International Journal of Computer Applications Technology and Research
Volume 13–Issue 01, 01 – 06, 2024, ISSN:-2319–8656
DOI:10.7753/IJCATR1301.1001

Mitigating Factors Affecting Secure Interoperability of Medical Systems Using DLTs in Healthcare

Dorothy G. Bundi
Department of Computer Science& Information Technology
Kabarak University
P.O Private Bag 20157,
Kabarak Nakuru, Kenya

Stephen M. Mutua
Department of Computer Science
Meru University of Science and Technology
P.O BOX 972-60200 Meru, Kenya

Simon M. Karume
Department of Computer Science& Information Technology
Kabarak University
P.O Private Bag 20157,
Kabarak Nakuru, Kenya

Abstract: The need for more people in the world to connect with one another via use of networked computerized distributed information systems is on the rise in different sector as well as in the medical sector. With many medical information systems being complex and private owned, networking such systems to aid interoperability in order to allow secure sharing of the electronic medical records remains a challenge. This calls for secure connections of different medical system platforms that will aid easy and timely sharing of electronic medical records across different medical facilities. Distributed ledger technologies such as enhanced blockchain is one of the such technologies that when implemented in the healthcare sector have ability to support secure sharing of electronic medical records. The study used exploratory and a survey-based descriptive research design. Information was gathered through both a literature review and a questionnaire survey involving a sample of twenty (20) companies specializing in the development of medical systems software. For this survey, two (2) domain experts from each company were purposefully selected as respondents, totaling forty (40) respondents. The response rate was substantial, with seventeen (17) companies participating, contributing a total of thirty-four (34) domain experts, representing an 85% response rate. The aim of the study was to explore the factors that are hindering secure interoperability and sharing of electronic medical records across different medical systems. The findings revealed that technical factors like data formats, syntax, organization and protocols are the factors affecting structural interoperability levels while data meaning, models codification schemes and data definition standardization are the factors affecting semantic interoperability. Other factors include financial, organizational, human, cultural, security and privacy. The study proposes integration of Distributed Ledger Technologies (DLTs) into the medical systems to mitigate the factors that affect secure interoperability of medical systems and to enhance secure sharing of electronic medical records (EMRs) across medical systems.

Keywords: Interoperability, Structural interoperability, Semantic interoperability, Medical System, Distributed Ledger Technologies, Healthcare, Blockchain

1. INTRODUCTION

Information communication technology (ICT) is critical and valuable to health sector in our society. ICT systems support medical processes by storing, processing, and communicating critical and sensitive data and information [1]. Distributed ledger technologies (DLTs) like Blockchain technology have been penetrating every aspect of ICT and its use has been growing rapidly in recent years with the use of cryptocurrencies in the finance sector [2].

In developing countries, the adoption of ICT in healthcare has proliferated over the years and continues to increase [3], however the health sector has not been fully automated to use eHealth or medical systems some health institutions are still using the manual processes [4]. According to [5] some of the major challenges that affect the full use and hinder full potential of eHealth or medical information systems in the healthcare industry include: the fragmented patients data that is distributed in different hospitals databases across various healthcare facilities, inconsistent patients health or medical records which makes it difficult to track, access and manage patients data, untimely access to patients records, lack of medical systems interoperability and lack of data security in transmission of patient data across different medical systems platforms and geographies which compromises its privacy and security . These eHealth information systems challenges can be solved by use of distributed ledger technology which will allow the sharing of patients' data, electronic medical records and information across different eHealth and medical systems platforms and medical facilities [6].

Distributed ledger or a shared ledger or distributed ledger technology (DLT) is a technological infrastructure and protocols that allows users to simultaneous access, record, validate, share, and synchronize data and transactions updating across a networked database in a distributed network consisting of numerous participants [7]. It can also be understood as a range of technologies with comparable structures but can be executed in various ways with different rules. DLT uses cryptography to securely store data, cryptographic signatures and keys to allow access only to authorized users. The technology also creates an immutable database, which means information, once stored, cannot be deleted and any updates are permanently recorded for posterity [8]. These unique features of DLTs make them suitable for the applications in the healthcare sector.

Healthcare Information and Management Systems Society [9] defines interoperability as the ability of different information systems, devices and applications (systems) to access, exchange, integrate and cooperatively use data in a coordinated manner, within and across organizational, regional and national boundaries, to provide timely and seamless portability of information and optimize the health of individuals and populations globally. Medical data exchange system architectures, application interfaces and standards should be designed in a manner that enable data to be accessed and shared securely across different healthcare facilities despite their medical enterprise system platforms [10].

Interoperability of medical systems is categorized into four levels namely: Foundational which is Level 1: this level establishes the inter-connectivity requirements needed for one system or application to securely communicate data to and receive data from another. Structural which is Level 2: which defines the format, syntax and organization of data exchange including at the data field level for interpretation. Semantic Level 3: this level provides for common underlying models and codification of the data including the use of data elements with standardized definitions from publicly available value sets and coding vocabularies, providing shared understanding and meaning to the user. Lastly, Organizational level which is Level 4: this level includes governance, policy, social, legal and organizational considerations to facilitate the secure, seamless and timely communication and use of data both within and between organizations, entities and individuals. These components enable shared consent, trust and integrated end-user processes and workflows [9].

Interoperability of information systems has evolved over the years, starting with the use of middleware in web services using technologies like firewall and protocols like hypertext transfer protocol (HTTPs) to support sharing of electronic medical records via the web across different health facilities located in different geographical areas but this is faced with a challenge since this type of web configurations inhibits smooth communication of different middleware making interoperability impossible [11]. Other web systems use XML and JSON as marshalling technology for packaging parameters to be communicated over the internet in a technology neutral format [12]. These technologies have still not fully address the structural and semantic interoperability levels which remains unsolved due to use of distinct data formats, protocols and standards which still remains to be software platform and vendor dependent [13]. This paper suggests the use of distributed ledger technology (DLT) based systems to solve the challenges of structural and semantic interoperability levels of medical systems.

Using DLTs different medical systems, devices and applications can securely access, exchange, integrate and cooperatively use medical data in the process of coordinating and organizing electronic medical records (EMRs). DLTs supported medical system interoperability will aid different medical systems and medical devices from different vendors and manufacturers to securely share and exchange electronic medical records between applications, databases and other computer information systems.

2. MATERIALS AND METHODS

This study applied survey-based descriptive and exploratory research design. Exploratory research was carried out through reviewing existing literature on factors affecting secure interoperability of medical systems in the healthcare sector that was published between the periods of (2017 - 2023) years. The study cited the factors that affect secure interoperability of medical systems at structural and semantic interoperability levels. A survey-based descriptive research design was employed to gather information from domain experts, specifically medical system software developers in Kenya. Forty (40) questionnaires were distributed to twenty (20) medical systems software development companies in Kenya and subsequently Thirty-four (34) responded by filling and returning the questionnaire, providing data from two experts in each of the seventeen (17) out of the twenty (20) purposive sampled medical system software development companies in Kenya, which was 85% response rate. The subsequent sections show the steps and process that followed during the review of existing literature.

I. Research Questions Addressed

RQ: What are the factors that are affecting secure interoperability of medical systems at structural and semantic interoperability levels?

II. Inclusion and Exclusion Criteria

This literature review only includes research that address the issue of interoperability of medical system with a focus on structural and semantic interoperability levels. Additionally, studies on the application of DLTs by the medical systems in healthcare sector and the studies from the years 2017 to 2023 are the ones included for the review. Review type research, discussions, uses and applications of DLTs in other sectors, non-relevant publications and any work that are not empirical are excluded.

III. Data Sources

The literature review included the review of ten electronic databases and electronic libraries. The libraries reviewed include; IEEE Xplore, Google Scholar, PubMed – NCBI, Elsevier Science Direct, Mendeley, PNAS, Springer link, Web of Science (WoS), Medline EBSCO, and ACM Digital Library.

The researcher conducted the advanced search for the relevant publications from the electronic libraries and databases using the query string(s) defined below:

(Distributed ledger OR Distributed Ledger Technologies OR “DLTs”) AND (medical systems OR healthcare OR eHealth OR e-health OR health* OR health systems* OR medical information systems OR *health information systems* OR medical*)

The researcher constructed the search string based on the research domain and the defined research question.

Due to a lack of advanced search options for some libraries and databases like Google Scholar, Mendeley, PNAS and Springer Link, they returned many non-related results that were not meeting the inclusion - exclusion criteria. Therefore, the researcher only included the first 100 most relevant results from these four databases. This search in the online digital libraries was conducted in January 2023. The researcher intentionally made the search query as broad as possible in order to consider as many results related to the systematic research questions as possible. The summary of the search in all databases and libraries returned 4777 results and the results returned for each database search are presented in Table 1.

Table 1 Summary of Search Results

Database / Library	**Number of Results**	**Number of Suitable results after detailed screening**
IEEE Xplore	17	10
Google Scholar	3562(100)	12
PubMed – NCBI	30	5
Elsevier Science Direct	18	8

Mendeley	167(100)	7
PNAS	202(100)	2
Springer link	745 (100)	1
Web of Science (WoS)	10	2
Medline EBSCO	20	4
ACM Digital Library	6	1

IV. Selection of Studies

The selection process started with 501 publications gathered from online digital databases and digital libraries. Based on the inclusion-exclusion criteria, the publications were either included in the review or not and a total of 52 papers were reviewed. The researcher was interested in how the distributed ledger technology (DLT) is used in providing structural and semantic interoperability of medical systems in the healthcare sector and finding out what are the factors that are affecting secure interoperability of medical systems at structural and semantic interoperability levels. Later the researcher suggests the use and integration of DLTs to mitigate the challenges identified.

3. DISCUSSION

The study revealed that today, most healthcare organizations have adopted electronic medical records (EMR) technology. A decade ago, EMR adoption in hospitals hovered around 73%. Now, roughly 98% of hospitals are using a government-certified EMR. While the increased adoption is a step toward achieving interoperability, it also reveals a new challenge. There are hundreds of EMR systems on the market today, each with its own unique set of technical specifications[14].

Different medical systems used by different health facilities use different data formats, specifications, and semantics, further fragmenting patient information and complicating health information exchange. Due to the varying data standards, former attempts to promote interoperability have been ineffective. For example, electronic medical records (EMRs) - a primary source of healthcare data - produce disparate and non-standardized data, making it difficult to access, share and analyze patient information across systems [15].

The findings indicate that distribute ledger technology research in healthcare is increasing and it is mostly used for data sharing, managing health records and access control [16]. The findings indicated that 78% of the most commonly used DLT in the medical sector is Blockchain. This is used with aim to provide security and privacy of electronic medical records.

The findings further revealed that the most challenges related to interoperability of medical systems are financial costs at 74% of the revealed articles, Technical challenges which includes the system designs, data structures and architectural accounted for 48% of these challenges. The findings further shown that 31% was due to identifying and implementing standards. Unrealistic end user expectations accounted for 26% and patients matching 21%. The results are shown in the figure 1.

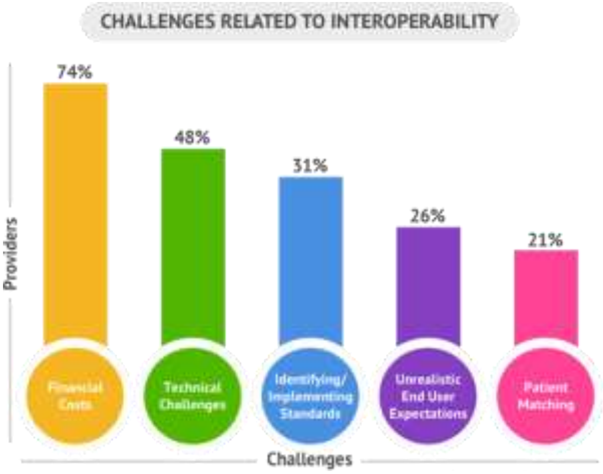
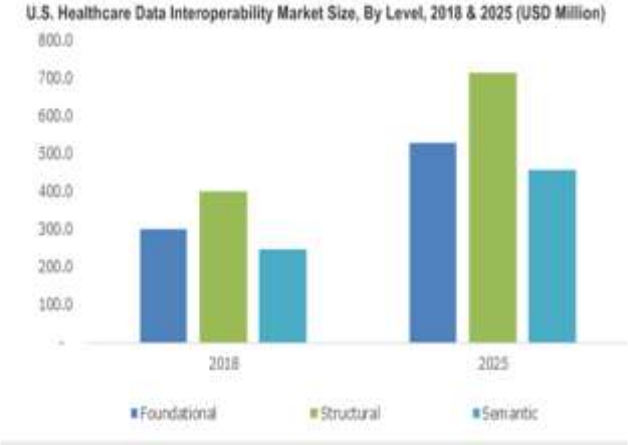


Figure 1: Challenges related to interoperability

Additionally, 55% of the reviewed articles revealed that common research problems addressed in the area of DLTs dealt with structural designs in the form of frameworks, architectures or models of Blockchain which is one of the DLT types [17]. 94% of the reviewed articles also show that technical details about the used DLT elements are not given in most of the analyzed publications and that most research does not present any prototype implementation or implementation details on medical systems and secure access and sharing of electronic medical systems [18], [19]. Often even with a prototype implementation, no details about DLT elements are given, hence the need to conduct a research on DLT prototypes with the aim of providing interoperability of medical systems. Some of the key methodologies and methods used in this area include the exploratory, descriptive and systematic literature review (SLR).

Current trends of DLT research in healthcare from the reviewed articles 82% indicate that it is mostly used for data distribution, health records and access control, but rarely for other scenarios, such as providing interoperability of medical systems that are design and developed by different vendors [20], [21]. Therefore, much potential for DLTs is still unexploited. The findings as published by Global market insights [22], show that U.S. healthcare data interoperability market size by level forecast between the year 2018 to 2025 revealed that structural and semantic interoperability level



Source : www.gminsights.com

factors are the highest contributors and deterrents of medical system interoperability as shown in Figure 2.

Figure 2: U.S. Health Data Interoperability Levels Indicators by Global Market Insights

In addition, Emergen Research [23] report also support that in the year 2021, structural and semantic interoperability levels lead in the solution in healthcare market in US billions as shown in the figure 3.

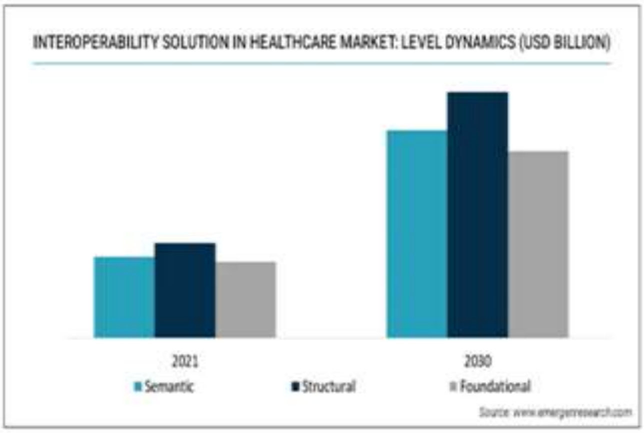


Figure 3: Interoperability Solutions in Healthcare Market by Emergen Research

The study shows that some of the challenges that hinder interoperability of medical systems include structural based factors like data formats, syntax, organization and protocols of the enterprise medical systems which are vendor and platform dependent. These factors affect the structural level of medical systems interoperability. Medical data can be inputted into medical systems in many diverse formats which includes text, numeric, string, special characters, multimedia, which is encoded to be understood by only the medical practitioners and specific to a health facility medical system. Some of these data formats are universal and others are single system based [24]. Data Syntax is defined as a set of rules defining the way in which data is put together with appropriate identifiers, delimiters, separator character(s), and other non-data characters to form messages [25].

Data organization is based on different database structures and models that are implemented by different health facilities to support their medical systems. Data in the databases can be organized and modeled in form of relational databases, hierarchical databases, network based databases, NoSQL databases and object oriented databases [26]. These data organizations and models will vary if the systems are centralized, distributed, cloud or IOTs and AI based [24]. System protocols are a set of procedures and technological measures to ensure secure and efficient operation of information within an organization [27]. These protocols determine how different medical systems are going to exchange data and manage access of the electronic medical records.

Semantic interoperability is the ability of computer systems to exchange data, with unambiguous meaning [28]. It is a requirement not only for medical data be shared between different systems or applications, but for them to be understood. Semantic interoperability refers to the transmission of the meaning of data [29]. Some of the semantic based factors that affect medical systems interoperability include data meaning, models codification schemes and data definition standardization are the factors affecting medical interoperability at the semantic interoperability level. Data codification allow system users to reduce large quantities of information into a form that can be more easily handled, especially by computer information systems [30]. In healthcare data codification needs to be done in a more systematic manner to ensure similar interpretation of the coded data and avoid misinterpretation which in turn can lead to misdiagnosis. Coded medical data stored in the medical systems are used by many entities outside the health facility for a variety of purposes including research, insurance of patients, public health, development of health policy, quality and safety monitoring patients [31]. Data Standards are information artefacts developed in community-driven consensus processes that specify uniform features, criteria, methods, processes and practices for a certain domain [32]. Healthcare standards offer health information technology (IT) developers, EMR vendors, and healthcare organizations the means to ensure medical systems and devices can exchange data successfully.

To address these structural and semantic interoperability issues, stakeholders should embrace the use of DLTs to aid secure sharing of electronic medical records. The DLT based medical system will automate workflows, minimize document errors, and, most importantly, collect, store, and deliver medical information in a way that is private, secure, and follows all industry and HIPAA protocols. Adopting health data standards in a consistent and comprehensive manner will be key to enabling meaningful healthcare interoperability at all levels. Consequently, a data architecture and data structures that works for one health facility may not work for another health facility, hence there is need also to consider a technology that will aid data interoperability of medical systems at different interoperability levels. Since data is encrypted as it is stored in different databases, integration of DLTs to aid medical systems to share data should be considered as a solution to solve the structural and semantic interoperability challenges across medical systems.

Consequently, the results from the survey based descriptive study concurred with the literature review findings. The medical system software developers indicated that some of the factors that affect interoperability of medical systems can be classified as semantic, technical, organizational, legal/regulatory, security and privacy, human, financial, and cultural aspects as shown in Figure 4.

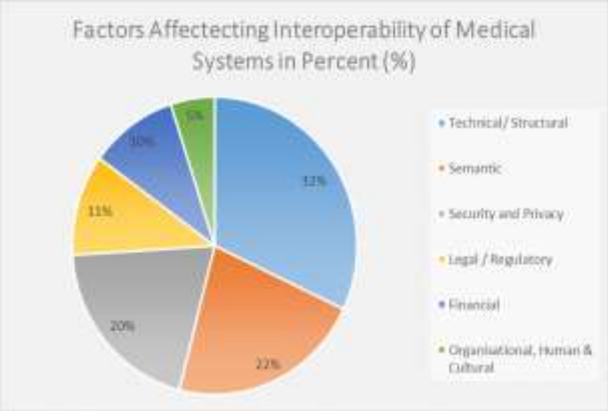


Figure 4: Categories of Factors Affecting Interoperability of Medical Systems

In this research, participants identified technical factors as the most significant. A notable 32% of respondents believe that technical aspects exert the most considerable influence on the interoperability of medical systems. This classification includes essential components like data standards, interoperability protocols, data integration, scalability, and technical infrastructure, all recognized as key contributors to the broader challenge of achieving overall medical systems interoperability.

In this study, semantic factors emerged as the second most influential considerations, with 22% of participants highlighting terminology, vocabulary, data mapping, and ontologies as primary elements. This suggests that maintaining a uniform application of medical terminology and coding systems is crucial for ensuring shared meaning of data across various systems. Additionally, the creation of mappings between diverse coding systems or vocabularies facilitates the translation of data among systems with differing terminologies. Moreover, the utilization of ontologies and knowledge graphs proves beneficial in representing intricate medical concepts and relationships, ultimately supporting semantic interoperability.

Similarly, legal and regulatory factors were recognized as another obstacle to the attainment of interoperability in medical systems, as indicated by 11% of survey respondents. Healthcare regulations, exemplified by HIPAA in the healthcare sector, impose stringent requirements regarding the storage and sharing of patient data. Regulatory bodies may lag in establishing clear standards for interoperability, leading to potential challenges. The absence of such standards can impede innovation and introduce uncertainty for healthcare organizations. Varied regulations across regions and countries add complexity to compliance, thereby obstructing the seamless sharing of data. Consequently, finding a nuanced equilibrium between ensuring compliance with these regulations and promoting interoperability becomes a critical challenge in achieving interoperability. In a similar vein, financial considerations, marked by 10% of respondents as a concern, also pose a barrier to achieving interoperability in medical systems.

Organizational, human, and cultural factors, collectively representing 5% of responses, were identified as notable influences on medical system interoperability. Organizational aspects encompass healthcare policies and regulations, necessitating compliance with standards like HIPAA and ISO, which can impact the exchange of patient data across organizations.

Cultural factors, constituting the final 5%, include resistance to change, a prevalent sentiment in the healthcare industry due to its traditionally conservative nature, potentially impeding the adoption of new technologies and interoperable systems. Healthcare professionals may express reluctance toward embracing change, even in the face of potential benefits. medical systems.

4. CONCLUSION AND FUTURE WORK

This study investigated factors that affect interoperability of medical systems. Distributed ledger technologies presents a decentralized network and is regarded as having great potential for use in healthcare sector, because of the sensitive nature and need for privacy and security of data being processed and managed. DLTs also when used in medical systems has capability of providing system interoperability, trust, timely access to data when needed, solving the issue of data fragmentation and security of patients' electronic medical data.

The aim of the study was to carry out a literature review and survey-based study with the goal to revealing the factors that affect interoperability of medical systems. The highlight of these factors are data formats, syntax, organization and protocols. Consequently, the semantic based factors, technical, organizational, legal/regulatory, security and privacy, human, financial, and cultural factors were cited by the medical system software developers as key categories of factors that hinder interoperability of medical systems.

Further, data meaning, models codification schemes and data definition standardization are the specific factors affecting medical interoperability at the semantic interoperability level. To achieve the study objectives, the researcher defined research questions and using the predefined methodology the researcher narrowed down the analyzed literature to 52 publications. These were then further analyzed and 10 relevant online databases for publications published between 2017 and 2023 searched. The researcher collected data as prompted by the research question and assessed the publications using the predefined assessment criteria.

The study findings indicate that distributed ledger technology research and its employment in eHealth, and healthcare is increasing. Current trends of DLTs research in healthcare indicate that it is mostly used for data sharing, health records and access control, but rarely for other scenarios, such as providing medical system interoperability of medical systems located at various health facilities across different geographical areas. Therefore, much potential for DLTs is still unexploited in relation to solving interoperability challenges of medical systems. Future work can consider designing and developing frameworks and models for integrating DLT into medical systems with an aim to address interoperability challenges.

5. ACKNOWLEDGMENTS

Our thanks to the medical systems software developers who have contributed towards this study.

6. REFERENCES

[1] M. Holderried, A. Hoeper, F. Holderried, and T. Kraus, "ISQUA18-1576Mobile Digital Information and Communication Technology in Healthcare: Patients Attitude and Quality Management Aspects," *Int. J. Qual. Heal. Care*, vol. 30, no. suppl_2, pp. 8–8, Sep. 2018, doi: 10.1093/intqhc/mzy167.07.

[2] A. Anjum, M. Sporny, and A. Sill, "Blockchain Standards for Compliance and Trust," *IEEE Cloud Comput.*, vol. 4, no. 4, 2017, doi: 10.1109/MCC.2017.3791019.

[3] T. O. Afolaranmi *et al.*, "Referral System : An Assessment of Primary Health Care Centres in Plateau State , North Central Nigeria," *world J. Res. Rev.*, no. 1, pp. 82–86, 2018.

[4] W. J. Gordon and C. Catalini, "Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability," *Computational and Structural Biotechnology Journal*, vol. 16. 2018. doi: 10.1016/j.csbj.2018.06.003.

[5] Capgemini, "Healthcare." Dec. 15, 2018. [Online]. Available: https://www.capgemini.com/service/healthcare-life-sciences/healthcare/

[6] G. Wolfond, "A Blockchain Ecosystem for Digital Identity: Improving Service Delivery in Canada's

Public and Private Sectors," *Technol. Innov. Manag. Rev.*, vol. 7, no. 10, pp. 35–40, 2017, doi: 10.22215/timreview/1112.

[7] S. Troy, "distributed ledger technology (DLT)," 2021. https://www.techtarget.com/searchcio/definition/distributed-ledger (accessed Jun. 13, 2022).

[8] R. Leonulous, "Various types of Distributed Ledger Technology | DataDrivenInvestor," 2020. https://www.datadriveninvestor.com/2020/12/04/various-types-of-distributed-ledger-technology/ (accessed May 03, 2021).

[9] HIMSS, "Interoperability in Healthcare | HIMSS," 2022. https://www.himss.org/resources/interoperability-healthcare (accessed Nov. 07, 2022).

[10] D. G. Katehakis and A. Kouroubali, "A Framework for eHealth Interoperability Management," *J. Strateg. Innov. Sustain.*, vol. 14, no. 5, pp. 51–61, 2019, doi: 10.33423/jsis.v14i5.2521.

[11] M. Guclu, C. Bakir, and V. Hakkoymaz, "A New Scalable and Expandable Access Control Model for Distributed Database Systems in Data Security," *Sci. Program.*, vol. 2020, 2020, doi: 10.1155/2020/8875069.

[12] Z. Zhou, C. Sun, J. Lu, and F. Lv, "Research and implementation of mobile application security detection combining static and dynamic," *Proc. - 10th Int. Conf. Meas. Technol. Mechatronics Autom. ICMTMA 2018*, vol. 2018-Janua, pp. 243–247, 2018, doi: 10.1109/ICMTMA.2018.00065.

[13] D. A. Clunie, "DICOM Format and Protocol Standardization—A Core Requirement for Digital Pathology Success," *Toxicol. Pathol.*, vol. 49, no. 4, pp. 738–749, 2021, doi: 10.1177/0192623320965893.

[14] Leadership for it Security and Privacy Across HHS, "Electronic Medical Records in Healthcare," *Dep. Heal. Hum. Serv.*, pp. 1–35, 2022, [Online]. Available: https://www.hhs.gov/sites/default/files/2022-02-17-1300-emr-in-healthcare-tlpwhite.pdf

[15] V. Patel, "A framework for secure and decentralized sharing of medical imaging data via blockchain consensus," 2018, doi: 10.1177/1460458218769699.

[16] M. Hölbl, M. Kompara, A. Kamišalić, and L. N. Zlatolas, "A systematic review of the use of blockchain in healthcare," *Symmetry (Basel).*, vol. 10, no. 10, 2018, doi: 10.3390/SYM10100470.

[17] P. Lafourcade and M. Lombard-Platet, "About blockchain interoperability," *Inf. Process. Lett.*, vol. 161, p. 105976, 2020, doi: 10.1016/j.ipl.2020.105976.

[18] B. Debnath, R. Roychoudhuri, and S. K. Ghosh, "E-Waste Management – A Potential Route to Green Computing," *Procedia Environ. Sci.*, vol. 35, pp. 669–675, 2016, doi: 10.1016/j.proenv.2016.07.063.

[19] B. C. Ghosh, T. Bhartia, S. K. Addya, and S. Chakraborty, "Leveraging public-private blockchain interoperability for closed consortium interfacing," *Proc. - IEEE INFOCOM*, vol. 2021-May, May 2021, doi: 10.1109/INFOCOM42981.2021.9488683.

[20] E. Union *et al.*, "Understanding the landscape of Distributed Ledger Technologies/Blockchain: Challenges, opportunities, and the prospects for standards," *Underst. Landsc. Distrib. Ledger Technol. Challenges, Oppor. Prospect. Stand.*, vol. 8, no. August, pp. 1–17, 2020, doi: 10.7249/rr2223.

[21] P. Zhang, J. White, D. C. Schmidt, G. Lenz, and S. T. Rosenbloom, "FHIRChain: Applying Blockchain to Securely and Scalably Share Clinical Data," *Comput. Struct. Biotechnol. J.*, vol. 16, pp. 267–278, Jan. 2018, doi: 10.1016/j.csbj.2018.07.004.

[22] Gminsights.com, "Healthcare Data Interoperability Market Analysis 2019-2025 Report," *Global Markets Insights*, 2019. https://www.gminsights.com/industry-analysis/healthcare-data-interoperability-market (accessed Feb. 27, 2023).

[23] Emergen Research, "Interoperability Solution in Healthcare Market Trend | Healthcare Interoperability Solutions Industry Forecast 2021-2030," *Emergen Research*, 2022. https://www.emergenresearch.com/industry-report/interoperability-solutions-in-healthcare-market (accessed Feb. 27, 2023).

[24] R. Lozano *et al.*, "Measuring progress from 1990 to 2017 and projecting attainment to 2030 of the health-related Sustainable Development Goals for 195 countries and territories: a systematic analysis for the Global Burden of Disease Study 2017," *Lancet*, vol. 392, no. 10159, pp. 2091–2138, 2018, doi: 10.1016/S0140-6736(18)32281-5.

[25] A. S. Yadav, S. Shikha, S. Gupta, and D. S. Kushwaha, "The efficient consensus algorithm for land record management system," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1022, no. 1, 2021, doi: 10.1088/1757-899X/1022/1/012090.

[26] D. Ola, "Ola David Egbe Department of Computer Science University of Port Harcourt Choba," no. November, 2021.

[27] I. Albarki, M. Rasslan, A. M. Bahaa-eldin, and M. Sobh, "ScienceDirect Robust Robust Hybrid-Security Hybrid-Security Protocol Protocol for for HealthCare HealthCare Systems Systems," *Procedia Comput. Sci.*, vol. 160, no. 2018, pp. 843–848, 2019, doi: 10.1016/j.procs.2019.11.001.

[28] B. H. de Mello *et al.*, "Semantic interoperability in health records standards: a systematic literature review," *Health Technol. (Berl).*, vol. 12, no. 2, pp. 255–272, Mar. 2022, doi: 10.1007/S12553-022-00639-W/FIGURES/4.

[29] M. W. L. Moreira, J. J. P. C. Rodrigues, A. K. Sangaiah, J. Al-Muhtadi, and V. Korotaev, "Semantic interoperability and pattern classification for a service-oriented architecture in pregnancy care," *Futur. Gener Comp Sy*, vol. 89, pp. 137–147, Dec. 2018, doi: 10.1016/j.future.2018.04.031.

[30] P. Bakibinga, E. Kamande, L. Kisia, M. Omuya, D. J. Matanda, and C. Kyobutungi, "Challenges and prospects for implementation of community health volunteers' digital health solutions in Kenya: a qualitative study," *BMC Health Serv. Res.*, vol. 20, no. 1, p. 888, 2020, doi: 10.1186/s12913-020-05711-7.

[31] Berkeley, "Secure Coding Practice Guidelines | Information Security Office," 2019. https://security.berkeley.edu/secure-coding-practice-guidelines (accessed Dec. 05, 2019).

[32] S. Schulz, R. Stegwee, and C. Chronaki, "Standards in Healthcare Data Chapter 3 Standards in Healthcare Data," no. January, 2019, doi: 10.1007/978-3-319-99713-1.

International Journal of Computer Applications Technology and Research
Volume 13–Issue 01, 07 - 08, 2024, ISSN:-2319–8656
DOI:10.7753/IJCATR1301.1002

Reinventing Cyber Security with AI

Dr.A.Jeyalakshmi
Associate Professor
Sri Ramakrishna College of Arts and Science
Coimbatore, Tamil Nadu, India

Abstract: In the digital age, data is the new gold, and a valuable asset, it needs to be safeguarded. Cyber security has always been a critical concern for individuals and businesses alike, but as technology advances, so do the threats that seek to compromise the data. In response to this escalating issue, artificial intelligence (AI) is stepping up to the plate, offering innovative solutions that are reinventing the way to protect the data. This paper provides a concise overview of AI implementations of various cyber security using artificial technologies and evaluates the prospects for expanding the cyber security capabilities by enhancing the defense mechanism

Keywords: Artificial Intelligence, Intelligent Agents, Neural networks, Smart Cyber Security methods.

1. INTRODUCTION

Cyber security is important because it encompasses everything that relates to protecting our data from cyber attackers who want to steal this information and use it to cause harm[1][2][3]. This can be sensitive data, governmental and industry information, personal information, personally identifiable information (PII), intellectual property, and protected health information (PHI). Therefore, they are obviously vulnerable to cyber attacks. A cyber attack is an attack launched from one or more computers against cyber attacks is either to disable the target computer, or take the services offline, or get access to the target computer's data[4]. In response to the issues, artificial intelligence tools are commonly implemented to deal with cyber threats. Artificial intelligence (AI) has helped more organizations to improve the security posture effectively and reduce the breach risks. Machine learning and artificial intelligence are the essential tools in technology for information security as it helps companies and individuals to check and analyze the threats posed to the organization [5].

2. LITERATURE REVIEW

Many more research works have been reported for cyber security threats, predicting the cyber threats with machine learning and deep learning algorithms in AI.Vipin Kumar [1] used a simple k-means clustering approach on NSL-KDD dataset to perceive the accuracy for intrusion detection. K-means, an unsupervised algorithm, is used for classification and defines an unlabeled class to which the clustering is performed. Rahman Ali, et al.,[5] reported A systematic literature review of existing classification algorithms, applied to the area of detection of cyber security attacks is presented and it is concluded that Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT) and Artificial Neural Network (ANN) are the most frequently used classifiers. Jie Chen. et al.,[6] suggested that AI algorithms are mainly applied in cyber security to predict the threats using Machine learning and deep learning.

3. RESEARCH METHODOLOGY

3.1 Cyber Security Challenge

Cyber security is essential for protecting digital assets, including sensitive personal and financial information, intellectual property, and critical infrastructure. The most difficult challenge to cyber security is adapting to a remote workforce. With more and more companies around the world turning to remote work, there are new risks in cyber security that have emerged. Companies must now invest in solutions that protect their systems from attacks outside their networks. The most common cyber threats are phishing, malware, and ransom ware. Phishing is a type of online fraud that involves attackers sending fake emails or websites that look legitimate in order to trick victims into entering personal or financial information. Due to most of the organizations gets challenges in financial loss, reputational damage, and even physical harm.

3.2 AI based Cyber Security Process

3.2.1 Threat Detection

AI algorithms have the capability to analyze huge volumes of data in real time objects, identify patterns, monitor network traffic ,user behaviour, and system logs that could signal a potential security breach.

3.2.2 Predictive Analysis

Predictive analysis is a statistical method which is used to gather data from historic data.AI algorithms to predict anomalies, identify patterns and create forecasts. Predict future threats and attacks and create safety borders for them.

3.2.3 Zero Trust Architecture:

The principle of ZTA is "Never Trust, Always Verify". AI assists in continuously monitoring and analyzing user behaviour, devices, and network traffic to ensure trustworthiness. If an unusual or suspicious activity is detected, AI can swiftly trigger security measures to restrict access until trust is re-established.

3.2.4 BlockChain Technology

Recently, [7] crypto currencies have popularly increased in the market. These are processed based on block chain technology and provide an innovative technical solution for secure transactions and saving the money. Block chain can be used to enable medical records and help in security management by identifying criminal identity loopholes in the system. [7] With block chain technology, verification keys wouldn't be required anymore. If someone tries to hack the data, the system analyzes the whole mass of data chains. Even if one data node is left uninterrupted by the hacker, the entire system can be restored successfully.

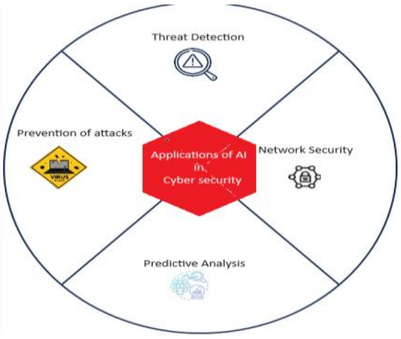


Fig 1 : Applications of Artificial Intelligence in Cyber Security

4. AI BASED CYBER SECURITY CHALLENGES:

AI in cyber security is a double edged sword which enhances and secures threats in all direction. Though, some challenges also exist. The lack of resources to build and maintain AI Systems in cyber security. Vulnerability of AI systems to attacks, infiltration, and manipulation by adversaries also, Inconsistency and privacy concerns around data laws, policies, and regulations.

5. CONCLUSION:

Artificial Intelligence is a fast growing technology in the current era for improving digital security.AI gives a needed analysis and threat identification that can be used by security professionals to minimize breach risk and enhance security posture. Also, As many harmful threats can be detected before any damage, security experts will have more response time to fight against these malicious attacks. Though AI is a valuable asset in cyber security, its limitations-such as data dependencies, false alarms and lack of transparency-should be carefully considered when integrating AI into security strategies.

6. REFERENCES

1. Vipin Kumar, Himadri Chauhan, Dheeraj Panwar,”K-Means Clustering Approach to Analyze NSL-KDD Intrusion Detection Dataset”,International Journal of Soft computing and engineering,2013,3,4, 2013, ISSN: 2231-2307.
2. Torres, J.M., Comesaña, C.I., García-Nieto, P.J. Machine learning techniques applied to cybersecurity.International journal of machine learning and cybernetics. 2019, 1–14.
3. Hashemi, H., Azmoodeh, A., Hamzeh, A., Hashemi, S. Graph embedding as a new approach for unknown malware detection. Journal of computer virology and hacking techniques, 2017, 13, 153–166.
4. Ab Razak, M.F., Anuar, N.B., Othman, F., Firdaus, A., Afifi, F., Salleh, R. Bio-inspired for features optimization and malware detection. Arabian journal of science and engineering,. 2018, 43, 6963–6979.
5. Rahman Ali, Asmat Ali, Fark Hund Iqbal, Asad Masood Khattak and Saiqa Aleem,” A Systematic Review of Artificial Intelligence and Machine Learning Techniques for Cyber Security”, International conference on big data and security, August 2020,584-593.
6. Jie Chen,Dandan Wu,Ruiyun Xie, “Artificial intelligence algorithms for cyberspace security applications: a technological and status review”, Frontiers of Information Technology and Electronic Engineering, August 2023,24,1117-1142.
7. Gaurav Belani, **“**The Use of Artificial Intelligence in Cybersecurity: A Review”, IEEE Computer society,2021.

International Journal of Computer Applications Technology and Research
Volume 13–Issue 01, 09 - 18, 2024, ISSN:-2319–8656
DOI:10.7753/IJCATR1301.1003

Alert Correlation Model Based on Hybrid Machine Learning Techniques to Enhance the Performance of NIDS

Joseph Mbugua
Garissa University
Kenya

Enoch Mogendi
Garissa University
Kenya

ABSTRACT

There obstacles in developing an effective intrusion detection systemin this modern digital world. This work proposes a three level model in developingNIDS that offers multiple types of correlations. In the first level, several feature selection techniques are integrated existing feature selection techniques Correlation Feature Selection, Information Gain and Chi square to find the best set of features used in this work. The second level enhances the structural based alert correlation model based on Expectation and Maximization (EM) to improve the quality of alerts and detection capability by grouping alerts with common attributes. Then an anomaly classification module is designed in the third level based on fusion of five heterogeneous classifiers Support Vector Machine (SVM), Instance based Learners (IBL), Random Forest, J48, and Bayes Net using Voting as a Multi-Classifier.

The NSL KDD dataset is used in this experiment. The overall detection rate is 99.9%, false error rate 0.1% and execution rate of 1340.7 seconds. This shows that HAC is effective and practical in providing complete correlation even on high dimensionality, large scaled and low quality dataset used in intrusion detection system.

Keywords: Alert Correlation, Machine Learning, Model, Performance, Intrusion Detection.

Introduction

The advancement of modern computers, network and internet has led to their widespread adoption and application in organizations' critical systems. These organizations are susceptible to intrusions and malicious activities that attempt to compromise the proprietary business plans (integrity and confidentiality) loss of critical business data and disruption of services (availability) of system resources (Alkhpor & Alserhani, 2023). Intrusion detection is a system for detecting intrusions and hence works as the major defensive mechanism in a network environment (Albasheer et al., 2022; Alsoufi et al., 2021; Kiruki, Muketha, & Kamau, 2023). It's main goal is to automatically monitor network traffic and classify them as normal or suspicious activities and inform the security analyst or response system to take appropriate action before the intrusion compromises the network.

Alert Correlation (AC) takes the generated alerts, process and produce compact reports on the security status of the network under surveillance(Albasheer et al., 2022; Alkhpor & Alserhani, 2023).

There are four main techniques proposed in alert correlation focusing on analyzing intrusion alerts produced by computer networks to improve detection and prediction ability in NIDS. In Structural-based AC (SAC), alerts are correlated based on similarity of attributes. that it cannot discover the causal relationships among alerts(Ho, Hua, Siraj, & Din, 2017). The Causal-based AC (CAC) analysis finds the relationship between alert types in the alert stream to discover alert attributes that have the greatest impact on the relationship between intrusion alerts. Research by(Diehl & Ramirez-Amaro, 2023; Makhlouf, Zhioua, & Palamidessi, 2020; Wang et al., 2022) have showed that the technique can discover unknown alerts but it is expensive to build a complete attack database. The Statistical-based AC (STAC) defines normal behavior by collecting data relating to the behavior of legitimate users over a period of time. The work by (Boero et al., 2017)indicates that good performance of Statistical-based AC strongly depends on good parameters setting which is very difficult to estimate. The goal of data mining and machine learning technique is to produce a model expressed as an executable code which can be used to perform data mining tasks such as classification, prediction or other similar task(Kayode Saheed, Idris Abiodun, Misra, Kristiansen Holone, & Colomo-Palacios, 2022; Liu & Lang, 2019; Mari, Zinca, & Dobrota, 2023; Othman, Ba-Alwi, Alsohybe, & Al-Hashida, 2018; Saranya, Sridevi, Deisy, Chung, & Khan, 2020).

The aim of this work is to design alert correlation model for Improving performance of network intrusion detection based on hybrid machine learning techniques. It will determine the optimum features based on hybrid feature selection techniques, enhance the structural based alert correlation model using unsupervised machine learning techniques and enhance the causal-based alert correlation model using supervised machine learning techniques.

Literature Review

The research(Mbugua, Thiga, & Siror, 2019)a comparative analysis on performance of three different ensemble methods, bagging, boosting and stacking is performed in order to determine the algorithm with high detection accuracy and low false positive rate. Three different experiments on NSL KDD data set are conducted and their performance evaluated based on accuracy, false alarms and computation time. The overall performance of the different types of classifiers used proved that ensemble machine learning classifiers outperformed the single classifiers with high detection accuracy and low false rates.

A new feature selection model (Chahira, 2020)proposed is based on hybrid feature selection techniques (information gain, correlation, chi squere and gain ratio) and Principal Component Analysis (PCA) for feature reduction. This study employed data mining and machine learning techniques on NSL KDD dataset in order to explore significant features in detecting network intrusions. The experimental results showed that the proposed model improves the detection rates and also speed up the detection process.

Theresearch(Chahira & Kiruki, 2022)compares four unsupervised learning algorithms namely Self-organizing maps (SOM), K-means, Expectation and Maximization (EM) and Fuzzy C-means (FCM) to select the best cluster algorithm based on Clustering Accuracy Rate (CAR), Clustering Error (CE) and processing time. The result inferred that the proposed model based on hybrid feature selection, PCA and EM is effective in terms of Clustering Accuracy Rate (CAR) and processing time for The NSL-KDD Dataset

Methodology

This research addresses the issues of improving the quality of alerts that are generated by multiple NIDSs and recognizing the attack strategy from the unrelated alerts. It is executed through a series of experiments and testing to achieve the goal of objectives of the research. This approach is preferred as the main method due to certain characteristics, such as performance measures, dataset evaluations and the usability of the results.

Proposed Hybrid-Based Alert Correlation Model

The five processing levels includes.

a) Feature selection Extracts the optimum features from synthetic dataset based on ensemble feature selection methods

b) Dimension Reduction uses PCA to reduce the dimensionality of the alerts for optimal correlation performance.

c) Unsupervised Learning Algorithm clusters alerts into groups/attack steps to discover the structural correlation among the alerts.

d) Post-Clustering Algorithms improve the quality of alerts by filtering out the unwanted low quality alerts (redundant, false positives and low-risk alerts).

e) Ensemble Supervised Learning Algorithm classifies alerts into classes/attack stages to discover the causal correlation among the alerts.

f) Statistical Correlation Tests calculate the strength of dependencies among the alerts attributes to discover the statistical correlation,

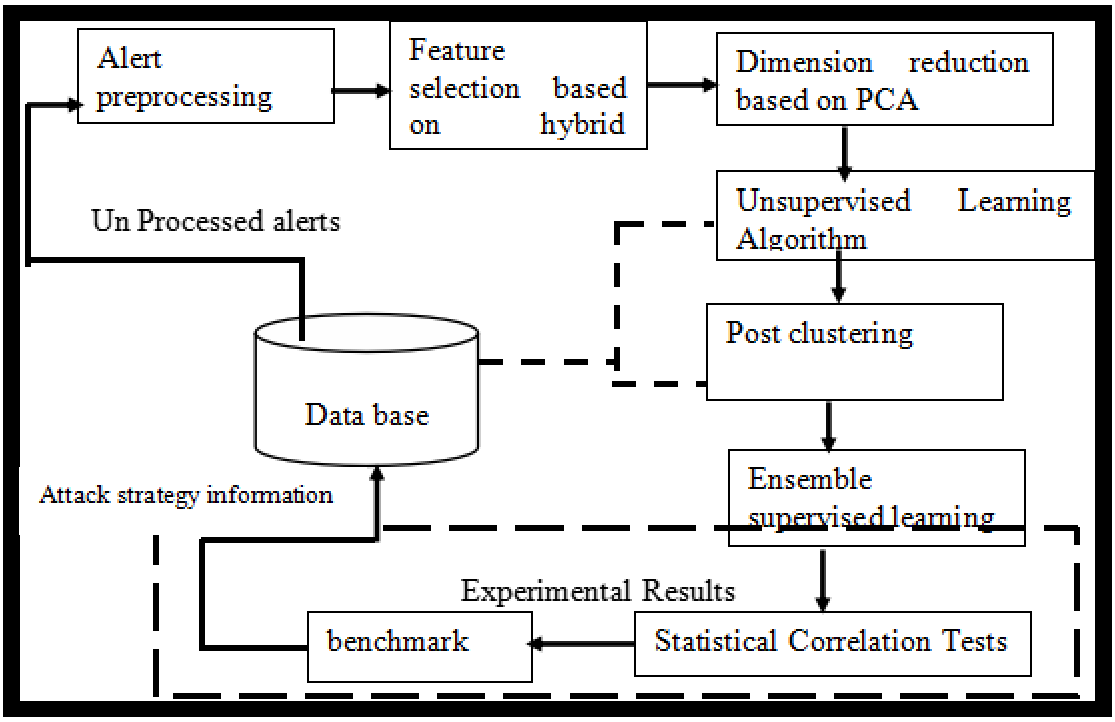

Alert preprocessing
Feature selection based on hybrid
Dimension reduction based on PCA
Unsupervised Learning Algorithm
Un Processed alerts
Post clustering
Data base
Ensemble supervised learning
Attack strategy information
Experimental Results
benchmark
Statistical Correlation Tests

Figure 1: Hybrid-Based Alert Correlation Model

Step 1 **Ensemble-based Multi-Filter Feature Selection (EMFFS) Method**

In this phase, irrelevant and less important features are removed. An ensemble for feature evaluation and feature selection algorithms were invoked to select the set of relevant features. a novel feature selection model is proposed based on hybridizing feature selection techniques (information gain, correlation feature selection and chi square). The experiment, select attribute set based on the repetition of attribute from four scheme. Existing FS that are employed in experiments are 1) Correlation Feature Selection (CFS) based evaluator with Best-first searching method, 2) Information Gain (IG) based Attributes Evaluator with ranker searching method, and 3) Chi Squared Eval and Ranker searching method we obtained

Each algorithm evaluated each class dependent dataset created resulted in a relevant set of features for each particular class. The researcher considered only features that are selected by ten folds (like., k = 10). On the other hand, features that not selected by any algorithm were irrelevant and removed from the list. Output of this phase is a reduced set of common relevant features that were ranked by its relevance value for each attack class.

In the proposed model, these algorithms select the best features set for all attack types in NSL KDD dataset (DOS, PROBE, R2L, U2R, and NORMAL. NSL-KDD, 2014 which contains simulated attack scenarios in a protected environment an off-site server. KDD”99 testing set includes 37 attack types that are included in the testing set. The optimum features selected using the hybrid feature selection technique include: duration, src bytes, dst bytes, logged_in, same_srv_rate, diff_srv_rate, dst_host_count, dst_host_count, dst_host_srv, diff_host_rate, dst_host_srv_rerror_rate. Protocal_type, service, attck. Detailed experiment process and results are disscussed in (Chahira, 2020)

Step 2 Enhanced Structural-Based Alert Correlation Method

The detection component of NIDSs generates a massive amount of alerts and can overwhelm the security experts. An automated and intelligent clustering system is important to reveal their structural correlation by grouping alerts with common attributes. The aim of this objective is to enhance the Structural-based AC model using machine learning technique to improve the quality of alerts and identify attack strategy. A novel hybrid clustering model is developed based on normalization, discretization and Improved Unit Range (IUR) technique to preprocess the dataset, EMFFS, Principal Component Analysis (PCA), SAC and proposed Post-Clustering algorithms is implemented to reduce the alerts dimensionality and optimize the performance and unsupervised learning algorithm to aggregate similar alerts and to reduce the number of alerts. In the proposed model the performance of various unsupervised learning techniques like Self-organizing maps (SOM), Expectation Maximization, K-means, hybrid clustering and Fuzzy c-means (FCM) is compared. The output are comtained in (Chahira & Kiruki, 2022)

Mode	FCM				K Means				SOM				EM			
	CE	ER	AR	TI	CE	ER	AR	TI	CE	ER	AR	TI	CE	ER	AR	TI
HFS	74	17.5	82.6	1.3	57	13.4	86.6	4.4	135	31.8	68.2	4.2	45	10.6	89.4	1.9
PCA	133	31.3	68.6	3.6	141	33.3	66.2	5.2	170	40.1	60.0	6.5	86	20.3	79.7	2.7
IPCA	67	15.8	84.2	4.8	46	10.9	89.2	6.2	112	26.4	73.6	7.4	41	9.7	90.3	4.6

Figure 2: Clustering Performance based on Self-organizing maps (SOM), Expectation Maximization, K-means and Fuzzy c-means (FCM) The result inferred that the proposed model based on hybrid feature selection, PCA and EM is effective in terms of clustering accuracy and processing time for this dataset.

Step 3 Enhanced causal-based alert correlation model.

In the third phase, the output from the second phase which is the results from the hybrid clustering model (PCA and EM) is fed as input to the Multiple IDS Unit (MIU), and the output is the local decision (yi) derived from running different learning algorithms on the same data set. This section has five IDSs, each utilizing a unique algorithm is used independently for detecting a certain class of attack with improved accuracy, while performing moderately on the other classes. The five different types of IDS algorithms used are Support Vector Machines (SVM), IBK, Random Forest, J48, and Bayes Net and different results obtained and five outputs (local decisions) y1, y2, to y5 are obtained.The output from each IDS in MIU, considered as local decision (yi), is passed onto the multi classifier component based on majority voting rule and makes the final decision. Each classifier has a weight to denote the contributions of the classifier to the voting system. For each class to be identified, a weighted sum of base learners can be calculated. The output from each classifier is taken to the decision unit, and the global decision is taken based on the majority voting rule. If majority outputs from the MIU unit suggest Attack, then the decision unit decides that the input traffic is of ATTACK type; else it is NOT ATTACK. Detailed experiment process and results are disscussed in (Mbugua et al., 2019)

Experimentation, Results and Discussion

In the experiment, we apply full dataset as training set and 10-fold cross validation for the testing purposes. The available dataset is randomly subdivided into 10 equal disjoint subsets and one of them is used as the test set and the remaining sets are used for building the classifier. In this process, the test subset is used to calculate the output accuracy while the N1 subset is used as a test subset and to find the accuracy for each subset. The process is repeated until each subset is used as test set once and to compute the output accuracy of each subset. The final accuracy of the system is computed based on the accuracy of the entire 10 disjoint subsets.

The experiments will be conducted on MIT Lincoln's Lab's DARPA 2000 Scenario Specific

The performance of the proposed intrusion detection system is evaluated with the help of confusion matrix. The conducted experiments will be evaluated according to four performance measures which are defined below:

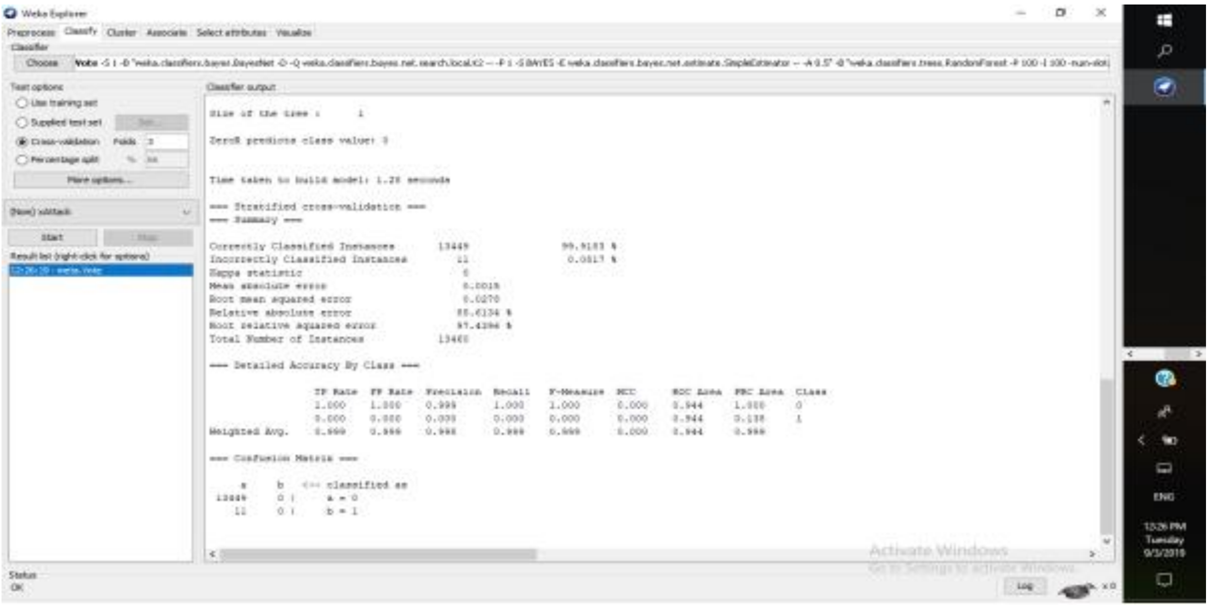
i. TPR: TP/(TP+FN), also known as detection rate (DR) or sensitivity or recall.

ii. The False Alarm Rate (FAR) is the rate of the misclassified to classified records,

iii. Precision (P): TP/(TP+FP) is defined as the proportion of the true positives against all the positive results.

iv. Total Accuracy (TA): (TP+TN)/(TP+TN+FP+FN) is the proportion of true results (both true positives and true negatives) in the population.

v. F-measure: 2PR/(P+R) is the harmonic mean of precision and recall.



```
Vote -S 1 -B "weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5" -B "weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots
```

```
Size of the tree :      1

ZeroR predicts class value: 0


Time taken to build model: 1.28 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       13449               99.9183 %
Incorrectly Classified Instances        11                0.0817 %
Kappa statistic                          0
Mean absolute error                      0.0015
Root mean squared error                  0.0270
Relative absolute error                 88.6134 %
Root relative squared error             97.4396 %
Total Number of Instances            13460

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.999      1.000    1.000      0.000    0.944     1.000     0
                 0.000    0.000    0.000      0.000    0.000      0.000    0.944     0.138     1
Weighted Avg.    0.999    0.999    0.998      0.999    0.999      0.000    0.944     0.998

=== Confusion Matrix ===

     a     b   <-- classified as
 13449     0 |     a = 0
    11     0 |     b = 1
```

Figure 3: Performance of Proposed Hybrid-Based Alert Correlation Model

The value of TP, P, R, FM and ROC are 0.998, 0.99, 0.99 and 0.94 that close to ‘1’ indicates excellent performance and below ‘0.5’ indicates average or bad performance. A smaller value of FP (close to zero) shows good performance since the amount of false classification is very small. The time taken to build the model is only 1340.7seconds.

Conclusion

The approach based on classifier combination achieve effective attack detection as the combination of multiple evidences usually exhibits higher accuracies, like. lower false positives, than individual decisions. In addition, the generalization capabilities of pattern recognition algorithms allow for the detection of novel attacks that is not provided by rule-based signatures.

References

Albasheer, H., Siraj, M. M., Mubarakali, A., Tayfour, O. E., Salih, S., Hamdan, M., … Kamarudeen, S. (2022). Cyber-Attack Prediction Based on Network Intrusion Detection Systems for Alert Correlation Techniques: A Survey. *Sensors*, *22*(4), 1–15. https://doi.org/10.3390/s22041494

Alkhpor, H. K., & Alserhani, F. M. (2023). Collaborative Federated Learning-Based Model for Alert Correlation and Attack Scenario Recognition. *Electronics (Switzerland)*, *12*(21). https://doi.org/10.3390/electronics12214509

Alsoufi, M. A., Razak, S., Siraj, M. M., Ali, A., Nasser, M., & Abdo, S. (2021). Anomaly Intrusion Detection Systems in IoT Using Deep Learning Techniques: A Survey. *Lecture Notes on Data Engineering and Communications Technologies*, *72*(May), 659–675. https://doi.org/10.1007/978-3-030-70713-2_60

Boero, L., Cello, M., Marchese, M., Mariconti, E., Naqash, T., & Zappatore, S. (2017). Statistical fingerprint-based intrusion detection system (SF-IDS). *International Journal of Communication Systems*, *30*(10). https://doi.org/10.1002/dac.3225

Chahira, J. M. (2020). Model for Intrusion Detection Based on Hybrid Feature Selection Techniques. *International Journal of Computer Applications Technology and Research*, *09*(03), 115–124. https://doi.org/10.7753/ijcatr0903.1005

Chahira, J. M., & Kiruki, J. K. (2022). Model for Enhancing Performance of Network Intrusion Detection based on Hybrid Feature Selection and Unsupervised Learning Techniques. *International Journal of Computer Applications Technology and Research*, *11*(08), 341–350. https://doi.org/10.7753/ijcatr1108.1008

Diehl, M., & Ramirez-Amaro, K. (2023). A causal-based approach to explain, predict and prevent failures in robotic tasks. *Robotics and Autonomous Systems*, *162*, 104376. https://doi.org/10.1016/j.robot.2023.104376

Ho, H., Hua, W., Siraj, M., & Din, M. M. (2017). Integration of PSO and K-Means Clustering Algorithm for Structural-Based Alert Correlation Model. *International Journal of Innovative Computing*, *7*(2), 34–39.

Kayode Saheed, Y., Idris Abiodun, A., Misra, S., Kristiansen Holone, M., & Colomo-Palacios, R. (2022). A machine learning-based intrusion detection for detecting internet of things network attacks. *Alexandria Engineering Journal*, *61*(12), 9395–9409. https://doi.org/10.1016/j.aej.2022.02.063

Kinanu Kiruki, J., Muchiri Muketha, G., & Kamau, G. (2023). a Novel Alert Correlation Technique for Filtering Network Attacks. *International Journal of Network Security & Its Applications*, *15*(03), 33–47. https://doi.org/10.5121/ijnsa.2023.15303

Kiruki, J. K., Muketha, G. M., & Kamau, G. (2023). Metrics for Evaluating Alerts in Intrusion Detection Systems. *International Journal of Network Security & Its Applications*, *15*(01), 15–37. https://doi.org/10.5121/ijnsa.2023.15102

Liu, H., & Lang, B. (2019). Machine learning and deep learning methods for intrusion detection systems: A survey. *Applied Sciences (Switzerland)*, *9*(20). https://doi.org/10.3390/app9204396

Makhlouf, K., Zhioua, S., & Palamidessi, C. (2020). *Survey on Causal-based Machine Learning Fairness Notions*. *Proceedings of ACM Conference (Conference'17)* (Vol. 1). Association forComputing Machinery. Retrieved from http://arxiv.org/abs/2010.09553

Mari, A. G., Zinca, D., & Dobrota, V. (2023). Development of a Machine-Learning Intrusion Detection System and Testing of Its Performance Using a Generative Adversarial Network. *Sensors*, *23*(3). https://doi.org/10.3390/s23031315

Mbugua, J., Thiga, M., & Siror, J. (2019). A Comparative Analysis of Standard and Ensemble Classifiers on Intrusion Detection System. *International Journal of Computer Applications Technology and Research*, *8*(4), 107–115. https://doi.org/10.7753/ijcatr0804.1005

Othman, S. M., Ba-Alwi, F. M., Alsohybe, N. T., & Al-Hashida, A. Y. (2018). Intrusion detection model using machine learning algorithm on Big Data environment. *Journal of Big Data*, *5*(1). https://doi.org/10.1186/s40537-018-0145-4

Saranya, T., Sridevi, S., Deisy, C., Chung, T. D., & Khan, M. K. A. A. (2020). Performance Analysis of Machine Learning Algorithms in Intrusion Detection System: A Review. *Procedia Computer Science*, *171*(2019), 1251–1260. https://doi.org/10.1016/j.procs.2020.04.133

Wang, L., Adiga, A., Chen, J., Sadilek, A., Venkatramanan, S., & Marathe, M. (2022). CausalGNN: Causal-Based Graph Neural Networks for Spatio-Temporal Epidemic

Forecasting. *Proceedings of the 36th AAAI Conference on Artificial Intelligence, AAAI 2022*, *36*(Cdc), 12191–12199. https://doi.org/10.1609/aaai.v36i11.21479

Wu, M., & Moon, Y. (2019). Alert Correlation for Cyber-Manufacturing Intrusion Detection. *Procedia Manufacturing*, *34*, 820–831. https://doi.org/10.1016/j.promfg.2019.06.197

Yu, M., & Zhang, X. (2023). AlertInsight: Mining Multiple Correlation For Alert Reduction. *Computer Systems Science and Engineering*, *46*(2), 2447–2469. https://doi.org/10.32604/csse.2023.037506