Open Access Research Journal of
Multidisciplinary Studies
Journals home page: https://oarjpublication/journals/oarjms/
ISSN: 2783-0268 (Online)

(REVIEW ARTICLE)

Enhancing healthcare data security through machine learning: A data mining perspective

Sivakumar Venkataraman [1,*] and Subitha Sivakumar [2]

[1] *Department of Computer Science, Botho University, Gaborone, Botswana.*
[2] *Department of Quality Assurance, Programme and Institutional Accreditation, ABM University College, Gaborone, Botswana.*

Open Access Research Journal of Multidisciplinary Studies, 2026, 11(01), 107-114

Publication history: Received on 12 February 2026; revised on 22 March 2026; accepted on 24 March 2026

Article DOI: https://doi.org/10.53022/oarjms.2026.11.1.0024

Abstract

Globally the health industry has grown rapidly and turn out to be problematic day by day. Building a strong and secure healthcare system is challenging and it may lead to systematizing the health marketplace. At present, healthcare organizations can offer their stakeholders better and lower-priced services of implementing the Electronic Health Record (EHR) that rolled out the manual-based systems. This transaction made possible by innovations in ICTs. Data mining has made a big change in computer industries with the support of Data Mining. The value of information can be increased by supporting enormous data volumes with implementation of machine learning algorithms to efficiently use the data. In order to provide regular health services, personal information and health-related data must be recorded. These data are strongly related to user privacy and confidential information. Data impairment may result from improper disclosure, loss of data integrity, or inaccessibility. Data miners should have a fundamental awareness of healthcare information privacy, information security, and network security to minimize the risk of harm to people, their organizations, or themselves. The concepts, elements, and guidelines for controlling the security and privacy of healthcare information utilized for data mining are examined in this paper.

Keywords: Data Mining; Electronic Health Record (HER); Healthcare Environment; Information and Communication Technology (ICT); Information Security; Machine Learning; Network Security

1. Introduction

Huge amounts of healthcare data become available for important insights as the delivery system for healthcare uses information technology. Information technology is typically adopted by healthcare institutions to lower costs and increase effectiveness and quality. Due to the growth in digital health care systems, the treatments are made simpler and more open, because of that it made our years more enjoyable.

The healthcare organizations generate a large amount of information about individuals and their conditions, health insurance companies, medication and treatment schedules for various diseases, medical services, and other subjects. In order to make decisions about management, planning, or treatments, the healthcare community is increasingly required to convert the vast amounts of healthcare data already available into value-added data by spotting previously undetected patterns and connections between the data.

Doctors and researchers utilize health data to uncover information that is implicitly contained in a single patient's health record or in a large collection of patient health records. Due to the reliance on a very high level of confidence in the patient-physician relationship, these new applications of clinical data may have an impact on healthcare. Health data not only covers patient medical information, but it also covers the other as aspects of patient and family demographic details like age, date of birth, gender, marital status, contact number, address, email, credit card number, income, and

* Corresponding author: Sivakumar Venkataraman.

other personal details. Compilations of this data have been made available to researchers and business. Such collections are characterized as de-identified because the data don't include the patient's name, address, phone number, or social security number. There doesn't appear to be much of a risk to the patient's privacy as a result. By comparing this data with publicly available databases, methods like data mining may be able to connect a person to a certain diagnosis. This information is gathered and transmitted via the internet or kept in a centralized database, where it can be accessed by researchers or healthcare experts. Security is one of the main problems, and it's crucial when working with such sensitive data.

1.1. Data Mining

Useful information can be created from unstructured data using data mining. This helps to generate the pattern from the large amount of data which supports the industries to know about the customers' activities in developing the market strategies which are more effective in increasing the sales. Datamining is a process which needs resourceful data gathering methods, Data ware housing, and system processing. Datamining is a tool that businesses may utilize a number of items, determine the goods and facilities their clients attracted in acquiring frauds and spam penetrating [3].

Knowledge Discovery in Databases (KDD) and data mining are frequently used interchangeably; the latter is actually simply one phase in the process. KDD's primary objective is to extract from huge data sets meaningful and frequently undiscovered information. KDD is widely used in every industry that stands to gain from the analysis of massive amounts of data, including business analysis, marketing research, and scientific studies. Cybersecurity experts use it to recognize and break new attacks, while cybercriminals use it to devise new attack vectors [2].

1.2. Electronic Health Record (EHR)

HER is system collects the patients and people health data stored electronically. An digital representation of a patient health record that is updated over time by the health care worker is known as an electronic health record (EHR). It may include any important demographics, progress notes, concerns, prescriptions, vital signs, previous medical history, vaccines, laboratory data, and radiology reports that are necessary for that person's care under a specific physician [5].

1.3. Healthcare Environment

The Healthcare environment term refers to the working conditions for healthcare personnel. The physical geographic location of the workplace and its immediate environs are included in the totality of the elements, factors, and circumstances. The healthcare environment is constantly of utmost importance in the medical industry, and health care systems are expanding their services to make patients feel fresher, healthier, and more secure in order to maintain a healthy atmosphere [7].

1.4. Information and Communication Technology (ICT)

A wide range of technological resources and techniques are referred to collectively as information and communication technology (ICT), and they are used to transmit, save, generate, share, or exchange information. ICT is the infrastructure and components that enable modern computing. No single accepted definition of ICT, the phrase is generally understood to refer to all hardware, software, applications, and frameworks that together allow people and businesses to interact in the digital world [15].

1.5. Information Security

The process of guarding against unauthorized access to, alteration of, or theft of digital data over the course of its existence, or from a dataset or database, is known as information security. According to [12]'s definition of "Information Security," data and information systems need to be guarded against unauthorized access, use, disclosure, interruption, alteration, and destruction in order to ensure integrity, confidentiality, and availability.

1.6. Machine Learning

Machine Learning is a technique converging on data and algorithm to think on how people learn, aids in knowledge acquisition, and eventually improves the accuracy of the system. The phrase "machine learning" (ML) describes a group of methods or algorithms that let computers automatically create data-driven models and build models by methodically spotting patterns in statistically significant data [10], [16].

1.7. Network Security

Preventing unauthorized methods, exploitation, and thievery is a fundamental element in network infrastructure of network security. A secure infrastructure must be constructed to ensure that customers, applications, systems, and users can function safely. It can be referred to as a set of rules and configurations designed to protect the reliability, accessibility, and integrity of computer networks and data using both software and hardware technologies [13].

2. Study on Healthcare Environment: Privacy and Security

The ability of the user to operate, access, and control their personal information is often referred to as privacy. The security system is what keeps such data from being taken into the wrong hands as a result of a breach, leak, or hacking. Due to the rapid advancement of technology, security and privacy are extremely dynamic and fast-paced studies domains. For telehealth solutions to be successfully implemented by patients and healthcare providers, security and privacy are essential.

Security and privacy are the main issues that need to be resolved in the implementation of wearables as a component of health care [9]. The personal information stored on the relatively small-sized healthcare gadgets is vast and readily hackable or stolen. One way to reduce security and privacy concerns is to monitor the activity of the device. To minimize the risk of harm to patients, their company, or themselves, data miners should have a fundamental awareness of healthcare information privacy and security [14].

To maintain security, it is important to consider the security components that are employed, the security principles and policies in place, and Confidentiality, Integrity, Availability, and Accountability. Security principles are influenced by a variety of factors, including accountability principles, awareness principles, multidisciplinary principles, proportionality principles, timeliness principles, reassessment principles, and equity principles.

2.1. Data mining in Security

The goal of data mining is to find solutions to pattern matching issues. Data miners are experts in using specialized technologies to filter through large data sets looking for patterns (and anomalies). Once the data has been extracted, it may be used to forecast future trends, enabling organizations to take proactive, well-informed decisions based on massive data sets.

- The data mining process typically involves these five steps:
- Organizations collect data and distribute it to data warehouses.
- Businesses either keep their own servers in-house, use the cloud, or share open source to store and manage their data.
- After accessing the data, management groups, IT specialists, and business analysts decide how to organize it.
- Information is organized by using the application software's in accordance with customer demands.
- The data is presented by the end user in a way that is simple to communicate, like charts or tables.

Data mining has several uses in the field of security, including both physical and cyber security [4]. Attacks on buildings and the destruction of vital infrastructures such as power grids and telecommunications systems are examples of dangers to national security. Cyber security is the focus of protecting computer and network systems from malware caused by malicious software, such as viruses and Trojan horses. Solutions like Intrusion Detection System (IDS) and Auditing [11] are also provided through data mining.

Industry now employs data mining in their arsenal of cybersecurity tools. To find unexpected patterns and behaviours, for instance, abnormality detection techniques could be applied. Network analysis may be performed to identify the criminals of the viruses. Different cyberattacks may be grouped together by classification, and attack detection may then be performed using the profiles. Based in part on knowledge gained about terrorists through phone and email conversations, prediction may be used to identify potential future attacks.

Other applications of data mining include auditing and intrusion detection systems (IDS). Building defense measures like firewalls, authentication tools, and virtual private networks is the conventional approach to defending computer systems from cyberattacks. These systems, virtually invariably have weaknesses. Due to this, there is now a need for intrusion detection, a security technology that monitors systems and detects computer threats as a supplement to traditional security measures.

2.2. Data Mining for Intrusion Detection System (IDS)

Analyzing audit reports with data mining can help detect anomalous trends and expose intrusions. Operating systems, networks, databases, servers, and web clients are all susceptible to malicious intrusion. Data mining is starting to be included into commercial cyber security systems, and consumers are beginning to profit from it for cyber security and intrusion detection, even if it is unlikely that individuals would use it to prevent breaches into their personal computers. [8] Hackers often utilize network-based attacks to obtain access to an organization's entire network.

Intrusion detection is the process of watching and examining the events occurring in a computer system or network for signs of intrusions—attempts to circumvent a computer or network's security mechanisms and endanger the confidentiality, integrity, or accessibility of information resources. According to the type of assault, the number of network connections involved, the source of the attack, the environment, and automation, intrusions are categorized. A hardware and software system called the Intrusion Detection System (IDS) is made to find intrusions.

An IDS is a tool or program that monitors traffic, detects unusual behavior, and reports its findings to an administrator, but it is not able to prevent it from occurring. The system defends against internet attacks on confidentiality, integrity, and access to data and information systems [1]. The network has evolved rapidly, so it has the potential for hostile attacks.

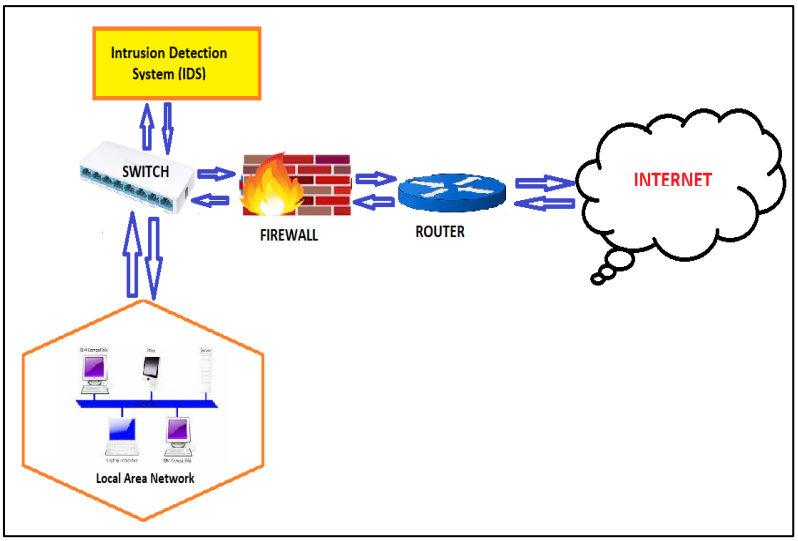

Figure 1 Intrusion Detection System (IDS)

3. Intrusion Detection and Prevention Using Data Mining

A high level of security measures is required for information to be exchanged between a user and a client using contemporary network technologies in a secure and reliable manner. Data mining methods are mostly used for intrusion detection, which is a data-centric process. Intrusion Detection System will also notice unusual activities and learn from previous invasions to perform better in the future [17]. It can be used for several things, including spotting suspicious behaviour, spotting fraud and abuse, spotting terrorist activity, and using lie detection to look into crimes.

The list of applications for data mining techniques for intrusion detection is below.

3.1. Creating a new IDS model using data mining methods

The IDS model's data mining method has a greater performance rate and less false alarm rates. Data mining methods can be used for both the methods like anomaly-based detection as well as signature-based detection.

Signature-based Detection: When using signature-based detection, a classifier can be created to locate acknowledged intrusions by dividing training data into "normal" and "intrusion" categories. Cost-sensitive modelling, association rule mining, and software for explaining algorithms have all been researched in this field.

Anomaly-based detection: Anomaly-based complete detections, creates model of typical behaviour, and instantly recognizes significant departures from it. Methods include statistical techniques as well as software for clustering, outlier analysis, and class algorithms. The solutions employed must be effective, scalable, and capable of handling excessive volume, dimensionality, and heterogeneity of community information.

3.2. Studying stream data

Analysis of stream data entails continuous data analysis, but due to sophisticated calculations and lengthy processing times, data mining is typically employed on static data rather than streaming data. It is more important than ever to carry out intrusion detection inside the environment of the records stream because of the dynamic nature of various intrusion and malicious outbreaks. Therefore, it is important to consider the typical activity sequences that are observed together, identify sequential patterns, and identify outliers. Real-time intrusion detection also requires more datamining techniques to find the growth of clusters and developing the dynamic class model on the record streams.

3.3. Distributed data mining

It is used in the analysis of random data in different databases. There are many unique areas from which intrusions can be launched, with many different destinations as their target. To find those dispersed attacks, community data from several network locations might be investigated using distributed data mining techniques.

3.4. Visualization tools

These technologies are used to display any aberrant patterns found and to show the data as graphs. These techniques could show relationships while highlighting patterns, clusters, and outliers. In order to allow security analysts to inquire about network data or the results of an intrusion detection, intrusion detection structures must really include a graphical user interface.

4. Data Mining used in Intrusion Detection Systems

For information to be transmitted between a user and a client in a secure and reliable manner using modern network technologies, a high level of security measures is necessary. The system's security is ensured by an intrusion detection system in the event that conventional technologies fail. Data mining is the process of obtaining meaningful information from a large volume of data. It supports both supervised and unsupervised learning algorithms. Since intrusion detection is primarily a data-centric operation, IDS can use data mining methods to both uncover anomalous activity and learn from past incursions to perform better going forward. By boosting segmentation, it lets firms plan in real-time and save time by searching the much-expanded database and gathering only reliable information. This can be used for several things, including spotting suspicious behaviour, spotting fraud and abuse, spotting terrorist activity, and using lie detection to look into crimes.

The list of applications for data mining techniques for intrusion detection is below.

4.1. Creating a new IDS model using data mining algorithms:

The IDS model's data mining method has a greater efficiency rate and fewer false alarms. Data mining methods can be used for anomaly-based detection as well as signature-based detection. A classifier can then be built to find acknowledged intrusions in the signature-based detections, where training data is classed as either "normal" or "intrusion".

Software for clarifying algorithms, association rule mining, and cost-sensitive modelling have all been studied in relation to this area. Anomaly-based complete detection creates models of typical behaviour and instantly recognizes significant departures from it. Methods include statistical techniques as well as software for clustering, outlier analysis, and class algorithms. The solutions employed must be effective, scalable, and capable of handling excessive volume, dimensionality, and heterogeneity of community information.

4.2. Data analysis for streams

Analysis of stream data entails continuous data analysis, but due to sophisticated calculations and lengthy processing times, data mining is typically employed on static data rather than streaming data. It is more important than ever to carry out intrusion detection inside the environment of the records stream because of the dynamic nature of intrusions and malicious attacks. A single act can be ordinary on its own yet be seen as malicious if it is considered a part of a series

of actions. Therefore, it's crucial to consider the typical activity sequences that are observed together, identify sequential patterns, and identify outliers.

4.3. Mining of distributed data

It is used to examine random data that is intrinsically dispersed over many databases, making it challenging to combine data processing. There are many unique areas from which intrusions can be launched, with many different destinations as their target. To find those dispersed attacks, community data from several network locations might be investigated using distributed data mining techniques.

4.4. Tools for visualization

In order to help the user understand the data visually, these technologies are used to show the data as graphs. The ability to view any discovered aberrant patterns is another use for these technologies. These techniques could show relationships while highlighting patterns, clusters, and outliers. Security analysts must be able to inquire about network data or the results of an intrusion detection using a graphical user interface provided by intrusion detection structures.

The various datamining approaches are used in Intrusion Detection Systems and all of them are having advantages and disadvantages. Some of the methods are

4.4.1. Classification

Creates a tuple classification. Although it could be used to identify specific attacks, past sample studies documented in the literature show that it has a significant rate of false alarms. By using fine-tuning methods like boosting, this issue might be lessened.

4.4.2. Association

Illustrates the connections between tuples. When several tuples display previously unknown relationships, abnormalities may be discovered. Grouping: Assembles tuples that share characteristics based on predefined criteria. It can be applied to categorization-style broad analysis or to find outliers that might or might not be signs of an attack.

5. The Factors Influencing Healthcare Security in Cost

The following key variables affect the security and privacy of its healthcare data:

- The current software price in relation to the program used.
- The impact of the software platform on the current workforce, particularly the "silent" FTEs that are frequently overlooked (e.g., members of the business intelligence, nursing, and legal teams).
- The price of hiring outside companies to deal with data anomalies, reaction times that are slow, and fines for breaking the law.
- Most crucially, the expense of losing clients because of the erosion of client confidence brought on by a data breach.

6. Conclusion

The objective of this study is to apply data mining techniques to boost healthcare environment security. The transition from paper to electronic health records has considerably assisted efforts to use patient data to enhance numerous facets of the healthcare industry. Thanks to the adoption of electronic health records, healthcare professionals are now able to share their knowledge across all corners of the industry in order to eradicate medical errors and improve patient care and happiness. Data mining has a lot of potential for both conventional and cyber security. Huge data sets can be analysed using data mining, and insights and results can be gleaned and put into practice utilizing AI and machine learning. There are many security approaches that might be used to prevent unauthorized access through the organization's network, but privacy and security concerns pose the biggest and most significant obstacle in the healthcare environment. In addition to a few other security techniques, this study found Intrusion Detection System (IDS) for firewall categories, hostile activities, or policy violations.

Recommendations

The following are the recommendations proposed from the study to strengthen data security in healthcare environments.

- Integrate Machine Learning Models in IDS
  - Healthcare Organizations should implement machine learning techniques with IDS which helps to identify and protect unauthorized users to access sensitive information. Technics like signature based and anomaly based should be implemented concurrently to have ideal results.
- Adopt Real-time stream Data Monitoring
  - This method helps to detect ongoing intrusion attempts dynamically, when integrate with the healthcare systems.
- Regular Audits and Assessments
  - Healthcare organizations should conduct regular audits to identify the vulnerability in hardware and software infrastructure, which helps to improve compliance with data protection regulations.
- Distributed Datamining Technologies
  - Across different places the healthcare data are spread, the implementation of distributed datamining methods will provide a holistic view of potential security risks and breaches.
- Enhance Awareness and training
  - IT Professionals and cyber security professionals should educate the best practices and importance of cyber security data protection when handling healthcare data.

Future works

The following works are suggested for future research,

- Build new hybrid IDS model
  - Future research can be focused on combining machine learning, deep learning and reinforcement learning techniques to improve the accuracy and reduce false positive.
- Security Framework for IOT healthcare systems
  - As wearable and IoT medical devices become more prevalent, future work should focus on creating specialized lightweight security frameworks that can protect these data-rich devices from cyber threats.
- Blockchain integration for secure healthcare systems
  - Integrating the blockchain technology into healthcare systems will provide decentralized and tamper proof records to strengthen trust and security
- AI Governance in Healthcare systems
  - Future research can be focused on the governance framework to safeguard AI implementation in healthcare systems considering the transparency, accountability, and fairness.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Abhinav , K., Jeberson, W., and Klinsega , J. (2017). Intrusion Detection System Based on Data Mining Techniques. Oriental Journal of Computer Science and Technology, 10(2), pp. 491–96.

[2] www.computerscijournal.org, https://www.computerscijournal.org/vol10no2/intrusion-detection-system-based-on-data-mining-techniques/.

[3] Ahmad, B. (December 2017). Intrusion detection with tree-based data mining classification techniques by using kdd dataset. European Journal of Computer Science and Information Technology, https://eajournals.org/ejcsit/vol-5-issue-6-december-2017/intrusion-detection-tree-based-data-mining-classification-techniques-using-kdd-dataset/. Accessed 23 Jan. 2025.

[4] Alshawi, A. (2016). Applying Data Mining Techniques to Improve Information Security in the Cloud: A Single Cache System Approach. Scientific Programming, pp 1-5, DOI.org (Crossref), https://doi.org/10.1155/2016/2385654.

[5] Bhavani, T. M., Latifur , K., Mehedy , M., and Kevin, H. W. (2008). Data Mining for Security Applications. International Conference on Embedded and Ubiquitous Computing, Shanghai China. pp. 585–89. IEEE Xplore, https://doi.org/10.1109/EUC.2008.62.

[6] Clemens , S. K., Brenna , S., Hannah , V., and Alexandra , N. (2017). Security Techniques for the Electronic Health Records. Journal of Medical Systems, https://doi.org/10.1007/s10916-017-0778-4.

[7] Dena , T. A., Savio , M., and Lakshman, T. S. (2020). Design for eHealth and telehealth. Design for Health, edited by Arathi Sethumadhavan and Farzan Sasangohar, Academic Press, pp. 67–86. ScienceDirect, https://doi.org/10.1016/B978-0-12-816427-3.00004-X. DOI:10.1007/s11277-020-07040-8.

[8] Ferdinando, F., and Claudia, C. A. (2012). The Oxford Handbook of Environmental and Conservation Psychology edited by Susan D. Clayton, 1st ed., Oxford University Press, 2012, pp. 295–315. DOI.org (Crossref), https://doi.org/10.1093/oxfordhb/9780199733026.013.0016.

[9] Gulshan , K., and Krishan , K. (2014). Network security – an updated perspective. Systems Science and Control Engineering, 2(1), pp. 325–34. DOI.org (Crossref), https://doi.org/10.1080/21642583.2014.895969.

[10] Indrakumari, R., Poongodi, T., Suresh, S., and Balamurugan, B. (2020). The growing role of Internet of Things in healthcare wearables, edited by Valentina Emilia Balas et al., Academic Press, 2020, pp. 163–94. ScienceDirect, https://doi.org/10.1016/B978-0-12-819593-2.00006-6

[11] Jafar , A., Anand , N., and Akshi , K. (2018). Machine Learning from Theory to Algorithms: An Overview. Journal of Physics: Conference Series. https://doi.org/ 10.1088/1742-6596/1142/1/012012.

[12] Joshua , N. O. (2021). A Model for Auditing Smart Intrusion Detection Systems (IDSs) and Log Analyzers in Cyber Physical Systems (CPSs). Cybersecurity Threats with New Perspectives, IntechOpen, 2021. www.intechopen.com, https://doi.org/10.5772/intechopen.94569.

[13] Lei, X., Jian , W., Jian , Y., and Yong , R. (2014). Information Security in Big Data: Privacy and Data Mining. IEEE Access, 1149-1176, https://doi.org/10.1109/ACCESS.2014.2362522.

[14] Pranav, C. O., and Satish, T. R. (2015). Information Privacy and Security in Data Mining. International Journal of Engineering Sciences and Management Research, Accessed 23 Jan. 2025.

[15] Priyank , S., Kreena , M., and Shikha , S. (2013). Network Security. International Journal of Scientific and Research Publications.

[16] Ritu, C., Harleen, K., and Victor, C. (2021). An Optimized Integrated Framework of Big Data Analytics Managing Security and Privacy in Healthcare Data. Wireless Personal Communications 117(2):1-22.

[17] Samuel , H., Tigists , T. S., Selamawit , B., Kirubel , M., Anteneh, G., Daniel , B., and Shegaw , A. M. (2022). Utilization of information and communication technology (ICT) among undergraduate health science students: a cross-sectional study. BMC Medical Education, https://doi.org/10.1186/s12909-022-03296-9.

[18] Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science, https://doi.org/10.1007/s42979-021-00592-x.

[19] Solane , D., and Mohd. , O. N. (2015). Using Data Mining Algorithms for Developing a Model for Intrusion Detection System (IDS). Procedia Computer Science, 46-51, https://doi.org/10.1016/j.procs.2015.09.145.