagement features (retweet and favourite count), 19 user profile features, fact verification score, and bias score for low resource languages, such as Hindi and Bengali tweets. The study found that the feature representations extracted from Hindi and Bengali languages were highly transferable across Indic languages. Another study by Guibon *et al.* (2019) performed statistical text analysis and used a feature stacking approach on a dataset of vaccination-related fake news in English and French.

Model-oriented approach

The model-oriented approach in fake news detection focuses on developing new models for detecting fake news using machine learning algorithms, NLP techniques, and network analysis methods. This approach aims to improve the accuracy, robustness, and scalability of fake news detection models. Research conducted by Abonizio *et al.* (2020) evaluated the performance of four machine learning algorithms (*K*-Nearest Neighbours, SVM, RF, and Extreme Gradient Boosting) using Lang-IND features. Oliveira *et al.* (2020) proposed a one-class SVM model that grouped training samples into one class, with samples that did not fit into that class being placed into a new class (fake news). Kesarwani *et al.* (2020) used a *K*-Nearest Neighbour model to classify instances of fake news and found that the model achieved maximum accuracy when the value of *K* was between 15 and 20. Faustini & Covões (2020) performed a fake news detection study on multiple platforms and languages, using four machine learning algorithms, namely *K*-Nearest Neighbour, RF, Gaussian Naïve Bayes and SVM.

http://doi.org/10.18080/jtde.v11n3.789

In the recent years, the use of deep learning classifiers has gained popularity for identifying fake news. Deep learning can be considered a subset of machine learning that uses multi-layered neural networks to build complex connections between the inputs and outputs. Research works, such as Ivancová *et al.* (2021) and Abdelminaam *et al.* (2021), have trained and compared different neural network architectures for fake news detection, such as one dimensional Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM), or modified versions of LSTM and Gated Recurrent Unit (GRU). Other research works, such as Hande *et al.* (2021) and Hayawi *et al.* (2022), proposed a fake news detection model using a deep-learning technique called Bidirectional LSTM (Bi-LSTM), which is a sequence of two LSTMs with one taking the input in a forward direction and the other in a backward direction. These models also used pre-trained transformer embeddings, such as BERT, XLNet, RoBERTa, or Glove embeddings, as input features.

Evolutionary methods, like Particle Swarm Optimization (PSO) and Salp Swarm Algorithm (SSA), have also emerged as options for reducing features for fake news detection (Al-Ahmad *et al.*, 2021). Choudhury & Acharjee (2022) proposed machine learning classifiers, such as SVM, Naïve Bayes, Logistic Regression (LR), and RF, as fitness functions in a genetic algorithm, while using TF-IDF as features input and confusion matrix to calculate evaluation metrics, such as precision, recall, and F1-score. Note that precision shows the true positive rate, while recall shows the measure of how many true positive samples the model can predict correctly out of all the true positive samples in the data. Finally, F1-score combines precision and recall, making it useful for analysing unbalanced datasets.

Application-oriented approach

The application-oriented approach in fake news detection focuses on the practical applications of fake news detection models. This includes developing systems that can automatically detect and flag fake news on social media platforms, or integrating fake-news detection models into news aggregators and search engines. This approach aims to improve the usability and effectiveness of fake news detection systems in real-world scenarios. The application-oriented approach focuses on two main areas: the diffusion of fake news; and interventions to address it. Research on fake news diffusion examines how false information spreads on social media platforms, with early studies showing that false information spreads differently than reliable information. To mitigate the effects of fake news, interventions can be implemented proactively, such as removing user accounts and labelling false news, or reactively targeting specific groups of users or the entire network, when the spread of fake news is known or unknown. Galal *et al.* (2021) suggested that reactive intervention methods involve launching campaigns to counter fake news by targeting a specific group of individuals when the affected

http://doi.org/10.18080/jtde.v11n3.789

users are known, or targeting the entire network when the affected users are not identified. Kim *et al.* (2018) proposed frameworks for reducing the spread of fake news by flagging it for fact-checking and lowering its visibility in users' feeds.

Multilingual Fake News Detection Challenges

The main challenge in detecting fake news in languages other than English is the lack of datasets and NLP tools for those languages (De *et al.*, 2021). Research works have been done for detecting fake news in low-resource languages such as Arabic (Jardaneh *et al.*, 2019; Maakoul *et al.*, 2020), Spanish (Pizarro, 2020), Portuguese (Faustini & Covões, 2019), Indonesian (Al-Ash *et al.*, 2019; Rusli *et al.*, 2020; Prasetyo *et al.*, 2019), Slovak (Ivancová *et al.*, 2021), Chinese (Yang *et al.*, 2021), Bengali (Mugdha *et al.*, 2020) and Bangladeshi (Hussain *et al.*, 2020). However, creating datasets for these languages can be difficult and time-consuming (Kong *et al.*, 2020). Studies in Kong *et al.* (2020), Li *et al.* (2020) and Abonizio *et al.* (2020) proposed to build models that can detect fake news in multiple languages and can create a multi-language fake news dataset, which is a challenging task. Note that it can be difficult to differentiate between fake news and real news when the language is not the mother tongue (Sutter *et al.*, 2017). According to Shu *et al.* (2017), there are several characteristics of this problem that make it uniquely challenging for automated detection. One of these characteristics is that fake news is intentionally written to mislead readers, making it difficult to detect based solely on the content. Additionally, fake news can be diverse in terms of topics, styles, and media platforms, and may use diverse linguistic styles and sarcasm to distort the truth. For example, fake news may use true evidence in the wrong context to support a false claim.

Black Box vs White Box Models

It is worth noting that the above-mentioned categories are not mutually exclusive and often overlap with each other. Future research may involve combining different aspects of these categories to develop more effective fake news detection systems. In addition, as the technology and the way people consume information is changing rapidly, research in this field will have to adapt to that and keep updating the methods and techniques accordingly. This section gives a brief review of the existing developments in fake news detection from various perspectives. In particular, data acquisition, feature generation, and machine-learning models are applied accordingly. To date, there have been many research works working on Lang-IND fake news detection, such as Zervopoulos *et al.* (2022) and Imaduwage *et al.* (2022).

However, the issue with the machine learning and deep learning techniques is that they are "black box" in nature. This means that we do not understand how the decision or classification

http://doi.org/10.18080/jtde.v11n3.789

progresses. In other words, they do not offer much knowledge and transparency of how a particular set of news is labelled as legitimate or fake news. On the other hand, "white box" models are easy to interpret because they are based on patterns, rules, and decision trees (Loyola-González, 2019). Kong *et al.* (2023) proposed a two-stage evolutionary approach to generate a white box model for Lang-IND fake news detection. However, the transparency of white box models comes at the expense of reduced accuracy when compared to black box models (Fung *et al.*, 2021). It is also worth noting that the definition of lower accuracy is fuzzy. For example, improperly setting black box model parameters might result in lower accuracy than a white box model.

In this paper, we compare the performance of black box models and white box models for the Fake.my-COVID19 dataset, which is a COVID-19 bilingual Twitter dataset. We consider three black box models, i.e., SVM, K-Nearest Neighbour (KNN), and RF, and three white box models, i.e., LR, Decision Tree (DT), and Genetic Programming (GP).

Support Vector Machine (SVM)

SVM is a type of supervised learning algorithm that can be used for classification or regression tasks. The goal of an SVM is to find the best boundary (or "hyperplane") that separates the different classes in the training data. Let (x, y) be the pair of (features, label); the optimization problem in SVM is given by:

$$\min_{w,b} \|w\|^2 + c\sum_{i}^{m} \xi_i$$

subject to

$$y^{(i)}\left(w^T x^{(i)} + b\right) \geq 1 - \xi_i,\ i = 1, 2, \ldots, m$$

$$\xi_i \geq 0,\ i = 1, 2, \ldots, m$$

where w is the weights, b is bias, m is the number of data samples, ξ_i is the slack variable to allow some misclassification, and c is a penalty parameter with range between 0.5 and 1.0. Note that a larger c discourages misclassification.

SVM can also be used for non-linear classification by introducing the kernel trick, which maps the original data into a high-dimensional space where a linear boundary can be found. There are three types of kernel functions in SVM: linear, polynomial and Radial Basis Function (RBF). Using a linear kernel means that the SVM will find a linear decision boundary in the input space, which can be useful in situations where the data is linearly separable. The linear kernel is the simplest kernel function, which can be faster and use less memory compared to

http://doi.org/10.18080/jtde.v11n3.789

other more complex kernel functions, like polynomial or RBF kernel. In RBF kernel, the mapping function is given by:

$$K(x_1, x_2) = \exp\left(-\gamma \|x_1 - x_2\|^2\right)$$

where γ is a control parameter. If the γ value is small, more data samples are clustered.

K-Nearest Neighbour (KNN)

KNN is one of the most basic machine learning algorithms. It categorizes data based on a distance metric. The distance can be Euclidean distance, Minkowski distance, or Manhattan distance. It can be considered as a lazy learner since it does not "learn" until the test example is provided. As a result, whenever we have a new datum to classify, we use the training data to discover its k-nearest neighbours. Keep in mind that K is frequently an odd number to avoid ties.

Random Forest (RF)

RF is an ensemble machine learning algorithm which combines many DTs. RF can be used for classification as well as regression. A brief overview of DT will be given below. In principle, each DT will be trained on a distinct dataset, resulting in shorter depth and thereby avoiding overfitting. As a result, the DT will produce a large number of outputs. The majority of the voted classes (for classification) or the average of the individual results (for regression) will be used to make the decision.

Logistic Regression (LR)

LR can be formed by generating a multiple linear regression at the first stage, as follows:

$$z = \sum_i a_i x_i$$

where a_i are the weights and x_i are the features. The result is then input to a logistic (sigmoid) function, as follows:

$$y = \frac{1}{1+e^{-z}}$$

Decision Tree (DT)

DT is a tree-like structure that consists of a root node, decision nodes, and terminal nodes. It begins with a single node, known as the root, which reflects the initial decision. Then, branches are formed to depict the possible outcomes. Each branch leads to a new node that represents the following decision depending on the previous option. This procedure is repeated until a final result or choice is obtained, which is represented by a terminal node. Each node in the decision tree provides information that aids in deciding which branch to take, such as criteria

http://doi.org/10.18080/jtde.v11n3.789

or requirements that must be met. It is clear that the structure represents transparency, so that we can understand how the decision is taken. DT uses entropy as well as Gini entropy to split the data. There are a few algorithms that are commonly used for building a DT, such as ID3, C4.5, and Classification and Regression Trees (CART) (Javed *et al.*, 2022).

Genetic Programming (GP)

GP is an evolutionary algorithm used to find the best solution of a given problem. GP can be formed using a tree-like structure where arithmetic operators can be used as the nodes. The result is a mathematical expression that represents the solution of the problem. The arithmetic operators can be basic operators and complex operators. Similar to DT, an algorithm is used to generate the GP tree. Basically, the algorithm uses the operations similar to a Genetic Algorithm (GA): reproduction (copy the program without modifications), mutation (modify a part of the program), and crossover (two programs are selected to generate a new program).

Fake.my-COVID19 Dataset and Its Features

A public dataset is required to open the opportunity for other researchers to develop knowledge. Consider, for example, a more established domain, such as generalized optimization problems (Wong & Ming, 2019) and malware detection (Wong *et al.*, 2021), where both domains have community-based public datasets that allow for more comprehensive benchmarking, thereby enabling systematic progress. Motivated by Alameri & Mohd (2021), who encouraged work on fake news detection in Malay news, we created a Malay-English COVID-19 fake news tweets dataset that can be publicly accessed at https://github.com/z3fei/Malaysia-COVID-19-Tweet-ID/tree/main/Fake.my-COVID19. To the best of our knowledge, a fake news dataset in Malay language had not been available before we published our dataset. This has been created to contribute to the low resource Malay language. Note that bilingual mode is presented because most Malaysian people speak both English and Malay (Albury, 2017)

An initiative was undertaken to construct and refine a data collection system, which aimed at procuring COVID-19 related updates shared within Malaysia through Twitter's Standard search API. The designated timeframe for this collection spanned 1 September 2021 to 31 March 2022. Within this duration, Malaysia had initiated the administration of a third round of COVID vaccines to its healthcare frontliners and elderly people. The vaccination campaign was also extended to encompass adolescents aged between 12 and 17 years. The data collection effort yielded the accumulation of 251,216 tweets spanning a period of 231 days. From the data collected, 68% of tweets are in Malay, 28% in English, and 4% in other languages (Chinese, Tamil, etc.). We omitted the tweets containing languages other than Malay and English. In the

http://doi.org/10.18080/jtde.v11n3.789

data collection program, there are two important search criteria, namely keywords and locations (to ensure the collected tweets were the ones posted within Malaysia).

After the tweets were collected, we performed the annotation task. The process of annotation encompassed the identification of assertions put forth by users within the tweet content, subsequently cross-referencing these assertions against reliable fact-checking websites to determine their accuracy. This undertaking adhered to the guidelines outlined in Vogel & Meghana (2020) to ascertain binary classifications, distinguishing between claims classified as either fake or real. According to the guidelines, context of the tweets was considered carefully by excluding tweets that were sarcastic and/or humorous. Furthermore, the tweets that expressed general opinions regarding the vaccine, official news, and appointment details of vaccination centres were not considered as fake news. The method contributes to guaranteeing the precision of manual data annotation and upholding the elevated calibre of the dataset. Tweets that had been categorized were assigned binary labels: '1' for fake news and '0' for real news. Tweets categorized as fake typically encompassed unverifiable assertions, deceptive information, deliberate deception, or unfounded conspiracy theories lacking scientific support. To mitigate bias, the labels were initially assigned by an individual and subsequently validated by three other experts. In addition, a tweet should be marked as real news if it contained useful information on COVID-19, such as numbers, dates, vaccine progress, government policies, hotspots, etc. (Faustini & Covões, 2019). All tweets posted by government agencies, medical institutes, and official news media channels were also considered as real news (Guibon *et al.*, 2019).

Finding verifiable factual claims among the retrieved tweets was not an easy task. In Vogel & Meghana (2020), the authors suggested that tweets of less than five words should be removed. We used a filter that eliminated tweets containing fewer than 20 characters, ensuring their exclusion. Additionally, tweets comprised solely of emojis, emoticons, or greetings were disregarded. To refine the scope, a language filter was integrated to exclusively consider tweets in Malay and English, which constituted approximately 96% of the amassed tweets.

During data collection, certain keywords were often used in fake news tweets, such as *beksin* and *baksin*. The actual word for *beksin* or *baksin* is *vaksin* (Malay) or vaccine (English). The anti-vaxxers who posted these tweets misspelled the word on purpose to avoid authorities screening for the actual word (i.e., vaccine). There were other words used instead of COVID, misspelt on purpose, such as *kovid*, *kobid*, *convid*, *konvid*. Other than that, some words recorded in fake news tweets are Adverse Events Following Immunization (AEFI), ivermectin, *haram* (Eng: illegal), flu, fake, scam, deltacron, *bunuh* (Eng: kill), *cipta* (Eng: create), *racun* (Eng: poison), etc. The same user would not just post one false claim and stop spreading false claims in many weeks to come. The users that had been identified as a fake news spreader or

http://doi.org/10.18080/jtde.v11n3.789

anti-vaxxers were put into a list. All tweets posted by them would be read and further checked on their claims made against trusted sources. The Fake.my-COVID19 dataset consists of 3,068 tweets, where 1,422 tweets are fake news and the remaining tweets are real news. The extracted features are shown in Table 1.

Table 1. List of 25 Lang-IND Features

Index	Lang-IND Feature	Description
1	*cnt_sentences*	Number of sentences in tweet
2	*cnt_words*	Number of words in tweet
3	*cnt_uniquewords*	Number of unique words in tweet
4	*tweet_length*	Number of characters in tweet
5	*cnt_uniquechars*	Number of unique characters in tweet
6	*avg_words_sent*	Average number of words per sentence
7	*avg_word_length*	Average number of characters per word
8	*TTR*	Number of Type-Token Ratio (TTR) in tweet. TTR is defined as the total number of unique words divided by total words.
9	*hashtags*	Number of hashtag symbols in tweet
10	*hashtags_ratio*	Number of hashtag symbols per sentence
11	*urls*	Number of URLs in tweet
12	*urls_ratio*	Number of URLs per sentence
13	*emojis*	Number of emojis in tweet
14	*emojis_ratio*	Number of emojis per sentence
15	*puncs*	Number of punctuation marks in tweet
16	*puncs_ratio*	Number of punctuation marks per sentence
17	*exclam*	Number of exclamation marks in tweet
18	*exclam_ratio*	Number of exclamation marks per sentence
19	*question*	Number of question marks in tweet
20	*question_ratio*	Number of question marks per sentence
21	*quote*	Number of quotation marks in tweet
22	*quote_ratio*	Number of quotation marks per sentence
23	*uppercase*	Number of uppercase characters in tweet
24	*uppercase_ratio*	Number of uppercase characters per sentence
25	*alluppercase*	Number of all-uppercase words in tweet

Results and Discussion

Table 2 shows the testing mean accuracies of seven algorithms. For GP, we use two sets of functions: basic functions and extended functions. Basic functions include $\{+,-,\times,\div\}$ while, for extended functions, we use the combination of basic functions and $\{(.)^2, |.|, log1p(.), sign(.), \exp(.), \sqrt{.}\}$. The output of GP is a mathematical expression R which should be translated to the binary decision, i.e., fake or real news. The decision rule is given as follows: if R is less than 0.5, then it is real news. Otherwise, it is fake news.

http://doi.org/10.18080/jtde.v11n3.789

Table 2. Mean Accuracy Comparison (Testing)

Algorithm	Parameters	Accuracy (mean)
Black Box Models		
SVM	Kernel = RBF $c = 0.5$ $\gamma = 1$	0.5358
	Kernel = RBF $c = 1.0$ $\gamma = 1$	0.5309
	Kernel = linear $c = 0.5$	0.8339
	Kernel = linear $c = 1.0$	0.8453
KNN	$K = 5$	0.8020
RF	Max. tree depth = 8 Max. population = 100	0.8607
White Box Models		
LR		0.8423
DT		0.8219
GP	Function = basic Depth = 4	0.8228
	Function = basic Depth = 6	0.8269
	Function = basic Depth = 8	0.8228
	Function = complex Depth = 4	0.8208
	Function = complex Depth = 6	0.8350
	Function = complex Depth = 8	0.8310
Optimized GP	Function = complex Depth = 8	0.8482

From Table 2, it can be seen that the SVMs with RBF kernel show very bad accuracy. As this is a binary classification task, we can say that the model does not learn and the decision is random. SVM with linear kernel gives better results. For the black box models, RF gives the best result of 86% accuracy. Regarding GP, it can be seen that using complex functions does not significantly improve the accuracy. Moreover, an improved accuracy can be achieved by optimizing the expression obtained from GP. It is interesting to show that we can further improve the mathematical expression R by adding some weights to the variables. The weights can be optimized by using, for example, Adaptive Differential Evolution (ADE). ADE has been used to optimize the raw GP equations as presented by Kong *et al.* (2023) and Wong *et al.* (2023). In this paper, we only optimize the GP with depth 8 and complex function.

Finally, we can see that white box models return lower accuracies compared with RF. However, we can see that the difference is not significant, i.e., less than 2%. With the

http://doi.org/10.18080/jtde.v11n3.789

transparency capability of the white box models, it would be advantageous to use white box models. Future challenges might involve further enhancement of the white box models.

Concluding Remarks

Detecting fake news in mixed languages in a limited number of characters in social media messages can be complex. In addition, there is a lack of work in fake news detection for low resource languages, such as Malay. We have created and published a Malay-English fake news Twitter dataset to support research on this topic. Furthermore, two approaches of machine learning models, i.e., black box and white box models, have been discussed and compared. Mathematical expressions can be generated by using white box models, making the model clearer to the users. Simulation results show that white box models (in terms of optimized GP) result in lower accuracy. However, the difference is insignificant; therefore, white box models with their transparency capabilities are preferred.

Acknowledgement

An earlier version of this paper was presented at ICDATE 2023, Malaysia, July 2023. We would like to express our gratitude to the editor for providing another opportunity to publish the extended version of this research work.

References

Abdelminaam, D. S., Ismail, F. H., Taha, M., Taha, A., Houssein, E. H., & Nabil, A. (2021). CoAID-DEEP: An Optimized Intelligent Framework for Automated Detecting COVID-19 Misleading Information on Twitter. *IEEE Access*, *9*, 27840–27867. https://doi.org/10.1109/ACCESS.2021.3058066

Abonizio, H. Q., Morais, J. I., Tavares, G. M., & Barbon Junior, S. (2020). Language-Independent Fake News Detection: English, Portuguese, and Spanish Mutual Features. *Future Internet*, *12*, 1–18. https://doi.org/10.3390/fi12050087

Al-Ahmad, B., Al-Zoubi, A., Abu Khurma, R., & Aljarah, I. (2021). An Evolutionary Fake News Detection Method for COVID-19 Pandemic Information. *Symmetry*, *13*, 1091. https://doi.org/10.3390/sym13061091

Alam, F., Shaar, S., Dalvi, F., Sajjad, H., Nikolov, A., Mubarak, H., Martino, G. D. S., Abdelali, A., Durrani, N., Darwish, K., Al-Homaid, A., Zaghouani, W., Caselli, T., Danoe, G., Stolk, F., Bruntink, B., & Nakov, P. (2020). Fighting the COVID-19 Infodemic: Modeling the Perspective of Journalists, Fact-Checkers, Social Media Platforms, Policy Makers, and the Society. arXiv preprint arXiv:2005.00033. https://doi.org/10.48550/arXiv.2005.00033

Alameri, S. A., & Mohd, M. (2021). Comparison of Fake News Detection Using Machine Learning and Deep Learning Techniques. 3rd International Cyber Resilience Conference (CRC). https://doi.org/10.1109/CRC50527.2021.9392458

Al-Ash, H. S., Putri, M. F., Mursanto, P., & Bustamam, A. (2019). Ensemble Learning Approach on Indonesian Fake News Classification. 3rd International Conference on Informatics and Computational Sciences (ICICoS). https://doi.org/10.1109/ICICoS48119.2019.8982409

Albury, N. J. (2017). Mother Tongues and Languaging in Malaysia: Critical Linguistics Under Critical Examination. *Language in Society*, *46,* 567–589. https://www.jstor.org/stable/26847179

Choudhury, D., & Acharjee, T. (2022). A Novel Approach to Fake News Detection in Social Networks Using Genetic Algorithm Applying Machine Learning Classifiers. *Multimedia Tools and Applications*, *82*, 9029–9045. https://doi.org/10.1007/s11042-022-12788-1

Cui, L., & Lee, D. (2020). CoAID: COVID-19 Healthcare Misinformation Dataset. arXiv preprint arXiv:2006.00885. https://doi.org/10.48550/arXiv.2006.00885

Das, S. D., Basak, A., & Dutta, S. (2021). A Heuristic-driven Ensemble Framework for COVID-19 Fake News Detection. International Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation (pp. 164–176). https://doi.org/10.48550/arXiv.2101.03545

De, A., Bandyopadhyay, D., Gain, B., & Ekbal, A. (2021). A Transformer-based Approach to Multilingual Fake News Detection in Low-resource Languages. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *21*, 1–20. https://doi.org/10.1145/3472619

Domenico, G. D., Sit, J., Ishizaka, A., & Nunan, D. (2021). Fake News, Social Media and Marketing: A Systematic Review. *Journal of Business Research*, *124*, 329–341. https://doi.org/10.1016/j.jbusres.2020.11.037

Faustini. P., & Covões, T. (2020). Fake News Detection in Multiple Platforms and Languages. *Expert Systems with Applications*, *158*, 1–17. https://doi.org/10.1016/j.eswa.2020.113503

Faustini, P., & Covões, T. (2019). Fake News Detection Using One-class Classification. 8th Brazilian Conference on Intelligent Systems (BRACIS). https://doi.org/10.1109/BRACIS.2019.00109

Ferreira Caceres, M. M., Sosa, J. P., Lawrence, J. A., Sestacovschi, C., Tidd-Johnson, A., Rasool, M. H. U., Gadamidi, V. K., Ozair, S., Pandav, K., Cuevas-Lou, C., Parrish, M., Rodriguez, I., & Fernandez, J. P. (2022). The Impact of Misinformation on the COVID-19 Pandemic. *AIMS Public Health*, *9*(2), 262–277. https://doi.org/10.3934/publichealth.2022018

Fung, P. L., Zaidan, M. A., Timonen, H., Niemi, J. V., Kousa, A., Kuula, J., Luoma, K., Tarkoma, S., Petäjä, T., Kulmala, M., & Hussein, T. (2021). Evaluation of White-box Versus Black-box Machine Learning Models in Estimating Ambient Black Carbon Concentration. *Journal of Aerosol Science*, *152*, 105694. https://doi.org/10.1016/j.jaerosci.2020.105694

Galal, S., Nagy, N., & El-Sharkawi, M. E. (2021). CNMF: A Community-Based Fake News Mitigation Framework. *Information*, *12*(9), 376. https://doi.org/10.3390/info12090376

Grossman, G. M., & Helpman, E. (2023). Electoral Competition with Fake News. *European Journal of Political Economy*, *77*, 1–12. https://doi.org/10.1016/j.ejpoleco.2022.102315

Guibon, G., Ermakova, L., Seffih, H., Firsov, A., & Noé-Bienvenu, G. (2019). Multilingual Fake News Detection with Satire. International Conference on Computational Linguistics and Intelligent Text Processing (pp. 392–402). https://doi.org/10.1007/978-3-031-24340-0_29

Gupta, M., Dennehy, D., Parra, C. M., Mäntymäki, M., & Dwivedi, Y. K. (2023). Fake News Believability: The Effects of Political Beliefs and Espoused Cultural Values. *Information & Management*. *60*, 1–12. https://doi.org/10.1016/j.im.2022.103745

Hande, A., Puranik, K., Priyadharshini, R., Thavareesan, S., & Chakravarthi, B. R. (2021). Evaluating Pretrained Transformer-based Models for COVID-19 Fake News Detection. 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 766–772). https://doi.org/10.1109/ICCMC51019.2021.9418446

Hayawi, K., Shahriar, S., Serhani, M. A., Taleb, I., & Mathew, S. S. (2022). ANTi-Vax: A Novel Twitter Dataset for COVID-19 Vaccine Misinformation Detection. *Public Health*, *203*, 23–30. https://doi.org/10.1016/j.puhe.2021.11.022

Hu, L., Wei, S., Zhao, Z., & Wu, B. (2022). Deep Learning for Fake News Detection: A Comprehensive Survey. *AI Open*, *3*, 133–155. https://doi.org/10.1016/j.aiopen.2022.09.001

Hussain, M G., Hasan, M. R., Rahman, M., Protim, J., & Hasan, S. A. (2020). Detection of Bangla Fake News Using MNB and SVM Classifier. arXiv preprint arXiv:2005.14627. https://doi.org/10.48550/arXiv.2005.14627

Imaduwage, S., Kumara, P. P. N. V., & Samaraweera, W. J. (2022). Importance of User Representation in Propagation Network-based Fake News Detection: A Critical Review and Potential Improvements. 2nd International Conference on Advanced Research in Computing (ICARC) (pp. 90–95). https://doi.org/10.1109/ICARC54489.2022.9754103

Imbwaga, J. L., Chittaragi, N., & Koolagudi, S. (2022). Fake News Detection Using Machine Learning Algorithms. Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing (IC3-2022). https://doi.org/10.1145/3549206.3549256

Ivancová, K., Sarnovský, M., & Maslej-Krcšňáková, V. (2021). Fake News Detection in Slovak Language Using Deep Learning Techniques. IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI). http://dx.doi.org/10.1109/SAMI50585.2021.9378650

Jardaneh, G., Abdelhaq, H., Buzz, M., & Johnson, D. (2019). Classifying Arabic Tweets Based on Credibility Using Content and User Features. Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT). https://doi.org/10.1109/JEEIT.2019.8717386

Javed Mehedi Shamrat, F. M., Ranjan, R., Hasib, K. M., Yadav, A., & Siddique, A. H. (2022). Performance Evaluation Among ID3, C4.5, and CART Decision Tree Algorithm. In Ranganathan, G., Bestak, R., Palanisamy, R., & Rocha, Á. (eds). *Pervasive Computing*

and Social Networking. *Lecture Notes in Networks and Systems*, *317*. Springer, Singapore. https://doi.org/10.1007/978-981-16-5640-8_11

Jiang, T., Li, J. P., Haq, A. U., & Saboor, A. (2020). Fake News Detection Using Deep Recurrent Neural Networks. 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). https://doi.org/10.1109/ICCWAMTIP51612.2020.9317325

Jiang, T., Li, J. P., Haq, A. U., Saboor, A., & Ali, A. (2021). A Novel Stacking Approach for Accurate Detection of Fake News. *IEEE Access*, 9, 22626–22639. https://doi.org/10.1109/ACCESS.2021.3056079

Kar, D., Bhardwaj, M., Samanta, S., & Azad, A. P. (2020). No Rumours Please! A Multi-indic-lingual Approach for COVID Fake-tweet Detection. 2021 Grace Hopper Celebration India (GHCI) conference. https://doi.org/10.1109/GHCI50508.2021.9514012

Kesarwani, A., Chauhan, S. S., & Nair, A. R., (2020). Fake News Detection on Social Media Using K-Nearest Neighbours Classifier. International Conference on Advances in Computing and Communication Engineering (ICACCE). https://doi.org/10.1109–/ICACCE49060.2020.9154997

Kim, J., Tabibian, B., Oh, A., Schoelkopf, B., & Gomez-Rodriguez, M. (2018). Leveraging the Crowd to Detect and Reduce the Spread of Fake News and Misinformation. arXiv preprint arXiv:1711.09918. https://doi.org/10.48550/arXiv.1711.09918

Kong, J. T. H., Wong, W. K., Juwono, F. H., & Apriono, C. (2023). Generating Fake News Detection Model Using a Two-stage Evolutionary Approach. *IEEE Access*, *11*, 85067–85085. https://doi.org/10.1109/ACCESS.2023.3303321

Kong, S. H., Tan, L. M., Gan, K. H., & Samsudin, N. H. (2020). Fake News Detection Using Deep Learning. 2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE). https://doi.org/10.1109/DSAA49011.2020.00088

Li, Y., Jiang, B., Shu, K., & Liu, H. (2020). Mm-covid: *A Multilingual and Multimodal Data Repository for Combating COVID-19 Disinformation*. arXiv preprint arXiv:2011.04088. https://doi.org/10.48550/arXiv.2011.04088

Lin, J., Tremblay-Taylor, G., Mou, G., You, D., & Lee, K. (2019). Detecting Fake News Articles. 2019 IEEE International Conference on Big Data (Big Data) (pp. 3021–3025). http://dx.doi.org/10.1109/BigData47090.2019.9005980

Loyola-González, O. (2019). Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses from A Practical Point of View. *IEEE Access*, 7, *154096–154113*. https://doi.org/10.1109/ACCESS.2019.2949286

Maakoul, O., Boucht, S., Hachimi, K., & Azzouzi, S. (2020). Towards Evaluating the COVID'19 Related Fake News Problem: Case of Morocco. 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS). https://doi.org/10.1109/ICECOCS50124.2020.9314517

Melo, T., & Figueiredo, C. M. (2020). A First Public Dataset from Brazilian Twitter and News on COVID-19 in Portuguese. *Data in brief*, *32*, 106179. https://doi.org/10.1016–/j.dib.2020.106179

Memon, S. A., & Carley, K. M. (2020). Characterizing COVID-19 Misinformation Communities Using a Novel Twitter Dataset. arXiv preprint arXiv:2008.00791. https://doi.org/10.48550/arXiv.2008.00791

Mugdha, S. B. S., Ferdous, S. M., & Fahmin, A. (2020). Evaluating Machine Learning Algorithms for Bengali Fake News Detection. 23rd International Conference on Computer and Information Technology (ICCIT). https://doi.org/10.1109–/ICCIT51783.2020.9392662

Murayama, T., Wakamiya, S., Aramaki, E., & Kobayashi, R. (2021). Modeling the Spread of Fake News on Twitter. *PLOS ONE*, *16*(4), e0250419. https://doi.org/10.1371–/journal.pone.0250419

Nordberga, P., Kävrestada, J., & Nohlberg, M. (2020). Automatic Detection of Fake News. 6th International Workshop on Socio-Technical Perspective in IS Development (STPIS'20). https://ceur-ws.org/Vol-2789/paper23.pdf

Oliveira, N. R., Medeiros, D. S., & Mattos, D. M. (2020). A Sensitive Stylistic Approach to Identify Fake News on Social Networking. *IEEE Signal Processing Letters*, *27*, 1250–1254. http://dx.doi.org/10.1109/LSP.2020.3008087

Patwa, P., Sharma, S., Pykl, S., Guptha, V., Kumari, G., Akhtar, M. S., Ekbal, A., Das, A., & Chakraborty, T. (2021). Fighting an Infodemic: COVID-19 Fake News Dataset. In Chakraborty, T., Shu, K., Bernard, H. R., Liu, H., & Akhtar, M.S. (eds), Combating Online Hostile Posts in Regional Languages during Emergency Situation. CONSTRAINT 2021. *Communications in Computer and Information Science*, *1402*. Springer, Cham. https://doi.org/10.1007/978-3-030-73696-5_3

Pizarro, J. (2020). Profiling Bots and Fake News Spreaders at Pan'19 and Pan'20: Bots and Gender Profiling 2019, Profiling Fake News Spreaders on Twitter 2020. IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA). https://doi.org/10.1109/DSAA49011.2020.00088

Prasetyo, A., Septianto, B. D., Shidik, G. F., & Fanani, A Z. (2019). Evaluation of Feature Extraction TF-IDF in Indonesian Hoax News Classification. International Seminar on Application for Technology of Information and Communication (iSemantic). https://doi.org/10.1109/ISEMANTIC.2019.8884291

Probierz, B., Stefański, P., & Kozak, J. (2021). Rapid Detection of Fake News Based on Machine Learning Methods, *Procedia Computer Science*, *192*, 2893–2902. https://doi.org–/10.1016/j.procs.2021.09.060

Rocha, Y. M., Moura, G. A., Desidério, G. A., Oliveira C. H., Lourenço, F. D., & Figueiredo Nicolete, L. D. (2023). The Impact of Fake News on Social Media and Its Influence on Health During The COVID-19 Pandemic: A Systematic Review. *Journal of Public Health*, *31*, 1007–1016. https://doi.org/10.1007/s10389-021-01658-z

Rusli, A., Young, J. C., & Iswari, N. M. S. (2020). Identifying Fake News in Indonesian via Supervised Binary Text Classification. IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology, pp. 86–90. https://doi.org /10.1109/IAICT50021.2020.9172020

Shahi, G. K., & Nandini, D. (2020). FakeCovid –A Multilingual Cross-domain Fact Check News Dataset for COVID-19. arXiv preprint arXiv:2006.11343. https://doi.org–/10.36190/2020.14

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD explorations newsletter*, *19*(1), 22–36. https://doi.org/10.48550/arXiv.1708.01967

Sutter, G. D., Cappelle, B., Clercq, O. D., Loock, R., & Plevoets, K. (2017). Towards A Corpus-based, Statistical Approach to Translation Quality: Measuring and Visualizing Linguistic Deviance in Student Translations. *Linguistica Antverpiensia, New Series–Themes in Translation Studies*, *16*, 16–25. https://doi.org/10.52034–/lanstts.v16i0.440

Törnberg, P. (2018). Echo Chambers and Viral Misinformation: Modeling Fake News as Complex Contagion. *PLOS ONE*, *13*(9), e0203958. https://doi.org/10.1371–/journal.pone.0203958

Verma, P. K., Agrawal, P., Amorim, I., & Prodan, R. (2021). Welfare: Word Embedding Over Linguistic Features for Fake News Detection. *IEEE Transactions on Computational Social Systems*, *8*(4), 881–893. https://doi.org/10.1109/TCSS.2021.3068519

Veselý, K., Karafiát, M., Grézl, F., Janda, M., & Egorova, E. (2012). The Language-Independent Bottleneck Features. IEEE Spoken Language Technology Workshop (SLT). http://dx.doi.org/10.1109/SLT.2012.6424246

Vogel, I., & Meghana, M. (2020). Detecting Fake News Spreaders on Twitter From A Multilingual Perspective. 7th International Conference on Data Science and Advanced Analytics (DSAA). http://dx.doi.org/10.1109/DSAA49011.2020.00084

Wang, Y., Hou, Y., Che, W., & Liu, T. (2020). From Static to Dynamic Word Representations: A Survey. *International Journal of Machine Learning and Cybernetics*, *11*, 1611–1630. https://doi.org/10.1007/s13042-020-01069-8

Waszak, P., Kasprzycka-Waszak, W., Kubanek, A. (2018). The Spread of Medical Fake News in Social Media – The Pilot Quantitative Study. *Health Policy and Technology*, *7*(2), 115–118. https://doi.org/10.1016/J.HLPT.2018.03.002

Wong, W. K., Juwono, F. H., & Apriono, C. (2021). Vision-based Malware Detection: A Transfer Learning Approach Using Optimal ECOC-SVM Configuration. *IEEE Access*, *9*, 159262–159270. https://doi.org/10.1109/ACCESS.2021.3131713

Wong, W. K., Juwono, F. H., Nuwara, Y., & Kong, J. T. H. (2023). Synthesizing Missing Travel Time of P-Wave and S-Wave: A Two-Stage Evolutionary Modeling Approach. *IEEE Sensors Journal*, *23*(14), 15867–15877. http://dx.doi.org/10.1109–/JSEN.2023.3280708

Wong. W., Ming, C. I. (2019). A Review on Metaheuristic Algorithms: Recent Trends, Benchmarking and Applications. 7th International Conference on Smart Computing Communications (ICSCC). https://doi.org/10.1109/ICSCC.2019.8843624

Yang, C., Zhou, X., Zafarani, R. (2021). Checked: Chinese COVID-19 Fake News Dataset. *Social Network Analysis and Mining*, *11*(1), 1–8. https://doi.org/10.1007/s13278-021-00766-8

Zervopoulos, A., Alvanou, A. G., Bezas, K., & Papamichail, A. (2022). Deep Learning for Fake News Detection on Twitter Regarding the 2019 Hong Kong Protests. *Neural Computing and Applications*, *34*(1), 969–982. https://doi.org/10.1007/s00521-021-06230-0

Zhou, X., & Zafarani, R. (2019). Network-based Fake News Detection: A Pattern-driven Approach. *SIGKDD Explorations Newsletter*, *21*(2), 48–60. https://doi.org–/10.1145/3373464.3373473

Phishing Message Detection Based on Keyword Matching

Keng-Theen Tham
Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia

Kok-Why Ng
Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia

Su-Cheng Haw
Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia

Abstract: This paper proposes to use the Naïve Bayes-based algorithm for phishing detection, specifically in spam emails. The paper compares probability-based and frequency-based approaches and investigates the impact of imbalanced datasets and the use of stemming as a natural language processing (NLP) technique. Results show that both algorithms perform similarly in spam detection, with the choice between them depending on factors such as efficiency and scalability. Accuracy is influenced by the dataset configuration and stemming. Imbalanced datasets lead to higher accuracy in detecting emails in the majority class, while they struggle to classify minority-class emails. In contrast, balanced datasets yield overall high accuracy for both spam and ham email identification. This study reveals that stemming has a minor impact on algorithm performance, occasionally decreasing in accuracy due to word grouping. Balancing the dataset is crucial for improving algorithm performance and achieving accurate spam email detection. Hence, both probability-based and frequency-based Naïve Bayes algorithms are effective for phishing detection using balanced datasets. The frequency-based approach, with a balanced dataset and stemming, achieves a balanced performance between recall and precision, while the probability-based method with a balanced dataset and no stemming prioritises overall accuracy.

Keywords: keyword matching, phishing detection, Naïve Bayes, natural language processing, stemming

http://doi.org/10.18080/jtde.v11n3.776

Introduction

In today's digital age, communication is predominantly conducted through email and Short Message Service (SMS). These channels serve as essential means for various purposes, including business transactions involving substantial amounts of money and important notifications from subscribed companies or government entities. However, this convenience also attracts malicious individuals who attempt to deceive others by impersonating reputable entities, a practice commonly known as phishing. According to Cveticanin (2023), a study from Verizon shows that one-third of the data breaches in 2018 were caused by phishing attacks. An article from Comparitech stated that financial services are the biggest targets for phishing attacks, based on statistics provided by the Anti-Phishing Working Group (AWPG) (Cook, 2023). In the statistics provided by the AWPG, Software as a Service and webmail were the second most targeted in phishing attacks.

Desolda *et al.* (2022) stated that factors contributing to successful phishing attacks include a lack of knowledge, distraction, fatigue, pressure and a lack of awareness. Jari (2022) mentioned that the human factors leading to phishing victimisation are reciprocation, consistency and commitment, social proof, liking, authority and scarcity, where these terms are described as follows.

- Reciprocation
  Attackers may send phishing emails that appear to offer something valuable or urgent, such as free items or financial opportunities. This is to encourage recipients to click a malicious link or provide personal information.
- Consistency and commitment
  Attackers may craft emails that mimic legitimate organizations or services, relying on the recipients' previous commitments to those entities. This attack can be done by sending fake account verification or password reset emails.
- Social proof
  Phishing emails may contain fake testimonials or user reviews, giving the impression that others have already followed suit. This allows the recipients to be more comfortable taking the desired action, even if the social proof is fabricated.
- Liking
  Attackers may pose as individuals or entities that the recipients may like or trust, such as friends, colleagues or family members. Through this action the attacker can increase the likelihood that the recipients will engage with their malicious content.
- Authority
  Attackers may impersonate authoritative figures or trusted institutions, such as banks or government agencies. This misrepresentation of authority can convince the recipient that they are interacting with a credible source and will be more likely to comply with the attackers' requests.
- Scarcity
  Phishing emails may create a sense of urgency or scarcity to encourage recipients into

http://doi.org/10.18080/jtde.v11n3.776

taking immediate action, such as claiming that an account will be suspended or an opportunity will expire soon, pushing recipients to act hastily without thinking.

This is also supported by Frauenstein & Flowerday's (2020) study, wherein these factors are also known as the six key principles of persuasion or the six principles of influence. Lin *et al.* (2019) found that older women were the most susceptible to phishing emails, compared to other demographics, and younger users were less susceptible over time, while older users' susceptibility remained the same over the study period. They also concluded that weapons of influence work differently on distinct groups of people, such as younger generations are susceptible to scarcity while older generations are susceptible to reciprocation. Lastly, they proved that the current security training and warning solutions are not suitable for most users. Since different demographics are susceptible to different types of weapons of influence, the current 'one-size-fits-all' security training and warning solutions are less effective.

As most communication nowadays is online, there is a huge increase in spam messages received by users (Tay, 2023). These messages are a waste of time and can cause potential security risks, especially in the case of phishing messages. Although there are built-in spam detection methods in some messaging platforms, not all platforms receive the same treatment. This will lead to users being victims of phishing messages. Therefore, this paper proposes to use a Naïve Bayes-based spam detection method to detect phishing and spam messages received by users. The existing spam detection approaches that solely rely on word frequency may fail to capture the underlying probabilities associated with words occurring in spam emails. This limitation diminishes the overall accuracy of spam detection systems and leaves room for improvement. Therefore, this research also aims to compare the performance of the Naive Bayes classifier using probability-based features against the traditional frequency-based approach in terms of accuracy. Stemming is a technique commonly used in natural language processing (NLP) to reduce words to their base or root form, which can aid in spam detection by capturing the essence of the words without considering their specific variations. However, it remains unclear whether incorporating stemming in spam detection algorithms significantly improves the accuracy of detection compared to approaches that do not use stemming. Hence, one of the objectives of this paper is to compare the accuracy of spam detection methods in different conditions, such as in imbalanced datasets and NLP techniques.

Literature Review

For detecting phishing attacks, Adebowale *et al.* (2019) proposed a method using related features of images, frames and text of both genuine and fraudulent websites to detect a website's legitimacy. They found that although many phishing websites make their websites look similar to the original, there are still many distinctive features that can differentiate the two, such as spelling errors, long URL addresses and image alterations. The features used by

http://doi.org/10.18080/jtde.v11n3.776

their methods include page ranking of the website, the length of the website's URL, identifying abnormal URLs and if the website is using any URL-shortening services.

In addition, features such as if the website submits information to any personal email and the layout similarity of the website are critical in detecting if a website is illegitimate. Combined with the Adaptive Neuro-Fuzzy Inference System model and the Sugeno fuzzy model, the approach yielded an accuracy of 98.3%, showing success as an integrated solution for detecting web phishing. According to the authors, their approach is based on a scheme proposed by Aburrous *et al.* (2010) and a similar scheme proposed by Barraclough & Sexton (2015), both of which suggested using fuzzy techniques to detect phishing.

Aljofey *et al.* (2022) proposed a similar approach that uses machine learning and hybrid feature sets, such as the URL character sequence, hyperlink features and term frequency-inverse document frequency (TF-IDF) character-level features from the plaintext and noisy part of the web page's Hypertext Markup Language (HTML). Classification algorithms, such as eXtreme Gradient Boosting, random forest (RF), logistic regression and AdaBoost classifiers, were used to train the proposed approach. It managed to meet the requirements of real-time detection, third-party independence and high detection efficiency. However, the approach has limitations, such as its dependence on the English language and its inability to detect attached malware because it is incapable of reading and processing external files from a website. Nonetheless, the proposed approach achieved great results, reaching 96.76% accuracy, 98.28% precision and an F1-score of 96.38%.

Table 1 lists the pros and cons of the existing phishing methods.

Table 1. Literature on phishing detection

Method	Author	Pros	Cons
Phishing detection based on hyperlinks using the K-nearest neighbour algorithm	Nurul & Isredza (2021)	Achieved accuracy of 97.80% and 99.60% with 2 datasets consisting of 500 URLs, respectively.	Can only detect phishing attempts related to COVID-19. Can only detect emails that contain a URL. Currently not executed on online websites.
Phishing detection using deep learning technique	Mughaid *et al.* (2022)	Datasets contain distinctive features; thus, the result is more accurate. Achieved 97.7% accuracy using the neural network algorithm.	Feature selection needs more improvement. Lack of an automated tool to extract new features from new raw emails.
Phishing detection through a Bayesian algorithm	Baykara & Gurel (2018)	Manually add spam keywords and URLs. Contains Graphic User Interface.	Program able to directly connect to Gmail inbox; therefore, sensitive emails may be read. Only works with Gmail.
Phishing detection through machine learning approach	Mohamed *et al.* (2022)	Achieved accuracy of 95.18% using neural network	Only works for emails containing URLs.

http://doi.org/10.18080/jtde.v11n3.776

Method	Author	Pros	Cons
Phishing detection in Short Message Service (SMS) based on multiple correlation algorithms	Sonowal (2020)	Achieved accuracy of 98.40% via the Kendall ranking algorithm with AdaBoost classifiers. Lessened the number of features by 61.53%.	Time consuming.
Phishing detection through multi-layer perceptron (MLP) and random forest (RF) classification algorithms	Dalia *et al.* (2021)	Achieved accuracy of 99.46% using MLP and RF.	
SMS spam detection using content-based features and averaged neural network	Sheikhi *et al.* (2020)	Achieved accuracy of 98.8% using an averaged neural network with selected features.	Needs more records for better classification accuracy. Lack of standard sizable dataset.
Spam detection in SMS using machine learning through text mining	Julis & Alagesan (2020)	Achieved accuracy of 98% using support vector machine. Naïve Bayes has the fastest prediction time.	Cannot add additional filtering techniques or change current aspects.
SMS spam detection using the term frequency-inverse document frequency (TF-IDF) and RF algorithms.	Amir *et al.* (2019)	Accurately classifies 97.50% and achieves 98% precision using TF-IDF with RF.	Performance is lacking. Trained data are not up to standard.

Proposed Methodology

In this paper, two approaches for spam detection in Naïve Bayes were built and compared: probability-based and frequency-based approaches. Both methods were tested with and without stemming to determine if stemming affects the accuracy of Naïve Bayes in detecting spam.

Raw data

In this paper, spam and ham emails from the Spam Assassin Corpus were collected for training and testing. In this case, spam emails refer to emails that are unwanted, such as advertisements or scams, while ham emails refer to emails that are intended or are safe and legitimate emails. A total of 4,638 spam and ham emails were used for training. Based on these emails, two datasets were created. The first dataset consisted of all the collected emails, which included 1,989 spam emails and 2,649 ham emails. The second dataset consisted of 501 spam emails and 501 ham emails. The creation of the second dataset was to determine if imbalanced data would affect the accuracy of spam detection. Here, 598 emails from the Corpus were used for testing: 299 spam emails and 299 ham emails. A summary of the datasets is shown in Table 2.

http://doi.org/10.18080/jtde.v11n3.776

Table 2. Dataset summary

Dataset	Spam Email	Ham Email	Total
Imbalanced, stemmed + not stemmed	1,989	2,649	4,638
Balanced, stemmed + not stemmed	501	501	1,002
Testing	299	299	598

Processing data

To process the data in JavaScript Object Notation (JSON) format, each email was first parsed to extract the actual message from the email. After parsing the email, the message underwent a basic NLP process, which included:

- Word tokenisation
- Stop word removal
- Remove empty tokens
- Remove duplicate words
- Remove numbers and single-letter words.

When creating the dataset for testing the stemming approach, the message underwent an additional process—stemming. For this process, the Porter–Stemmer algorithm was applied.

After NLP, the message, which was now an array of words, was saved in JSON format, which includes information such as the word, its spam frequency, ham frequency, spam probability and ham probability. The spam and ham probabilities were calculated using the following formulas:

$$\textbf{\textit{Spam Probability}} = \frac{\textit{Spam Frequency}}{\textit{Spam Frequency+Ham Frequency}} \quad (1)$$

$$\textbf{\textit{Ham Probability}} = \frac{\textit{Spam Frequency}}{\textit{Spam Frequency+Ham Frequency}} \quad (2)$$

Classification algorithm

Two classification algorithms were created in this paper: Naïve Bayes with frequency-based features and with probability-based features. Before classifying an email, it went through the same basic NLP process as outlined above.

Naïve Bayes with frequency-based features

In Naïve Bayes with frequency-based features, the probabilities of an email being ham or spam without considering its features will first be calculated. These probabilities are known as prior ham and prior spam and their formulas are as follows:

$$\textbf{\textit{Prior Spam Probability}} = \frac{\textit{Total Number of Spam Emails}}{\textit{Total Number of Emails}} \quad (3)$$

$$\textbf{\textit{Prior Ham Probability}} = \frac{\textit{Total Number of Ham Emails}}{\textit{Total Number of Emails}} \quad (4)$$

http://doi.org/10.18080/jtde.v11n3.776

To verify the probability that an email is spam or ham, the spam and ham scores are calculated based on the following formulas:

$$\boldsymbol{Spam\ Score = Prior\ Spam\ Probability * Feature1\ Spam\ Probability} * \ldots \quad (5)$$

$$\boldsymbol{Ham\ Score = Prior\ Ham\ Probability * Feature1\ Ham\ Probability} * \ldots \quad (6)$$

When a decimal number is repeatedly multiplied by other decimal numbers, the value gradually decreases towards 0. This phenomenon is known as 'decimal underflow' or 'floating-point overflow', and it is solved by using the logarithmic exponential technique. The basic idea of this technique is to perform calculations in the logarithmic domain instead of the original decimal domain. This is because by working with logarithms, which are additive, instead of exponentials, which are multiplicative, the calculation will be more stable and accurate, especially when dealing with extreme values. After performing the necessary operations in the logarithmic domain, the result is exponentiated back to the original domain. However, due to the nature of the spam and ham probability being too small, a slight modification must be applied to the technique. Instead of performing the exponential function first and applying the logarithmic function later, the logarithmic function must be applied first, as performing the exponential function on the original feature's probability will not result in considerable change.

Using the above technique, the new formulas would be:

$$\boldsymbol{Spam\ Score = Exp(Log(Prior\ Spam\ Probability) + Log(Feature1\ SP) * \ldots)} \quad (7)$$

$$\boldsymbol{Ham\ Score = Exp(Log(Prior\ Ham\ Probability) + Log(Feature1\ HP) * \ldots)} \quad (8)$$

***SP** = Spam probability; ***HP** = Ham probability

To calculate each feature's spam and ham probabilities, the following formulas are used:

$$\boldsymbol{Feature\ Spam\ Probability = \frac{Feature\ Spam\ Frequency}{Total\ Number\ of\ Spam\ Frequency}} \quad (9)$$

$$\boldsymbol{Feature\ Ham\ Probability = \frac{Feature\ Ham\ Frequency}{Total\ Number\ of\ Ham\ Frequency}} \quad (10)$$

Using the above formulas, if a feature has never been registered as spam or ham, the feature's spam or ham probability will be 0, which is also known as the zero probabilities problem. Therefore, when calculating the score, the score will be undefined, as log(0) returns undefined. In this case, Laplace smoothing is employed to solve this problem. Laplace smoothing adds a constant value, usually 1, to the observed frequencies of each feature. Here, the probability estimate will never be 0, and it also evenly distributes the probability mass across all features. Hence, the new formulas would be:

$$\boldsymbol{Feature\ Spam\ Probability = \frac{Feature\ Spam\ Frequency + 1}{Total\ Number\ of\ Spam\ Frequency + Total\ Number\ of\ Features}} \quad (11)$$

http://doi.org/10.18080/jtde.v11n3.776

$$\textbf{Feature Ham Probability} = \frac{\textbf{Feature Ham Frequency} + 1}{\textbf{Total Number of Ham Frequency} + \textbf{Total Number of Features}} \quad (12)$$

To calculate the spam and ham probabilities of an email, the formulas are:

$$\textbf{Spam Probability} = \frac{\textbf{Spam Score}}{\textbf{Spam Score} + \textbf{Ham Score}} * \mathbf{100}\% \quad (13)$$

$$\textbf{Ham Probability} = \frac{\textbf{Ham Score}}{\textbf{Spam Score} + \textbf{Ham Score}} * \mathbf{100}\% \quad (14)$$

Naïve Bayes with probability-based features

In Naïve Bayes with probability-based features, the probabilities of an email being ham or spam without considering its features will first be calculated. These probabilities are known as prior ham and prior spam and their formulas are as follows:

$$\textbf{Prior Spam Probability} = \frac{\textbf{Total Number of Spam Emails}}{\textbf{Total Number of Emails}} \quad (15)$$

$$\textbf{Prior Ham Probability} = \frac{\textbf{Total Number of Ham Emails}}{\textbf{Total Number of Emails}} \quad (16)$$

To calculate the probability that an email is spam or ham, the spam and ham scores will first be computed using the formulas stated previously, with slight modification. The modification applied here is that instead of performing the logarithmic operation first, the exponential operation is performed first. This is because, here, each feature's probability value is significantly higher compared to the values calculated in the frequency-based algorithm. Therefore, the original logarithm exponential technique can be used in this case:

$$\textbf{Spam Score} = \textbf{Log}(\textbf{Exp}(\textbf{Prior Spam Probability}) + \textbf{Exp}(\textbf{Feature1 SP}) * \dots) \quad (17)$$

$$\textbf{Ham Score} = \textbf{Log}(\textbf{Exp}(\textbf{Prior Ham Probability}) + \textbf{Exp}(\textbf{Feature1 HP}) * \dots) \quad (18)$$

***SP** = Spam probability; ***HP** = Ham probability

To calculate each feature's spam and ham probabilities, the following formulas are used:

$$\textbf{Feature Spam Probability} = \frac{\textbf{Feature Spam Frequency}}{\textbf{Feature Spam Frequency} + \textbf{Feature Ham Frequency}} \quad (19)$$

$$\textbf{Feature Ham Probability} = \frac{\textbf{Feature Ham Frequency}}{\textbf{Feature Spam Frequency} + \textbf{Feature Ham Frequency}} \quad (20)$$

Using the above formulas, if a feature has never been registered as spam or ham, the feature's spam or ham frequency will be 0, which will also lead to the zero probabilities problem. Therefore, when calculating the score, the score will be undefined, as log(0) returns undefined. In this case, Laplace smoothing is employed to solve this problem, as above. Hence, if the spam or ham probability is 0, the probability will be a small value that will not greatly affect the result, which was 0.00001 here.

To calculate the spam and ham probabilities of an email, the following formulas are used:

http://doi.org/10.18080/jtde.v11n3.776

$$\textbf{\textit{Spam Probability}} = \frac{\textit{Spam Score}}{\textit{Spam Score} + \textit{Ham Score}} * \mathbf{100\%} \quad (21)$$

$$\textbf{\textit{Ham Probability}} = \frac{\textit{Ham Score}}{\textit{Spam Score} + \textit{Ham Score}} * \mathbf{100\%} \quad (22)$$

Results, Analysis, and Discussions

To test the accuracy of the algorithms, a testing dataset that contains 299 spam emails and 299 ham emails from the Spam Assassin Corpus was used. The testing was performed on the probability-based and the frequency-based algorithms. Both algorithms were evaluated under different conditions: whether stemming was applied, and whether a balanced dataset was used. For each combination of the algorithm, the dataset and the use of stemming, the number of correctly and incorrectly classified spam and ham emails was recorded. These results were then used to calculate the accuracy of each algorithm under specific conditions. The results of the testing are shown in Table 3, where the accuracy for each situation is recorded. The accuracy is shown as the percentage of correctly classified emails out of the total number of emails in the spam and ham categories.

Based on the results for the frequency-based approach, a comparable performance to the probability-based algorithm was seen across different dataset scenarios. Both the stemmed and non-stemmed versions of the algorithm obtained high spam accuracies of 96.66% and 95.99%, respectively, in the unbalanced dataset. This suggests that a significant number of spam emails were appropriately identified by the algorithm. The ham accuracy numbers were still high, showing that ham emails were also correctly classified. The frequency-based method worked remarkably well in the balanced dataset case, similar to the probability-based approach. With high spam accuracies of 96.99% and 97.32%, respectively, both the stemmed and non-stemmed versions successfully classified the spam emails. The ham accuracies remained constant at 100%, indicating that ham emails can be correctly identified.

Table 3. The accuracy values for the probability-based and frequency-based algorithms under different datasets, stemming and spam/ham emails

Algorithm	Dataset Used	Stemming	Spam/Ham	Correct	Wrong	Accuracy
Probability-based	Imbalanced	Yes	Spam	182	117	60.87%
Probability-based	Imbalanced	Yes	Ham	299	0	100.00%
Probability-based	Imbalanced	No	Spam	227	72	75.92%
Probability-based	Imbalanced	No	Ham	299	0	100.00%
Probability-based	Balanced	Yes	Spam	295	4	98.66%
Probability-based	Balanced	Yes	Ham	296	3	99.00%

http://doi.org/10.18080/jtde.v11n3.776

Algorithm	Dataset Used	Stemming	Spam/Ham	Correct	Wrong	Accuracy
Probability-based	Balanced	No	Spam	295	4	98.66%
Probability-based	Balanced	No	Ham	298	1	99.67%
Frequency-based	Imbalanced	Yes	Spam	289	10	96.66%
Frequency-based	Imbalanced	Yes	Ham	293	6	97.99%
Frequency-based	Imbalanced	No	Spam	287	12	95.99%
Frequency-based	Imbalanced	No	Ham	292	7	97.66%
Frequency-based	Balanced	Yes	Spam	290	9	96.99%
Frequency-based	Balanced	Yes	Ham	299	0	100.00%
Frequency-based	Balanced	No	Spam	291	8	97.32%
Frequency-based	Balanced	No	Ham	299	0	100.00%

Based on the results for the probability-based approach, the accuracy for identifying spam emails was low, at 60.87%, when working with an unbalanced dataset and using stemming. This demonstrates that the algorithm had trouble correctly classifying spam emails. However, the algorithm successfully identified ham emails, with a ham accuracy of 100%. The approach somewhat improved in spam detection, reaching 75.92%, when stemming was not performed on the unbalanced dataset. Regardless of whether stemming was employed, the algorithm exhibited great accuracy in recognising spam and ham emails when utilising a balanced dataset. The algorithm was capable of 98.83% and 99.17% overall accuracy in stemmed and non-stemmed scenarios, respectively, based on the results.

To assess each algorithm's performance, metrics such as precision, recall, F1-score and accuracy were used (Harikrishnan, 2021). For this assessment, true positives were spam emails that were correctly identified, true negatives were ham emails that were correctly identified, false positives were ham emails that were incorrectly identified as spam emails and false negatives were spam emails that were incorrectly identified as ham emails. Each metric is explained below.

Precision

Precision was used to measure the proportion of correctly classified spam emails out of all emails that were classified as spam. It indicates the reliability of the algorithm's identification of spam emails. Precision is calculated using the following formula:

$$\boldsymbol{Precision} = \frac{\boldsymbol{True\ Positives}}{\boldsymbol{True\ Positives + False\ Positives}} \tag{23}$$

http://doi.org/10.18080/jtde.v11n3.776

Recall

Recall was used to measure the proportion of correctly classified spam emails out of all spam emails. It indicates the algorithm's ability to identify all spam emails. Recall is calculated using the following formula:

$$\boldsymbol{Recall} = \frac{\boldsymbol{True\ Positives}}{\boldsymbol{True\ Positives + False\ Negatives}} \tag{24}$$

F1-Score

The F1-score was used to combine precision and recall into a single value, providing a balanced measure of the model's performance. It is calculated using the following formula:

$$\boldsymbol{F1\ Score} = \boldsymbol{2} * \frac{\boldsymbol{Precision} * \boldsymbol{Recall}}{\boldsymbol{Precision + Recall}} \tag{25}$$

Accuracy

Accuracy measures the overall correctness of the model's prediction. It calculates the proportion of correctly identified emails out of the total number of emails tested. Accuracy is calculated using the following formula:

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{True\ Positives + True\ Negatives}}{\boldsymbol{True\ Positives + False\ Positives + True\ Negative + False\ Negatives}} \tag{26}$$

The performance metrics of each method are presented in Table 4.

From Table 4, it can be observed that the accuracy of the algorithms varied depending on the usage of the dataset and stemming. When using an imbalanced dataset, where there was a large number of ham emails compared to spam emails, the algorithms tended to achieve higher accuracy in detecting the ham emails. This is because the accuracy is heavily influenced by the majority class. For example, the word 'transaction' appeared 1,000 and 50 times in 2,000 spam emails and 700 ham emails, respectively. If the word 'transaction' is detected in a new email, the probability will favour spam instead of ham due to the dataset used. Therefore, the algorithm and dataset used will have a high accuracy when predicting the majority class but a poor accuracy when predicting the minority class. Stemming, as a NLP technique, had a minor impact on the performance of the algorithms. In certain scenarios, stemming decreased the accuracy of the algorithms. This may be because words can have different meanings when reduced to their base or root forms. For example, the words 'occupant' and 'occupation' can be stemmed into 'occup', which may cause these two different words to be grouped into a single category.

http://doi.org/10.18080/jtde.v11n3.776

Table 4. Performance metrics for the probability-based and frequency-based Naïve Bayes algorithms in spam email classification

Algorithm	Dataset	Stemming	True Positive	True Negative	False Positive	False Negative	Precision	Recall	F1-Score	Accuracy (%)
Probability-based	Imbalanced	Yes	182	299	0	117	1	0.608	0.756	80.44
Probability-based	Imbalanced	No	227	299	0	72	1	0.759	0.863	87.96
Probability-based	Balanced	Yes	295	296	3	4	0.99	0.986	0.988	98.83
Probability-based	Balanced	No	295	298	1	4	0.996	0.987	0.992	99.17
Frequency-based	Imbalanced	Yes	289	293	6	10	0.98	0.966	0.973	97.33
Frequency-based	Imbalanced	No	287	292	7	12	0.976	0.96	0.968	96.83
Frequency-based	Balanced	Yes	290	299	0	9	1	0.97	0.985	98.5
Frequency-based	Balanced	No	291	299	0	8	1	0.973	0.986	98.66

Both algorithms achieved a higher accuracy, precision, recall and F1-score when working with balanced datasets compared to imbalanced datasets. Using a balanced dataset helps the algorithms to learn more effectively from both types of emails, which increases the performance. Therefore, it can be concluded that balancing the dataset or collecting a balanced dataset could lead to more accurate spam email detection. Comparing the probability-based and frequency-based algorithms, both approaches achieved high accuracy when using a balanced dataset. This indicates that both algorithms are effective in spam email classification tasks.

The probability-based algorithm was significantly weaker when working with an imbalanced dataset. Analysing the result for the probability-based algorithm using an imbalanced dataset, the algorithm achieved 100% accuracy in classifying ham emails but only 60–75% accuracy when classifying spam emails. This is because the algorithm is more biased towards the majority class. Thus, the algorithm will be more likely to predict that an email is ham, leading to a higher number of true negatives and a higher ham accuracy. This also causes a higher number of false negatives, leading to a lower number of true positives and a lower spam accuracy. However, the imbalanced dataset did not greatly affect the accuracy of the frequency-based algorithm compared to the probability-based algorithm. This is because both algorithms calculate the spam probability of a feature in a slightly different way. The probability-based algorithm calculates the spam and ham probabilities by considering the frequency of both spam and ham emails. The frequency-based algorithm, on the other hand, calculates the spam and ham probabilities based on the frequency within all spam emails in a dataset. Since the probability-based algorithm takes both spam and ham frequencies into

http://doi.org/10.18080/jtde.v11n3.776

account, when there is an imbalance, the dominant presence of ham emails influences the calculation of the spam probability. The frequency-based algorithm relies only on a feature's frequency within all spam emails and, therefore, is less likely to be influenced by the dataset's imbalance.

Two combinations stood out in terms of the F1-score, which reflects the balance between precision and recall. The first combination achieved an F1-score of 0.985 using the frequency-based algorithm with a balanced dataset and stemming. With this combination, spam and ham emails can be distinguished with high accuracy. The second combination achieved a remarkable F1-score of 0.992 using the probability-based algorithm with a balanced dataset and no stemming. This combination shows that both spam and ham emails can be classified with an elevated level of recall and precision. The probability-based algorithm with a balanced dataset and no stemming, however, surpassed the others, obtaining an accuracy of 99.17%, considering overall accuracy as the main factor. The accuracy of the probability-based algorithm using a balanced dataset and stemming was 98.83%, which was the second-highest overall accuracy among all combinations.

The ideal combination will be determined by the needs and priorities of the phishing detection system. For a balanced performance between recall and precision, the frequency-based approach with a balanced dataset and stemming is a great option. On the other hand, if overall accuracy is the main concern, the probability-based method with a balanced dataset and no stemming would be the best option.

Conclusion

The performance of the spam detection algorithms varied depending on the dataset configuration and the presence of stemming. The results indicated that using a balanced dataset led to a higher accuracy, precision, recall and F1-score for both the probability-based and frequency-based approaches. Balancing the dataset proved to be crucial in achieving accurate spam email detection. Stemming had a minor impact on the algorithms' performance, sometimes even decreasing the accuracy due to the grouping of words with different meanings into a single category.

Comparing the two algorithms, both the probability-based and frequency-based approaches demonstrated effectiveness in spam email identification when using a balanced dataset. However, the probability-based algorithm showed weakness when dealing with an imbalanced dataset, with a lower accuracy in classifying spam emails compared to ham emails. The frequency-based algorithm, on the other hand, was less affected by dataset imbalances, as it relied on the frequency of features within all spam emails.

http://doi.org/10.18080/jtde.v11n3.776

Based on the results, the ideal combination for creating a strong spam detection system would depend on the specific requirements and priorities of the system. If a balanced performance between recall and precision is desired, the frequency-based approach with a balanced dataset and stemming could be chosen. However, if overall accuracy is the main concern, the probability-based method with a balanced dataset and no stemming would be a preferable option.

Acknowledgements

A version of this paper was presented at the 3rd International Conference on Computer, Information Technology and Intelligent Computing (CITIC 2023), held in Malaysia on 26–28 July 2023.

References

Aburrous, M., Hossain, M. A., Dahal, K. & Thabtah, F. (2010). Intelligent phishing detection system for e-banking using fuzzy data mining. *Expert Systems with Applications, 37*(12), 7913–7921. https://doi.org/10.1016/j.eswa.2010.04.044

Adebowale, M. A., Lwin, K. T., Sánchez, E. & Hossain, M. A. (2019). Intelligent web-phishing detection and protection scheme using integrated features of images, frames and text. *Expert Systems with Applications*, *115*, 300–313. https://doi.org/10.1016/j.eswa.2018.07.067

Aljofey, A., Jiang, Q., Rasool, A., Chen, H., Liu, W., Qu, Q. & Wang, Y. (2022). An effective detection approach for phishing websites using URL and HTML features. *Scientific Reports, 12*(1). https://doi.org/10.1038/s41598-022-10841-5

Amir Sjarif, N. N., Mohd Azmi, N. F., Chuprat, S., Sarkan, H. M., Yahya, Y. & Sam, S. M. (2019). SMS spam message detection using term frequency-inverse document frequency and random forest algorithm. *Procedia Computer Science, 161*, 509–515. https://doi.org/10.1016/j.procs.2019.11.150

Barraclough, P. & Sexton, G. (2015). *Phishing website detection fuzzy system modelling* [Paper presentation]. 2015 Science and Information Conference (SAI). https://doi.org/10.1109/sai.2015.7237323

Baykara, M. & Gurel, Z. Z. (2018). *Detection of phishing attacks* [Paper presentation]. 2018 6th International Symposium on Digital Forensic and Security (ISDFS). https://doi.org/10.1109/isdfs.2018.8355389

Cook, S. (2023, 21 June). *50+ Phishing statistics, facts and trends 2017–2018*. Comparitech. https://www.comparitech.com/blog/vpn-privacy/phishing-statistics-facts/

Cveticanin, N. (2023, 14 July). *Phishing statistics & how to avoid taking the bait*. Dataprot. https://dataprot.net/statistics/phishing-statistics/

Dalia, S. A., Hanan, A. A. A. A. & Ishraq, K. A. (2021). Effective phishing emails detection method. *Turkish Journal of Computer and Mathematics Education, 12*(14), 4898–4904. https://turcomat.org/index.php/turkbilmat/article/view/11456

Desolda, G., Ferro, L. S., Marrella, A., Catarci, T. & Costabile, M. F. (2022). Human factors in phishing attacks: A systematic literature review. *ACM Computing Surveys*, *54*(8), 1–35. https://doi.org/10.1145/3469886

Frauenstein, E. D. & Flowerday, S. (2020). Susceptibility to phishing on social network sites: A personality information processing model. *Computers & Security, 94*, 101862. https://doi.org/10.1016/j.cose.2020.101862

Harikrishnan N B. (2021, 13 December). *Confusion matrix, accuracy, precision, recall, F1 score*. Medium. https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63cd

Jari, M. (2022). *An overview of phishing victimization: Human factors, training and the role of emotions* [Paper presentation]. 12th International Conference on Computer Science and Information Technology. https://doi.org/10.5121/csit.2022.121319

Julis, M. & Alagesan, S. (2020). Spam detection in SMS using machine learning through text mining. *International Journal of Scientific & Technology Research, 9*. Available at https://www.ijstr.org/final-print/feb2020/Spam-Detection-In-Sms-Using-Machine-Learning-Through-Text-Mining.pdf

Lin, T., Capecci, D. E., Ellis, D. M., Rocha, H. A., Dommaraju, S., Oliveira, D. S. & Ebner, N. C. (2019). Susceptibility to spear-phishing emails. *ACM Transactions on Computer–Human Interaction, 26*(5), 1–28. https://doi.org/10.1145/3336141

Mohamed, G., Visumathi, J., Mahdal, M., Anand, J. & Elangovan, M. (2022). An effective and secure mechanism for phishing attacks using a machine learning approach. *Processes, 10*(7), Article 1356. https://doi.org/10.3390/pr10071356

Mughaid, A., AlZu'bi, S., Hnaif, A., Taamneh, S., Alnajjar, A. & Elsoud, E. A. (2022). An intelligent cyber security phishing detection system using deep learning techniques. *Cluster Computing*, *25*, 3819–3828. https://doi.org/10.1007/s10586-022-03604-4

Nurul, A. A. & Isredza, R. A. H. (2021). COVID-19 phishing detection based on hyperlink using K-nearest neighbor (KNN) algorithm. *Applied Information Technology and Computer Science, 2*(2), 287–301. Available from https://publisher.uthm.edu.my/periodicals/index.php/aitcs/article/view/2317

Sheikhi, S., Taghi Kheirabadi, M. & Bazzazi, A. (2020). An effective model for SMS spam detection using content-based features and averaged neural network. *International Journal of Engineering, Transactions B: Applications, 33*(2), 221–228. http://dx.doi.org/10.5829/ije.2020.33.02b.06

Sonowal, G. (2020). Detecting phishing SMS based on multiple correlation algorithms. *SN Computer Science, 1*(6). https://doi.org/10.1007/s42979-020-00377-8

Tay, Y. H., Ooi, S. Y., Pang, Y. H., Gan, Y. H., & Lew, S. L. (2023). Ensuring Privacy and Security on Banking Websites in Malaysia: A Cookies Scanner Solution. *Journal of Informatics and Web Engineering*, 2(2), 153-167. https://doi.org/10.33093/jiwe.2023.2.2.12

Improving Phishing Email Detection Using the Hybrid Machine Learning Approach

Naveen Palanichamy
Faculty of Computing and Informatics, Multimedia University, Malaysia
Yoga Shri Murti
Faculty of Computing and Informatics, Multimedia University, Malaysia

Abstract: Phishing emails pose a severe risk to online users, necessitating effective identification methods to safeguard digital communication. Detection techniques are continuously researched to address the evolution of phishing strategies. Machine learning (ML) is a powerful tool for automated phishing email detection, but existing techniques like support vector machines and Naive Bayes have proven slow or ineffective in handling spam filtering. This study attempts to provide a phishing email detector and reliable classifier using a hybrid machine classifier with term frequency-inverse document frequency (TF-IDF) and an effective feature extraction technique (FET) on a real-world dataset from Kaggle. Exploratory data analysis is conducted to enhance understanding of the dataset and identify any conspicuous errors and outliers to facilitate the detection process. The FET converts the data text into a numerical representation that can be used for ML algorithms. The model's performance is evaluated using accuracy, precision, recall, F1 score, receiver operating characteristic (ROC) curve and area under the ROC curve metrics. The research findings indicate that the hybrid model utilising TF-IDF achieved superior performance, with an accuracy of 87.5%. The paper offers valuable knowledge on using ML to identify phishing emails and highlights the importance of combining various models.

Keywords: machine learning, phishing email detection, hybrid classification.

Introduction

In the modern era, email has become the primary mode of official communication, both in government and non-government sectors. It is widely recognised as an effective, fast and cost-efficient method of communication. However, with the ever-increasing use of email, the risk of receiving unsolicited and malicious content such as spam, fraud and phishing emails has also risen. According to a study by Kolmar (2023) a staggering 333.2 billion emails are sent

http://doi.org/10.18080/jtde.v11n3.778

worldwide every day. This highlights the high prevalence of email usage and the potential for encountering unwanted content.

In the first quarter of 2023, Vade, a leading email security firm, detected a significant surge in phishing emails. They identified 562.4 million phishing emails, an increase of 284.8 million compared to the previous quarter. This represented a 102% quarter-on-quarter increase, making it the highest Q1 total since 2018 (Vade Secure, n.d.). These phishing attacks typically begin with victims receiving emails that impersonate trustworthy entities, such as Wallet Connect, a program that links mobile cryptocurrency wallets to decentralised applications. By posing as a legitimate source through electronic communication, these emails deceive individuals into divulging sensitive information like passwords, credit card details and personal data. The email urges recipients to verify their wallets to prevent account suspension.

Machine learning (ML) has the potential to revolutionise the detection and prevention of phishing attacks. By leveraging large datasets and advanced algorithms, ML offers the promise of more precise and efficient phishing detection compared to traditional methods. This can play a vital role in safeguarding individuals and organisations from the detrimental consequences of phishing attacks.

In the field of phishing email detection, the research problem lies in improving the accuracy and efficiency of ML algorithms for the classification and detection of phishing emails. While previous studies have explored various methods, including email clustering and traditional detection techniques, the results have not yielded a highly accurate and time-friendly solution. Therefore, there is a crucial need to enhance the performance of ML algorithms in effectively identifying and categorising phishing emails.

This paper aims to propose a novel hybrid ML classification approach to enhance the detection of phishing emails for better protection against associated risks. The study explores the benefits of using multiple combinations of ML algorithms to deliver improved performance, increased robustness, enhanced interpretability and effective real-time predictions.

The subsequent sections of this manuscript are meticulously arranged to provide a comprehensive understanding of the study. The related studies section discusses research done in the same field. The research methodology section provides a detailed account of how the study was conducted. The evaluation of findings and discussion section explains the research setup and presents the results of the study. The last section concludes the paper with a concise and well-organised summary of the study.

http://doi.org/10.18080/jtde.v11n3.778

Related Studies

The literature review for this project consists of a summary or analysis of journals, conference papers and other sources relevant to the discussion of phishing email detection using ML algorithms. The following section discusses the feature extraction techniques (FETs) used in the reviewed papers. Measurements that assess the performance of detection are discussed next. In the last section, the papers that the authors reviewed are discussed in detail.

Machine learning

In their study, Gallo *et al.* (2021) used supervised ML to analyse suspicious emails for phishing attempts. The project encountered challenges that included the need to manually evaluate all received emails, which required human intervention, a lack of protection against targeted attacks on specific individuals and limitations associated with supervised learning. To assess performance, the authors relied on manual identification of reported emails. However, this approach made it impossible to evaluate emails that had not been reported. They evaluated various ML algorithms that included Naïve Bayes (NB), Nearest Neighbours, Linear support vector machine (SVM), RBF SVM, decision tree (DT), random forest (RF), AdaBoost and MLP Neural Net, based on precision, recall and F-score. The results showed that RF, utilising 36 characteristics, achieved a maximum precision of 95.2%, recall of 91.6% and an F-score of 93.3%. The authors propose that future research explore unsupervised ML for the same objective.

Additionally, Karim *et al.* (2020) propose an anti-spam framework that evaluates emails based on their domain and headers. They experimented with six different clustering algorithms and concluded that their proposed approach, named 'Optics', achieved an average of 3.5% better efficiency compared to the spectral and k-means algorithms. In a study by Chandra *et al.* (2019), classification and regression methods were implemented on an email spam filter dataset, with a specific focus on identifying spear phishing methods. Ensemble methods, such as boosting, bagging, stacking and voting, were incorporated into existing algorithms to improve the classification results. Through evaluation, they discovered that the K-nearest neighbours (KNN) algorithm provided efficient predictions and demonstrated effective implementation compared to the other available algorithms.

Jawale *et al.* (2018) in their paper propose a hybrid spam filtering algorithm that achieves higher accuracy by combining both filter models. While NB demonstrates faster classification, it requires a small dataset and exhibits lower accuracy performance. The SVM exhibits the highest level of accuracy performance, albeit at a slower classification speed, and it requires a substantial dataset. Their paper combined NB and SVM to leverage the strengths of both

http://doi.org/10.18080/jtde.v11n3.778

algorithms while mitigating their respective limitations. This combination achieves 99.44% accuracy surpassing the result obtained when implementing this algorithm separately. Recall and precision ratios are used to measure the spam filtering performance. They assessed spam detection by using a separate algorithm to compare its performance with the proposed combined hybrid spam detection algorithm.

The study by Vazhayil *et al.* (2018) focuses on developing phishing detection models through a non-sequential approach using classical ML classifications. The pre-processed data is converted into numeric values using a term-document matrix (TDM). The numerical data is subsequently fed ML algorithms, such as DT, KNN, logistic regression (LR), NB, RF, AdaBoost and SVM. Despite the study's unbalanced dataset, RF outperformed other methods of ML in terms of classification accuracy, leading to high categorisation rates. The authors also point out that integrating data from outside sources can improve the suggested phishing detection architecture without the need for feature selection, which typically requires domain expertise.

Raza *et al.* (2022) focus on the different techniques for classifying spam emails using ML algorithms. The authors emphasise that a supervised ML approach is the most adopted method, with the research mainly focused on bag of words (BoW) and body text features. The paper highlights the need for focusing on different features, the use of multi-algorithm systems, real-time spam classification and low false positive rates. The authors found that multi-algorithm systems tend to yield superior results compared to relying on a single algorithm. Among these, NB and SVM emerge as the most commonly used ML algorithms in this field.

In their paper, Kontsewaya *et al.* (2020) made dataset of 5728 emails from Kaggle, which consisted of both spam and non-spam (ham) emails. The authors used tokenisation to split the sentences into words and count vectorisation (CV) methods to convert words to numbers for feature extraction. The authors employed F-measure, accuracy, precision, recall and the receiver operating characteristic (ROC) area to assess the model quality. They used six alternative models to determine whether an email was spam or not, and all but the NB classifier underwent hyperparameter optimisation using GridSearchCV. Although LR was the most effective algorithm, the NB classifier had the highest level of accuracy, reaching 99%.

In a paper by Wijaya *et al.* (2016), a hybrid model combining LR and DT was proposed for spam email classification. LR was used to filter noisy data before feeding it to the DT, as DTs are sensitive to noise and can perform poorly in its presence. A false negative threshold was introduced to enhance true positive results and improve DT induction. The hybrid model (LRFNT+DT) achieved an accuracy of 91.7% on a spam-based dataset containing 57 attributes

http://doi.org/10.18080/jtde.v11n3.778

and 4061 messages. The authors concluded that LR could enhance the performance of DTs by mitigating the effect of noisy data.

Lastly, Form *et al.* (2022) proposed a method for detecting phishing emails using hybrid features and compared it to the hybrid approach (HA) from a previous study. The method utilised nine features, including domain sender, blacklist words, IP address, symbols and unique sender, selected based on common phishing techniques. SVM was employed as the classifier. The proposed method outperformed the HA with a higher accuracy of 97.25% and a lower error rate of 2.75%, compared to a 95.50% accuracy and a 4.50% error rate for HA. However, the proposed method was found to be time-consuming and had limitations with the blacklist keyword feature. Table 1 provides a summary of the ML classifiers used.

Table 1. Summary of machine learning classifiers used

No.	Authors	RF	DT	SVM	NB	LR	K-Means	KNN	DL
1.	Gallo *et al.*, 2021	/	/						
2.	Karim *et al.*, 2020						/		
3.	Toolan *et al.*, 2022		/						
4.	Raza *et al.*, 2022		/	/	/		/	/	
5.	Fang *et al.*, 2019								/
6.	Vazhayil *et al.*, 2018	/	/	/	/	/		/	
7.	Chandra *et al.*, 2019	/		/	/			/	
8.	Kontsewaya *et al.*, 2020	/	/	/	/	/		/	
9.	Ablel-Rheem *et al.*, 2020		/		/				
10.	Jawale *et al.*, 2018			/	/				
11.	Wijaya *et al.*, 2016		/			/			
12.	Form *et al.*, 2022			/					

Feature extraction techniques

Table 2 summarises the feature extraction/selection techniques used by the authors in the research papers reviewed. Along with the commonly employed techniques, there are some additional methods mentioned as well. Among the papers listed in Table 2, the term frequency-inverse document frequency (TF-IDF) was the most frequently used FET, employed by five authors (Gallo *et al.*, 2021; Karim *et al.*, 2020; Vazhayil *et al.*, 2018; Chandra *et al.*, 2019; Wijaya *et al.*, 2016). Two papers (Raza *et al.*, 2022; Toolan *et al.*, 2022) used BoW techniques. CV was used by three authors (Jawale *et al.*, 2018; Form *et al.*, 2022; Kontsewaya *et al.*, 2020). Information gain (IG) was employed by three papers (Fang *et al.*, 2019; Vazhayil

http://doi.org/10.18080/jtde.v11n3.778

et al., 2018; Jawale *et al.*, 2018). One paper (Raza *et al.*, 2022) did not specify any particular feature extraction or selection technique. IG was employed by four papers (Toolan *et al.*, 2022; Fang *et al.*, 2019; Ablel-Rheem *et al.*, 2020; Jawale *et al.*, 2018). In addition to the mentioned techniques, several authors used alternative methods. Gallo *et al.* (2021) employed a wrapper method, while Karim *et al.* (2020) utilised principal component analysis (PCA), Laplacian score (LS) and multi-cluster-based feature selection (MCFS). Form *et al.* (2022) incorporated a hybrid feature selection approach involving blacklist keywords. Vazhayil *et al.* (2018) explored TDM, along with singular value decomposition (SVD) and non-negative matrix factorisation (NMF), as feature extraction methods. Wijaya *et al.* (2016) utilised a threshold-based approach for feature selection. Overall, the most widely used FET among the papers in Table 2 was TF-IDF, while CV and IG were also commonly employed. Some authors utilised additional techniques, such as wrapper methods, PCA, LS, MCFS, hybrid feature selection and threshold-based selection.

Table 2. Summary of the feature extraction/selection techniques used by different authors

No.	Authors	TF-IDF	BoW	CV	IG	Other
1.	Gallo *et al.*, 2021	/				wrapper
2.	Karim *et al.*, 2020	/				PCA LS MCFS
3.	Toolan *et al.*, 2022		/		/	
4.	Raza *et al.*, 2022		/			
5.	Fang *et al.*, 2019				/	
6.	Vazhayil *et al.*, 2018	/				TDM SVD NMF
7.	Chandra *et al.*, 2019	/				
8.	Kontsewaya *et al.*, 2020			/		
9.	Ablel-Rheem *et al.*, 2020				/	
10.	Jawale *et al.*, 2018			/	/	
11.	Wijaya *et al.*, 2016	/				threshold
12.	Form *et al.*, 2022			/		hybrid feature blacklist keywords

Performance metrics

Table 3 summarises evaluation metrics that are commonly used in related research papers. Among the listed papers, several evaluation metrics were commonly used. Accuracy, precision,

http://doi.org/10.18080/jtde.v11n3.778

recall and F-measure were the most frequently employed metrics used by v authors, including Fang *et al.* (2019), Chandra *et al.* (2019), Kontsewaya *et al.* (2020), Ablel-Rheem *et al.* (2020), and Form *et al.* (2022). The confusion matrix was used by Fang *et al.* (2019), Chandra *et al.* (2019) and Wijaya *et al.* (2016), while area under the ROC curve (AUC) and ROC were employed by Chandra *et al.* (2019) and Form *et al.* (2022). Table 3 highlights the common evaluation trend of focusing on metrics related to accuracy, precision, recall and F-measure, which are essential for assessing the performance of phishing email detection models. These metrics provide insights into the effectiveness of the models by correctly classifying phishing emails and minimising false positives and false negatives.

Discussion of reviewed papers

The review of recent literature has uncovered certain limitations on the applicability of learning algorithms for phishing detection. The complexity of phishing emails designed to seem like legitimate emails presents challenges for accurate classification. This project aims to find efficient individual algorithms that have a high accuracy in the detection of phishing and then combine them to create an effective hybrid algorithm. The ML literature review shows that one of the drawbacks of using SVMs for phishing detection is the potentially time-consuming process of training and testing the models. However, the combination of NB and SVM has proven to be effective in spam classification when compared to other ML methods Jawale *et al.* (2018) integrated NB and SVM to leverage the strengths of both algorithms and to minimise their drawbacks. This combination yielded an accuracy of 99.44%, surpassing the results achieved when using each algorithm separately. Previous studies have predominantly relied on DTs, NB, SVM and RF for their experiments. A few papers used deep learning (DL) techniques, such as recurrent convolutional neural networks, as demonstrated in the study by Fang *et al.* (2019), which achieved notably high accuracy. NB, RF and DTs are the most frequently used and recommended ML methods in prior studies, as they perform well and are efficient in terms of time.

Based on feature extraction/selection techniques most research in the field has utilised feature extraction methods such as TF-IDF, BoW and CV. A different approach, such as the hybrid feature extraction approach described in Form *et al.* (2022), also exhibited limitations like reliance on blacklist keywords and increased time consumption. Consequently, this project will use TF-IDF and CV as FETs. Accuracy, precision, recall and F1 score are regularly used as performance assessment metrics in the papers reviewed, as well as in this paper. The performance will also be visualised using ROC and AUC measurements. Another limitation was that most of the researchers relied on spam-based datasets for their projects, as phishing

http://doi.org/10.18080/jtde.v11n3.778

emails were not sufficient for experimental purposes. As a result, this project focuses on using phishing email datasets for all the experiments.

Table 3. Summary of evaluation metrics

No.	Authors	Accuracy	Precision	Recall	F-measure	Confusion Matrix	AUC	ROC
1	Gallo *et al.*, 2021		/	/	/		/	
2	Karim *et al.*, 2020	/						
3	Toolan *et al.*, 2022			/				
4	Raza *et al.*, 2022	/						
5	Fang *et al.*, 2019	/	/	/	/	/		
6	Vazhayil *et al.*, 2018	/	/	/	/			
7	Chandra *et al.*, 2019	/	/	/	/		/	/
8	Kontsewaya *et al.*, 2020	/	/	/	/			/
9	Ablel-Rheem *et al.*, 2020		/	/	/	/		
10	Jawale *et al.*, 2018		/	/				
11.	Wijaya *et al.*, 2016				/	/		
12.	Form *et al.*, 2022	/		/		/		

Research Methodology

The overall flow of the proposed approach is shown in Figure 1. The proposed flow provides a systematic and structured framework for building and evaluating an ML classifier to detect phishing emails.

Data collection

The dataset used for this project was obtained from Kaggle, a platform that allows users to find, publish and explore datasets in a web-based data science environment. Two types of phishing datasets were collected, namely a large unbalanced and pre-made dataset and a balanced raw phishing email dataset. Dataset 1 (Akashsurya156, 2020) was used to compare the performance of ML algorithms with those previously used in ML algorithms in the reviewed papers. Dataset 2 (Hall, n.d.) was used to construct a hybrid phishing detection model using the proposed method in this project.

http://doi.org/10.18080/jtde.v11n3.778

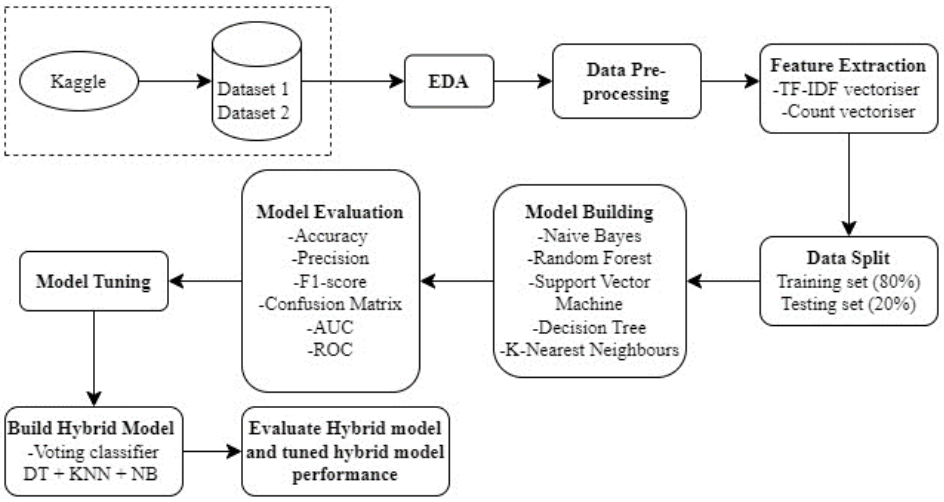


Figure 1. Proposed project methodology

Exploratory data analysis

Exploratory data analysis (EDA) techniques are applied to the dataset to examine the data before continuing to the next steps. EDA helps identify obvious errors, detect outliers and better understand the relationships between the variables. Dataset 1 exhibits an unbalance, while Dataset 2 is balanced. EDA is employed to eliminate unnecessary columns. The specified columns were then renamed. Following this, label encoding was used to assign a label of 1 to phishing emails and a label of 0 to legitimate emails. Any missing or duplicated values in the dataset were subsequently addressed and removed. The email texts were tokenised and grouped according to the total number of characters, number of words and number of sentences.

Data pre-processing

Datasets were in CSV file format. To perform data pre-processing, the required libraries must be imported—libraries like NumPy, pandas, Nltk, sklearn, matplotlib and other necessary ML model libraries. Next feature scaling and feature extraction steps were conducted to prepare the model to perform effectively without noises in the dataset. The feature extraction steps are as follows:

Lowercase

Converting all the characters to lowercase eliminates irrelevant elements or noise from the data. This approach ensures consistent flow during the mining and aids in natural language processing (NLP) tasks. It simplifies the entire dataset.

Tokenisation

Tokenisation involves dividing larger pieces of text into smaller units, called tokens. It is an important step in NLP as it allows for the generation of a meaningful representation of text

http://doi.org/10.18080/jtde.v11n3.778

data for analysis and manipulation. This project employs word tokenisation, which entails that text is broken down into individual words.

Removing special characters

Removing special characters from text is a common pre-processing step in NLP analysis. Special characters are punctuation marks, symbols and other alphanumeric characters that do not give a significant meaning to the text. By removing these characters, the text can be cleaned and made more suitable for further processing.

Removing stop words and punctuation

Stop words are common words in a language that have little meaning on their own. They are often used to connect other words in a sentence. Some common English stop words include 'the', 'an' and 'and'. They are removed from the text corpus because they do not contribute much meaning to the text and slow the processing.

Stemming

Stemming is a process that simplifies words to their base or root form, called a stem. The purpose of stemming is to reduce words while retaining their meaning to make the analysis process easier when comparing the words in a body of text. Porter stemming algorithms are used in this project. Porter stemming is a widely used method for text pre-processing in tasks related to NLP and text analysis.

Feature extraction

The objective of this method is to transform unstructured text input into a numerical representation that ML algorithms can comprehend and use. The two most often used approaches for extracting features from a dataset of phishing emails are TF-IDF and CV.

Count vectoriser

Conversely, CV is a simple technique that converts text data into a numerical representation based on the frequency of words in a document. Each document is represented by a row, while each term from the vocabulary is represented by a column in a matrix. The values in the matrix indicate the frequency of each word in each document. Table 4 illustrates the theoretical concept of CV, while Table 5 demonstrates its practical implementation (Ganesan, 2019). The sentence is 'cents, two cents, old cents, new cents: all about money'.

Table 4. Theoretical implementation of count vectorisation

	about	all	cent	money	new	old	two
doc	1	1	4	1	1	1	1

http://doi.org/10.18080/jtde.v11n3.778

Table 5. Practical implementation of count vectorisation

	0	1	2	3	4	5	6
doc	1	1	4	1	1	1	1

TF-IDF

TF-IDF is a powerful tool that was developed by Karen Spärck Jones. It is a well-known method for information retrieval and NLP to evaluate the significance of a word in a particular document when compared to a collection of documents. It considers both the frequency of the word in the document, referred to as the term frequency, and the inverse document frequency, which indicates how often the word appears in the collection. The combination of these two measures provides the TF-IDF score, which can be used to assess the relevance of a word to a specific document.

Term frequency (TF) refers to how often a term or syllable appears in a document relative to all other terms in that document. The TF-IDF equation is displayed in Equation (1).

$$TF = \frac{\text{number of times the term appears in the document}}{Total\ number\ of\ terms\ in\ the\ document} \quad (1)$$

How frequently a word appears across the whole corpus of documents that contain the phrase is determined by its IDF. IDF is determined by dividing the number of documents containing the phrase by the logarithm of the total number of documents. The equation to compute IDF is shown in Equation (2).

$$IDF = log(\frac{number\ of\ the\ documents\ in\ the\ corpus}{Number\ of\ documents\ in\ the\ corpus\ contain\ the\ term}) \quad (2)$$

The last step in calculating the TF-IDF value for each word in the document is the TF-IDF score. TF and IDF ratings for each syllable are multiplied to obtain the TF-IDF of a term. This number shows the relative importance of the word in each document relative to the total body of documents or r. Equation (3) displays the TD-IDF equation.

$$TF - IDF = TF * IDF \quad (3)$$

Splitting the dataset

The dataset is divided into two parts, with 80% designated as the training set and 20% as the testing set. This split is commonly used in ML because it uses a large enough sample for training the model and separate, independent samples for evaluating its performance. This split also indicates how effectively the model will adapt to new and unseen data, drawing on the learned relationship patterns from the training phase.

http://doi.org/10.18080/jtde.v11n3.778

Machine learning algorithms

This section discusses the implementation of the NB, SVM, RF, DT and KNN ML classifiers. The models are discussed in the following sections.

Naïve Bayes

The NB algorithm based on IBM (n.d.) calculates the probability of each feature given each class and uses these probabilities to determine the likelihood of each feature set for each class label. It assumes that all features are independent of each other when given the class label, thereby simplifying the likelihood calculation. This algorithm assumes that the features follow a Gaussian distribution and is commonly used for tasks like text classification and phishing detection. Before tokenising the texts, the dataset is cleaned using NB techniques. The probability of each word (A) is then determined, and the likelihood of an email is represented by Equation (4).

$$P(A) = \frac{\frac{a\hat{}spam}{s}}{W(\frac{a\hat{}ham}{h}+\frac{a\hat{}spam}{s})} \quad (4)$$

where P(A): is a probability for a word.

a^{spam}: Appearance time of word in spam mail.

a^{ham}: Appearance time of word in spam email.

s: Number of spam emails.

h: Number of ham mails.

w: weight, i.e., TF-IDF, which is calculated by taking the probability of spam to ham.

The composite probability for the message is then calculated after the spam probability, P(A1), has been determined.

Support Vector Machine

SVM is a popular ML method (Dhiraj, 2019) used for both classification and regression tasks. Its main objective is to create a hyperplane, which is a boundary that separates data points into distinct groups and assigns each group to a specific class. The hyperplane is represented by a line, as depicted in Equations (5) and (6):

$$y = a * x + b \quad (5)$$

$$a.x + b - y = 0 \quad (6)$$

Assume that X = (x, y) and W = (a, 1). In vector notation, we create a hyperplane, as in Equation (7):

http://doi.org/10.18080/jtde.v11n3.778

$$W * X + b = 0 \quad (7)$$

In the context of this study, where x represents the input characteristics, the weight value is denoted by W, and the bias term is represented by b. The study employs a linear kernel SVM, which aims to find the optimal hyperplane for separating opinion data into two binary classes. The process of identifying the ideal hyperplane involves considering the outermost data points from the two classes and incorporating them into the determination of the best hyperplane.

Random forest

RF is a popular ensemble learning algorithm used for regression and classification tasks, as referenced from IBM (n.d.). It combines multiple DTs to make predictions, resulting in more accurate results. RF can also assess feature importance, aiding in feature selection. However, the RF training phase can be computationally expensive due to the creation of multiple DTs. In RF and other DT-based algorithms, the Gini impurity index measures the impurity within a set of instances based on their class labels. The goal is to find the feature that minimises the Gini index, indicating higher homogeneity among class labels within the instances. Mathematically, the Gini impurity index can be calculated using Equation (8).

$$\text{Gini Index} = 1 - \sum_{i=1}^{n} (Pi))^2 \quad (8)$$

$$= 1 - [(P+)^2 + (P-)^2]$$

Decision Tree

A DT is a versatile algorithm commonly employed in various applications such as diagnosis (Saini, 2021), segmentation and phishing risk assessment. It uses a tree-like structure and a series of rules to predict the value of a target variable. Each node in the tree represents a test on an input feature, while the branches represent possible outcomes. The leaves of the tree correspond to the model's final predictions. The tree is built by repeatedly dividing the data into smaller subsets based on the most influential features of the target variable. This process, as shown in Equation (9), continues until the subset is homogeneous, meaning that all instances in the subset have the same target variable value.

$$\text{E(S)} = \sum_{i=1}^{c} - \ pi \ log2 \ \text{pi} \quad (9)$$

K-nearest neighbours

KNN classification (Harrison, 2018) predicts the class label of a new sample based on the class labels of its KNN in the training data. The value of K, determined by the user, determines the number of neighbours considered. KNN makes no assumptions about the data distribution and is easy to use. However, making predictions with larger datasets takes longer because it requires finding the KNN for each new sample. The algorithm calculates the distance between the new data point and all other instances using a chosen distance metric, such as Euclidean

http://doi.org/10.18080/jtde.v11n3.778

distance, and assigns the class label based on the majority vote among the K most similar neighbours. The Euclidean method, a standard distance measurement technique, is shown in Equation (10).

$$D(x, y)=\sqrt{\sum_{i=1}^{n}(yi - xi)^2} \quad (10)$$

Performance evaluation

This study evaluates the model's performance using confusion matrix, accuracy, precision, recall, F1 score, ROC curve and AUC. Each of these metrics will be discussed in the following sections.

Confusion matrix

Based on Bhandari (2023), the performance of an algorithm can be assessed using a particular table structure called a confusion matrix, also known as an error matrix by Karl Pearson. This method is referred to as a matching matrix in unsupervised learning, whereas in supervised methods, it is commonly known as a confusion matrix. The examples in a predicted class are represented by each column of the matrix, whereas the occurrences in an actual class are represented by each row. In Equation (11), the confusion matrix is depicted as false positives (FP), false negatives (FN), true positives (TP) and true negatives (TN). TP + TN + FP + FN represent the total number of tuples. Table 6 shows the confusion matrix.

Table 6. Confusion matrix

	Predicted Positive	**Predicted Negative**
Actual Positive	TP (True Positive)	FN (False Negative)
Actual Negative	FP (False Positive)	TN (True Negative)

(11)

where:

True positives (TP) are positive tuples that the classifier successfully categorised; TP represents the quantity of real positives.

True negatives (TN) are the negative tuples that the classifier correctly categorised; the total number of genuine negatives is TN.

False positives (FP) are negative tuples that were mistakenly classified as positive; the number of false positives, FP, will be used.

False negatives (FN) are positive tuples that were incorrectly classified as negatives. The number of false negatives, FN, will be used.

Accuracy

Accuracy gauges how well an algorithm predicts actual values. Equation (12) derived from BYJU'S (n.d.):

http://doi.org/10.18080/jtde.v11n3.778

$$Accuracy = \frac{TP+TN}{(TP+TN+FP+FN)} \quad (12)$$

Precision

Precision is used to assess the accuracy of the predicted TP in relation to the TP identified in the ground truth. The precision calculated in Equation (13), which was derived from BYJU'S (n.d.), is particularly relevant when dealing with imbalanced datasets where the minority class holds more significance or interest.

$$Precision = \frac{TP}{(TP+FP)} \quad (13)$$

Recall

Recall is a metric that assesses how accurately a model makes positive predictions. It calculates the number of correct positive predictions made by the model out of all the positive cases. Recall reveals the situations in which the model incorrectly identified affirmative cases, as in Equation (14).

$$Recall = \frac{TP}{(TP+FN)} \quad (14)$$

F1 score

The evaluation statistic known as the F-score (Shafiq *et al.* 2022), is produced using the harmonic mean of recall and precision, which is calculated as in Equation (15).

$$F1\ score = \frac{2*P*R}{P+R} \quad (15)$$

where P = precision; R = recall.

ROC curve

The accuracy of a binary classifier is visually represented by the ROC curve. As outlined by Narkhede (2018), this technique is used to show how well a classifier can distinguish between positive and negative instances. The ROC curve depicts the false positive rate (FPR) on the x-axis and the true positive rate (TPR) on the y-axis.

$$TPR = \frac{TP}{(TP+FN)};\ FPR = \frac{FP}{(FP+TN)} \quad (16)$$

AUC

The AUC is a metric of a binary classifier's ability to distinguish between two classes. As explained by Narkhede (2018), it offers a concise interpretation of the ROC curve, which visually illustrates the classifier's performance. AUC is computed by applying the equation shown in Equation (17). It provides a single numerical value that summarises the classifier's

http://doi.org/10.18080/jtde.v11n3.778

propensity to discern between positive and negative samples, and it is frequently employed in assessing ML models.

$$AUC = \int_0^1 TPR\ dFPR \tag{17}$$

Soft voting classifier

The (soft) voting classifier approach used for a hybrid combination model is a method of ML that combines the predictions drawn from various models. The architecture of the soft voting classifier is shown in Figure 2. In this case, the soft voting approach entails assigning a probability or confidence score to each model's prediction. These scores are then combined using a weighted average, taking the models' performance and reliability into account. By considering the confidence scores, the method can provide more nuanced and accurate predictions compared to a simple majority voting scheme.

Evaluation of Findings and Discussion

This section displays the results that were achieved by the proposed methods. The first section discusses the time taken to test and train, along with the accuracy and precision of the algorithms. The second section covers the evaluation of TF-IDF and CV as feature extraction techniques. The final section highlights the performance results of both the hybrid model and the tuned hybrid model. Overall, the evaluation of findings provides valuable information that can be used to make informed decisions in the ML process.

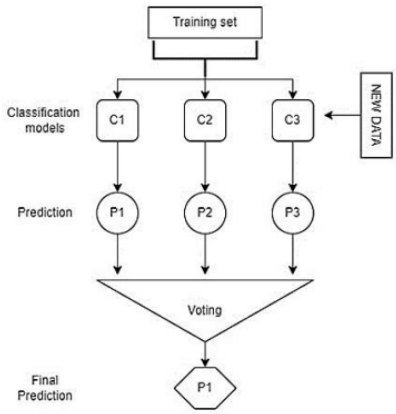


Figure 2. Voting classifier architecture for a hybrid model

Results of machine learning algorithms

Table 7 shows the results of the comparison between different ML techniques used in phishing detection. The models were trained and tested on dataset 1, and the time taken for both training and testing was recorded. The train accuracy and test accuracy of each model are also reported. Considering the result shown in Table 7, the DT model has the fastest training and

http://doi.org/10.18080/jtde.v11n3.778

testing time. It also achieved a high training accuracy of 100% and a respectable test accuracy of 99.104%. Therefore, DT appears to be the fastest and most efficient model among the other models. Following that, RF performs well in terms of training and testing accuracy. While KNN and SVM have high accuracy, the training and testing time for these two models are significantly longer compared to DT and RF.

Table 7. Machine learning technique evaluation (time and accuracy)

Model	Time taken to train	Time taken to test	Training accuracy	Testing accuracy
NB	0.2 s	0.1 s	96.754	96.669
KNN	0.5 s	33.9s	98.576	98.316
RF	50.4 s	1.1 s	99.999	99.451
SVM	3 m 59.5 s	2 m 43.9 s	98.187	98.130
DT	2.9 s	0.0 s	100.0	99.104

According to Table 8, all the models demonstrate relatively low misclassification rates, ranging from 0.55% for RF to 3.331% for NB. This suggests that the models can effectively reduce the number of incorrectly classified instances.

Table 8. Detection accuracy and misclassification rate

ML	Accuracy rate %	Misclassification rate %
NB	96.669	3.331
SVM	98.13	1.87
RF	99.45	0.55
DT	99.093	0.907
KNN	98.316	1.684

Figure 3 provides a visual representation of the model's performance, focusing on accuracy and precision. Among the models, NB and SVM exhibit the lowest precision, while in terms of accuracy, all models performed admirably, with RF showing the highest performance.

http://doi.org/10.18080/jtde.v11n3.778

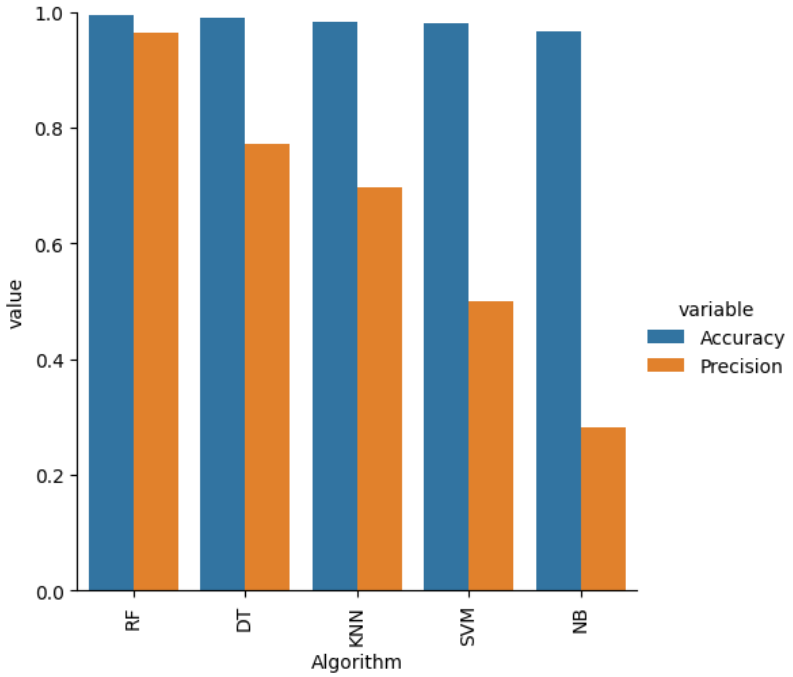


Figure 3. Accuracy and precision of machine learning classification algorithms

Results of TF-IDF vectoriser and count vectoriser

Referring to Table 9, we can assess the performance of various ML algorithms based on two feature representations, namely TF-IDF vectorisation and CV. The results highlight the variation in performance based on the feature representation used. NB consistently performs well with both feature representations, achieving high accuracy and precision scores (1.0) in all cases. KNN and SVM perform better in terms of accuracy with TF-IDF compared to CV. RF demonstrates consistent accuracy scores of 0.750 for both feature representations. In terms of precision, CV yields a slightly higher score of 0.700 compared to TF-IDF (0.6667). Among all models, only DT exhibits slightly higher accuracy with CV. The choice of feature representation, whether TF-IDF or CV, has a notable effect on the performance of ML algorithms.

Table 9. Accuracy and precision of feature extraction

Model	Accuracy TF-IDF	Accuracy CV	Precision TF-IDF	Precision CV
NB	0.9375	0.875	1.000	1.000
KNN	0.813	0.500	1.000	0.500
RF	0.750	0.750	0.6667	0.700
SVM	0.750	0.625	1.000	0.583
DT	0.6875	0.750	0.800	0.750

http://doi.org/10.18080/jtde.v11n3.778

As indicated in Table 9, TF-IDF vectorisation proves to be more effective for phishing email detection compared to CV. This aligns with the general preference for TF-IDF in text analysis tasks, where the emphasis is on capturing the importance and distinctiveness of terms. Therefore, considering the tabulated results and the overall advantages of TF-IDF in highlighting important terms, TF-IDF emerges as the preferred choice for phishing email detection and other text analysis tasks.

The result of the proposed model

The proposed hybrid model combines three algorithms: hyperparameter tuned of NB, DT and KNN, with TF-IDF vectorisation. Table 10 shows the hybrid combination algorithms both before and after hyperparameter tuning.

Table 10. Performance of the proposed hybrid model

Model	Accuracy	Precision	Recall	F1 score
NB+DT+KNN	0.8125	1.0	0.625	0.769
Tuned Hybrid	0.938	1.0	0.875	0.933

By combining the NB, KNN and DT algorithms, the model offers several advantages. These algorithms can leverage their strengths and mitigate their weaknesses. DT can capture complex relationships and provide interpretability. KNN can identify similar instances and benefit from the local patterns in the data. NB can model probabilistic dependencies between features and the target variable. Therefore, this combination potentially achieves more robustness, interpretability and efficiency of these algorithms, collectively improving the overall performance in detecting phishing emails.

Before hyperparameter tuning, the hybrid model achieves an accuracy of 0.8125, indicating the proportion of correctly classified instances. The precision score is 1.0, reflecting the model's ability to correctly identify positive instances. The recall score is 0.625, representing the model's ability to correctly identify all positive instances. The F1 score, which combines precision and recall, is 0. 769.

After hyperparameter tuning, the hybrid model exhibits significant performance improvement. The accuracy increases to 0.938, indicating a higher proportion of correctly classified instances. The precision score remains at 1.0, demonstrating the model's consistent ability to correctly identify positive instances. The recall score improves to 0.875, indicating a higher rate of correctly identifying all positive instances. The F1 score shows a notable improvement, reaching 0.933, which indicates a better balance between precision and recall.

Figure 4 shows the AUC score and ROC curve for both the hybrid model and the tuned hybrid model.

http://doi.org/10.18080/jtde.v11n3.778

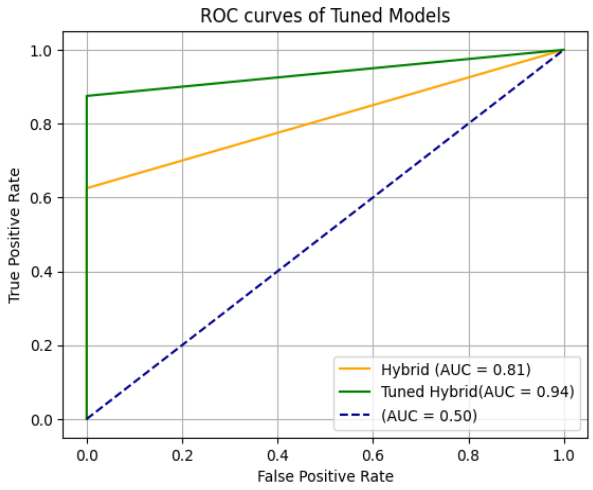


Figure 4. AUC score and ROC curve for the hybrid and tuned hybrid models

Overall, the hyper-parameter tuning process enhances the performance of the hybrid model, resulting in improved accuracy, recall and F1 score. The model shows promising potential for phishing email detection with its high precision and effective combination of multiple algorithms. This suggests that the hybrid model with tuning has successfully enhanced the model's ability to make accurate predictions and effectively capture the positive instances. Therefore, the results support the use of the tuned hybrid model as this paper proposed.

One of the major advantages of the proposed hybrid model with TF-ID is improved accuracy. By combining multiple models through a voting classifier, the hybrid model can often achieve better accuracy compared to using a single model alone. This is because the voting classifier considers the predictions of multiple models, which can provide a more robust and diverse prediction compared to a single model.

However, the proposed hybrid model also has several disadvantages, one of which is the increased computational cost. The hybrid model can be more computationally intensive compared to using a single model, due to the need to train multiple models and make predictions from each. This can lead to longer training and testing times, which can be a hindrance for larger and more complex datasets.

Another disadvantage of the hybrid model is its complexity. The hybrid model is more complex to interpret compared to a single model, as it involves multiple models and requires an understanding of how they interact with each other. This complexity can lead to greater difficulty in diagnosing and resolving issues with the model.

Conclusion

In conclusion, ML techniques have been proven to be effective in detecting phishing attacks. The results of the various models examined in this study showed that different algorithms have their strengths and weaknesses. NB, KNN, RF, SVM and DTs are some of the commonly used

http://doi.org/10.18080/jtde.v11n3.778

algorithms in the field of phishing detection. The comparison of the TF-IDF and CV methods showed that in most models, TF-IDF demonstrated higher accuracy and precision. The proposed hybrid model, which combined NB, DT and KNN with parameter tuned and combined using a voting classifier, also showed promising results with an accuracy of 0.938, precision of 1.0, recall of 0.875, F1 score of 0.933 and ROC AUC score of 0.94. The hybrid model leverages the strengths of the individual models and combines them to make predictions. This can result in better performance compared to individual models, as well as reduced overfitting and improved generalisation capabilities.

Acknowledgements

A version of this paper was presented at the third International Conference on Computer, Information Technology and Intelligent Computing, CITIC 2023, held in Malaysia from 26–28 July 2023.

References

Ablel-Rheem, D. M., Ibrahim, A. O., Kasim, S., Almazroi, A. A., & Ismail, M. A. (2020). Hybrid Feature Selection and Ensemble Learning Method for Spam Email Classification. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(1.4), 217–223. https://doi.org/10.30534/ijatcse/2020/3291.42020

Akashsurya156. (2020). Phishing Email Collection. Kaggle. https://www.kaggle.com/datasets/akashsurya156/phishing-paper1

Bhandari, A. (2023, March 13). Understanding & Interpreting Confusion Matrices for Machine Learning (Updated 2023). https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning

BYJU'S. (n.d.). *Accuracy And Precision - Definition, Examples, Need for Measurement*. BYJUS. https://byjus.com/physics/accuracy-precision-measurement/

Chandra, J. V., Challa, N., & Pasupuleti, S. K. (2019, October). Machine Learning Framework to Analyze Against Spear Phishing. *International Journal of Innovative Technology and Exploring Engineering*, *8*(12). https://doi.org/10.35940/ijitee.l3802.1081219

Dhiraj, K. (2019, June 14). Top 4 Advantages and Disadvantages of Support Vector Machine or SVM. Retrieved from https://dhirajkumarblog.medium.com/top-4-advantages-and-disadvantages-of-support-vectormachine-or-svm-a3c06a2b107

Fang, Y., Zhang, C., Huang, C., Liu, L., & Yang, Y. (2019). Phishing Email Detection Using Improved RCNN Model with Multilevel Vectors and Attention Mechanism. *IEEE Access*, 7, 56329–56340. https://doi.org/10.1109/ACCESS.2019.2913705

Form, L. M., Chiew, K. L., Sze, S. N. & Tiong, W. T. (2022, September 25). Phishing Email Detection Technique by Using Hybrid Features. 2015 9th International Conference on IT in Asia (CITA) (p. 5). https://doi.org/10.1109/cita.2015.7349818

Gallo, L., Maiello, A., Botta, A., & Ventre, G. (2021). 2 Years in the Anti-Phishing Group of a Large Company. *Computers and Security*, *105*, 102259. https://doi.org/10.1016/j.cose.2021.102259

Ganesan, K. (2019, December 5). 10+ Examples for Using CountVectorizer. *Kavita Ganesan, Ph.D*. https://kavita-ganesan.com/how-to-use-countvectorizer

Hall, C. (n.d.). Phishing Email Data by Type. www.kaggle.com. https://www.kaggle.com/datasets/charlottehall/phishing-email-data-by-type

Harrison, O. (2018, September 10). Machine Learning Basics with the K-Nearest Neighbors Algorithm. *Medium; Towards Data Science*. https://towardsdatascience.com/machine-learning-basics-with-the-knearest-neighbors-algorithm-6a6e71d01761

IBM. (n.d.). What Is Random Forest? Retrieved from https://www.ibm.com/topics/random-forest#:~:text=Random%20forest%20is%20a%20commonly,both%20classification%20and%20regression%20problems

IBM. (n.d.). What Are Naïve Bayes Classifiers? Retrieved from https://www.ibm.com/topics/naive-bayes#:~:text=The%20Na%C3%AFve%20Bayes%20classifier%20is

Jawale, D. S., Diksha, S., Jawale, K. R., & Shinkar, K. R. (2018). Hybrid Spam Detection Using Machine Learning. *International Journal of Advance Research, Ideas and Innovations in Technology*, 4(2), 1–6. https://www.ijariit.com/manuscript/hybrid-spam-detection-using-machine-learning

Karim, A., Azam, S., Shanmugam, B., & Kannoorpatti, K.(2020). Efficient Clustering of Emails into Spam and Ham: The Foundational Study of a Comprehensive Unsupervised Framework. *IEEE Access*, 8, (pp. 154759–154788). https://doi.org/10.1109/access.2020.3017082

Kolmar, C. (2023, March 30). 75 Incredible Email Statistics [2023]: How Many Emails Are Sent Per Day? Retrieved from https://www.zippia.com/advice/how-many-emails-are-sent-per-day

Kontsewaya, Y., Antonov, E., & Artamonov, A. (2020). Evaluating the Effectiveness of Machine Learning Methods for Spam Detection. *Procedia Computer Science*, *190*, 479–486. Retrieved https://doi.org/10.1016/j.procs.2021.06.056

Narkhede, S. (2018, June 26). Understanding AUC – ROC Curve. *Medium; Towards Data Science*. https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

Raza, M., Jayasinghe, N. D., & Muslam, M. M. (2022). A Comprehensive Review on Email Spam Classification Using Machine Learning Algorithms. 2021 *International Conference on Information Networking (ICOIN)*, (pp. 1–6). https://doi.org/10.1109/icoin50884.2021.9334020

Saini, A. (2021, August 29). Decision Tree Algorithm – A Complete Guide. *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2021/08/decision-tree-algorithm

Shafiq, M., Ng, H., Yap, T. T. V., & Goh, V. T. (2022). Performance of Sentiment Classifiers on Tweets of Different Clothing Brands. Journal of Informatics and Web Engineering, 1(1), 16-22.

Toolan, F., & Carthy, J. (2022). Feature Selection for Spam and Phishing Detection. 2010 eCrime Researchers Summit, Dallas, TX, USA. (pp. 1–12). https://doi.org/10.1109/ecrime.2010.5706696

Vade Secure. (n.d.). Q1 2023 Phishing and Malware Report: Phishing Increases 102% QoQ. https://www.vadesecure.com/en/blog/q1-2023-phishing-and-malware-report-phishing-increases-102-qoq

Vazhayil, A., Harikrishnan, N. B., Vinayakumar, R., & Soman, K. P. (2018). Phishing Email Detection Using Classical Machine Learning Techniques. In *Proceedings of the 1st AntiPhishing Shared Pilot at 4th ACM International Workshop on Security and Privacy Analytics (IWSPA, 2018)*, (pp. 1–8). Arizona. https://ceur-ws.org/Vol-2124/paper_11.pdf

Wijaya, A., & Bisri, A. (2016). Hybrid Decision Tree and Logistic Regression Classifier for Email Spam Detection. 2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE) (p. 4). https://doi.org/10.1109/iciteed.2016.7863267

Big Data Analytics in Tracking COVID-19 Spread Utilizing Google Location Data

Yaw Mei Wyin
Institute of Sustainable Energy, Universiti Tenaga Nasional, Kajang, Malaysia

Prajindra Sankar Krishnan
Institute of Sustainable Energy, Universiti Tenaga Nasional, Kajang, Malaysia

Chen Chai Phing
Institute of Sustainable Energy, Universiti Tenaga Nasional, Kajang, Malaysia

Tiong Sieh Kiong
Institute of Sustainable Energy, Universiti Tenaga Nasional, Kajang, Malaysia

Abstract: According to mobility data that records mobility traffic using location trackers on mobile phones, the COVID-19 epidemic and the adoption of social distance policies have drastically altered people's visiting patterns. However, rather than the volume of visitors, the transmission is controlled by the frequency and length of concurrent occupation at particular places. Therefore, it is essential to comprehend how people interact in various settings in order to focus legislation, guide contact tracking, and educate prevention initiatives. This study suggests an effective method for reducing the virus's propagation among university students enrolled on-campus by creating a self-developed Google History Location Extractor and Indicator software based on actual data on people's movements. The platform enables academics and policymakers to model the results of human mobility and the epidemic condition under various epidemic control measures and assess the potential for future advancements in the epidemic's spread. It provides tools for identifying prospective contacts, analyzing individual infection risks, and reviewing the success of campus regulations. By more precisely focusing on probable virus carriers during the screening process, the suggested multi-functional platform makes it easier to decide on epidemic control measures, ultimately helping to manage and avoid future outbreaks.

Keywords: Contact networks, human mobility simulation, epidemic control policy

http://doi.org/10.18080/jtde.v11n3.771

Introduction

Human mobility is a key factor in the analysis of an infectious disease's dissemination, especially for highly contagious diseases like COVID-19 (Gatto *et al.*, 2020; Kiang *et al.*, 2021; Venkatramanan *et al.*, 2021; Wang *et al.*, 2020; Wu *et al.*, 2020). The most efficient way to decrease the disease's rate of transmission of an outbreak is to identify potential patients early and to implement preventative measures as soon as possible. With the spread of devices with precise localization capabilities and strong wireless networks, it is now possible to track human movement over long periods of time and across a wide area. Many databases that track people's travels are now available to researchers, allowing for more in-depth analysis and epidemic spread modelling at a finer spatial and temporal granularity. Traditional sources of mobility data, where many types of motions can be gathered, include GPS and Google Location History (Ruktanonchai *et al.*, 2018).

The Google History Location Extractor software platform is suggested in this article. The platform estimates on-campus student contact behaviours through GPS trajectories, simulates on-campus student mobility in response to public policies to assess their efficacy in slowing the epidemic's spreading process, and simulates the spread of the epidemic among actual students' trajectory datasets. The platform allows data mining and exploration on the risks of infection spread, including the identification of potential secondary contacts. It consists of two key functions: a probabilistic model of infectious disease transmission at the individual level; and an individual infection risk exploration.

By simulating how people behave in response to public policies and determining how effectively each policy functions, our platform can help governments make educated judgements. In order to better understand the potential COVID-19 spread among university students, this study suggests the creation of a Google History Location Extractor program as a tool for data analysis. It will provide interactive visualization features and capabilities.

The three functions of our platform—exploring regional and individual infection risk; simulating public policy using trajectory replacement; and modelling probabilistic individual-level infectious disease transmission—are all described in depth in the sections that follow. We also evaluate related prior research in these sections. The next sections of the study deal with mobility pattern modelling, implementation, outcomes, and a discussion of our findings. We summarize the main points and stress the importance of our platform in our conclusion.

Related Works

Building a network to represent person-to-person interactions can help us understand how diseases spread throughout a society. These networks may be created using a variety of

http://doi.org/10.18080/jtde.v11n3.771

approaches, such as surveys, statistical methods, census data, and movement data (Chang *et al.*, 2021; Kumar *et al.*, 2021; Maheshwari & Albert, 2020; Muller *et al.*, 2020; Rechtin *et al.*, 2020; Schlosser *et al.*, 2020; Soures *et al.*, 2020; Yi *et al.*, 2021). As a result of the widespread use of location tracking programs on mobile phones, information about people's locations and the amount of time they spend in particular places is documented in a huge amount of data. New systems, such as Google Community Movement Reports (Google, 2020), offer tools and aggregated movement data in response to COVID-19. To protect people's privacy, data is gathered and rendered anonymous.

People's behaviour quickly changed when social distance measures were implemented across the country. The effects of these measures on state-by-state mobility trends have been extensively reported in scholarly publications and press pieces (Dave *et al.*, 2020; Lasry *et al.*, 2020; McMinn & Talbot, 2020; Pan *et al.*, 2020; Weill *et al.*, 2020). These assessments frequently use measures like visit duration, visit frequency, or travel distance to firms to gauge traffic (Klise *et al.*, 2021). These measurements offer helpful traffic indications, but they do not adequately depict the interactions that happen there (Klise *et al.*, 2021). It is crucial to take individual clustering and concurrent visits to certain venues into consideration in order to comprehend the potential of disease transmission within a community.

In the early twenty-first century, a number of extensions of epidemiological models were developed after the development of the traditional epidemiological SIR model. The SIS model was created by researchers (Parshani, Carmi & Havlin, 2010) to simulate epidemic diseases without immunity by eradicating the "recovered" population and leaving the diseased population merely susceptible. The "E" (exposed) abbreviation stands for the incubation period, which is the time when a person is infected but not yet contagious, according to the SEIR model by Prem *et al.* (2020). Due of its relevance to COVID-19, the SEIR model is employed in this work to simulate the pandemic at the individual level. Other epidemic models can easily be adapted to this simulation model.

Numerous researches examining the connection between human movement and the transmission of infectious illnesses have been published since the COVID-19 epidemic, which is exceedingly contagious, even while the illness is incubating. For instance, to simulate the spread of COVID-19, the study by Kraemer *et al.* (2020) used real-time human movement data from Baidu Inc. According to the research, isolation and other local control measures are more successful in stopping the spread of illness than travel restrictions that apply just locally. Similar to this, Wang *et al.* (2020) and Chiang *et al.* (2022) used Google movement data to examine the connection between COVID-19 spread and human movement. To further forecast the spread of diseases among diverse ethnic and socioeconomic groups, Chang *et al.* (2021) developed a meta-population SEIR model.

http://doi.org/10.18080/jtde.v11n3.771

This study was motivated by recent studies on the use of visual analytics in the modelling and control of epidemics. For instance, Guo (2007) suggests a visual analytic tool that combines extraordinarily huge geographical interaction data in order to find patterns that aid decision-making during an epidemic. Using an interactive interface, researchers in Afzal *et al.* (2020) and Afzal *et al.* (2011) compare and assess the effectiveness of various control regimes. Additionally, Dunne *et al.* (2015) illustrate how illness spreads in response to demographic factors, such as population density, allowing users to see both long- and short-term patterns of epidemic spread. In contrast to other studies, this one identifies high-risk students based on the entire campus student trajectory and offers a more detailed epidemic simulation at the individual level.

Mobility Patterns Modelling

In this study, on-campus students' mobility and network data, including their latitude and longitude coordinates, visited sites, and close proximity (within a 1-metre range) between two mobility data devices over a 14-day period, were analysed. The researchers also used computer modelling to look at movement patterns and determine how likely it was that COVID-19 would spread.

Direct contact model

In this work, a direct contact paradigm for COVID-19 was utilised from Ghayvat *et al.* (2021). Based on Ghayvat *et al.* (2021), a mobility matrix $\boldsymbol{Mob}^{\boldsymbol{mobility}}$was created using data from a worldwide mobility network, with the elements of the matrix being represented as $\boldsymbol{M}_{\boldsymbol{p,q}}^{\boldsymbol{mobility}}$. The COVID-19 subject's daily movements are shown by the $\boldsymbol{M}_{\boldsymbol{p,q}}^{\boldsymbol{mobility}}$, as well as the number of people who passed the 1-m barrier and drew nearer to the COVID-19 subject. Additionally, it shows the COVID-19 subject's movements during relocation at a specified site $\boldsymbol{c}$ at a certain time $\boldsymbol{t}$ and day $\boldsymbol{d}$, as well as during the incubation period (maximum $\boldsymbol{d}$ for COVID-19 is set to 15 days) (Gupta *et al.*, 2020). Equation (1) may be used to compute the direct contact $\boldsymbol{DC}(\boldsymbol{t},\boldsymbol{d})$ (Ghayvat *et al.*, 2021).

$$\boldsymbol{DC}_{(\boldsymbol{t},\boldsymbol{d})} = \frac{\left[\sum_{d=1}^{15}(C19_{p}^{mobility} x \sum_{n=1}^{N}\left(NS_{n,p}^{mobility}\right)\right] x A_{area}}{D_{C19_p,NS_{n,p}}} \quad (1)$$

where

$\boldsymbol{C19}_{\boldsymbol{p,q}}^{\boldsymbol{mobility}}$= COVID-19 Mobility at site p;

$\boldsymbol{NS}_{\boldsymbol{n}}$= A set of healthy people who travelled close to a COVID-19 subject for a predetermined minute at site p;

$\boldsymbol{n}$= 1, 2, 3…. N; N = number of people;

$\boldsymbol{d}$= day, 1, 2, 3…. 15;

http://doi.org/10.18080/jtde.v11n3.771

t= when the COVID-19 subject and a nearby, healthy person is close to the established threshold limit of 1 m;

n= total number of people who had contact with the COVID-19 subject;

A_{area}= the COVID-19 subject's average mobility and the total number of nearby healthy people depending on the radius $r_{p,q}$ at site p;

$D_{C19_p,NS_{n,p}}$= distance between a subject with COVID-19 and nearby healthy people.

The above listed criteria are subject to change with time, space, and place. The proposed methodology (Ghayvat *et al.*, 2021) makes use of easily accessible mobility network data to analyze instances of direct interaction with COVID-19 using a probabilistic technique. This method is useful for tracking contact instances in a particular area at a certain time and day. However, by taking into consideration those who have been close to a COVID-19 patient for a certain period of minutes or more, the COVID-19 Pandemic Direct Contact Model can be improved. In order to do this, an Indirect Contact Model from Ghayvat *et al.* (2021) was also integrated, allowing for the identification of those who may have been exposed to COVID-19 through indirect contact.

Indirect contact model

A probabilistic COVID19 pandemic indirect contact model is used (Ghayvat *et al.*, 2021) to identify people who have indirectly come into touch with COVID-19, in order to analyse the dynamics of indirect COVID-19 transmission. In order to do this, Ghayvat *et al.* (2021) created an indirect mobility matrix, $iM^{mobility}$, with $M_{p,q}^{mobility}$ as its constituent, utilising information from the global mobility network. $M_{p,q}^{mobility}$ displays the routine behaviours of healthy persons who have spent longer than the predetermined number of minutes around COVID-19 exposed individuals (Dunne *et al.*, 2015). Additionally, it takes into account their relocation moves between the regions p and q, as documented during a 15-day period beginning at time t, day d, and the $d_{final\ day}$-1. Using Equation (2), the indirect infection suspicion $ID_{(t,d)}$ may be determined (Ghayvat *et al.*, 2021).

$$ID_{(t,d)} = \frac{\left[\sum_{d=1}^{15}(iC19_p^{mobility} x \sum_{n=1}^{N}(iNS_{n,p}^{mobility})\right] x A_{area}}{D_{iC19_p,NS_{n,p}}} \quad (2)$$

where

$iC19_{p,q}^{mobility}$= suspected COVID19 subject Mobility at location p;

iNS_n= a list of nearby healthy people that have travelled near a probable COVID-19 subject for predefined duration of minutes at location p;

$D_{iC19_p,iNS_{n,p}}$= distance between suspected COVID-19 subject and the neighbouring healthy individuals.

http://doi.org/10.18080/jtde.v11n3.771

The Haversine formula, which uses the latitude and longitude information of the positions of two persons to get the distance between them in kilometres, is used to calculate the distance between two people, ($D_{C19_{p,}NS_{n,p}}$) and ($D_{iC19_{p,}iNS_{n,p}}$), at a certain time. The formula can be expressed as follows:

$$\boldsymbol{a} = sin^2\left(\frac{\Delta lat}{2}\right) + cos(lat1) \times cos(lat2) \times sin^2\left(\frac{\Delta long}{2}\right) \quad (3)$$

$$\mathbf{c} = 2 \times \arctan2(\sqrt{\mathrm{a}}, \sqrt{(1-\mathrm{a})}) \quad (4)$$

$$\mathbf{d} = \mathrm{R}.\mathrm{c} \quad (5)$$

Distance in km based on Pythagoras' theorem is given by the equations below:

$$\mathbf{x} = (\mathrm{long2} - \mathrm{long1}) \times \cos\left(\frac{\mathrm{lat1}+\mathrm{lat2}}{2}\right) \quad (6)$$

$$\mathbf{y} = \mathrm{lat2} - \mathrm{lat1} \quad (7)$$

$$\mathbf{d} = \mathrm{R}\left(\sqrt{\mathrm{x} \times \mathrm{x} + \mathrm{y} \times \mathrm{y}}\right) \quad (8)$$

where

R= Earth's radius (6371 km)
$\boldsymbol{lat1}, \boldsymbol{long1}$= latitude, longitude pair of 1st point
$\boldsymbol{lat1}, \boldsymbol{long1}$= latitude, longitude pair of 2nd point
$\Delta\boldsymbol{lat}$= difference between two latitudes
$\Delta\boldsymbol{long}$= difference between two longitudes
c= Axis interaction calculation
d= distance in kilometres

Implementation

Mobility data

Since infectious illnesses are spread by direct contact between people, epidemiologists can anticipate disease outbreaks by understanding the patterns of human movement. In order to respond to and recover from such crisis occurrences, it is essential to comprehend how individuals move through time and geography, which may be done with the use of mobile phone location data. Location history is one of the essential elements that characterizes these mobilities. Different techniques (Khalel, 2010) can be used to determine a person's location: cell site location method, GPS and Wi-Fi positioning.

To precisely pinpoint the user's location, Google Maps makes use of all three of these technologies. Users can use the Google Takeout service, which offers the data in JSON (JavaScript Object Notation) format to export this information. There are many different types of metadata in the exported JSON file. Some of the most important data points are outlined,

http://doi.org/10.18080/jtde.v11n3.771

which include heading, activity type, latitudeE7 (latitude multiplied by $1x10^7$), longitudeE7 (longitude multiplied by $1x10^7$), accuracy, timestampMs (timestamp in milliseconds) and altitude.

Along with place searches and route finding, Google Maps offers a variety of features. Through GPS location data, Google Timeline, a feature of Google Maps, continuously monitors user activity, including the locations visited and the mode of transportation utilized to get there. This unprocessed information gives the user's position at a certain time and date.

The location history information from the Android smart phones of 50 volunteer university students was examined in this study. Due to the pilot study's focus on the possible spread of COVID-19 among university students, the location data was gathered for 14 days, and only students who were staying on campus were included. The Google History Location Extractor and Indicator was used to examine the location records to spot possible COVID-19 spread. In order to check for potential COVID-19 dissemination scenarios, the study involved charting the positions of the students on a map with comprehensive data of the date, time, area, and distance between people. The Venn diagram used to examine the location data of two students is shown in Figure 1.

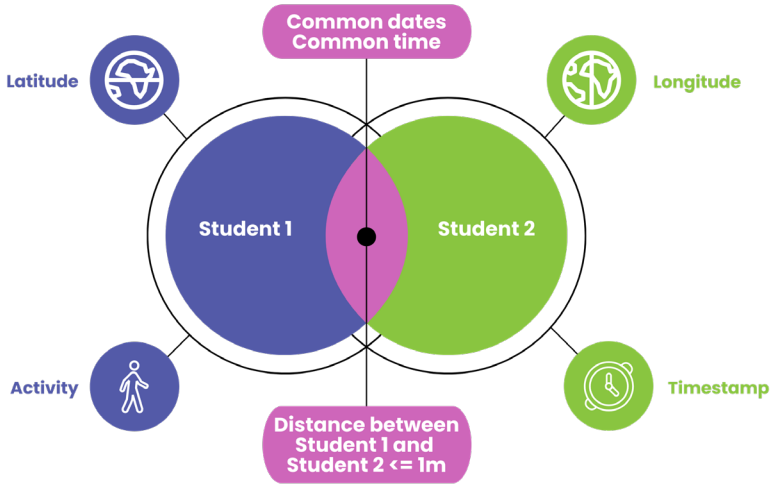


Figure 1. Venn diagram of location analysis for two-students

Software implementation

Using software to extract and visualize the downloaded data, this study examined the location history information of 50 students' JSON files. As shown in Figure 2, the program, dubbed "Google History Location Extractor", has a number of functions, such as data extraction, graphical user interface presentation, data storing in CSV or KML format, and mobility visualization on Google Maps. Additionally, for more in-depth study, the data may be exported to a Microsoft SQL Server database format. Figure 3 shows the flowchart for the entire

http://doi.org/10.18080/jtde.v11n3.771

software program, which consists of data collection, import, filtering, distance computation, and visualization.

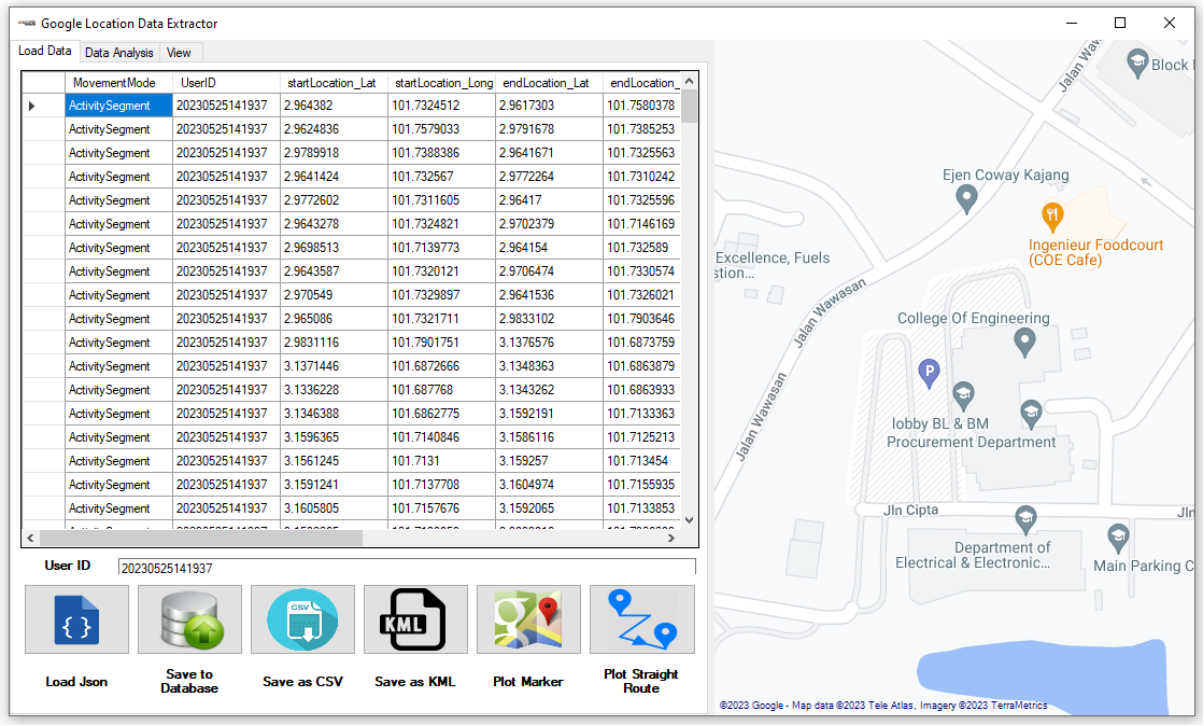


Figure 2. User interface of Google Historical Location Extractor software

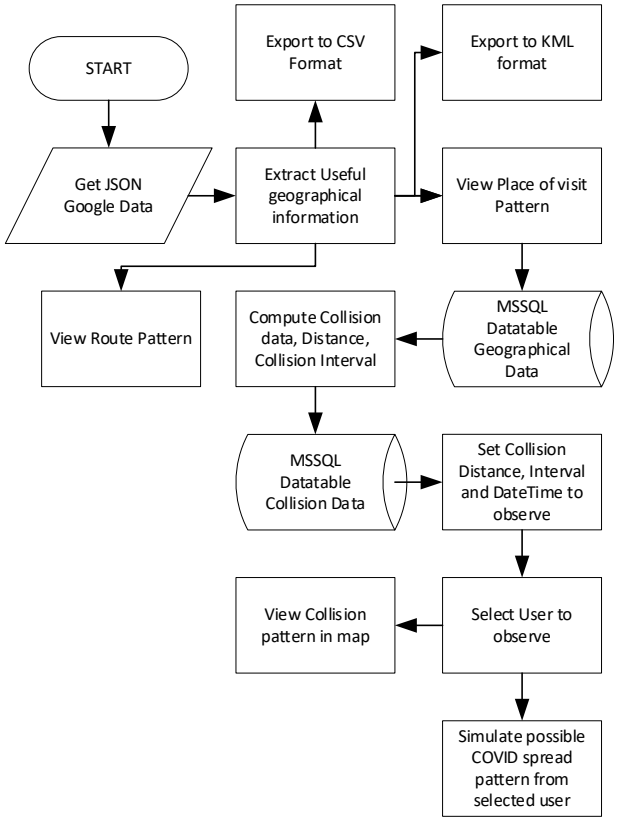


Figure 3. Entire system architecture

The process of this study started with 50 students submitting their Android smartphone location history during a 2-week period, which served as the source of the data for this

http://doi.org/10.18080/jtde.v11n3.771

research. The location history data was obtained in JSON format and converted to table format using the import technique in VB.net. Data was filtered according to accuracy. The common location sites were identified using the distance parameter, keeping only spots that are closer than one metre. Filtering speeds up the analytical process by removing irrelevant data. Lastly, a map is used to represent the findings, because it is an efficient way to convey the specifics and conclusions of the investigation. A visualization of the closeness between two students within a given distance is made using Google's Direction API.

Identifying hotspots through the use of mobility tracking data serves as a crucial tool in monitoring and mitigating the spread of contagious diseases among volunteers. While it is important to implement comprehensive infection monitoring strategies, hotspot identification plays a pivotal role in proactive surveillance and prevention.

Hotspot identification provides valuable insights into areas where the risk of disease transmission is heightened. By analyzing the density of mobility tracking data points, we can identify regions with a higher concentration of volunteers and potential interactions. This information serves as an early warning system, enabling researchers and health authorities to allocate resources effectively and implement targeted interventions to prevent future infections.

The identification of hotspots not only facilitates prompt responses but also aids in understanding the dynamics of disease transmission. By studying the characteristics of these hotspots, such as population density, demographic information, and environmental factors, we can gain valuable insights into the underlying mechanisms of disease spread. This knowledge can inform future prevention and control strategies, contributing to more effective infection monitoring and mitigation efforts.

Furthermore, hotspot identification provides a foundation for implementing complementary infection monitoring strategies. By focusing resources on these high-risk areas, researchers can prioritize regular testing protocols and enhance contact tracing efforts. This targeted approach ensures that potential infections are detected and contained at an early stage, minimizing the risk of further transmission within the volunteer group.

It is important to emphasize that hotspot identification should not be viewed as a standalone measure, but rather as an integral part of a comprehensive infection monitoring framework. By integrating hotspot identification with other strategies, such as regular testing, contact tracing, and preventive measures, researchers can establish a robust system that actively monitors and mitigates future infections. Hence, the identification of hotspots plays a crucial role in monitoring the spread of contagious diseases among volunteers. By leveraging mobility tracking data, we can proactively identify areas of heightened risk, enabling targeted

http://doi.org/10.18080/jtde.v11n3.771

interventions and resource allocation. Hotspot identification should be integrated with comprehensive infection monitoring strategies to enhance surveillance capabilities and facilitate effective prevention and control efforts.

Results

In this section, the SEIR model's deterministic compartment model for the spread of an infectious disease, which serves as the foundation for the investigation of infection risk and is further explained in the subsections that follow, will be used to develop the individual student detection method.

SEIR model

Depending on where a population is in the disease's course, the SEIR model (Kantor, 2021) splits them into four different categories. The susceptible population is the initial category and is at danger of contracting the illness. Given that SARS-CoV-2 is a novel virus, it is presumed that everyone who has never had the disease is vulnerable. The exposed population, which has been in contact with the virus but is not yet infectious, is the second category. The infected population, which has caught the virus and may spread it to others, makes up the third category. The population that has recovered from the illness and is no longer prone to the disease makes up the last group.

The following equations show how the compartment model works.

$$\frac{ds}{dt} = -\beta si \tag{9}$$

$$\frac{de}{dt} = \beta si - \alpha e \tag{10}$$

$$\frac{di}{dt} = \alpha e - \gamma i \tag{11}$$

$$\frac{dr}{dt} = \gamma i \tag{12}$$

where $\mathbf{s} + \mathbf{e} + \mathbf{i} + \mathbf{r} = \mathbf{1}$ is an invariant.

The rate processes are modelled as follows:

β is the model parameter, with units of 1/day.

s is the size of the susceptible population.

i is the size of the infective population.

αe is the rate at which the exposed population becomes infective, where e is the size of the exposed population. The average period of time in the exposed state is the incubation period of the disease, and equal to $\frac{1}{\propto}$.

http://doi.org/10.18080/jtde.v11n3.771

γi is the rate at which the infected population recovers and becomes resistant to further infection (r is the size of the recovered, resistant population). i is the size of the infective population. The average period of infectious state is $\frac{1}{\gamma}$.

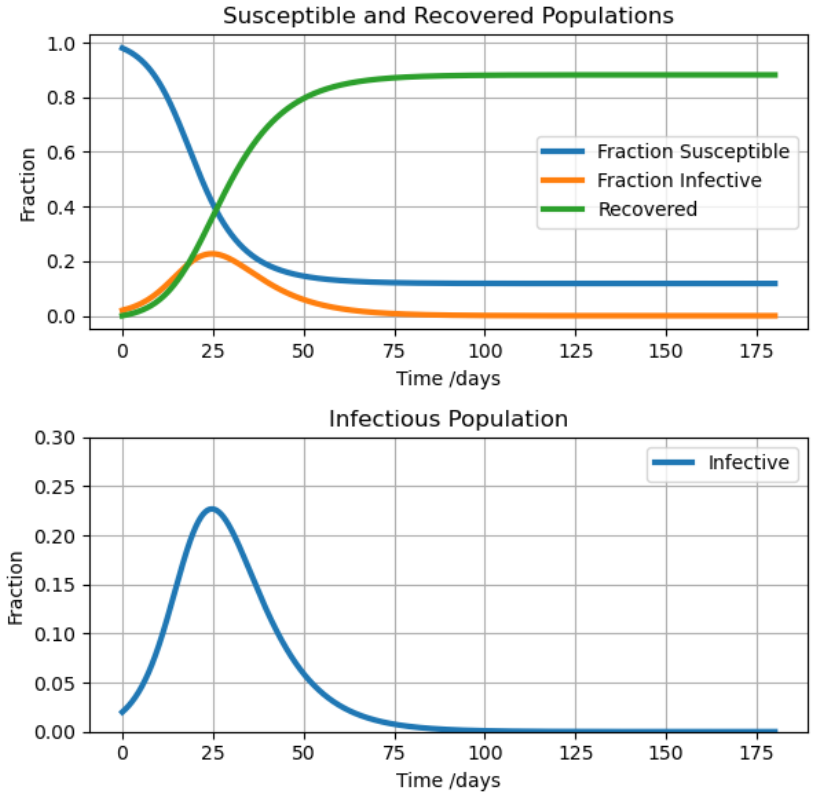


Figure 4. Simulation of the SIR model

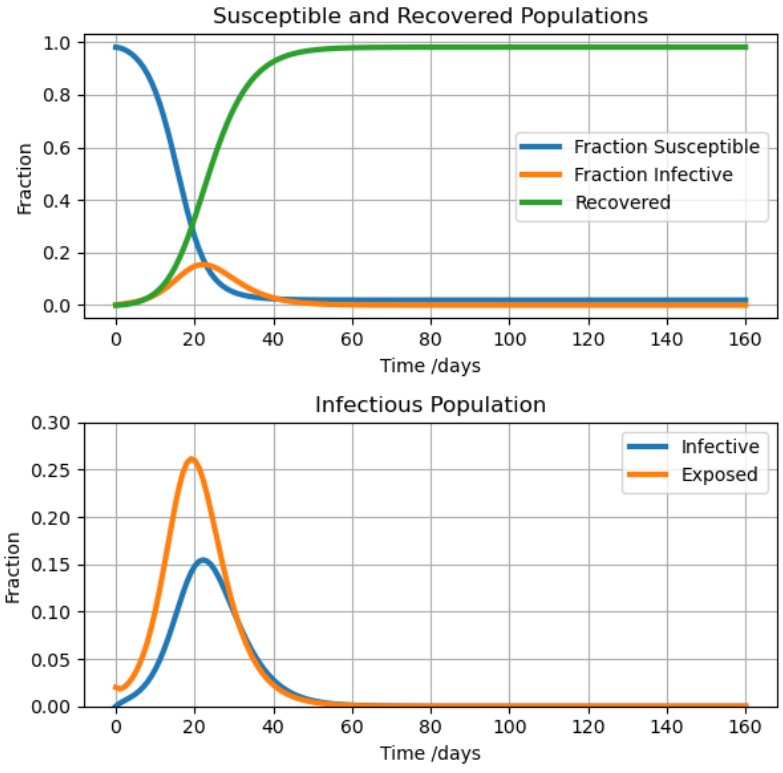


Figure 5. Simulation of the SEIR model

http://doi.org/10.18080/jtde.v11n3.771

Figure 5 depicts the introduction of an exposed population compartment, which appears to delay the spread of the illness but does not appear to reduce the number of afflicted students, in contrast to Figure 4, which is depicted without the exposed population compartment. The number of students is set at 50.

The purpose of campus regulations is to reduce the spread of the virus by social isolation techniques that work to stop infected people from infecting vulnerable others. A control parameter called μ is included to account for the efficacy of these measures in modelling: a value of 0 indicates no control; and a value of 1 indicates complete isolation of infectious persons. The goal of the model is to look at how social isolation policies may affect how an epidemic turns out.

The following diagram shows how the compartment model works (Kantor, 2021):

$$\textbf{Susceptible} \xrightarrow{(1-\mu)\frac{\beta SI}{N}} \textbf{Exposed} \xrightarrow{\alpha E} \textbf{Infectious} \xrightarrow{\gamma I} \textbf{Recovered}$$

The rate processes are modelled as follows:

The number $(\mathbf{1}-\boldsymbol{\mu})\frac{\boldsymbol{\beta SI}}{\mathbf{N}}$ represents the frequency with which an infected population transmits a disease to a vulnerable population $\boldsymbol{\mu}$, where $\boldsymbol{\mu} = \mathbf{0}$ denotes the lack of effective interventions and $\boldsymbol{\mu} = \mathbf{1}$ denotes the total cessation of disease transmission. It describes the efficacy of public health measures in preventing the spread of illness.

This produces a system of four equations (Kantor, 2021) after substitution:

$$\frac{ds}{dt} = -(1-\mu)\beta si \tag{13}$$

$$\frac{de}{dt} = (1-\mu)\beta si - \alpha e \tag{14}$$

$$\frac{di}{dt} = \alpha e - \gamma i \tag{15}$$

$$\frac{dr}{dt} = \gamma i \tag{16}$$

where $\mathbf{s}+\mathbf{e}+\mathbf{i}+\mathbf{r} = \mathbf{1}$ is an invariant.

Figure 6 shows how campus mobility restriction policies have slowed the spread of epidemics, decreased the percentage of infected students, reduced the demand on medical resources, decreased the number of students who eventually contract the disease, and saved lives for illnesses with non-zero mortality.

http://doi.org/10.18080/jtde.v11n3.771

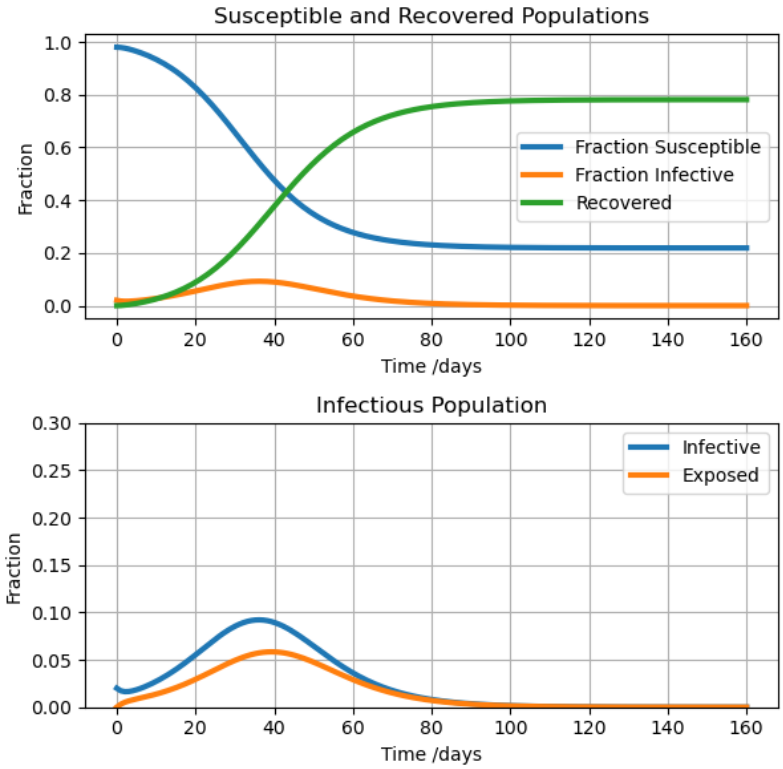


Figure 6. Simulation of the SEIR model with Control Parameter, μ

Infection risk exploration

The purpose of this project was to use data analysis to create a graphical user interface for identifying prospective encounters amongst students on campus based on their travel patterns. This was accomplished using a tracing program that classifies contacts into various danger levels according on the order of their encounters.

Using Google Mobility Data and Google Direction API, which show different combinations of latitude and longitude on a single map, Figure 7 shows the location and routes followed by two students. The whereabouts of two different students are mapped over the same days, showing that they often visited the same location. Common locations for many students are shown in Figure 8.

The proximity analysis parameters are shown in the Data Analysis of Figure 9, and the outcomes are shown in the resulting table at the bottom of the Figure. Date, time, latitude, longitude, visited place, and levels of communication are among the information in the table. In the table, Level 1 denotes direct touch; Level 2 and Level 3 denote indirect contact from Level 1 and Level 2, respectively. With the help of the settings, users may enter Start Date, Start Time, Interval (minute), Collision Distance (metre), Collision Interval (seconds), and Collision User (the user who will be chosen for proximity analysis).

http://doi.org/10.18080/jtde.v11n3.771

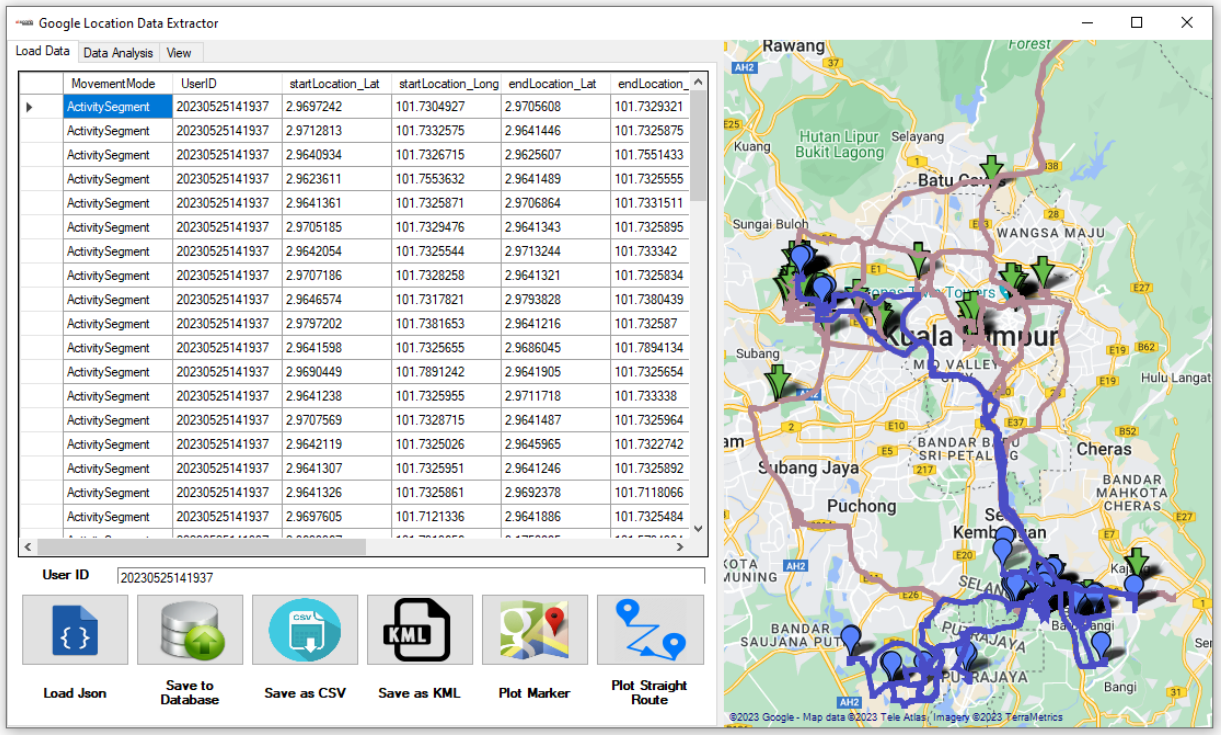


Figure 7. Location history of students displayed on the same map

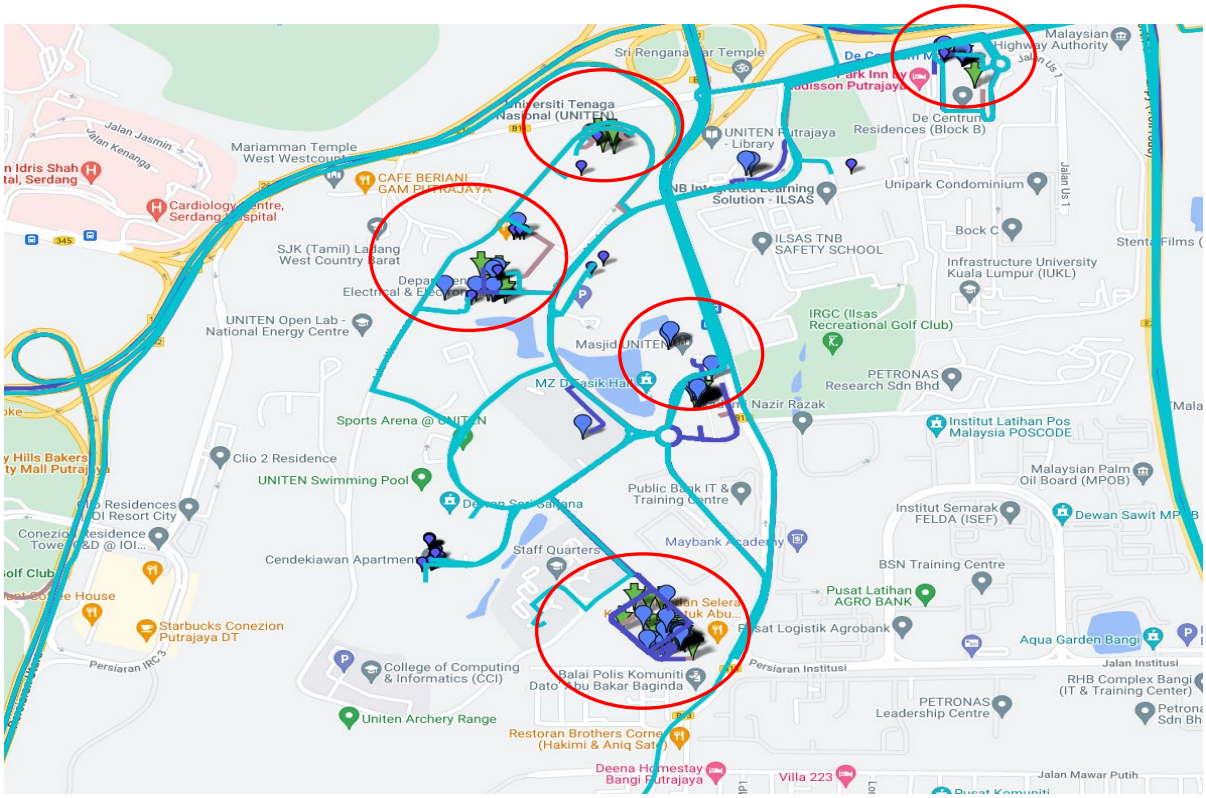


Figure 8. Common student locations on campus

Multiple vulnerable people (level 1 and 2) surround an infected person (chosen under the Collision User option) in Figure 9. Based on their level of contact with the infected user, these vulnerable people are divided into three tiers. Due to their close contact with the infected user, primary contacts (level 1) are at the highest risk of infection. Primary contacts have brought

http://doi.org/10.18080/jtde.v11n3.771

secondary and tertiary contacts (levels 2 and 3, respectively) into indirect touch with the infected user.

In order to prioritize quarantine and contact screening, particularly when resources are few, a precise classification of contacts may be highly helpful, given the exponential development in the number of connections with each level. The screening of primary contacts must be prioritized, and a full screening must be carried out as quickly as feasible. Additionally, because secondary and tertiary connections also pose a danger, screening for them should come right after screening for main contacts.

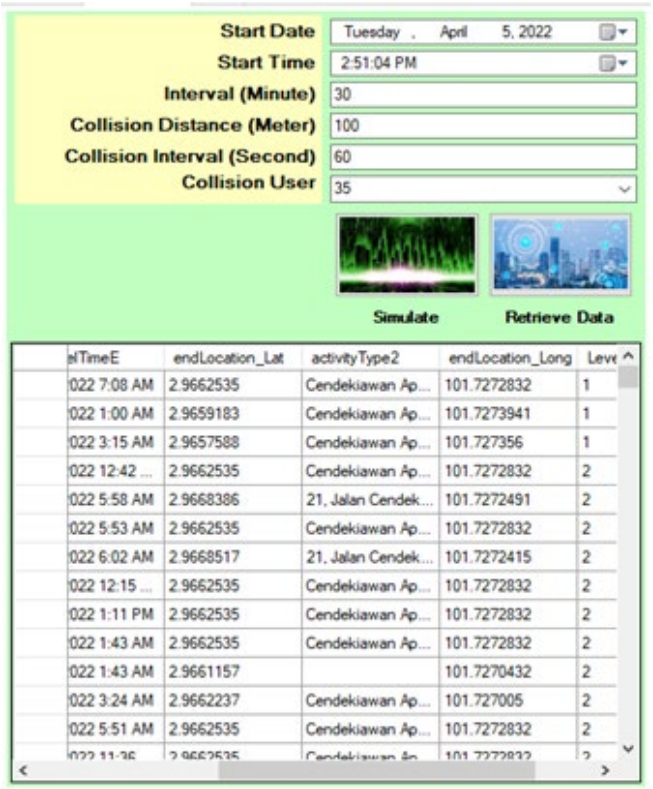


Figure 9. Graphical user interface settings for proximity analysis

Discussion

The purpose of this study was to learn more about the intricate relationships between students and how those relationships can affect how quickly a virus spreads. We identified probable transmission sites and the linkages between hotspots by examining the mobility networks of university students. A thorough description of the transmission network was produced as a consequence of the research, which made use of the data that was readily accessible regarding residency and involvement in on- and off-campus activities. The created visualization tool showed cross-paths that the system had found, indicating linkages that suggested recent touch. In order to manage the outbreak more effectively and stop further transmission within the university community, visualization and information from these links and networks could

http://doi.org/10.18080/jtde.v11n3.771

support the implementation of targeted mitigation activities, like isolating cases and quarantining contacts.

There are at least three restrictions on this work. First, certain data were missing or unknown, due to inadequate mobility data, and these data were not included in the network analysis. Second, meetings outside could not accurately represent transmission histories. There is a need for adherence to advised COVID-19 mitigation techniques, such as reducing social gathering sizes, social distance, wearing masks, improving hand cleanliness, and enhancing testing, to avoid the fast spread of COVID-19 in on- and off-campus university settings. The third drawback of this paper is the lack of data gathered from uninfected people or on adherence to mitigation techniques, such social isolation, mask usage, and hand cleanliness. On university and college campuses, promoting virtual activities, such as those connected to sizable student meetings, may assist to reduce the danger of transmission. Collaboration between university administrators, student organizations, and health authorities is essential to ensure that COVID-19 mitigation measures are followed.

Conclusion

This work proposes a data analysis approach that makes use of a GUI (graphical user interface) to anticipate social distance violations among students on university and college campuses and to detect common movement patterns. Important characteristics, like longitude, latitude, distance, and the time and date of contact within a specified range, are included in the GUI that is used to display analyzed information on a map. A student's past locations may be tracked and compared to those of others, making it possible to identify contacts who could be targeted for screening or quarantine of high-risk persons.

The GUI platform also incorporates region mining, which enables quick calculation of possible dangers and early detection of suspected diseases. These features give users, like the Centers for Disease Control and Prevention, an intuitive user interface to model potential outcomes and arrive at judgements.

The GUI has a number of benefits, such as the capacity to analyze massive quantities of student data concurrently, the ability to determine whether a student has visited a high-risk or polluted region, and its independence from extra hardware for analysis. However, it should be noted that using location services might reduce battery life of handsets and necessitate continual Internet access.

Overall, this data analysis technique has the potential to offer insightful information for monitoring and supressing the spread of COVID-19 on university and college campuses when

http://doi.org/10.18080/jtde.v11n3.771

paired with the GUI. The viability and efficacy of utilizing this technology as a component of an all-encompassing mitigation plan require more study.

Acknowledgements

This work is supported by IDRC Grant No. 109586 – 001 for Artificial Intelligence Framework for Threat Assessment and Containment for COVID-19 and Future Epidemics while Mitigating the Socioeconomic Impact to Women, Children and Underprivileged Groups.

A version of this paper was presented at the third International Conference on Computer, Information Technology and Intelligent Computing, CITIC 2023, held in Malaysia on 26–28 July 2023.

References

Afzal, S., Ghani, S., Jenkins-Smith, H. C., Ebert, D. S., Hadwiger, M., & Hoteit, I. (2020). A Visual Analytics Based Decision Making Environment for COVID-19 Modeling and Visualization. 2020 IEEE Visualization Conference (VIS), Salt Lake City, UT, USA, pp. 86–90. https://doi.org/10.1109/VIS47514.2020.00024

Afzal, S., Maciejewski, R., & Ebert, D. S. (2011). Visual analytics decision support environment for epidemic modeling and response evaluation. 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), Providence, RI, USA, pp. 191–200. https://doi.org/10.1109/VAST.2011.6102457

Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., & Leskovec, J. (2021). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, *589*, 82–87. https://doi.org/10.1038/s41586-020-2923-3

Chiang, W. H., Liu, X., & Mohler, G. (2022). Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. *International Journal of Forecasting*, *38*(2), 505–520. https://doi.org/10.1016/j.ijforecast.2021.07.001

Dave, D., McNichols, D., & Sabia, J. J. (2021). The Contagion Externality of a Superspreading Event: The Sturgis Motorcycle Rally and COVID-19. *Southern Economic Journal*, *87*(3), 769–807. https://doi.org/10.1002/soej.12475

Dunne, C., Muller, M. J., Perra, N., & Martino, M. (2015). VoroGraph: Visualization Tools for Epidemic Analysis. Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pp. 255–258. https://doi.org/10.1145/2702613.2725459

Guo, D. (2007). Visual analytics of spatial interaction patterns for pandemic decision support. *International Journal of Geographical Information Science*, *21*(8), 859–877, https://doi.org/10.1080/13658810701349037

Khalel, A. M. H. (2010). Position Location Techniques in Wireless Communication Systems (Dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:bth-4796

Gatto, M., Bertuzzo, E., Mari, L., Miccoli, S., Carraro, L., Casagrandi, R., and Andrea Rinaldo. (2020). Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency

containment measures. *Proceedings of the National Academy of Sciences*, USA, *117*(19), 10484–10491. https://doi.org/10.1073/pnas.2004978117

Ghayvat, H., Awais, M., Gope, P., Pandya, S., & Majumdar, S. (2021). Recognizing Suspect and Predicting the Spread of Contagion Based on Mobile Phone Location Data (COUNTERACT): A system of identifying COVID-19 infectious and hazardous sites, detecting disease outbreaks based on the internet of things, edge computing, and artificial intelligence. *Sustainable Cities and Society*, *69*, 102798. https://doi.org/10.1016/j.scs.2021.102798

Google. (2020). Google COVID-19 Community Mobility Reports, 2020. Available from https://www.google.com/covid19/mobility

Gupta, R., Bedi, M., Goyal, P., Wadhera, S., & Verma, V. (2020). Analysis of COVID-19 tracking tool in India: Case study of *Aarogya Setu* mobile application. *Digital Government: Research and Practice*, *1*(4), 28. https://doi.org/10.1145/3416088

Kantor, J. (2021). 3.9 modeling and control of a campus outbreak of Coronavirus Covid-19. Retrieved March 13, 2023, from https://jckantor.github.io/CBE30338/03.09-COVID-19.html

Kiang, M. V., Santillana, M., Chen, J. T., Onnela, J.-P., Krieger, N., Engø-Monsen, K., Ekapirat, N., Areechokchai, D., Prempree, P., Maude, R. J., & Buckee, C. O. (2021). Incorporating human mobility data improves forecasts of Dengue fever in Thailand. *Scientific Reports*, *11*, 923. https://doi.org/10.1038/s41598-020-79438-0

Klise, K., Beyeler, W., Finley, P., & Makvandi, M. (2021). Analysis of mobility data to build contact networks for COVID-19. *PLoS One*, *16*(4), e0249726. https://doi.org/10.1371/journal.pone.0249726

Kraemer, M. U. G., Yang, C. H., Gutierrez, B., Wu, C. H., Klein, B., Pigott, D. M., Open COVID-19 Data Working Group, du Plessis, L., Faria, N. R., Li, R., Hanage, W. P., Brownstein, J. S., Layan, M., Vespignani, A., Tian, H., Dye, C., Pybus, O. G., & Scarpino, S. V. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, *368*(6490), 493–497. https://doi.org/10.1126/science.abb4218

Kumar, N., Oke, J. & Nahmias-Biran, Bh. (2021). Activity-based epidemic propagation and contact network scaling in auto-dependent metropolitan areas. Scientific Reports, *11*, 22665. https://doi.org/10.1038/s41598-021-01522-w

Lasry, A., Kidder, D., Hast, M., Poovey, J., Sunshine, G., Winglee. K., *et al.* (2020). Timing of community mitigation and changes in reported COVID-19 and community mobility–four US metropolitan areas, February 26–April 1, 2020. *Morbidity and Mortality Weekly Report*, *69*(15), 451–457. https://doi.org/10.15585/mmwr.mm6915e2

Maheshwari, P., & Albert, R. (2020). Network model and analysis of the spread of Covid-19 with social distancing. *Applied Network Science*, *5*, 100. https://doi.org/10.1007/s41109-020-00344-5

McMinn, S., & Talbot, R. (2020). Mobile Phone Data Show More Americans Are Leaving Their Homes, Despite Orders. NPR, The Coronavirus Crisis. Available from https://www.npr.org/2020/05/01/849161820/mobile-phone-data-showmore-americans-are-leaving-their-homes-despite-orders

Muller, S. A., Balmer, M., Neumann, A., & Nagel, K. (2020). Mobility traces and spreading of COVID-19. medRxiv preprint. https://doi.org/10.1101/2020.03.27.20045302

Pan, Y., Darzi, A., Kabiri, A., Zhao, G., Luo, W., Xiong, C., Zhang, L. (2020). Quantifying human mobility behaviour changes during the COVID-19 outbreak in the United States. *Scientific Reports*, *10*(1):1–9. https://doi.org/10.1038/s41598-020-77751-2

Parshani, R., Carmi, S., & Havlin, S. (2010). Epidemic threshold for the susceptible-infectious-susceptible model on random networks. *Physical Review Letters*, *104*(25), 258701. https://doi.org/10.1103/PhysRevLett.104.258701

Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. E., Davies, N., Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Jit, M., & Klepac, P. (2020). The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study. *Lancet Public Health*, *5*(5), E261–E270. https://doi.org/10.1016/S2468-2667(20)30073-6

Rechtin, M., Feldman, V., Klare, S., Riddle, N., & Sharma, R. (2020). Modeling and Simulation of COVID-19 Pandemic for Cincinnati Tri-State Area. arXiv preprint: 200606021. https://doi.org/10.48550/arXiv.2006.06021

Ruktanonchai, N. W., Ruktanonchai, C. W., Floyd, J. R. & Tatem, A. J. (2018). Using Google Location History data to quantify fine-scale human mobility. *International Journal of Health Geographics*, *17*, 28. https://doi.org/10.1186/s12942-018-0150-z

Schlosser, F., Maier, B. F., Jack, O., Hinrichs, D., Zachariae, A., & Brockmann, D. (2020). COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proceedings of the National Academy of Sciences* (USA), *117*(52), 32883–32890. https://doi.org/10.1073/pnas.2012326117

Soures, N., Chambers, D., Carmichael, Z., Daram, A., Shah, D. P., Clark, K., Potter, L., & Kudithipudi, D. (2020). SIRNet: Understanding social distancing measures with hybrid neural network model for COVID-19 infectious spread. arXiv preprint, 200410376. https://doi.org/10.48550/arXiv.2004.10376

Venkatramanan, S., Sadilek, A., Fadikar, A., Barrett, C. L., Biggerstaff, M., Chen, J., Dotiwalla, X., Eastham, P., Gipson, B., Higdon, D., Kucuktunc, O., Lieber, A., Lewis, B. L., Reynolds, Z., Vullikanti, A. K., Wang, L., & Marathe, M. (2021). Forecasting influenza activity using machine-learned mobility map. *Nature Communications*, *12*, 726. https://doi.org/10.1038/s41467-021-21018-5

Wang, Y., Wang, Y., Chen, Y., & Qin, Q. (2020). Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. *Journal of Medical Virology*, *92*(6), 568–576. https://doi.org/10.1002/jmv.25748

Weill, J. A., Stigler, M., Deschenes, O., & Springborn, M. R. (2020). Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. *Proceedings of the National Academy of Sciences*, *117*(33), 19658–19660. https://doi.org/10.1073/pnas.2009412117

Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan,

China: a modelling study. *The Lancet*, *395*(10225), 689–697. https://doi.org/10.1016/S0140-6736(20)30260-9

Yi, H., Ng, S. T., Farwin, A., Pei Ting Low, A., Chang, C. M., & Lim, J. (2021). Health equity considerations in COVID-19: geospatial network analysis of the COVID-19 outbreak in the migrant population in Singapore. *Journal of Travel Medicine*, *28*(2), taaa159. https://doi.org/10.1093/jtm/taaa159

Utilizing Mobility Tracking to Identify Hotspots for Contagious Disease Spread

A Case Study of UNITEN Students Using Google Map Data

Yaw Mei Wyin
Institute of Sustainable Energy, Universiti Tenaga Nasional, 43000 Kajang, Malaysia

Prajindra Sankar Krishnan
Institute of Sustainable Energy, Universiti Tenaga Nasional, 43000 Kajang, Malaysia

Chen Chai Phing
Institute of Sustainable Energy, Universiti Tenaga Nasional, 43000 Kajang, Malaysia

Tiong Sieh Kiong
Institute of Sustainable Energy, Universiti Tenaga Nasional, 43000 Kajang, Malaysia

Abstract: A significant global health problem nowadays is the incidence of serious infectious illnesses. An extraordinary humanitarian crisis has been brought on by the current COVID-19 pandemic, which has spread around the world. The spread of new viruses has put established healthcare institutions under tremendous strain and created a number of pressing problems. It is important to predict the future movement and pattern of the illness in order to decrease infectious instances and maximize recovered cases. This research paper aims to utilize mobility tracking as a means to identify hotspots for contagious disease spread. The study focuses on collecting and analyzing mobility data from UNITEN students using Google Map data over a period of two weeks. The paper describes the data collection process, data pre-processing steps, and the application of the HDBSCAN algorithm for hotspot clustering. The results demonstrate the effectiveness of HDBSCAN in identifying hotspots based on the mobility data. The findings highlight the potential of mobility tracking for disease surveillance and provide insights for public health interventions and preventive measures.

Keywords: Hotspot, HDBSCAN, infection disease, COVID-19.

http://doi.org/10.18080/jtde.v11n3.775

Introduction

In particular for highly contagious illnesses like COVID-19, human mobility is a crucial consideration in the understanding of an infectious disease's transmission (Gatto *et al.*, 2020; Kiang *et al.*, 2021; Venkatramanan *et al.*, 2021; Wang *et al.*, 2020; Wu, Leung & Leung, 2020). Early patient identification and the implementation of prophylactic measures are the most effective ways to reduce the disease's rate of transmission during an outbreak. It is now feasible to follow people's movements over extended periods of time and across a large region because of the proliferation of gadgets with accurate localization capabilities and robust wireless networks.

Humans can contract infectious illnesses either directly or indirectly through time and distance. For instance, COVID-19 is regarded as illnesses that may be conveyed to vulnerable individuals directly when airborne droplets from an infected person's coughing or sneezing travel a short distance (Setti *et al.*, 2020; Devi *et al.*, 2022; Zarei *et al.*, 2021). As a result, effective transmissions between people require a close enough physical distance.

When contemplating the treatment of infectious illnesses, spatial dimensions are essential. An outbreak of an epidemic is defined from an epidemiological perspective as any temporally anomalous rise in the number of case-patients in a specific location (Porta, 2014; CDC, 2018). Effective disease management and prevention techniques depend on pinpointing areas where infectious diseases are most likely to spread. Contagious diseases pose significant threats to public health, requiring proactive strategies for containment and prevention. One crucial aspect of disease control is the identification of hotspots, geographic areas with a high concentration of disease transmission. Geographical regions or localities known as hotspots are those where an infectious illness is spreading particularly rapidly or heavily.

Researchers now have access to a variety of databases that monitor people's movements, enabling more thorough analysis and epidemic spread models at a finer geographical and temporal resolution. GPS and Google Location History are examples of conventional sources of mobility data where several sorts of motions may be collected (Ruktanonchai *et al.*, 2018).

This research paper explores the use of mobility tracking as a tool for hotspot identification. Mobility data, collected from UNITEN (Universiti Tenaga Nasional) students through Google Maps, provides valuable insights into individuals' movement patterns and their potential impact on disease spread. By analyzing this data, we aim to identify hotspots and understand the role of mobility in contagious disease transmission.

The software platform Google History Location Extractor is suggested in this article. The platform simulates the epidemic spreading using real-world UNITEN student trajectory

http://doi.org/10.18080/jtde.v11n3.775

datasets, estimates the contact behaviours using GPS trajectories, and simulates their mobility in response to public policies, to be able to assess how well they work to halt the epidemic's spread. The platform, which combines the probabilistic model of individual-level infectious disease transmission with the investigation of individual infection risks, also facilitates data research and mining on the risks of infection spreading, such as the identification of possible secondary contacts.

The next sections cover relevant research that has already been done and go into great depth on our movement tracking system as a tool for locating hotspots. The other sections of the study detail our findings and include the modelling of mobility patterns, the AI algorithm implementation procedure, outcomes, and results. We summarize the main points and emphasize the importance of our platform in our conclusion.

Related Works

Understanding disease transmission in a population may be done by building a network to reflect person-to-person interactions. Various techniques, including surveys, statistical methods, census data, and movement data (Chang *et al.*, 2021; Kumar, Oke & Nahmias-Biran, 2021; Maheshwari & Albert, 2020; Muller *et al.*, 2020; Rechtin *et al.*, 2020; Schlosser *et al.*, 2020; Soures *et al.*, 2020; Yi *et al.*, 2021), can be used to create these networks. People's positions and the amount of time they spend in certain areas are detailed in a vast quantity of data as a result of the widespread usage of location tracking applications on mobile phones. In response to COVID-19, new systems like Google Community Mobility Reports (Google, 2020) provide devices and aggregated movement data. Data is compiled and made anonymous to safeguard people's privacy.

In order to emphasize geographical patterns of illnesses and determine whether there is a pattern of disease incidence in a certain region, spatial analyses and statistics, such as spatial autocorrelation analysis, cluster analysis, and temporal analysis, are frequently utilised (Brownstein *et al.*, 2002; Tsai *et al.*, 2009; Pace, Barry & Sirmans, 1998; Ping *et al.*, 2004). Growing interest in the identification of disease clusters or "hotspots" for public health surveillance, particularly to better understand the rising prevalence of dengue fever, has been sparked by recent developments in spatial statistics in geographical information systems (Yeshiwondim *et al.*, 2009). The dynamics of disease propagation can be understood through the analysis of spatiotemporal patterns. So as to target disease monitoring and management in places and periods with increased risk for illness, it is helpful to detect geographical, temporal, and space-time clustering (Si *et al.*, 2008).

Previous research has investigated various methods for hotspot identification and their applications in disease spread analysis. Studies have utilized data from different sources, such

http://doi.org/10.18080/jtde.v11n3.775

as mobile phones, GPS tracking, and social media platforms, to understand human mobility patterns and their association with disease transmission. However, there are still limitations and gaps in the current research, including the need for more accurate and granular data, effective clustering algorithms, and comprehensive analysis techniques.

Methodology

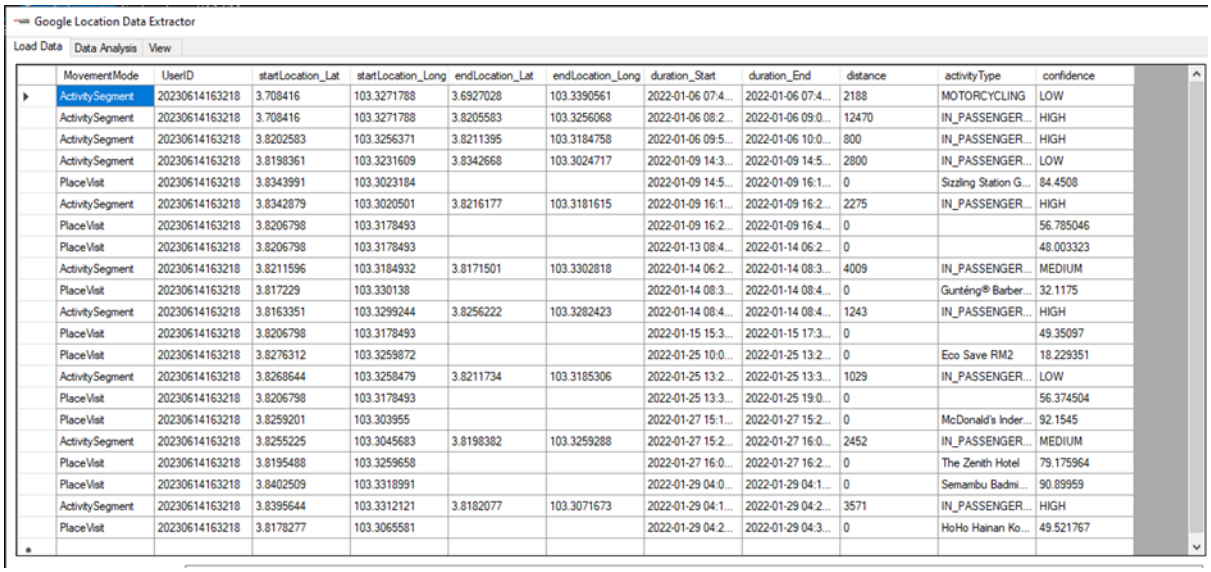
Epidemiologists can foresee disease outbreaks by comprehending the patterns of human movement, since infectious diseases are disseminated through direct human contact. Understanding how people move through time and space might help us respond to and recover from such crisis situations. This can be done with the use of mobile phone location data. One of the crucial components that defines these mobilities is location history. The cell site location method, GPS, and Wi-Fi positioning are some of the methods (Khalel, 2010) that may be used to locate someone.

All three of these technologies are used by Google Maps to locate the user's position with great accuracy. Users may export this data via the Google Takeout service, which provides the data in JSON (JavaScript Object Notation) format. The produced JSON file contains a wide variety of different kinds of metadata. The direction, activity type, latitudeE7 (latitude), longitudeE7 (longitude), precision (accuracy measure), timestampMS (timestamp in milliseconds), and altitude are some of the most crucial data items that are highlighted.

Google Maps has several capabilities in addition to place searches and route finding. Google Timeline, a function of Google Maps, constantly tracks user behaviour, including the places visited and the method of transportation used to get there, using GPS position data. This raw data reveals the user's location at a certain time and date.

The methodology employed in this study involved the utilization of mobility tracking and Google Maps data to identify hotspots for contagious disease spread among UNITEN students. At the beginning of the study, mobility data from 50 UNITEN students were collected using the Google Maps application in the mobile phone over two-week periods, as illustrated in Figure 1.

http://doi.org/10.18080/jtde.v11n3.775


Google Location Data Extractor

Load Data | Data Analysis | View

MovementMode	UserID	startLocation_Lat	startLocation_Long	endLocation_Lat	endLocation_Long	duration_Start	duration_End	distance	activityType	confidence
ActivitySegment	20230614163218	3.708416	103.3271788	3.6927028	103.3390561	2022-01-06 07:4...	2022-01-06 07:4...	2188	MOTORCYCLING	LOW
ActivitySegment	20230614163218	3.708416	103.3271788	3.8205583	103.3256068	2022-01-06 08:2...	2022-01-06 09:0...	12470	IN_PASSENGER...	HIGH
ActivitySegment	20230614163218	3.8202583	103.3256371	3.8211395	103.3184758	2022-01-06 09:5...	2022-01-06 10:0...	800	IN_PASSENGER...	HIGH
ActivitySegment	20230614163218	3.8198361	103.3231609	3.8342668	103.3024717	2022-01-09 14:3...	2022-01-09 14:5...	2800	IN_PASSENGER...	LOW
PlaceVisit	20230614163218	3.8343991	103.3023184			2022-01-09 14:5...	2022-01-09 16:1...	0	Sizzling Station G...	84.4508
ActivitySegment	20230614163218	3.8342879	103.3020501	3.8216177	103.3181615	2022-01-09 16:1...	2022-01-09 16:2...	2275	IN_PASSENGER...	HIGH
PlaceVisit	20230614163218	3.8206798	103.3178493			2022-01-09 16:2...	2022-01-09 16:4...	0		56.785046
PlaceVisit	20230614163218	3.8206798	103.3178493			2022-01-13 08:4...	2022-01-14 06:2...	0		48.003323
ActivitySegment	20230614163218	3.8211596	103.3184932	3.8171501	103.3302818	2022-01-14 06:2...	2022-01-14 08:3...	4009	IN_PASSENGER...	MEDIUM
PlaceVisit	20230614163218	3.817229	103.330138			2022-01-14 08:3...	2022-01-14 08:4...	0	Gunténg® Barber...	32.1175
ActivitySegment	20230614163218	3.8163351	103.3299244	3.8256222	103.3282423	2022-01-14 08:4...	2022-01-14 08:4...	1243	IN_PASSENGER...	HIGH
PlaceVisit	20230614163218	3.8206798	103.3178493			2022-01-15 15:3...	2022-01-15 17:3...	0		49.35097
PlaceVisit	20230614163218	3.8276312	103.3259872			2022-01-25 10:0...	2022-01-25 13:2...	0	Eco Save RM2	18.229351
ActivitySegment	20230614163218	3.8268644	103.3258479	3.8211734	103.3185306	2022-01-25 13:2...	2022-01-25 13:3...	1029	IN_PASSENGER...	LOW
PlaceVisit	20230614163218	3.8206798	103.3178493			2022-01-25 13:3...	2022-01-25 19:0...	0		56.374504
PlaceVisit	20230614163218	3.8259201	103.303955			2022-01-27 15:1...	2022-01-27 15:2...	0	McDonald's Inder...	92.1545
ActivitySegment	20230614163218	3.8255225	103.3045683	3.8198382	103.3259288	2022-01-27 15:2...	2022-01-27 16:0...	2452	IN_PASSENGER...	MEDIUM
PlaceVisit	20230614163218	3.8195488	103.3259658			2022-01-27 16:0...	2022-01-27 16:2...	0	The Zenith Hotel	79.175964
PlaceVisit	20230614163218	3.8402509	103.3318991			2022-01-29 04:0...	2022-01-29 04:1...	0	Semambu Badmi...	90.89959
ActivitySegment	20230614163218	3.8395644	103.3312121	3.8182077	103.3071673	2022-01-29 04:1...	2022-01-29 04:2...	3571	IN_PASSENGER...	HIGH
PlaceVisit	20230614163218	3.8178277	103.3065581			2022-01-29 04:2...	2022-01-29 04:3...	0	HoHo Hainan Ko...	49.521767

Figure 1. Extracted Data Display

In order to identify potential COVID-19 dissemination, the location records were examined using the Google History Location Extractor and Indicator. The study involves plotting the locations of the students on a map with complete information about the date, time, region, and distance between persons with the aim of looking for prospective COVID-19 distribution scenarios. Figure 2 displays the Venn chart that was used to analyse the location information for two students.

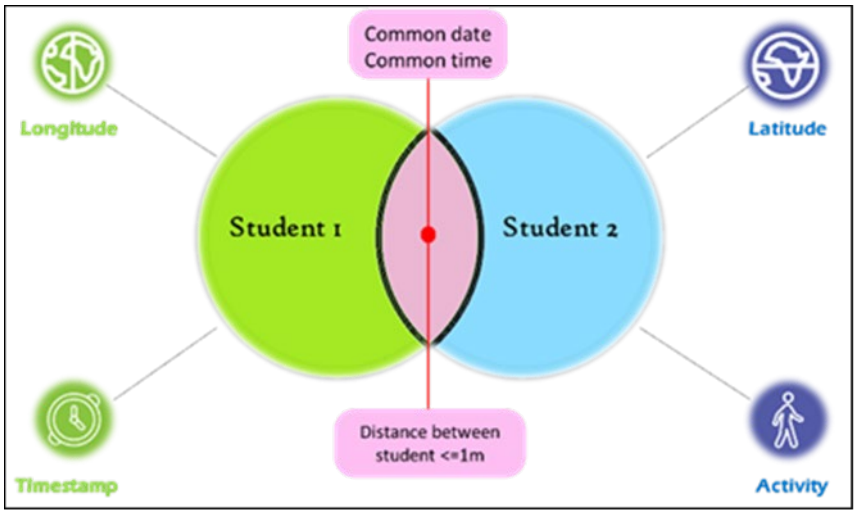


Figure 2. Venn diagram of location analysis for two students

This study looked at the JSON file containing the location history information of 50 students using tools to extract and visualise the downloaded data. The Google History Location Extractor program, which is depicted in Figure 3, performs a variety of tasks, including data extraction, graphical user interface presentation, data saving in CSV or KML format, and mobility visualisation on Google Maps. The data may also be exported to a Microsoft SQL Server database format for additional in-depth analysis. The full software program's flowchart,

http://doi.org/10.18080/jtde.v11n3.775

which includes data collecting, import, filtering, distance calculation, and visualization, is shown in Figure 4. The overall algorithm flow visualization is shown in Figure 5.

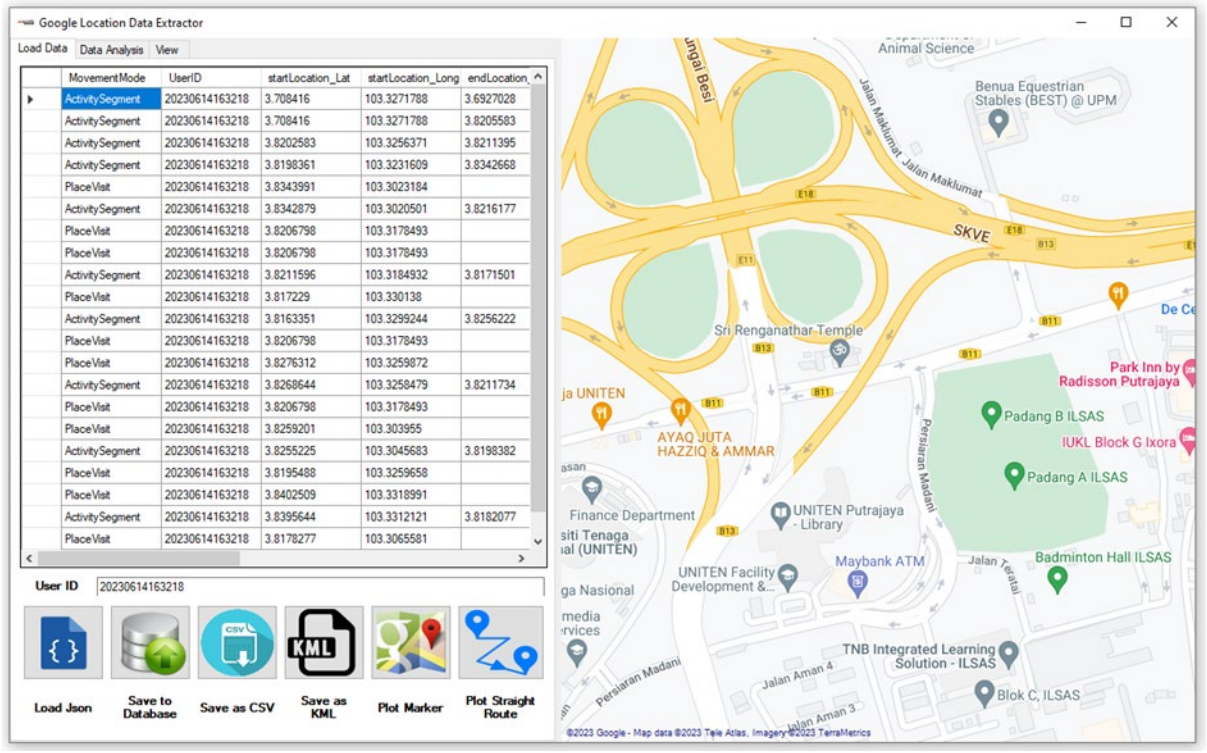


Figure 3. GUI of Google Historical Location Extractor software

Since it was simpler to manage using Google Timeline, the location history data was collected in JSON rather than KML (Sardianos, Varlamis & Bouras, 2018). Through the use of the import method in VB.net, the JSON file is transformed into a table format. The needed variable may be chosen, variable names can be changed, and the data type for variables can be specified while the data is still being imported. Latitude, longitude, accuracy, and timestamp are filtered from the raw data for this project and kept separately in tabular format. The mobility data was loaded from the 'MobilityData_49.csv' file into a Pandas DataFrame. To focus on a specific area of interest, the data was spatially filtered based on bounding box coordinates. The longitude and latitude ranges were defined to select the relevant data points.

Following the filtering process, the data underwent pre-processing steps. The 'duration_Start' and 'duration_End' columns were converted to datetime format using the pd.to_datetime() function, enabling the calculation of stay durations. The stay duration for each data point was determined by computing the difference between the 'duration_End' and 'duration_Start' timestamps and converting the result to minutes.

http://doi.org/10.18080/jtde.v11n3.775

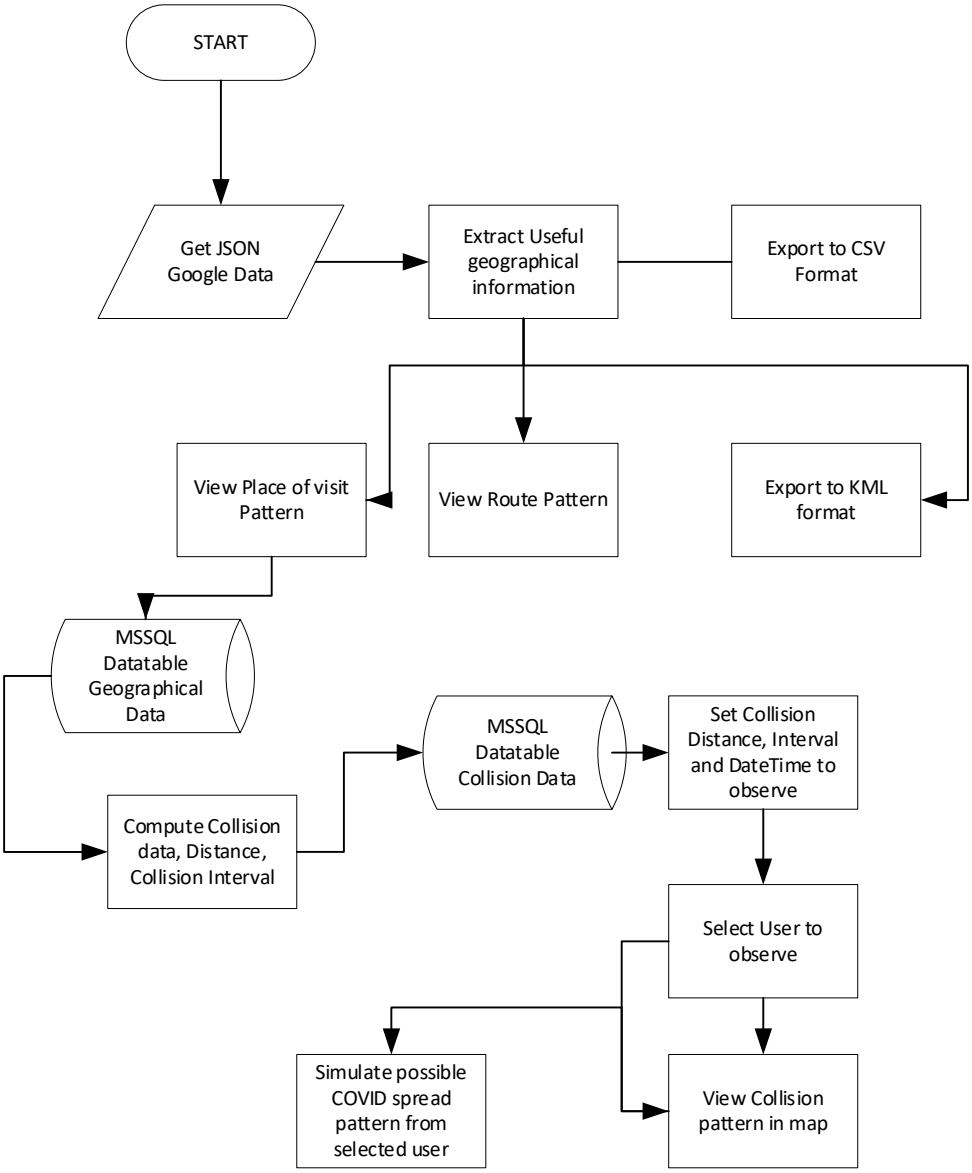


Figure 4. Entire system architecture

This study employs the HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) algorithm to identify hotspots for contagious disease spread using mobility tracking data. HDBSCAN is an extension of the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm, which is known for its ability to discover clusters of arbitrary shapes and handle noise effectively.

The HDBSCAN algorithm operates by defining clusters based on the density of data points. It starts by calculating the mutual reachability distance between each pair of points, which combines both the distance between points and their local densities. This distance measure allows the algorithm to identify regions of higher density, which are likely to represent meaningful clusters.

In the context of this research, the data points are denoted as $X = x_1, x_2, \ldots, x_n$, where x_i represents the coordinates of the i-th data point.

http://doi.org/10.18080/jtde.v11n3.775

- Density Calculation:
  The density of a data point x_i is calculated as follows:
  $D(x_i) = \Sigma_{j=1}^{n} \delta\big(dist(x_i, x_j), \varepsilon\big)$,
  where δ is the Kronecker delta function and $dist(x_i, x_j)$ represents the distance between data points x_i and x_j. ε is the specified radius within which neighboring points are considered.
- Mutual Reachability Distance:
  The mutual reachability distance between points x_i and x_j is computed as follows:
  $MRD(x_i, x_j) = max\left(RD(x_i, x_j), RD(x_j, x_i)\right)$,
  where $\mathrm{RD}(\mathrm{x_i}, \mathrm{x_j})$ represents the reachability distance from x_i to x_j, and $RD(x_j, x_i)$ represents the reachability distance from x_j to x_i.
- Reachability Distance:
  The reachability distance between points x_i and x_j is given by:
  $\mathrm{RD}(\mathrm{x_i}, \mathrm{x_j}) = \max\left(\mathrm{D}(\mathrm{x_j}), \mathrm{dist}(\mathrm{x_i}, \mathrm{x_j})\right)$,
  where $D(x_j)$ represents the density of data point x_j, and $dist(x_i, x_j)$ represents the distance between x_i and x_j.
- Constructing the Condensed Tree:
  The condensed tree is constructed by merging clusters based on their mutual reachability distances. Various techniques, such as minimum spanning tree or single linkage clustering, can be used to create the tree structure.
- Cluster Extraction:
  The final set of clusters is extracted from the hierarchy determined by the condensed tree. The appropriate level of density threshold is determined to define clusters, considering the minimum cluster size parameter.

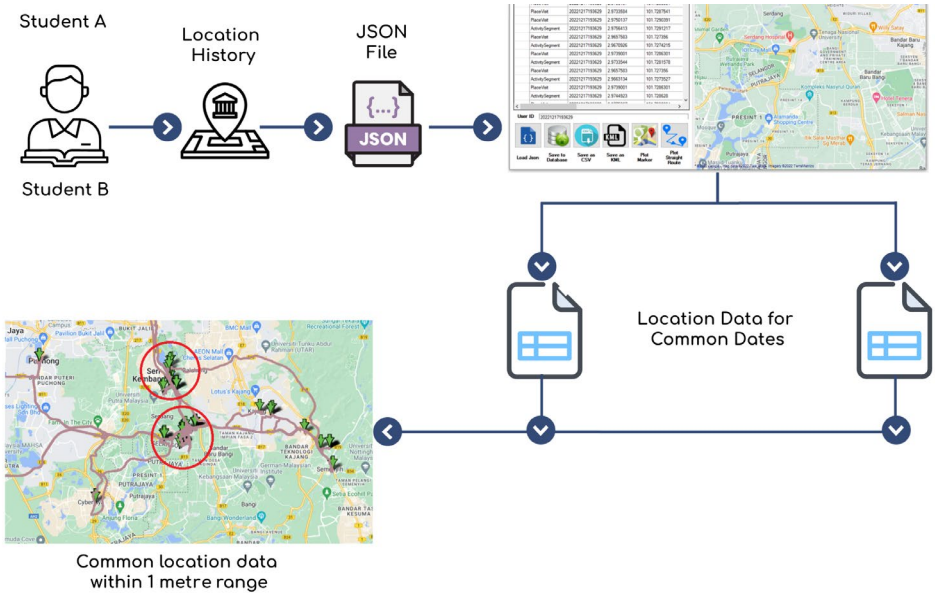


Figure 5. General flow of algorithm

In this research, mobility tracking data from Google Maps is first collected and filtered based on predefined bounding box coordinates to focus on the specific area of interest, the UNITEN campus as presented in Figure 6. The data is then pre-processed by converting the timestamps

http://doi.org/10.18080/jtde.v11n3.775

to datetime format and calculating the duration of stay at each location. These steps enable the data to be effectively utilized by the HDBSCAN algorithm.

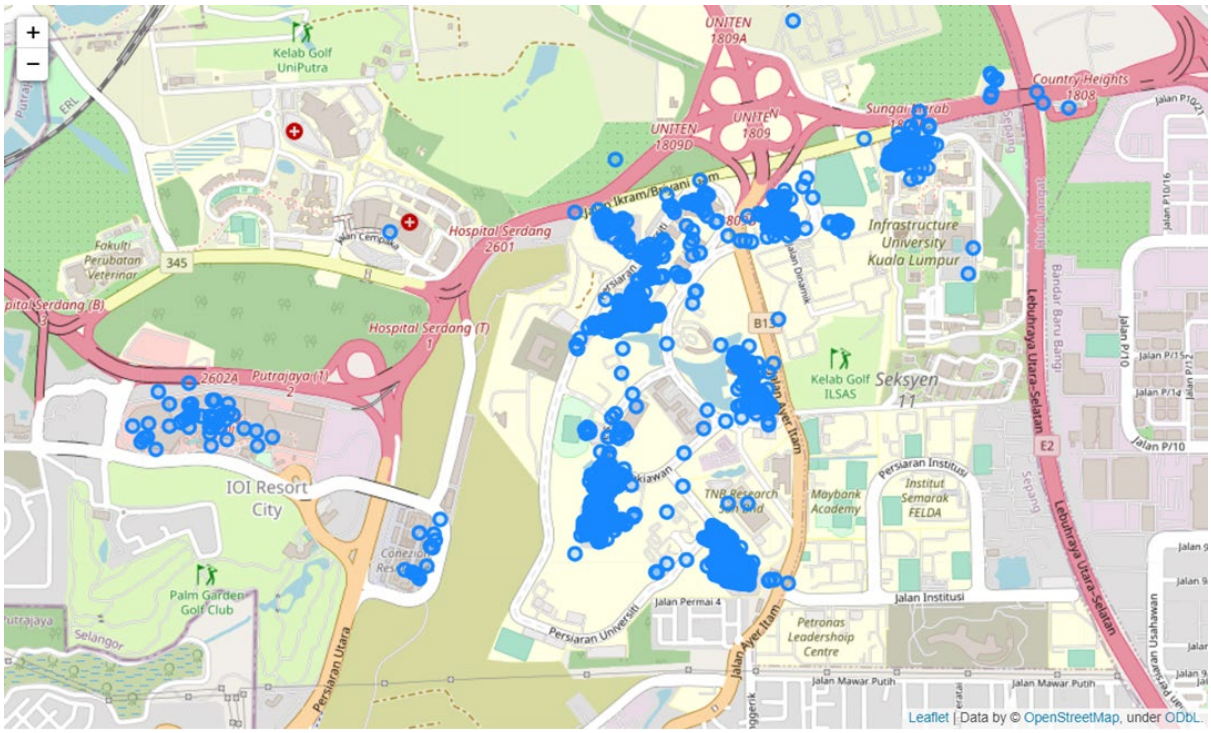


Figure 6. Students' GPS for area of interest after applying predefined bounding box

Next, the HDBSCAN algorithm is applied to the pre-processed mobility data, considering the coordinates of each data point. The algorithm takes into account the minimum cluster size parameter, which determines the minimum number of points required to form a cluster. This parameter ensures that only significant clusters, composed of a sufficient number of points, are identified as hotspots. The algorithm assigns each data point to a cluster or labels it as noise if it does not meet the density requirements.

HDBSCAN algorithm was used for hotspot identification because it is well-suited for clustering spatial data and requires the specification of parameters such as the minimum cluster size. The coordinates, represented by the 'startLocation_Lat' and 'startLocation_Long' columns, were used as input for the clustering algorithm. The resulting cluster labels were added as a new column ('Cluster') to the filtered data, facilitating the identification of hotspots.

After the clustering process, the identified clusters are analysed to determine the total number of hotspots within the UNITEN campus. The distribution of these hotspots across the study area is visualized using maps and charts. Additionally, the average time spent in each hotspot by all users is calculated and discussed, providing insights into the temporal dynamics of hotspot activity.

http://doi.org/10.18080/jtde.v11n3.775

Results and Discussion

The analysis of the mobility data and hotspot identification yielded several key findings. A total of 18 hotspots were identified within the study area, signifying areas with a high concentration of potential disease transmission, as shown in Figure 7. These hotspots provide valuable insights into the locations that require particular attention in terms of disease surveillance and control efforts.

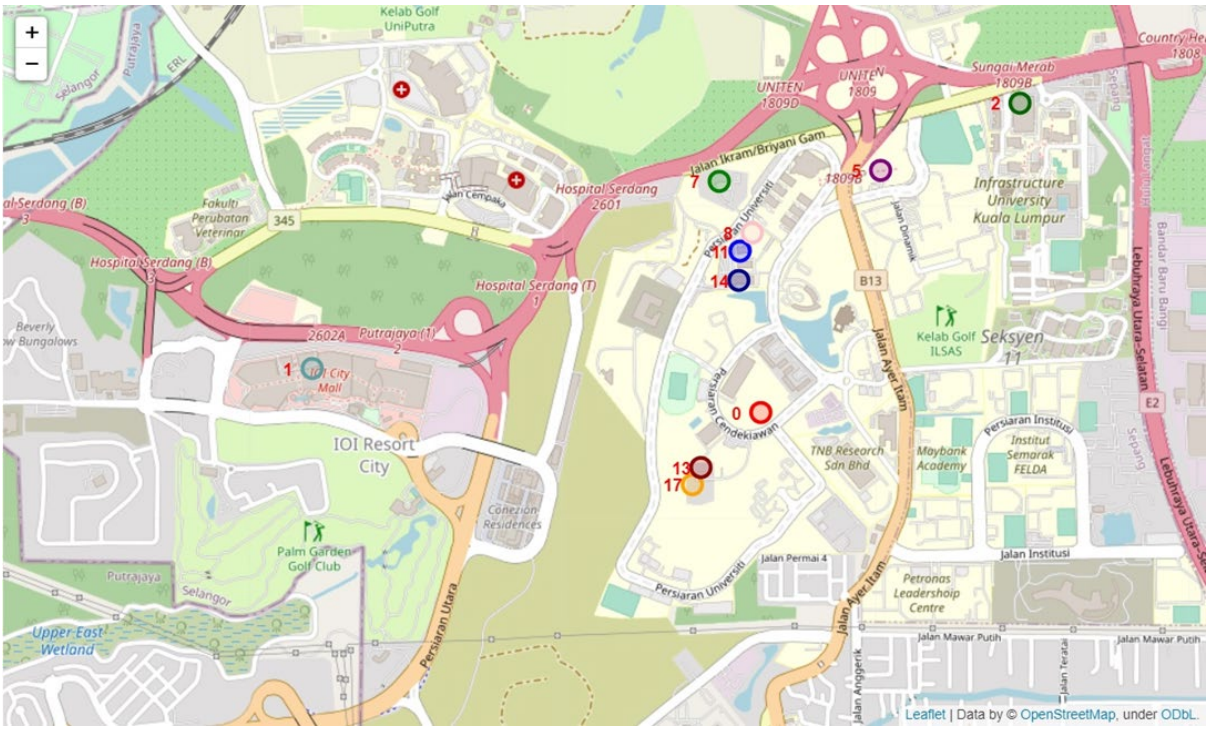


Figure 7. Hotspot identification

The average time spent in each hotspot by all users was also computed in Table 1. Figures 8 and 9 illustrate the sample of possible student collisions within the same day and same hour. This metric offers an understanding of the relative significance of specific locations in terms of duration. By identifying hotspots with longer average stay durations, public health authorities can prioritize these areas for targeted interventions and preventive measures.

Table 1. Average time spent (minutes) in each hotspot by all users for each cluster

Cluster	Average time spent in each hotspot by all users (minutes)
0	157.25
1	30.02
2	22.46
3	39.79
4	53.28
5	77.51
6	388.41
7	33.70
8	292.77
9	209.63

http://doi.org/10.18080/jtde.v11n3.775

Cluster	Average time spent in each hotspot by all users (minutes)
10	106.73
11	68.97
12	352.74
13	115.84
14	121.03
15	546.49
16	497.05
17	575.08

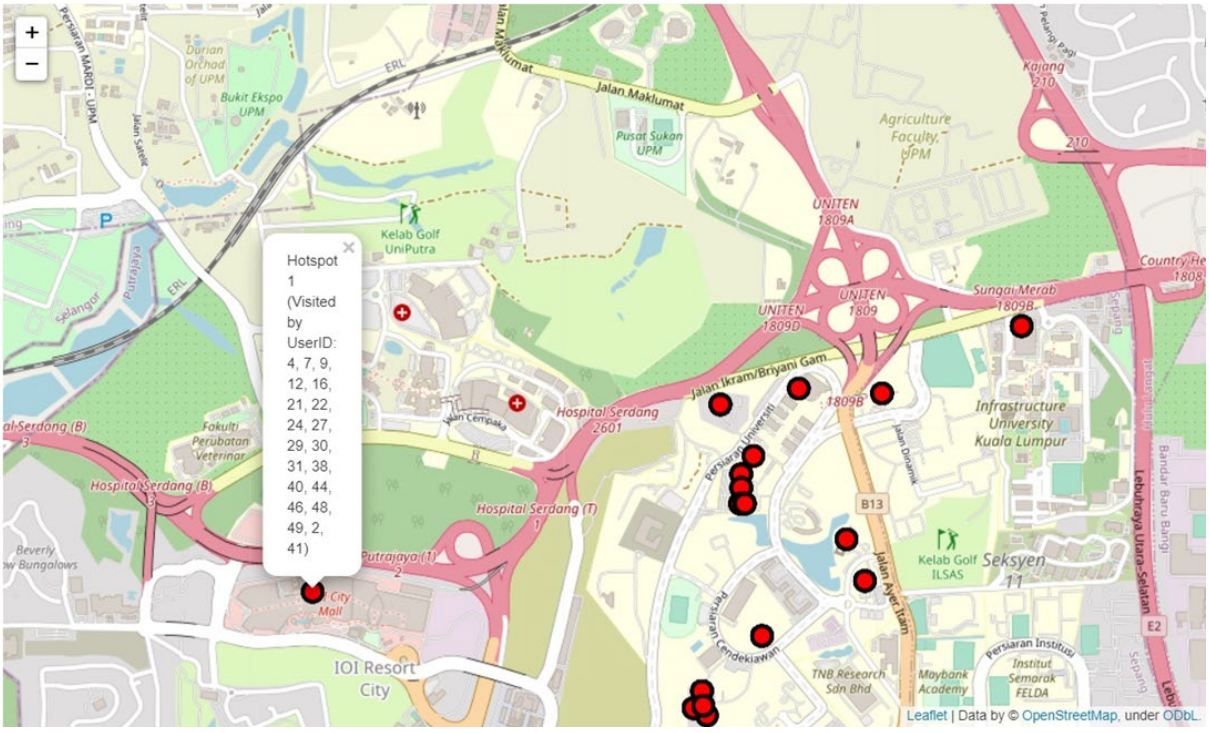


Figure 8. Sample of possible student collisions within the same day and same hour for hotspot 1

To visually represent the identified hotspots, a map was generated using the Folium library. Each hotspot was marked with a distinct colour and labelled with the corresponding hotspot number. This visualization allows for a clear understanding of the spatial distribution of hotspots and their relative proximity to each other.

Furthermore, the analysis focused on a specific user, UserID 1, and filtered the data for their visits from Monday to Sunday. By examining the unique hotspots visited by UserID 1, it was possible to identify the specific locations that this individual frequented. These visited hotspots were marked on the map using black circles filled in red, providing insights into the movement patterns and potential exposure risks for UserID 1, as illustrated in Figure 10. The same analysis was done for other UserIDs. Figure 11 is a sample of hotspots visited by UserID 2.

http://doi.org/10.18080/jtde.v11n3.775

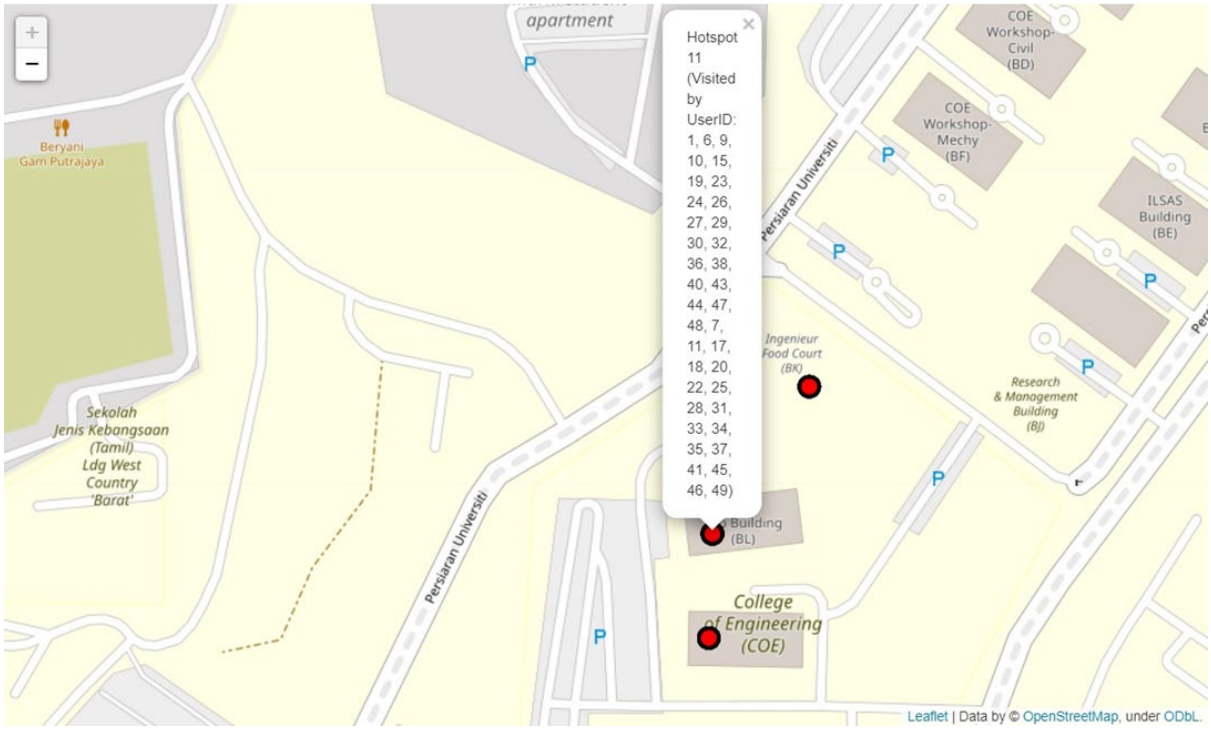


Figure 9. Sample of possible student collision within same day and same hour for hotspot 11

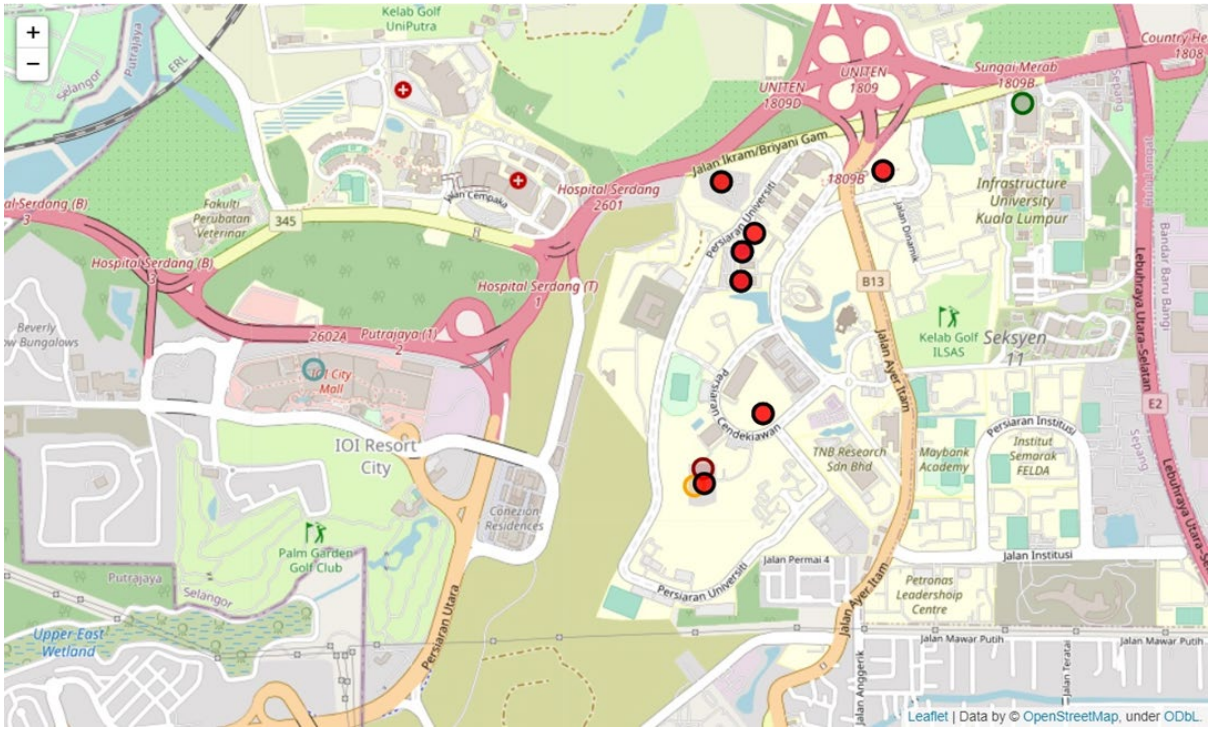


Figure 10. Hotspot visited by UserID 1

The combination of hotspot identification, average stay durations, and user-specific analysis offers a comprehensive understanding of the contagious disease spread patterns among UNITEN students. This information can inform public health strategies, interventions, and preventive measures tailored to mitigate the risk of disease transmission within the university community.

http://doi.org/10.18080/jtde.v11n3.775

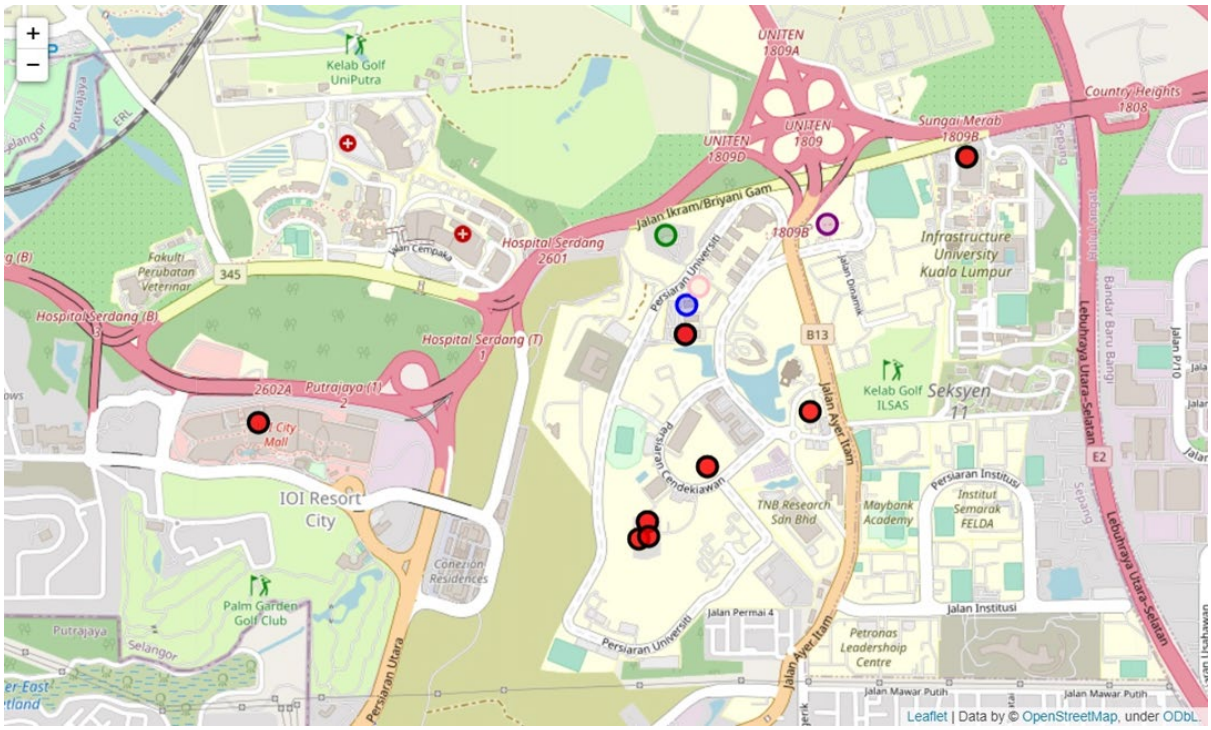


Figure 11. Hotspot visited by UserID 2

By applying the HDBSCAN algorithm to the mobility tracking data, this study leverages its ability to capture clusters of varying densities and irregular shapes. This allows for the identification of hotspots, representing areas of concentrated activity, which are crucial for understanding the potential spread of contagious diseases. Coordinates of HDBSCAN clustering with and without outliers are depicted in Figures 12 and 13.

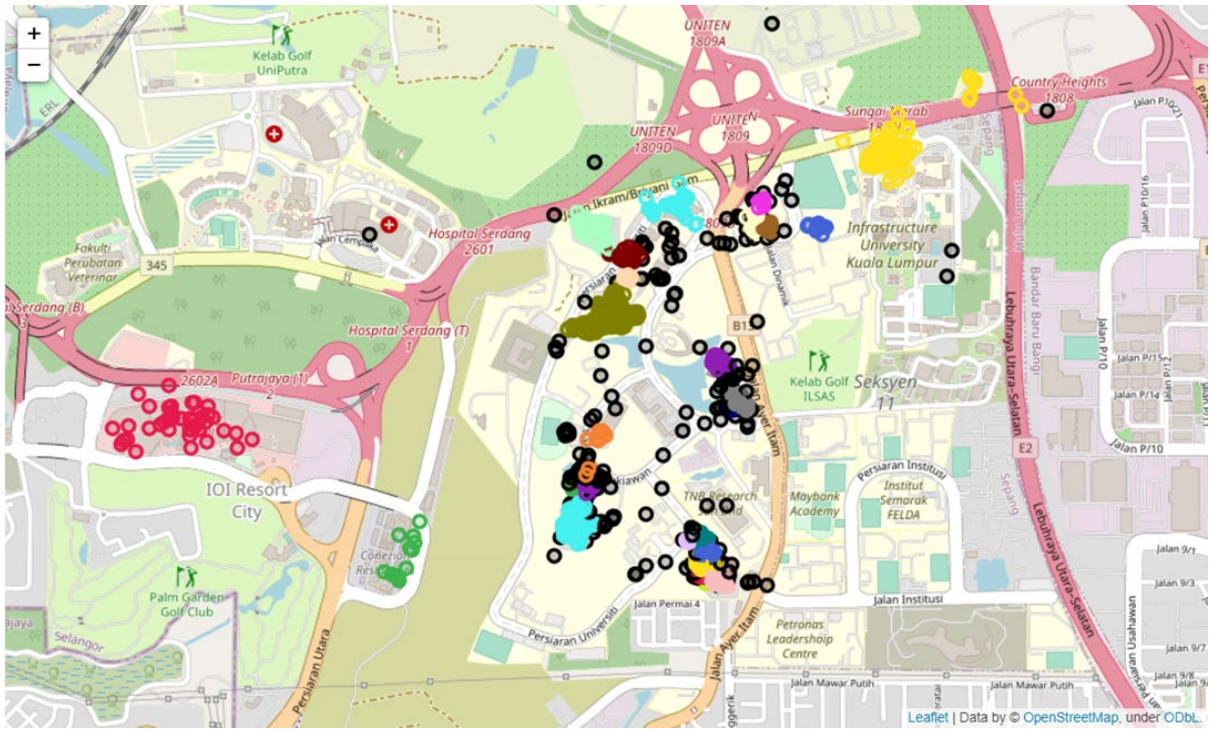


Figure 12. HDBSCAN clustering with outliers

http://doi.org/10.18080/jtde.v11n3.775

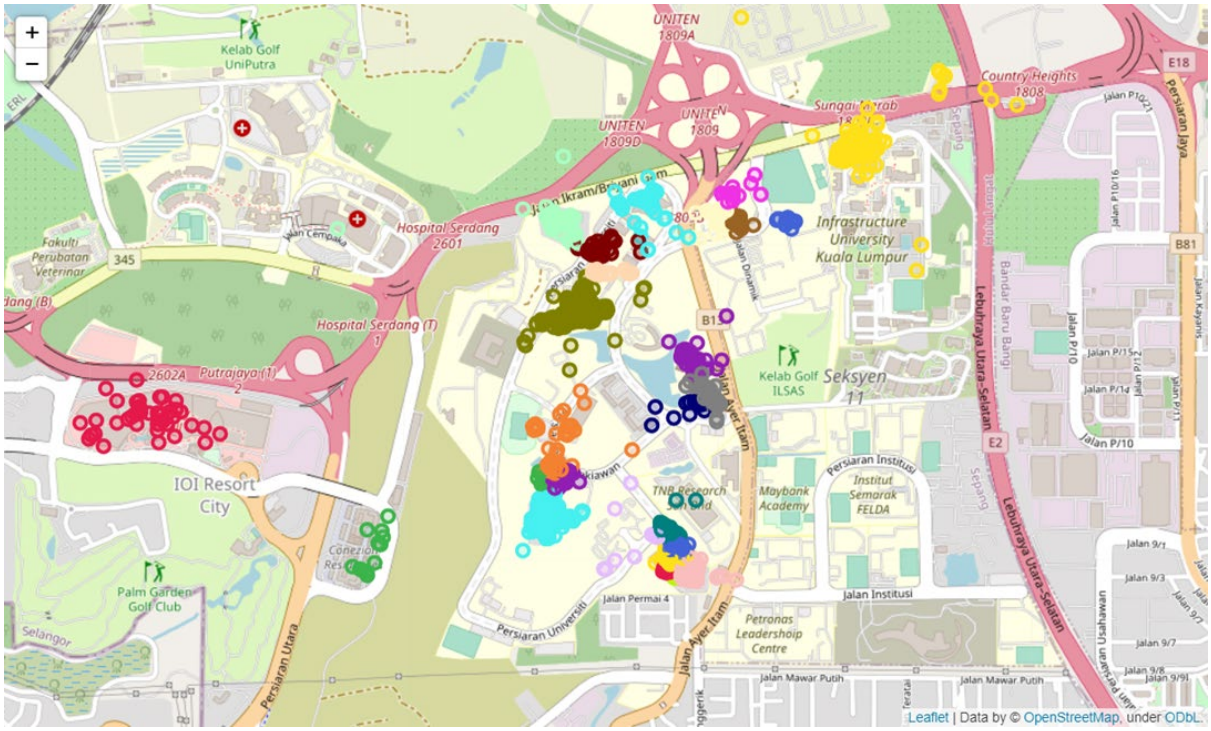


Figure 13. HDBSCAN clustering after removing outliers

In summary, the HDBSCAN algorithm is utilized in this research to effectively identify hotspots for contagious disease spread based on mobility tracking data. Its capability to handle density-based clustering and noise detection contributes to the accurate identification and characterization of hotspots within the UNITEN campus.

The findings of this research contribute to the understanding of contagious disease spread in relation to mobility patterns. The interpretation of the results in the context of disease transmission highlights the importance of hotspot identification for effective disease surveillance. The analysis of the effectiveness of mobility tracking and hotspot identification provides insights into the potential of these techniques for proactive public health interventions. The implications of the results include the development of targeted prevention strategies, resource allocation, and the implementation of timely interventions in hotspot areas.

Conclusion

This research has certain limitations that should be acknowledged. First, the data collection was limited to UNITEN students using Google Maps, which may not represent the entire population. Second, the algorithm used for hotspot clustering, HDBSCAN, has specific parameters that might affect the results. Future research should consider incorporating additional data sources to improve the accuracy and granularity of hotspot identification. Exploring different clustering algorithms and validation techniques could also enhance the robustness of the hotspot identification process.

http://doi.org/10.18080/jtde.v11n3.775

In conclusion, this research demonstrates the potential of mobility tracking to identify hotspots for contagious disease spread. The analysis of mobility data collected from UNITEN students using Google Maps provides valuable insights into hotspot distribution and the role of mobility in disease transmission. The findings contribute to the field of contagious disease spread analysis and hotspot identification, highlighting the significance of proactive surveillance and targeted interventions. This research has the potential to impact public health strategies by enabling more effective measures to prevent and control the spread of contagious diseases.

Acknowledgements

This work is supported by IDRC Grant No. 109586 – 001 for Artificial Intelligence Framework for Threat Assessment and Containment for COVID-19 and Future Epidemics while Mitigating the Socioeconomic Impact to Women, Children and Underprivileged Groups.

A version of this paper was presented at the third International Conference on Computer, Information Technology and Intelligent Computing, CITIC 2023, held in Malaysia on 26–28 July 2023.

References

Brownstein, J. S., Rosen, H., Purdy, D., Miller, J. R., Merlino, M., Mostashari, F., & Fish, D. (2002). Spatial analysis of West Nile virus: rapid risk assessment of an introduced vector-borne zoonosis. *Vector-Borne and Zoonotic Diseases*, *2*(3), 157–164. https://doi.org/10.1089/15303660260613729

CDC [US Centers for Disease Control and Prevention]. (2018) Managing HIV and Hepatitis C Outbreaks among People Who Inject Drugs - A Guide for State and Local Health Departments. Available from https://www.cdc.gov/hiv/pdf/programresources/guidance/cluster-outbreak/cdc-hiv-hcv-pwid-guide.pdf

Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., & Leskovec, J. (2021). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, *589*, 82–87. https://doi.org/10.1038/s41586-020-2923-3

Devi, S., Nagaraja, K. V., Thanuja, L., Reddy, M. V., & Ramakrishna, S. (2022). Finite element analysis over transmission region of coronavirus in CFD analysis for the respiratory cough droplets, *Ain Shams Engineering Journal*, *13*(6), 101766. https://doi.org/10.1016/j.asej.2022.101766

Gatto, M., Bertuzzo, E., Mari, L., Miccoli, S., Carraro, L., Casagrandi, R., & Rinaldo, A. (2020). Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proceedings of the National Academy of Sciences* (USA), *117*(19), 10484–10491. https://doi.org/10.1073/pnas.2004978117

Google. (2020). COVID-19 Community Mobility Reports, 2020. Available from https://www.google.com/covid19/mobility

Khalel, A. M. H. (2010). Position Location Techniques in Wireless Communication Systems (Dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:bth-4796

Kumar, N., Oke, J., & Nahmias-Biran, Bh. (2021). Activity-based epidemic propagation and contact network scaling in auto-dependent metropolitan areas. *Scientific Reports*, *11*, 22665. https://doi.org/10.1038/s41598-021-01522-w

Maheshwari, P., & Albert, R. (2020). Network model and analysis of the spread of Covid-19 with social distancing. *Applied Network Science*, *5*, 100. https://doi.org/10.1007/s41109-020-00344-5

Kiang, M. V., Santillana, M., Chen, J.T., Onnela, J.-P., Krieger, N., Engø-Monsen, K., Ekapirat, N., Areechokchai, D., Prempree, P., Maude, R. J., & Buckee, C. O. (2021). Incorporating human mobility data improves forecasts of Dengue fever in Thailand. *Scientific Reports*, *11*, 923. https://doi.org/10.1038/s41598-020-79438-0

Muller, S. A., Balmer, M., Neumann, A., & Nagel, K. (2020). Mobility traces and spreading of COVID-19. medRxiv preprint. https://doi.org/10.1101/2020.03.27.20045302

Pace, R. K., Barry, R., & Sirmans, C. F. (1998). Spatial Statistics and Real Estate. *The Journal of Real Estate Finance and Economics*, *17*, 5–13. https://doi.org/10.1023/A:1007783811760

Ping, J. L., Green, C. J., Zartman, R. E., & Bronson, K. F. (2004). Exploring spatial dependence of cotton yield using global and local autocorrelation statistics, *Field Crops Research*, *89*(2–3), 219–236. https://doi.org/10.1016/j.fcr.2004.02.009

Porta, M. (Ed.). (2014). *A Dictionary of Epidemiology*. Oxford University Press.

Rechtin, M., Feldman, V., Klare, S., Riddle, N., & Sharma, R. (2020). Modeling and Simulation of COVID-19 Pandemic for Cincinnati Tri-State Area. arXiv preprint: 200606021. https://doi.org/10.48550/arXiv.2006.06021

Ruktanonchai, N. W., Ruktanonchai, C. W., Floyd, J. R., & Tatem, A. J. (2018). Using Google Location History data to quantify fine-scale human mobility. *International Journal of Health Geographics*, *17*, 28. https://doi.org/10.1186/s12942-018-0150-z

Sardianos, C., Varlamis, I., & Bouras, G. (2018). Extracting User Habits from Google Maps History Logs. 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, 2018, 690–697, https://doi.org/10.1109/ASONAM.2018.8508442

Schlosser, F., Maier, B. F., Jack, O., Hinrichs, D., Zachariae, A., & Brockmann, D. (2020). COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(52), 32883–32890. https://doi.org/10.1073/pnas.2012326117

Setti, L., Passarini, F., De Gennaro, G., Barbieri, P., Perrone, M. G., Borelli, M., Palmisani, J., Di Gilio, A., Piscitelli, P., & Miani, A. (2020). Airborne Transmission Route of COVID-19: Why 2 Meters/6 Feet of Inter-Personal Distance Could Not Be Enough. *International Journal of Environmental Research and Public Health*, *17*(8), 2932. https://doi.org/10.3390/ijerph17082932

Si, Y. L., Debba, P., Skidmore, A. K., Toxopeus, A. G., & Li, L. (2008). Spatial and temporal patterns of global H5N1 outbreaks. In ISPRS 2008: Proceedings of the XXI congress:

Silk road for information from imagery: the International Society for Photogrammetry and Remote Sensing, 3-11 July, Beijing, China. Comm. II, WG II/1. Beijing: ISPRS, 2008. 69–74. International Society for Photogrammetry and Remote Sensing (ISPRS). http://www.isprs.org/proceedings/XXXVII/congress/2_pdf/1_WG-II-1/12.pdf

Soures, N., Chambers, D., Carmichael, Z., Daram, A., Shah, D. P., Clark, K., Potter, L., & Kudithipudu, D. (2020). SIRNet: Understanding social distancing measures with hybrid neural network model for COVID-19 infectious spread. arXiv preprint: 200410376. https://doi.org/10.48550/arXiv.2004.10376

Tsai, P. J., Lin, M. L., Chu, C. M., & Perng, C. H. (2009). Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006. *BMC Public Health*, *9*, 464. https://doi.org/10.1186/1471-2458-9-464

Venkatramanan, S., Sadilek, A., Fadikar, A., Barrett, C. L., Biggerstaff, M., Chen, J., Dotiwalla, X., Eastham, P., Gipson, B., Higdon, D., Kucuktunc, O., Lieber, A., Lewis, B. L., Reynolds, Z., Vullikanti, A. K., Wang, L., & Marathe, M. (2021). Forecasting influenza activity using machine-learned mobility map. *Nature Communications*, *12*, 726. https://doi.org/10.1038/s41467-021-21018-5

Wang, Y., Wang, Y., Chen, Y., & Qin, Q. (2020). Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. *Journal of Medical Virology*, *92*(6), 568–576. https://doi.org/10.1002/jmv.25748

Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*, *395*(10225), 689–697. https://doi.org/10.1016/S0140-6736(20)30260-9

Yeshiwondim, A. K., Gopal, S., Hailemariam, A. T., Dengela, D. O., & Patel, H. P. (2009). Spatial analysis of malaria incidence at the village level in areas with unstable transmission in Ethiopia. *International Journal of Health Geographics*, *8*, 5. https://doi.org/10.1186/1476-072X-8-5

Yi, H., Ng, S. T., Farwin, A., Pei Ting Low, A., Chang, C. M., & Lim, J. (2021). Health equity considerations in COVID-19: geospatial network analysis of the COVID-19 outbreak in the migrant population in Singapore. *Journal of Travel Medicine*, *28*(2), taaa159. https://doi.org/10.1093/jtm/taaa159

Zarei, M., Rahimi, K., Hassanzadeh, K., Abdi, M., Hosseini, V., Fathi, A., & Kakaei, K. (2021). From the environment to the cells: An overview on pivotal factors which affect spreading and infection in COVID-19 pandemic. *Environmental Research*, *201*, 111555. https://doi.org/10.1016/j.envres.2021.111555

Customer Churn Prediction through Attribute Selection Analysis and Support Vector Machine

Jia Yi Vivian Quek
Multimedia University

Ying Han Pang
Multimedia University

Zheng You Lim
Multimedia University

Shih Yin Ooi
Multimedia University

Wee How Khoh
Multimedia University

Abstract: An accurate customer churn prediction could alert businesses about potential churn customers so that proactive actions can be taken to retain the customers. Predicting churn may not be easy, especially with the increasing database sample size. Hence, attribute selection is vital in machine learning to comprehend complex attributes and identify essential variables. In this paper, a customer churn prediction model is proposed based on attribute selection analysis and Support Vector Machine. The proposed model improves churn prediction performance with reduced feature dimensions by identifying the most significant attributes of customer data. Firstly, exploratory data analysis and data preprocessing are performed to understand the data and preprocess it to improve the data quality. Next, two filter-based attribute selection techniques, i.e., Chi-squared and Analysis of Variance (ANOVA), are applied to the pre-processed data to select relevant features. Then, the selected features are input into a Support Vector Machine for classification. A real-world telecom database is used for model assessment. The empirical results demonstrate that ANOVA outperforms the Chi-squared filter in attribute selection. Furthermore, the results also show that, with merely ~50% of the features, feature selection based on ANOVA exhibits better performance compared to full feature set utilization.

Keywords: Churn Prediction, Attribute Selection, Machine Learning, Filter Methods, Support Vector Machine.

http://doi.org/10.18080/jtde.v11n3.777

Introduction

The term "churn" describes a scenario in which a customer discontinues a company's services. This can occur as a consequence of unforeseeable events. For example, the Covid 19 outbreak has led to businesses going above and beyond to entice customers to stay loyal (Johny & Mathai, 2017). Predicting customer churn is a good strategy to reduce customer churn. With the statistics, businesses can identify those potential churn customers as well as the reasons. Johnny & Mathai (2017) claimed that churn rate is affected by a variety of factors, including demographic details, such as age, gender, marital status, and location, and customer behaviour, such as frequency of interaction with service providers, monthly revenue, and total recurring charges. While predicting customer churn is useful and helpful for a business, it can be difficult due to a massive database. Thus, attribute selection is an important process in machine learning that aids in comprehending the complicated relationships between attributes and identifying the essential factors while eliminating irrelevant or redundant ones (Albulayhi *et al.*, 2022).

In this paper, a customer churn prediction model based on attribute selection analysis and Support Vector Machine is proposed. With an effective attribute selection technique, the most significant customer data attributes can be identified, leading to enhanced churn prediction performance while minimizing the feature dimension. In this work, exploratory data analysis and data preprocessing are first performed to understand the customer data. Then, the data is pre-processed for data quality improvement, which helps improve the classification performance. Next, two filter-based attribute selection techniques, i.e., Chi-squared and Analysis of Variance, are performed on the pre-processed data to analyse and select relevant features. Then, the selected feature sets are input into a Support Vector Machine for data classification. In this study, a real-world telecom database, i.e., Cell2Cell (2018), is used for model assessment.

The contributions of this study are listed as follows:

- A machine learning-based customer churn prediction framework is proposed for the telecommunication industry.
- The performance of filter-based attribute selection techniques, i.e., Chi-squared and Analysis of Variance, in identifying the most important attributes for churn prediction is examined.
- The performance of the proposed customer churn prediction system is assessed based on real- world telecommunication customer data.

http://doi.org/10.18080/jtde.v11n3.777

Related Work

There are numerous works for customer churn prediction in the telecommunication business. Since this study is using the Cell2Cell (2018) dataset, the literature review in this study focuses on the previous research that was conducted on the Cell2Cell dataset. In the literature, it had been demonstrated that Support Vector Machine produced a better model performance than other classification algorithms, such as Decision Tree and neural networks, in churn classification (Umayaparvathi & Iyakutti, 2016; Vaidya & Nigam, 2022).

Shuli Wu and Wei-Chuen Yau *et al.* (2021) explored several machine learning algorithms for predicting customer churn. In this work, the Synthetic Minority Oversampling Technique (SMOTE) was applied to the training set to address the issue of imbalanced data. Due to the large number of attributes in the dataset, a feature selection technique known as the Chi-squared test was performed to reduce the data dimensionality. The tested machine learning techniques adopted in this work were Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, AdaBoost, and Multi-Layered Perceptron (MLP). The empirical results showed that Multi-layer Perceptron demonstrated the best performance in terms of F1-score with 42.84%, whereas Random Forest was ranked the best in terms of accuracy with a score of 63.09% (Wu *et al.*, 2021).

Fujo *et al.* (2022) proposed Deep-BP-ANN using two feature selection techniques (i.e., Variance Thresholding and Lasso Regression). Furthermore, the authors also adopted the Random Oversampling (ROS) technique to solve the issue of imbalanced data in the Cell2Cell dataset. In this study, the performance was evaluated using a holdout set and 10-fold cross-validation. Different classifiers, such as Naïve Bayes, Logistic Regression, XG-Boost, and the KNN algorithm, were explored. From the experimental results, the XG-Boost algorithm surpassed the other machine learning algorithms (Fujo *et al.*, 2022).

Jain *et al.* (2022) performed churn prediction on a subset of the Cell2Cell database. This work mainly focused on the feature importance and feature engineering for churn data. Random Forest and Gradient Boosted Tree were examined for classification. From the empirical results, Gradient Boosted Tree performed better in terms of accuracy and sensitivity. In this work, additional new features based on specific rules were produced. It was shown that these new features achieved high importance for churn prediction (Jain *et al.*, 2022).

The Proposed System

There are four phases involved in the proposed system: (1) data collection and retrieval; (2) exploratory data analysis (EDA) and data preprocessing; (3) feature selection; and (4)

http://doi.org/10.18080/jtde.v11n3.777

model generation and training. Figure 1 illustrates the overview of the proposed customer churn prediction. The details of each phase will be explained in the following subsections.

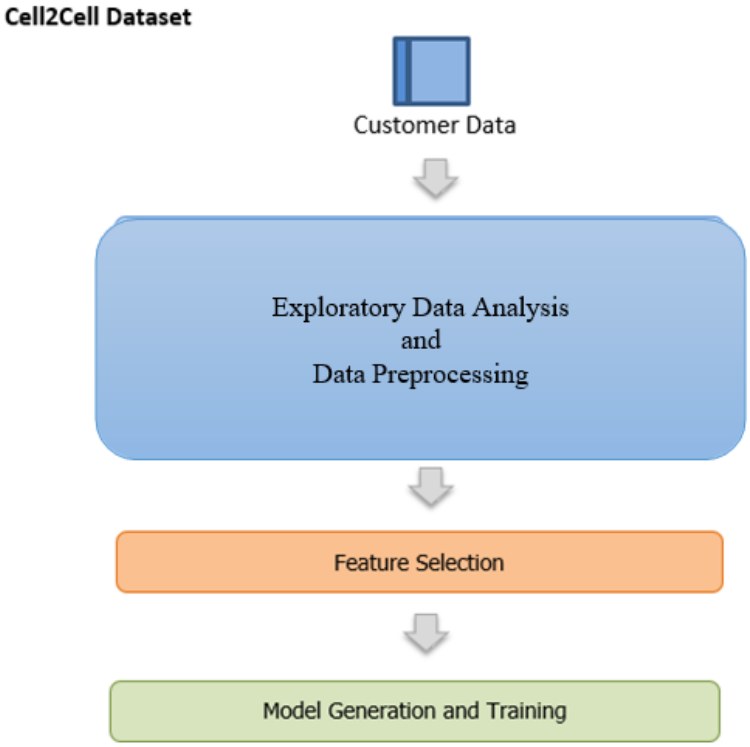


Figure 1. Flowchart of the proposed system

Customer data collection

Cell2Cell dataset is a telecommunication database containing 71047 instances and 58 attributes (Cell2Cell, 2018). The attributes include customer demographic information, product information, marketing-related data, service-related data, and payment-related data, etc., which includes numerical variables such as monthly revenue and roaming calls; binary variables, such as owned computer and new cell phone users; ordinal variables, such as credit rating and occupation, and so on. Table 1 records the details of the attributes. The attribute "Churn" is used as the class label.

Table 1. Details of the Attributes in the Cell2Cell dataset

Numerical Attributes	Data Format	Categorical Attributes	Data Format
CustomerID	int	Churn	Yes/No
Monthly Revenue	float	ServiceArea	string
MonthlyMinutes	float	ChildrenInHH	Yes/No
TotalRecurringCharge	float	HandsetRefurbished	Yes/No
DirectorAssistedCalls	float	HandsetWebCapable	Yes/No
OverageMinutes	float	TruckOwner	Yes/No
RoamingCalls	float	RVOwner	Yes/No
PercChangeMinutes	float	Homeownership	Known/Unknown
PercChangeRevenues	float	BuyViaMailOffers	Yes/No
DroppedCalls	float	RespondsToMailOffers	Yes/No

http://doi.org/10.18080/jtde.v11n3.777

Numerical Attributes	Data Format	Categorical Attributes	Data Format
BlockedCalls	float	OptOutMailings	Yes/No
UnansweredCalls	float	NonUSTravel	Yes/No
CustomerCareCalls	float	OwnsComputer	Yes/No
ThreewayCalls	float	HasCreditCard	Yes/No
ReceivedCalls	float	NewCellphoneUser	Yes/No
OutboundCalls	float	NotNewCellphoneUser	Yes/No
InboundCalls	float	OwnsMotorcycle	Yes/No
PeakCallsInOut	float	HandsetPrice	string
OffPeakCallsInOut	float	MadeCallToRetentionTeam	Yes/No
DroppedBlockingCalls	float	CreditRating	string
CallForwardingCalls	float	PrizmCode	Other/Suburban/ Town/Rural
CallWaitingCalls	float	Occupation	Other/Professional/ Crafts/Clerical/Self/ Retired/Student/ Homemaker
MonthsInService	int	MaritalStatus	Unknown/Yes/No
UniqueSubs	int		
ActiveSubs	int		
Handsets	float		
HandsetModels	float		
CurrentEquipmentDays	float		
AgeHH1	float		
AgeHH2	float		
RetentionCalls	int		
RetentionOffersAccepted	int		
ReferralsMadeBySubscriber	int		
IncomeGroup	int		
AdjustmentsToCreditRating	int		

Exploratory data analysis and data preprocessing

The Cell2Cell dataset is saved as a CSV file and the dataset variable type is a Pandas DataFrame, which is a two-dimensional structure used to examine a wide range of tabular data with associated labels. In this database, there are missing values and outliers, particularly in variables such as monthly revenue and total recurring charge. Moreover, some variables require data transformation to achieve the desired form for further processes. Therefore, data preprocessing is critical for cleaning and transforming these variables in order to improve prediction performance. The dataset is checked for missing values using the isnull() method. Instead of dropping the null values, the np.nan technique is used to clean the data by replacing those missing variables with zero. In the Cell2Cell database, there are 14 attributes with missing values, as recorded in Table 2.

After analyzing the data information, two attributes are eliminated from consideration for data learning and analysis in this study. The attributes are "CustomerID", which is the identifier for each client and is not useful in predicting churn behaviour, and "ServiceArea" attribute. The removal of “Service Area” is because it could cause a lot of noise in the dataset (Jain *et al.*,

http://doi.org/10.18080/jtde.v11n3.777

2022). Most values in this attribute are unique after the label encoding process, resulting in ~740 different category labels.

Table 2. Missing Variables in the Dataset

Attribute Name	Total Missing Values
Monthly Revenue	156
MonthlyMinutes	156
TotalRecurringCharge	156
DirectorAssistedCalls	156
OverageMinutes	156
RoamingCalls	156
PercChangeMinutes	367
PercChangeRevenues	367
ServiceArea	24
Handsets	1
HandsetModels	1
CurrentEquipmentDays	1
AgeHH1	909
AgeHH2	909

By utilizing log transformation, the skewness of numerical variables is decreased. Next, the values are further transformed to the range from 0 to 1. Next, the categorical variables are converted into numerical variables because the machine learning model only accepts numerical input data. Thus, a label encoding method, known as one-hot encoding, is applied to convert those categorical variables from string format into numerical variables. In the Cell2Cell database, the attribute of "Prizm Code" has four categories: other, suburban, town, and rural. After performing label encoding, the categories are represented in numerical formats of 0, 1, 2, and 3, respectively. Figure 2 illustrates the example of label encoding on the attribute of "Prizm Code".

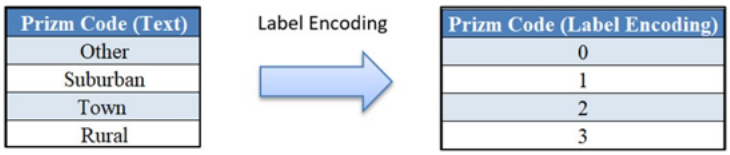


Figure 2. Label encoding on a categorical variable

In Exploratory Data Analysis, experimental results show that numerous strategies can be used to analyze data and improve the performance of predictive models (Zheng, 2022). A correlation heatmap is used to identify the correlations between variables, as it allows easier comparison between different pairs of variables. The colour palette in the legend represents the degree of correlation between the factors. In this study, we adopt the colour "Blues" to represent the heatmap; see Figure 3. The darker shade indicates a higher correlation, indicating that the variables tend to move in the same direction; whereas the lighter shade indicates a lower correlation, indicating that the variables are not closely related to one another. We can swiftly and easily determine which variables are most strongly correlated and

http://doi.org/10.18080/jtde.v11n3.777

which variables are independent by using a heatmap. Figure 4 illustrates the correlation heatmap between the attributes/variables of the Cell2Cell database. From the map, we can discover that there are a few attributes that are highly correlated. Attribute selection analysis will be performed to determine a subset of relevant and informative attributes for further processes.

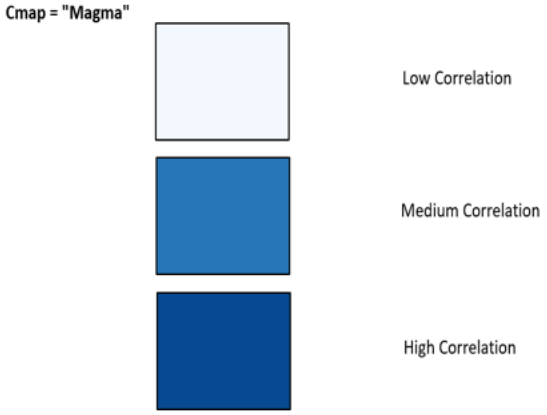


Figure 3. Correlation heatmap with different correlations identified using different colours

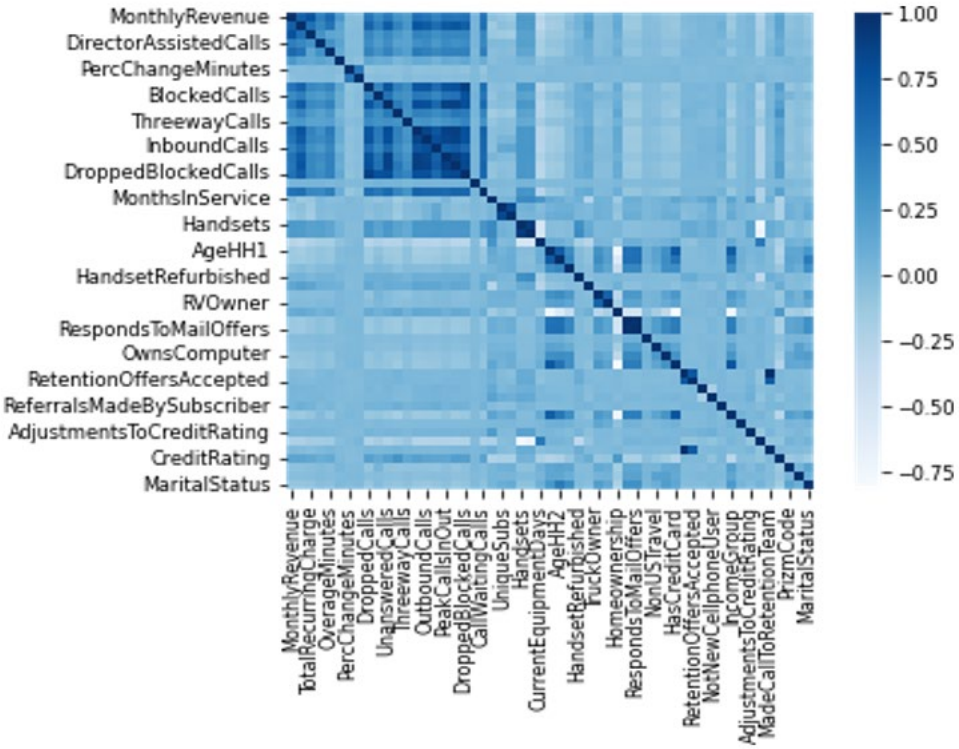


Figure 4. Correlation heatmap of Cell2Cell dataset

In some research, there are a few researchers who have neglected the issue of class imbalance, and the majority of the studies described above used historical data, primarily from Kaggle, and were conducted in wealthy countries. Such data, however, may not adequately reflect the issues in the real world. Besides that, the bulk of the research relied on a small number of datasets, which may have restricted the development and selection of better models that could

http://doi.org/10.18080/jtde.v11n3.777

define the overall issue (Seid & Woldeyohannis, 2022). In this study, it is observed that the churn rate of the Cell2Cell dataset provided by Kaggle is imbalanced, i.e., having more non-churn samples (i.e., majority class) than churn samples (i.e., minority class), with ~70% and ~30% in each class. Imbalanced data might lead to bias towards the majority class in the dataset. Thus, a data sampling technique, known as the Synthetic Minority Oversampling method (SMOTE), is used to balance the dataset's churn rate. This technique is an oversampling technique. By increasing the representation of the minority class, data sampling could balance the dataset and provide a more representative sample for training and testing the model. A balanced dataset is required for training the models, because it guarantees that the model is not biased towards one class and can make accurate predictions for both. SMOTE creates synthetic minority class observations by interpolating between minority class observations that already exist. This is accomplished by randomly selecting an observation from a minority class, then locating its k-nearest neighbours in the feature space. Next, one of the neighbours is randomly selected and a new observation is produced by interpolating between the selected observation and the randomly selected neighbour. Before implementing the SMOTE method, our training dataset contains 31265 samples with 9010 churn samples (28.8%) and 22255 non-churn samples (71.2%). After implementing the SMOTE method, our training dataset contains 44510 samples with 22255 churn samples (50%) and 22255 non-churn samples (50%).

Attribute selection analysis

Given the growing sample size of the database, predicting customer churn may not be simple. Research has stated that, even though the predictive models perform well at first, the addition of the proposed feature selection approach recursive feature elimination (RFE) results in a significant improvement in their performance. Regardless of the machine learning algorithms used to predict customer churn, a feature selection method should be included. However, establishing the best feature selection strategy for customer churn prediction in the telecoms business remains a difficult task (Naing *et al.*, 2022). As a result, one of the crucial methods to understand complicated attribute interactions is through attribute selection. Attribute selection analysis is able to discover crucial variables while removing those unnecessary and redundant ones. In other words, attribute selection can aid in the selection of the best representative features that might contribute to analyzing customer churn behaviour.

In this study, two filter-based approaches for attribute selection are examined: Chi-Squared Test and Anova (analysis of variance) Test method. The Chi-squared test is a statistical approach for comparing actual outcomes to predicted results. The primary goal of this test is to establish whether the difference between observed data (actual results) and predicted data

http://doi.org/10.18080/jtde.v11n3.777

(predictions) is due to chance or a real relationship between the variables under consideration. This method is frequently used to select highly related sample data, and then uses a minimum redundancy algorithm to further remove redundancy and select features (Wang & Zhou, 2021). When two features are unrelated, the actual and anticipated results are likely to be comparable, resulting in a lower Chi-squared score. A high Chi-squared score, on the other hand, indicates that the independence claim is unjustified, when there is a strong link between the variables. In other words, attributes with higher Chi-squared values are more reliant on the response variable and can be used to train the model. Figure 5 illustrates the Chi-squared score (representing the feature significance score in this study) of the Cell2Cell attributes, ranked from the highest score to the lowest score. In this study, different numbers of selected features (i.e., 10, 20, 30, 40 and 50) are examined. The results indicated that the top 30 features yielded better performance, so the top 30 significant features are selected for the subsequent processes (see Table 3). Note that the p-value is the area under the density curve of the Chi-squared distribution to the right of the value of the test statistic.

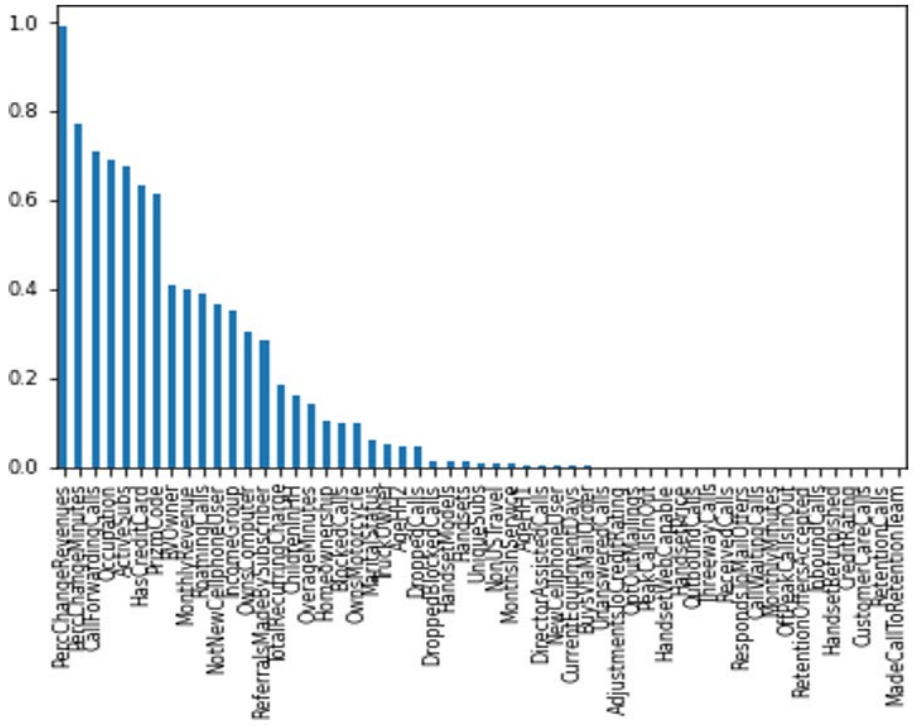


Figure 5. Chi-Squared Test Attribute Selection Results

The Anova Test, often known as the SelectKBest method, is a technique for choosing features based on their scores. It is used to perform the ANOVA statistical test and select the 30 most important features from the original dataset (Lazaros *et al.*, 2022). The technique deliberately excludes features with lower scores and selects the top *k* features with the highest scores. This is because it can help minimize data dimensionality while keeping the most important features; it is frequently used as a feature selection approach in machine learning for

http://doi.org/10.18080/jtde.v11n3.777

classification performance improvement. Table 4 records the top 30 significant features based on the Anova test. These features will be considered for the next processes.

Table 3. Top 30 Attributes selected by Chi-Squared Test Method

No.	Attribute Name	Scores	No.	Attribute Name	Scores
1	PercChangeRevenues	9.881e-01	16	ChildrenInHH	1.607e-01
2	PercChangeMinutes	7.693e-01	17	OverageMinutes	1.411e-01
3	CallForwardingCalls	7.072e-01	18	Homeownership	1.021e-01
4	Occupation	6.889e-01	19	BlockedCalls	9.925e-02
5	ActiveSubs	6.753e-01	20	OwnsMotorcycle	9.889e-02
6	HasCreditCard	6.324e-01	21	MaritalStatus	6.055e-02
7	PrizmCode	6.140e-01	22	TruckOwner	5.109e-02
8	RVOwner	4.100e-01	23	AgeHH2	4.711e-02
9	MonthlyRevenue	4.001e-01	24	DroppedCalls	4.405e-02
10	RoamingCalls	3.875e-01	25	DroppedBlockedCalls	1.444e-02
11	NotNewCellphoneUser	3.638e-01	26	HandsetModels	1.319e-02
12	IncomeGroup	3.488e-01	27	Handsets	1.220e-02
13	OwnsComputer	3.035e-01	28	UniqueSubs	8.466e-03
14	ReferralsMadeBySubscriber	2.844e-01	29	NonUSTravel	6.916e-03
15	TotalRecurringCharge	1.844e-01	30	MonthsInService	5.468e-03

Table 4. Top 30 Attributes selected by Anova Test Method

No.	Attribute Name	Scores	No.	Attribute Name	Scores
1	CurrentEquipmentDays	775.376	16	ReceivedCalls	121.937
2	MonthlyMinutes	433.309	17	HandsetPrice	101.028
3	TotalRecurringCharge	263.868	18	ThreewayCalls	96.762
4	CustomerCareCalls	224.992	19	CallWaitingCalls	89.230
5	HandsetModels	200.445	20	UniqueSubs	88.353
6	CreditRating	196.644	21	MonthsInService	75.534
7	OffPeakCallsInOut	186.874	22	DirectorAssistedCalls	73.334
8	PeakCallsInOut	175.178	23	MonthlyRevenue	72.644
9	MadeCallToRetentionTeam	165.937	24	AgeHH1	58.101
10	InboundCalls	165.011	25	DroppedBlockedCalls	55.591
11	Handsets	164.313	26	AdjustmentsToCredit-Rating	53.442
12	RetentionCalls	155.184	27	PercChangeMinutes	51.419
13	HandsetWebCapable	144.790	28	HandsetRefurbished	42.499
14	UnansweredCalls	135.294	29	RetentionOffers-Accepted	40.074
15	OutboundCalls	122.552	30	DroppedCalls	32.330

Classification model

The Support Vector Machine (SVM) classifier is a supervised machine learning approach which can classify both linear and nonlinear data. In other words, the SVM classifier can be categorized into:

- Linear SVM — operates on linearly separable data where the statistics that can be divided into groups by a single straight line;

http://doi.org/10.18080/jtde.v11n3.777

- Non-linear SVM — operates on non-linearly separable data where the statistics cannot be divided into groups by a straight line.

This classifier adopts kernel approaches to transform data from a low-dimensional space to a higher-dimensional space where a distinct separation can be made. It aims to find the best separation line (known as a hyperplane — a decision boundary) that could optimally separate the classes of data points. In this work, SVM Radial Basis Function is adopted. An SVM classifier is considered in this study due to the following advantages:

- It can work well for this study case where the sample number exceeds the number of feature dimensions;
- Its regularization parameter aids in preventing overfitting;
- It can handle non-linearly separable data (real-world data is often nonlinear). By utilizing kernel functions, the SVM classifier is able to transform the real-world nonlinear data into a higher-dimensional space where the data can be separated linearly, as illustrated in Figure 6.

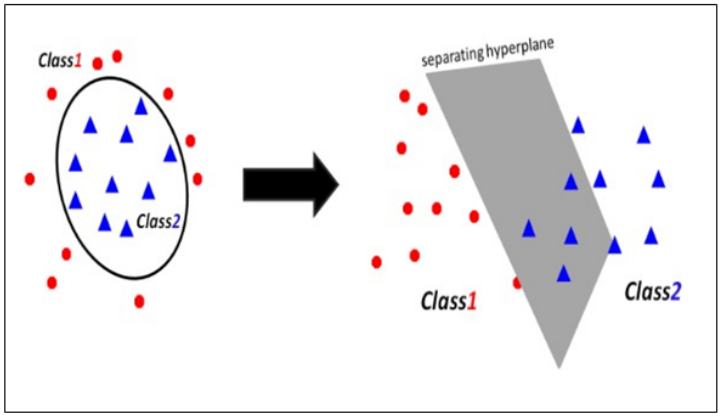


Figure 6. Nonlinear SVM: mapping data into the higher dimensional feature space (right). (The figure is extracted from the work of Mahmoodi *et al.* (2011))

Experimental Results and Discussion

In this study, the experiments are conducted using a train-test split protocol to evaluate the performance of the proposed customer churn prediction model. The Cell2Cell dataset is split into two separate subsets: a training set for training the model; and a testing set to assess the model's performance and its generalization ability to unseen data. After employing the train-test split protocol, there are 44510 training samples and 15315 testing samples from the database. We adopt several performance metrics for performance evaluation, such as precision, recall, and F1 score, as well as the confusion matrix. Precision, recall, and F1 score are formulated as below:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

http://doi.org/10.18080/jtde.v11n3.777

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

where *TP* is the number of true positive classifications; *FP* is the number of false positives; *FN* is the number of false negatives.

Before training the model, the top significant 30 attributes are selected for evaluating the model's performance. After selecting the attributes, the Support Vector Machine (SVM) model results are analyzed with three performance metrics, which are precision, recall, and F1 score, by using the two filter-based feature selection techniques, Chi-squared Test and Anova Test. Moreover, the achieved confusion matrix is also provided for reference purposes.

Table 5 records the classification performance (in terms of precision, recall, and F1 score) of the proposed customer churn prediction model with different attribute selection analyses. Figure 7 shows the performance for better illustration. From the obtained empirical results, we can observe that the proposed system using the Anova Test as attribute selection analysis performs better than that using the Chi-squared test. The former model achieves a precision score of 35.54%, a recall score of 62.39%, and an F1 score of 45.12%; whereas the latter model obtains a precision score of 33.35%, a recall score of 61.14%, and an F1 score of 43.16%. Furthermore, the experimental results show that the proposed model using a full feature set for classification attains a precision score of 35.79%, a recall score of 60.29%, and an F1 score of 44.91%.

Table 5. Classification Performance of the Proposed Customer Churn Prediction Model with Different Attribute Selection Analyses

Attribute Selection Analysis	Feature Dimension	Precision (%)	Recall (%)	F1-Score (%)
Chi-Squared	20	31.38	53.19	39.48
	30	33.35	61.15	43.16
Anova Test	20	33.1	63.64	43.55
	30	35.34	62.39	45.12
Full Feature Set	55	35.79	60.29	44.91

Figure 8 illustrates the confusion matrices of the proposed models (using the Chi-squared test and Anova test) with different feature selection dimensions. It is understood that the core objective of a customer churn prediction model is to identify customers who are very likely to leave a business (i.e., stopping purchasing/subscribing to the company's product or service), so that necessary actions can be taken to stop them from churning. From the figure, we can observe that the models can predict customer churn with a feature dimension, *K*, of 30, compared to that of 20. In the model with the Chi-squared test, 2699 out of 4414 churn customers are able to be identified by using K=30, compared to the model with K=20 having

http://doi.org/10.18080/jtde.v11n3.777

2348 churn customers detection. Similar to the models with the Anova test, K=30 allows for the identification of 2809 of 4414 churning customers, as opposed to 2754 with K=20.

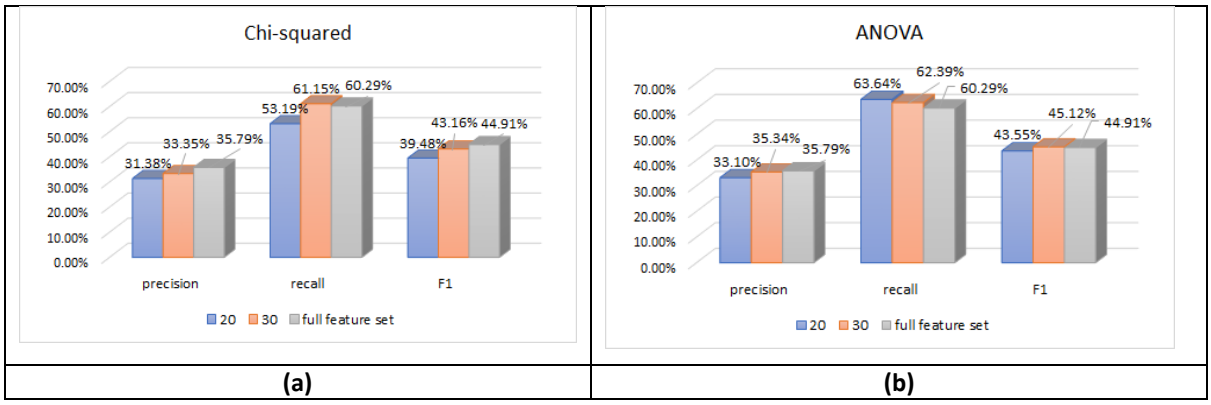


Figure 7. Performance of the proposed prediction model with different attribute selection analysis: (a) Chi-squared and (b) Anova

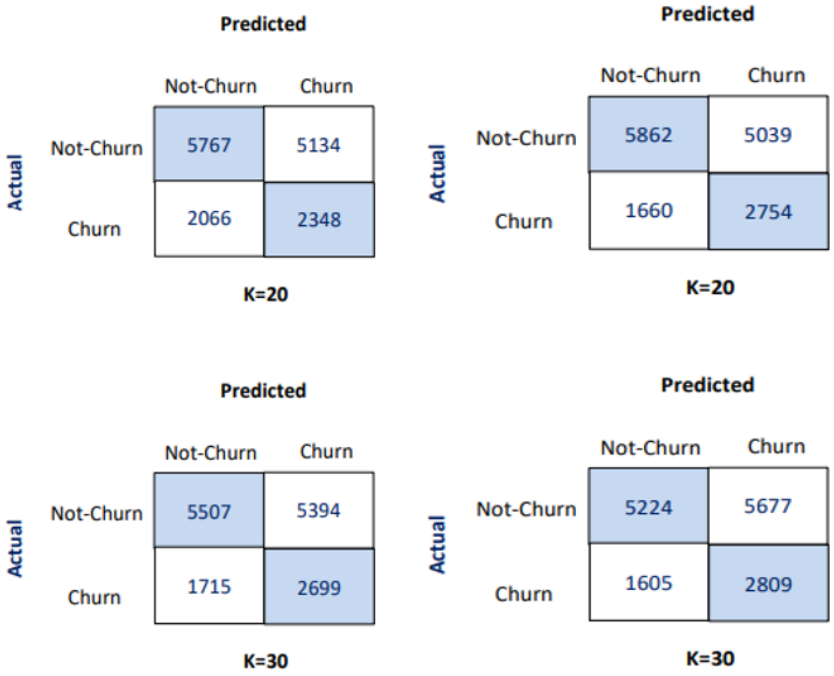


Figure 8. Confusion Matrix for Chi-squared Test (Left) and Anova Test (Right) with K values of 20 and 30

Conclusions

In this paper, a machine learning-based customer churn prediction model is proposed with filter-based attribute selection analysis and Support Vector Machine (SVM). In this research, a real-world telecommunication database, i.e., Cell2Cell dataset, is employed to assess the performance of the proposed churn prediction model. In order to better comprehend the customer data and prepare it for feature analysis, exploratory data analysis and data preprocessing are first carried out. These processes help increase classification performance by enhancing data quality. Next, relevant and representative features are selected using two

http://doi.org/10.18080/jtde.v11n3.777

filter-based attribute selection techniques, namely Chi-squared and Anova. The top 30 significant features are then passed into a Support Vector Machine to classify the data. From the obtained empirical results, it is observed that the proposed model with the Anova test attains a higher F1 score of 45.12% compared with the model with the Chi-squared test (with an F1 score of 43.16%) and the model with a full feature set (with F1 score of 44.91%). It is found that the Anova test is a better attribute strategy compared to the Chi-squared test. The finding shows the importance of using an adequate attribute selection strategy to comprehend complex attribute relations and obtain representative features for improved data classification. Furthermore, by eliminating irrelevant and redundant attributes, it helps in effective computation.

Acknowledgements

A version of this paper was presented at the third International Conference on Computer, Information Technology and Intelligent Computing, CITIC 2023, held in Malaysia on 26–28 July 2023.

References

Albulayhi, K., Abu Al-Haija, Q., Alsuhibany, S. A., Jillepalli, A. A., Ashrafuzzaman, M., & Sheldon, F. T. (2022). IoT intrusion detection using machine learning with a novel high performing feature selection method. *Applied Sciences*, *12*(10), 5015. https://doi.org/10.3390/app12105015

Cell2Cell. (2018). *Telecom Churn (Cell2Cell)*. [Online]. Available at https://www.kaggle.com/jpacse/datasets-for-churn-telecom

Fujo, S. W., Subramanian, S., & Khder, M. A. (2022). Customer churn prediction in telecommunication industry using deep learning. *Information Sciences Letters*, *11*(1), 24. http://dx.doi.org/10.18576/isl/110120

Jain, H., Khunteta, A., & Srivastava, S. (2022). Telecom Churn Prediction Using an Ensemble Approach with Feature Engineering and Importance. *International Journal of Intelligent Systems and Applications in Engineering*, *10*(3), 22–33. https://ijisae.org/index.php/IJISAE/article/view/2134

Johny, C. P., & Mathai, P. P. (2017). Customer churn prediction: A survey. *International Journal of Advanced Research in Computer Science*, 8(5), 2178-2181. http://www.ijarcs.info/index.php/Ijarcs/article/view/4079

Lazaros, K., Tasoulis, S., Vrahatis, A., & Plagianakos, V. (2022). Feature Selection For High Dimensional Data Using Supervised Machine Learning Techniques. *IEEE International Conference on Big Data (Big Data)*, 2022 (pp. 3891–3894). IEEE. https://doi.org/10.1109/BigData55660.2022.10020654

Mahmoodi, D., Soleimani, A., Khosravi, H., & Taghizadeh. M. (2011). FPGA Simulation of Linear and Nonlinear Support Vector Machine. *Journal of Software Engineering and Applications*, *5*(4), 320–328. http://dx.doi.org/10.4236/jsea.2011.45036

Naing, Y. T., Raheem, M., & Batcha, N. K. (2022). Feature Selection for Customer Churn Prediction: A Review on the Methods & Techniques applied in the Telecom Industry. *IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 2022 (pp. 1–5). IEEE. https://doi.org/10.1109/ICDCECE53908.2022.9793315

Seid, M. H., & Woldeyohannis, M. M. (2022). Customer Churn Prediction Using Machine Learning: Commercial Bank of Ethiopia. *International Conference on Information and Communication Technology for Development for Africa (ICT4DA)*, 2022 (pp. 1–6). IEEE. https://doi.org/10.1109/ICT4DA56482.2022.9971224

Umayaparvathi, V., & Iyakutti, K. (2016). A Survey on Customer Churn Prediction in Telecom Industry: Datasets, Methods and Metrics. *International Research Journal of Engineering and Technology*, *3*(4), 1065-1070. https://www.irjet.net/archives/V3/i4/IRJET-V3I4213.pdf

Vaidya, S., & Nigam, R. K. (2022). An Analysis of Customer Churn Predictions in the Telecommunications Sector. *International Journal of Electronics Communication and Computer Engineering*, *13*(4), 37–43. https://www.ijecce.org/index.php/component/jresearch/?view=publication&task=show&id=1382&Itemid=437

Wang, Y., & Zhou, C. (2021). Feature selection method based on chi-square test and minimum redundancy. In Emerging Trends in Intelligent and Interactive Systems and Applications: *Proceedings of the 5th International Conference on Intelligent, Interactive Systems and Applications (IISA2020)* (pp. 171–178). Springer International Publishing. http://dx.doi.org/10.1007/978-3-030-63784-2_22

Wu, S., Yau, W. C., Ong, T. S., & Chong, S. C. (2021). Integrated churn prediction and customer segmentation framework for telco business. *IEEE Access*, 9, 62118–62136. https://doi.org/10.1109/ACCESS.2021.3073776

Zheng, K. (2022). Identifying Churning Employees: Machine Learning Algorithms from an Unbalanced Data Perspective. *5th International Conference on Machine Learning and Machine Intelligence*, 2022 (pp. 14–22). https://doi.org/10.1145/3568199.3568202

Harry S. Wragge AM (1929-2023)

A Major Contributor to Australian Telecommunications

Peter Gerrand
Life Member, TelSoc

Abstract: Harry Stewart Wragge (23 November 1929–31 July 2023), Director of the Telecom (later Telstra) Research Laboratories (TRL) in Melbourne from 1985 to 1992, was a leading Australian research engineer. He catalysed the evolution of the public switched telephone network in Australia from its analogue, electromechanical form in the 1950s and 1960s to the digital, computer-controlled circuit-switched network of the 1980s and 1990s. His own research work was famous for his IST (integrated switching and transmission) project, building the first computer-controlled, integrated digital switch in the world to handle commercial telephone traffic (from 1974 to 1978). He also provided major support to telecommunications research at several Australian universities.

His many contributions were recognized by honours from the Australian Government, Melbourne and Monash universities, the Pearcey Foundation, the Telecommunications Society of Australia and the City of Frankston.

Keywords: Obituary, Australian telecommunications history, telecommunications research, University of Melbourne, yachting

The First Editior-in-Chief of ATR, Mr Harry Wragge, AM on the occasion of his receiving the Kernot Medal in 1990.

Figure 1. Harry Wragge in 1990 (Gerrand, 1996, p. 35)

http://doi.org/10.18080/jtde.v11n3.787

Introduction: Harry's Childhood and Adolescence

Harry Wragge's resourcefulness and sense of responsibility were formed in his character from an early age. At age ten, when his father had gone off to war, Harry became, in the recollection of his younger sister. Anne, the "man of the household", stepping up to help his mother run the family farm, *Carnmallam* (Hill, 2023).

Harry's father had married Lesley Sweatman in Caulfield, Victoria in September 1928, and their son Harry was born on 23 November in the following year. Harry and his sisters, Anne, Jean and Elizabeth, were brought up on the family farm, where he was home schooled until he was eight. He then went to Devon Meadows Primary School. But in 1940, while his father was away at war, *Carnmallam* burnt down, and his mother moved the family to a new home at Seaford (Hill, 2023).

Harry was sent to Scotch College, which he attended from 1942 to 1948. Two fellow pupils, Mac Cleland and John Cathcart, who became Harry's lifelong friends, recall that Harry stood out, noting that he had a "beloved, self-restored, vintage car" and, at one time, "a mysterious looking antenna protruding from the ink well on his desk and connected to a crystal set inside the desk". In his first year, Harry was dux of his year (Cleland, 2023).

In his third year at Scotch, Harry lost the sight of one eye as a result of a chemical explosion at school, leading to his absence from school for several months. But he soon caught up, especially in physics and mathematics, where he excelled. He joined the Signals section of the school cadets, where, Cleland believes, Harry "cemented his ambition to embark upon a career in telephony" (Cleland, 2023).

Harry completed his matriculation (Year 12) in 1948, and planned to repeat it in the hope of gaining a scholarship to university. But, as his sister Anne recalls, those plans changed when Harry went with a fellow train traveller to the PMG Department in St Kilda Road (Hill, 2023). He joined the PMG as a clerk, attending evening lectures in mathematics at the University of Melbourne during 1949 to improve his chances of winning a PMG Cadetship, which he achieved at the end of that year. In 1950, he enrolled for a B.Sc. as well as working at the PMG Workshops in South Melbourne. At the beginning of that year his cadetship was varied to enable him to study electrical engineering, which he commenced in 1951 (Wragge, 2004).

Engineering at Melbourne University in the 1950s

While studying for his BEE, Harry embraced life as an engineering student to the full. In his second year, he performed in the annual Engineers Revue (*Cranks & Nuts*, 1951) and, in his fourth and final year (Figure 2), he was Chairman of the Melbourne University Engineering

http://doi.org/10.18080/jtde.v11n3.787

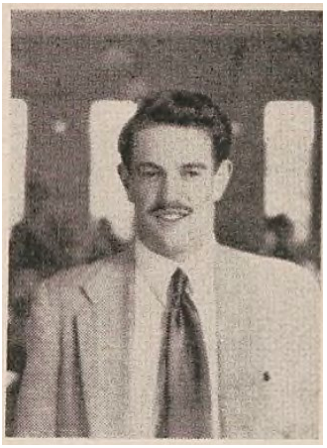


Figure 2. Harry Wragge in the fourth year of his BEE studies (Cranks & Nuts, 1953)

Students Club (MUESC) and the Chairman of Clubs and Societies on the Executive of the University's Student Representative Council (*Cranks & Nuts*, 1953). It was a heavy extra-curricular load.

Harry's situation attracted the attention of Professor Charles Moorhouse, then Dean of Engineering, as well as Head of Electrical Engineering. Moorhouse felt that Harry's extracurricular activities would detract from his final results, and that he should come back for a year's research and gain a Master of Engineering Science so as to get a good qualification. That started long negotiations with the PMG and the Commonwealth Public Service Board, resulting in the extension of Harry's cadetship to cover a further year's postgraduate research. In fact, the professor's concerns may have been groundless: Harry achieved First Class Honours and top place for his BEE degree (1954). But Moorhouse's efforts were rewarded by Harry obtaining his MEngSc degree in 1955 with Second Class Honours (Wragge, 2004).

Bill Brown, who studied electrical engineering at Melbourne seven years behind Harry, has told me that for Harry's MEngSc research he designed and built one of the first analogue computers in Australia. *Cranks & Nuts* (1954, p. 27) records that, at the 1954 Engineering School Exhibition, attended by an estimated 1,500 guests, "probably the most popular exhibit for the technically minded visitor was the analogue computer in Electrical Research". Bill Brown carried out his own MEngSc research on control systems in 1961 using Harry's computer[1] (Brown, 2023).

For a list of abbreviations used in this manuscript, see Table 1.

Table 1. Glossary of abbreviations

Abbreviation	Meaning
BEE	Bachelor of Electrical Engineering degree
CBD	Central Business District
CCITT	ITU's Consultative Committee(s) for International Telephony and Telegraphy
EEE	Electronic and Electrical Engineering
HKT	Hong Kong Telephone company
IST	Integrated Switching and Transmission project
ITU	International Telecommunication Union
MD	Managing Director
MEngSc	Master of Engineering Science degree
OTC	Overseas Telecommunications Commission of Australia

http://doi.org/10.18080/jtde.v11n3.787

Abbreviation	Meaning
PABX	Private Automatic Branch Exchange
PCM	Pulse Coded Modulation (of voice or other analogue signals)
PMG	PostMaster General's Department
TRL	Telecom Australia Research Laboratories; later Telstra Research Laboratories
VF	Voice Frequency

Early Research at the PMG Research Labs

Given his academic results and postgraduate degree, Harry had no difficulty being assigned to the PMG Research Laboratories in 1955 after completion of his university research. Within a year (1956), he was promoted to Divisional Engineer, VF Transmission ('Our Contributors', 1967).

The Research Labs, founded in 1922, were headquartered at 59 Little Collins Street in central Melbourne; but the research staff, numbering about 200 when Harry joined them, were accommodated in five buildings nearby in the CBD. Later, when the importance of Harry's work on digital switching and signalling systems warranted creation of a large Switching & Signalling Section, later extended to become a new Branch, the staff were accommodated at new premises at 140 Exhibition Street. In 1972, Harry was appointed Assistant Director (Research), head of the new Switching and Signalling Branch (Coxhill, 2007a).


H. S. WRAGGE

Figure 3. Harry Wragge at the PMG Research Labs in 1967 ('Our Contributors', 1967)

Early in Harry's career at the Research Labs, he spent time advocating and demonstrating the potential of transistor-based electronics for the PMG's equipment ('Our Contributors', 1967). At that time the PMG's national automatic telephony network was entirely based upon electromechanical step-by-step equipment. Its first technology upgrade, to crossbar switching in the 1960s (Moyal, 1984, p. 225), remained essentially an electromechanical technology.

Harry (Figure 3) began the process of knowledge transfer on the design of transistor circuits through a series of Research Laboratory Reports and papers in the PMG's house journal, the *Telecommunication Journal of Australia* (e.g., Wragge, 1960; Wragge & Wion, 1962). At the same time Harry was reviewing developments in the design of model electronic telephone exchanges in the major research laboratories around the world (Wragge, 1961). His next step was to form a small team to design and implement a 20-line experimental electronic telephone exchange within the Research Labs, completed by 1963. His publication (Wragge, 1963) describes the performance of a hard-

http://doi.org/10.18080/jtde.v11n3.787

wired electronically controlled and switched PABX, which his team had built and installed at the 10 Lonsdale St Annex of the PMG Research Laboratories. It handled real telephone traffic directed to or from it, from anywhere in the public switched telephone network.

Integrating Switching and Transmission

In the PMG's engineering organization, going back to its roots at Federation, the very different technologies required for 'switching', meaning the telephone exchanges of that era, and 'transmission', meaning the cables or radio links within the network, led to long-standing organizational separations between switching and transmission engineers in both the States and especially in Head Office. It is significant for Harry's subsequent success in integrating the two disciplines, at least within telephone exchanges, that he began his career at the PMG Research Laboratories working on the potential of electronic solutions for Voice Frequency transmission. Of these, Pulse Coded Modulation (PCM) systems were holding the greatest promise for supporting a large number of voice channels over a single transmission link.

Harry worked with Switching Planning engineer Blair Feenaghty in the late 1960s to ensure that the PMG (later Telecom Australia) adopted the European 32-channel standard for PCM systems rather than the North American 24-channel standard.

Blair Feenaghty:

> "That was quite a feat, as the transmission side of the business was very firmly wedded to the American standard. Indeed, the transmission engineers were very put out by the decision — how *dare* two switching engineers go against the wisdom of the transmission branch? But I was better at the economic analysis and Harry better at the technology, and so we won, to the ongoing benefit of Telecom and Australia" (Feenaghty, 2023).

Harry attended the second International Conference of Electronic Switching, held in Paris in 1966, followed by visits to Japanese research laboratories to check first hand on their progress with electronic telephone switches ('Our Contributors', 1967; Wragge, 1967). What he learned from this trip clearly influenced his 1968 design study, which proposed the funding of a new R&D project, the IST (integrated switching and transmission) project (Wragge, 1968). Its ultimate aim was to boost in-house expertise in preparation for the PMG's network planning, manpower planning and equipment purchasing, perhaps eight to ten years out. The key mechanism for acquiring in-depth engineering expertise would be to design and build a next-generation entirely electronic, stored program controlled, digital switch and install it in a live local telephone network, to demonstrate its viability.

The IST project was to become the single project most heavily identified with Harry, not just within the PMG and (from 1 July 1975) its successor Telecom Australia, but with its peer groups internationally. He recruited very capable engineers to design the hardware (Andy

http://doi.org/10.18080/jtde.v11n3.787

Domjan, Norm McLeod, Norman Gale, Michael Hunter) and specify the software (Fred Symons, Mel Ward, David King, Peter Gerrand) for this 'stored program controlled' (i.e., computer controlled) digital switch. He also recruited experienced PMG engineers from outside the Laboratories with knowledge of the current network (Greg Crew, Jim Vizard) to ensure that the model IST exchange and new signalling systems would meet operational needs when inserted as a transit switch between local telephone exchanges.

Some quotes from participants:

> "This was an extremely ambitious project, aimed at designing one of the world's first computer-controlled telephone exchanges, using digital technology. There were no textbooks available to help the designers" (Gerrand, 2007).
>
> "Our engineers and techs designed and built the system from the ground up, manufacturing circuit boards, racks and cabinets" (Crew, 2023).
>
> "The IST project was in competition with similar projects at Bell Labs in the US and the top telecommunications labs in Japan, Italy, France, Canada, Germany and the UK. And the Australian team was the first in the world to produce a computer-controlled digital switch that successfully handled live telephone traffic" (Gerrand, 2007).

The IST project was to have several outcomes. Firstly, by 1974 it achieved its objective of carrying live transit telephone calls in a local Melbourne suburban network, and stayed in situ until 1978 (Coxhill, 2007b). Secondly, working on the IST project provided several engineers with the expertise to accelerate their careers outside the laboratories into Telecom's mainstream network engineering department, and sometimes thence to very senior management positions (e.g., Mel Ward becoming Managing Director of Telecom Australia; and Greg Crew and Bill Craig reaching the senior ranks of Hong Kong Telecom). Harry's pioneering work with both the 1970s' IST project ("a notable 'first' for Australia") and its 1960 predecessor, Harry's experimental, all electronic PABX, were acknowledged when he was inducted to the Pearcey Hall of Fame (Pearcey Foundation, 2009).

Greg Crew:

> "They were happy and productive times, and Harry used all his many political skills to ensure the IST project was well supported, despite some opposition from the Engineering Department. [....] The IST switch was eventually moved to Windsor exchange and linked into the telephone network. By then I had joined HK Tel. and some years later Bill Craig also joined HKT. Eventually the IST switch was de-commissioned, around 1990, and the final phone connection was to Bill and me in Hong Kong, so we could join the celebration" (Crew, 2023).

Harry's expertise on future digital networks was enhanced by his participation from 1969 to 1981 in CCITT Special Study Group D meetings in Geneva. The CCITT Study Groups were then, as now, meetings of technical experts from the ITU's members: in that era largely PTT (government-owned postal, telegraph and telephone) administrations and monopoly private

http://doi.org/10.18080/jtde.v11n3.787

carriers, such as AT&T in the USA, together with experts from the major telecommunications manufacturers.

The CCITT Special Study Group D, for which Harry was Vice Chairman in 1976–1980, was a forum for sharing expertise on digital transmission, switching and signalling technologies. It provided him with the ability to gain an authoritative knowledge of world developments to take back to Telecom Australia, as well as providing inputs to Telecom Research Labs' (TRL's) own research. His visits to peer laboratories enabled him to build friendships with research directors in major organizations across Europe, Japan and North America (Wragge, Wragge & Wragge, 2023).

Promoting Australian Telecommunications Research

Meanwhile Harry was active in promoting Australian telecommunications research, as founding Editor-in-Chief (1967–1981) of a new journal, *Australian Telecommunications Research*, abbreviated to ATR. An initiative of the PMG Research Labs' fifth Director, Rollo Brett, ATR was established to promote Australian telecommunications research in general. Unlike other 'house journals', ATR was actively inclusive, inviting papers from universities, the CSIRO and manufacturing laboratories in Australia – and sometimes overseas. Volume 1 of ATR included papers by young researchers Mel Ward (later Managing Director of Telecom Australia) and J. L. (Jonathan) Parapak (later Secretary General of the Indonesian PTT). ATR folded in 1995, after the effects of industry competition within Australia in the early 1990s, together with career pressure on Australian academic researchers to publish in international journals, dried up the source of research papers for the journal (Gerrand, 1996).

The Analogue versus Digital Switching Controversy in the 1970s

In 1974, the PMG had installed its first SPC (stored program controlled) exchange, the Metaconta 10C exchange from BTM (Moyal, 1984, p. 318). This was carefully chosen to be installed deep within the network, as a trunk (transit) exchange in Sydney, to avoid any downside for customers if teething troubles arose. Once the 10C exchange had proven its value, the planners were keen to exploit the versatility and cost savings of both SPC and electronic switching across the network, beginning with local exchanges. The question was, should they make a small step by employing SPC analogue switching, or a bigger step by using SPC with digital switching, for which integrated circuits were proving a boon through miniaturisation of components?

Needless to say, Harry, having proven some of the advantages of SPC digital switching with his experimental IST exchange, and being well across world trends, was active with

http://doi.org/10.18080/jtde.v11n3.787

submissions and talks to Head Office network engineering staff, advocating the move to digital. In particular, he stressed the advantages of greater reliability and smaller bulk, leading to significant reduced accommodation, asset and operational costs. But, unlike the earlier success over the choice of European 32-channel PCM systems, he found his ideas blocked by a very conservative General Manager, Engineering, whom I shall call Mr X.

Mr X went so far as to cancel Harry's attendance at one of the International Switching Symposiums, at which he had been invited to give a keynote address. To the stupefaction of many of the delegates, Mr X used the keynote address to assure the conference that he foresaw no need for digital switching in the foreseeable future. For some years after this event, Telecom delegates to CCITT meetings were asked by amused representatives from Bell Labs and elsewhere as to whether Mr X had changed his mind.

In the meantime, Telecom Australia decided to invite tenders for an SPC switch for its local and tandem networks. Mr X's subordinates persuaded him to include in the specification an option to offer either digital or analogue switching after the first purchase. Ericsson's AXE exchange was selected. But the first exchange installed, required by Mr X's specification to be analogue, ended up a lonely orphan in the Telecom network. Fortunately, the flexible tender specification enabled Telecom to buy all subsequent AXE exchanges from Ericsson as digital switches. Harry was on the right side of history.

Researching Customer Needs

From 1979 to 1981, Harry moved sideways to head the Customer Systems and Facilities branch at TRL. This branch was unusual in employing psychologists and a geographer, as well as the usual mix of research engineers and scientists, with supporting technicians. It also provided Harry with another opportunity to collaborate with Blair Feenaghty, then head of Product Management in Head Office. They agreed that it was time to conduct some field experiments to establish the feasibility of providing a greatly increased suite of services for Telecom's customers, and to test their attractiveness in reality, as opposed to the results of customer surveys.

Feenaghty:

> "Hence the project known as FINCS: Field Investigation of New Customer Services. And so we became the 'Head Fincs' — Harry must have been an aficionado of The Wizard of Id! It was a good project, which provided some useful early information on what services would be possible and desirable" (Feenaghty, 2023).

http://doi.org/10.18080/jtde.v11n3.787

Responding to the Davidson Inquiry on Telecom's Future

From the 1970s, political pressure had built within the USA to break up the private sector monopoly of AT&T, which included both the long-distance networks and the regional local telephone companies. In 1984, AT&T divested itself of the Regional Bell Operating Companies. Inevitably, similar pressures arose in Australia, spearheaded by media mogul Kerry Packer (Moyal, 1984, pp. 338–339, 352, 379), as well as lobbying of the federal government by merchant bankers looking for the profits flowing from privatisation.

In response, the then Fraser Government set up a Commission of Inquiry, known as the Davidson Inquiry (1981–1982), to make recommendations on the extent to which the private sector could be more widely involved in telecommunications services in Australia — and hence the future of Telecom (Moyal, 1984, p. 380). Telecom put together a team of two experienced executives, George Hams[2] and Harry Wragge, together with a younger executive, Ken Loughnan, from Corporate Strategy, to produce Telecom's submissions and serve as the prime interface with the Davidson Inquiry.

Ken Loughnan:

"[I] have fond memories of working closely with both Harry and George Hams as the team in developing the Telecom response to the Davidson Inquiry in 1981–1982. Harry almost always had an alternate view, although I suspect that was partly his way of encouraging a 'young buck' to get to the bottom of an argument — his favourite phrase — always calmly delivered — 'well that's not necessarily so'. I learnt a lot from Harry…" (Loughnan, 2023).

The election of the ALP Hawke government in 1983 effectively killed off the Davidson recommendations.[3] However, Telecom's senior management had good reason to believe that increased competition and perhaps eventual privatization of the organization had only been deferred. Hence, much effort was expended by Telecom's senior management from 1985 onwards in preparing the organization for the introduction of competition.

Harry continued in Corporate Strategy as Assistant Director, Business Development until May 1985, when he returned to TRL as its new director, following Ed Sandbach's retirement.

Director of the Telecom Research Labs (1985–1992)

TRL under Harry, as under his predecessors, served the larger organization through both problem-solving and the transfer of expertise on new technologies. During Harry's time as director, technology transfer took place on the key technologies of optical fibre transmission, terrestrial and satellite radio technologies, ISDN switching, the Digital Radio Concentrator

http://doi.org/10.18080/jtde.v11n3.787

System (invented at TRL and deployed widely in the outback), geographical information systems, directory systems and packet switching, amongst others.

Sometimes the technology transfer took the form of TRL staff. Ian Campbell, Executive General Manager of Special Business Products in the mid-1980s, writes:

> "As Director, Research, Harry was a strong supporter within Telecom to me in Special Business Products, for the development, launch and deployment of mobile services. I knew nothing about the technologies and Harry was a valued sounding board. He donated radio systems expert, Dr Reg Coutts, then Head of Radio and Satellite Networks at TRL, from TRL to Mobiles in 1989" (Campbell, 2023).

Reg's first job was to advise on which new digital radio technology Telecom should recommend to the regulator for implementation, for the introduction of mobile competition in 1991 (Gerrand, 2021).

In 1989, Harry was made a Member of the Order of Australia for his services to telecommunications technology. His significant contributions to telecommunications research in Australian universities will be discussed below.

In 1991, the Australian Government merged Telecom with the Sydney-based Overseas Telecommunications Commission (OTC) as the Australian and Overseas Telecommunications Corporation (AOTC) in preparation for the introduction, in 1991, of competition in the long distance and mobile markets. The smaller OTC was the clear winner in this merger, with its Chairman, David Hoare, becoming chair of the combined entity AOTC, and many of OTC's senior executives appointed to head some of the former Telecom's major business units. In 1992, an American from AT&T, Frank Blount, was appointed to head AOTC. He rapidly appointed some of his former colleagues to fill the top marketing and information systems positions. AOTC traded within Australia as Telecom Australia and later became Telstra. On his retirement from Telstra, Blount boasted at having replaced Telecom's entire senior management team[4] (Ries, 1999).

Telecom's Managing Director, Mel Ward,[5] and Head of Corporate Strategy, Terry Cutler,[6] had seen the writing on the wall, and left the company before the merger. The days of the rest of Telecom's top management team, including Harry, were numbered. Often, a senior executive was moved sideways before being 'let go'.

As part of this chess game, Harry found himself transferred to Corporate Centre as Chief Technical Adviser to the CEO. It was, in this author's opinion, a face-saver for Harry but a job with little influence. Harry undoubtedly thought the same, because he decided to retire from Telstra in early 1993, aged only 63.

http://doi.org/10.18080/jtde.v11n3.787

But further honours were to flow to him in retirement: an Honorary Doctor of Engineering degree from Monash University in 2000; and a Centenary Medal from the Australian Government in 2001. In 2009, he was inducted into the Pearcey Foundation's Hall of Fame (Figure 4), in recognition of his outstanding services to the ICT (information and communication technology) sector (Pearcey, 2009).

Harry Wragge with Senator Stephen Conroy

Figure 4. Harry Wragge inducted into the Pearcey Hall of Fame by Senator Stephen Conroy (Pearcey, 2009)

Assistance to Australian Universities

Ever since graduating, Harry had maintained close ties to his alma mater, the University of Melbourne. He had lectured part-time in electronics there in 1965 (Packer, 1997, p. 35), and served as an external member of various advisory boards for over twenty years, including being President of Convocation in 1990–1991. In addition, he played a valuable occasional role in the accreditation of engineering courses around Australia as a member of the Institution of Engineers Australia's Accreditation Board – although conflict-of-interest protocols prevented him from reviewing the courses within the State of Victoria.

In 1987, the Williams Commission Report into engineering education in Australia was published. It provided a devastating critique of the Department of Electrical and Electronic Engineering (EEE) at the University of Melbourne, being critical of the department's poor performance in research and teaching, and the poor progression rates of students, given the quality of the intake of first-year students. In 1988, the new Vice Chancellor, Professor David Penington, intervened. He set up a Review Committee with an independent chair, two professors of Electrical Engineering from universities in NSW, two professors from faculties

http://doi.org/10.18080/jtde.v11n3.787

other than engineering, and two industry representatives. One of these was Harry Wragge, then head of TRL (Packer, 1997, pp. 34–35).

The final recommendations "sent shock waves through the rest of the University". One of the two EEE professors was sacked, and the other, perhaps unfairly, was required to spend at least half his time outside the department on industry consultancy (Packer, 1997, p. 35). However, in 1988, the Engineering Faculty learned that Rod Tucker, leading a research group in the new field of photonics at Bell Laboratories, was keen to return to Australia. With the help of Harry Wragge, David Penington was successful in attracting Tucker to head up the EEE department "with the promise of sufficient funding to set up a world class photonics laboratory" (Packer, 1997, p. 38).

Rod Tucker took up the positions of Professor and Head of EEE at the beginning of 1990. He redesigned the course, with strong support from industry leaders, reorganized the department, and started setting up the new photonics laboratory and recruiting new research and teaching staff, with funds from both the Vice Chancellor and Telecom Australia (Packer, 1997, pp. 41–42). Tucker and his fellow professors were able to transform the research and teaching calibre of the department to an outstanding level that has continued to the present day. Without Harry's crucial funding support back in 1990, Tucker "would not have returned to Australia and the University of Melbourne" (Tucker, 2023), and the revival of the EEE department would not have happened so quickly or so brilliantly.

It was fitting that the University of Melbourne's top engineering honour, the Kernot Medal, awarded annually since 1926 for "Distinguished Engineering Achievement in Australia", was conferred on Harry Wragge (Figure 1) in 1990 ('Champion of the cause', 1990).

Harry provided significant support to other Australian universities. From March 1987, he signed a five-year major research contract to establish the Teletraffic Research Centre at the University of Adelaide ('Teletraffic Research Centre', 1987). The Centre's commercial relationship with Telstra has continued through to the present ('Telstra', 2021).

From 1987, he provided seed funding via a research contract, and assistance from expert TRL staff, to Professor John Hullett's research team at Curtin University to take their QPSX queued packet-switching patent through to an industry prototype stage, and persuaded Telstra's investment arm to invest in it (Roberts, 1991). QPSX was floated on the Australian Stock exchange in December 2000 (Bolt, 2000).

As Director of TRL, co-located close to Monash University, he took initiatives to encourage closer cooperation. These included providing interim funding for Dr Fred Symons from TRL to take up a professorial role in telecommunications engineering in Monash's Department of Electrical and Computer Engineering from 1989 until Fred's retirement in 1996 (Gerrand,

http://doi.org/10.18080/jtde.v11n3.787

2007). Monash University awarded Harry an Honorary Doctor of Engineering in 2000 for his services to telecommunications research and education.

Harry as Family Man and Sailor

Harry married Shirley Ogilvie in 1957 at Scotch College Chapel. They bought land in Seaford and decided to build their own home there.

Figure 5. Harry Wragge yachting in the 1980s (courtesy of the Wragge family)

According to Harry's daughters, Sue, Jennie and Kate:

"The achievements that were most important to Dad were building the house, camping trips, and building our dinghies. Dad built the house with Mum, and they moved in with planks on floor joists, noggins as shelves in the kitchen and finished the house together. Together they created a place Mum never wanted to leave.

"One of the biggest things Dad did was get us involved in sailing. Our first season as a Family Member of Frankston Yacht Club was 1970-71. Dad built a Puffin Pacer from a plywood kit, and after Sunday morning cadets at Frankston he headed to Carrum where *he* sailed in the afternoon. The next winter Dad and some other fathers got together to build six Sabot dinghies at our place. *Inside* the house." (Wragge, Wragge & Wragge, 2023)

Harry was a member of the Frankston Yacht Club for fifty years. He was elected Commodore from 1978 to 1981, and later made a Life Member. The annual Open Harry Wragge Trophy Handicap race is named in his honour. In 2009, Harry was inducted into the City of Frankston's Hall of Fame as a local hero.

The Kindness of the Man

Blair Feenaghty:

> "[P]erhaps my most enduring memory of Harry is of the moral support he provided me several times over my career. When I arrived in Melbourne as a very green young engineer to pursue research for an MEngSc degree, he was welcoming and supportive beyond expectation.
>
> "He was quick to recruit people to support his professional interests. At his urging I agreed to become Secretary of the Electrical Branch of the Victorian Division of Engineers Australia, and later Chairman, something it would have been rather unlikely for me to do without his encouragement and support.
>
> "Later, on my first visit to Geneva with the Australian delegation, Harry took me under his wing to demonstrate the marvels of the Swiss railway system, how to secure a bed in Zermatt, and the glories of the Matterhorn. [...]
>
> "I am delighted that even in the last years of his life, when I met him at Shirley's funeral, Harry remembered that day in 1961 when we first met. A great and good man whose

http://doi.org/10.18080/jtde.v11n3.787

> memory will live with those of us lucky enough to have had him as part of our lives" (Feenaghty, 2023).

And, as the author of this obituary, I have been motivated to research it and write it, not just to record Harry's considerable accomplishments, but also to return a favour to someone who was very helpful to my own career. In the 1970s when he headed Switching and Signalling Branch, he gave two promising young engineers the opportunity to gain broader industrial and overseas experience by taking two years' leave without pay from the PMG, with our jobs guaranteed on our return. David King and I were able to spend two years working in the R&D laboratories of major manufacturing companies: David in Paris; and myself in Madrid. We not only gained valuable industrial experience, but also the long-lasting value of cultural immersion and second-language acquisition, which have enriched the lives of ourselves and our partners.

Conclusions

Harry Wragge was both a leading Australian research engineer and a major catalyst for the conversion of Australia's analogue, electromechanical telephony network of the 1950s and 1960s to the cheaper, more reliable and more versatile digital circuit-switched networks of the 1980s and 1990s. He was one of the most influential research leaders within Telecom Research Laboratories, both before and during his time as the Director of TRL (1985–1992).

He also played a vital role in supporting telecommunications research in several Australian universities. A first instance was his funding (initially with OTC) of the Teletraffic Research Centre at the University of Adelaide from 1986. A second was his key role in the revival of the EEE department at the University of Melbourne in 1989, enabling it through strategic funding support to become an eminent centre for phonics research under Professor Rod Tucker's leadership. A third was his strategic support for the QPSX project at Curtin University via a research grant in the late 1980s, helping it on its way to its float on the ASX in 2000.

In addition to being a beloved family man and a local hero at the Frankston Yachting Club, Harry's death this year at the ripe old age of 93 has evoked many testimonials from former colleagues, relating his kindnesses as a mentor and friend.

Acknowledgements

The author is grateful for the insightful information on Harry's life provided by his sister Anne Hill, his daughters Sue, Jennie and Kate Wragge, his nephew Chris Stewart, and his life-long friends Mac Cleland and John Cathcart.

On Harry's times as an undergraduate and postgraduate, I am grateful for access to historical documents provided by Dr Richard Gillespie, the Curator of the Faculty of Engineering and IT

at the University of Melbourne; and to my former MEngSc (Monash) supervisor, Dr W. A. (Bill) Brown; and equally to Emeritus Professor Rod Tucker AM for his recollections of Harry's key role in supporting the University of Melbourne's EEE department in 1989 and beyond.

On Harry's career at the PMG, Telecom and Telstra, I am grateful to have had my own first-hand knowledge extended by Harry's colleagues Ian Campbell, Greg Crew, Blair Feenaghty, Dr Jim Holmes, Ray Liggett and Ken Loughnan AO.

References

Black, S. (2023). George Edward Hams AM (1928–2023): A leader amongst Australian telecommunications engineers. *Journal of Telecommunications and the Digital Economy*, *11*(2), 252–261. https://doi.org/10.18080/jtde.v11n2.739

Bolt, C. (2000). Packer backs QPSX float. *Australian Financial Review*, December 8, 2000. Available at https://www.afr.com/politics/packer-backs-qpsx-float-20001208-k9v4b

Brown, W. A. (Bill). (2023). Email to Peter Gerrand, 4 August 2023.

Burry, M., Healy, T., & Spurling, T. (2020). A passionate contributor publicly and privately. *Sydney Morning Herald*, September 4, 2020. Available at https://www.smh.com.au/national/a-passionate-contributor-publicly-and-privately-20200904-p55set.html

Campbell, I. (2023). Email to Jim Holmes, 8 August 2023.

'Champion of the cause'. (1990). Kernot Memorial Medal: Harry Wragge. *Journal of the Institution of Engineers, Australia*, *62*(24), 38.

Cleland, M. (2023). Harry's Life Story: Eulogy for Harry Wragge, 8 August 2023.

Coxhill, R. (2007a). 'History of the Telecom Research Laboratories' at http://www.coxhill.com/trlhistory/

Coxhill, R. (2007b). 'Harry S Wragge ([at TRL from] 1985–Dec 1992)' at http://www.coxhill.com/trlhistory/history/wragge.htm

Cranks & Nuts. (1951). [Journal of the MUESC]. Available from the Curator, Faculty of Engineering and Information Technology, University of Melbourne.

Cranks & Nuts. (1953). [Journal of the MUESC] Club Committee, p. 20; MUESC Secretary's Report, p. 50; 4th Year Electrical Notes, p. 54; Rogues Gallery entry "Wragge, Harry S" within 4th Year Electrical Engineering, p. 66. Available from the Curator, Faculty of Engineering and Information Technology, University of Melbourne.

Cranks & Nuts. (1954). [Journal of the MUESC] Engineering School Exhibition, p. 27. Available from the Curator, Faculty of Engineering and Information Technology, University of Melbourne.

Crew, G. (2023). Email to Peter Gerrand, 10 August 2023.

Feenaghty, B. (2023). Email to Peter Gerrand, 9 August 2023.

Gerrand, P. (1996). Adios ATR. *Telecommunication Journal of Australia*, *46*(2), 34–36.

Gerrand, P. (2007). Fred Symons – Highlights of his Career. Available at http://www.coxhill.com/trlhistory/miscell/PeteraboutFred.pdf

Gerrand, P. (2021). Emeritus Professor Reginald Paul (Reg) Coutts (1949–2021). *Journal of Telecommunications and the Digital Economy*, *9*(3), 186–193. https://doi.org/10.18080/jtde.v9n3.448

Hill, A. (2023). 'Thoughts on Harry from Anne', a tribute to Harry Wragge from his sister, 8 August 2023.

Jellie, D. (2010). Pioneer in Communications: Melvyn Keith Ward, AO, Engineer, Businessman. *Sydney Morning Herald*, November 12, 2010. Available at https://www.smh.com.au/national/pioneer-in-communications-20101111-17pew.html

Loughnan, K. (2023). Email to Jim Holmes, 7 August 2023.

Moyal, A. (1984). *Clear Across Australia. A History of Telecommunications*. Melbourne: Nelson.

'Our Contributors'. (1967). H. S. Wragge. *Telecommunication Journal of Australia*, *17*(2), 171.

Packer, J. (1997). An official history of the Department of Electrical and Electronic Engineering. University of Melbourne: Department of Electrical and Electronic Engineering. Especially chapters 'Crisis' (pp. 33–36), 'Difficult Times' (pp. 37–39) and 'Renewal' (pp. 41–46).

Pearcey Foundation. (2009). 2009 Pearcey Hall of Fame: Harry Wragge AM. Available at https://www.pearcey.org.au/awards/national/pearcey-hall-of-fame/2009-pearcey-hall-of-fame/ (updated July 2023).

Ries, I. (1999). Farewell to Blount and 'eight years of triumph'. *Australian Financial Review*, February 17, 1999. Available at https://www.afr.com/politics/farewell-to-blount-and-eight-years-of-triumph-19990217-jladd

Roberts, P. (1991). QPSX signs up first overseas customer. *Australian Financial Review*, August 5, 1991. Available at https://www.afr.com/companies/qpsx-signs-up-first-overseas-customer-19910805-k4jam

'Teletraffic Research Centre at the University of Adelaide'. (1987). Information Transfer News, *Telecommunication Journal of Australia*, *37*(1), 68.

'Telstra'. (2021). Teletraffic Research Centre Case Study. Available at https://set.adelaide.edu.au/teletraffic-research-centre/case-studies/telstra

Tucker, R. (2023). Email to Peter Gerrand, 7 August 2023.

Wion, F. W. (1963). Transistor Circuits for Generation of Exchange Service Tones and Bell Ringing Current. *Telecommunication Journal of Australia*, *13*(6), 490–493.

Wragge, H. S. (2004). An extract supplied by Kate Wragge in August 2023 from a private memoir written by Harry Wragge for his family in 2004. Unpublished.

Wragge, S., Wragge, J., & Wragge, K. (2023). Tribute to Harry Wragge from his daughters, 8 August 2023.

Harry Wragge's publications

Wragge, H. S. (1960). The Design of Transistor Circuits. *Telecommunication Journal of Australia*, *12*(3), 151–163.

Wragge, H. S. (1961). Electronic Telephone Exchanges. *Telecommunication Journal of Australia*, *13*(1), 14–18.

Wragge, H. S., & Wion, F. W. (1962). An improved Transistor Ring and Tone Generator. *PMG Department Research Laboratory Report*, No. 5510. (quoted in Wion, 1963, p. 493).

Wragge, H. S. (1963). 20-Line Experimental Electronic Telephone Exchange. *Telecommunication Journal of Australia*, *13*(6), 479–451.

Wragge, H. S. (1967). Recent Developments in Electronic Telephone Exchanges. *Telecommunication Journal of Australia*, *17*(2), 84–89.

Wragge, H. S. (1968). Design Study – Proposed IST Project. *PMG Research Laboratory Report*, No. 6151, October 1968.

Wragge, H. S. (1969). Experimental PCM Switching System in the Melbourne Telephone Network. Proceedings, Conference on Switching Techniques for Telecommunications Networks, London, pp. 161–164.

Wragge, H. S. (2013). Obituary: John Henry (Jack) Curtis CB BE(Hons) BSc BA 1920–2013. *Journal of Telecommunications and the Digital Economy*. *1*(1), 11.1–11.3. https://doi.org/10.18080/jtde.v1n1.215

Endnotes

[1] Some decades later, when Bill Brown was a Professor at Monash University, he joined Harry in the accreditation of engineering courses in interstate universities on behalf of the Institution of Engineers.

[2] George Hams' obituary can be found at Black (2023).

[3] The previous Coalition government was unable to implement the Davison recommendations. Despite the Liberal Party's enthusiasm for privatisation, its Coalition ally, the National Country Party, also rejected the Davidson Inquiry's recommendations for introducing competition to Telecom (Moyal, 1984, p. 383)

[4] The sole exception was a Canadian, Doug Campbell, who had joined Telecom only 18 months before as Deputy Managing Director and was not seen as part of the former Telecom culture.

[5] Mel Ward's impressive career, including references to both his participation in the IST project at TRL and to his resistance, as MD, to a potential 'carve up of Telecom', is summarised in Jellie (2010).

[6] Dr Terry Cutler's brilliant career is well described by Burry, Healy & Spurling (2020).