Modeling and Analysing Healthcare Professionals' Security Practice using Data-Driven and Artificial Intelligence Approach:Systematic Review and a proposed framework *

Prosper Kandabongee Yeng[1], Livinus Obiora Nweke[1], Bian Yang[1], Einar Arthur Snekkenes[1]

Norwegian University of Science and Technology, Teknologiveien 22 2815 Gjøvik

Abstract

Blacklisting malicious activities in healthcare is challenging especially in the aspect of access control related security practices. This is for the fear of preventing legitimate accesses for therapeutic reasons. Preventing legitimate access will contravene the availability trait of the confidentiality, integrity, and availability (CIA) triad, and may result in worsening health conditions, leading to serious consequences including deaths. So, healthcare staffs are often provided with a wide range of accesses such as "Break the Glass" or "Self-Authorization" mechanism for emergency access. So, in modelling and analyzing security practice in healthcare, there is a need to examine accesses to healthcare data to determine if the accesses were in line with the security policy. To address the state-of-the art, a systematic review was conducted to pinpoint appropriate AI methods and data sources that can be used for effective modeling and analysis of healthcare staffs' security practice. A framework was also proposed based on the review, to provide a comprehensive approach towards effective modeling and analyzing security practice of healthcare staff in the aspect of access logs. Methods: Out of about 130 articles, which were initially identified in the context of human-

*Fully documented templates are available in the elsarticle package on CTAN.

*Prosper K. Yeng

Email address: `prosper.yeng@ntnu.no` (Einar Arthur Snekkenes)

URL: `www.elsevier.com` (Einar Arthur Snekkenes)

generated healthcare data for security measures in healthcare, 18 articles were found to meet the inclusion and exclusion criteria. A thorough assessment and analysis of the included article reveals that, KNN, Bayesian Network and Decision Trees (C4.5) algorithms were mostly applied on Electronic Health Records (EHR) Logs and Network logs with varying input features of healthcare staffs' security practices. Based on the review results, a framework was proposed towards implementation. What was found challenging is the performance scores of these algorithms which were not sufficiently outlined in the existing studies.

Keywords: Artificial Intelligence, Machine Learning, Healthcare, Security Practice

1. Introduction

Security practice in general, involves the required actions of various actors(such as the end users, organizational and national and international measures) in the security chain towards safeguarding the confidentiality, integrity and availability(CIA) of information systems and technology resources[1]. An aspect of security practice include authorization and access control of assets and resources such as patients records, in the healthcare sector [2]. Unlike other sectors, the healthcare sector cannot afford to implement stricter access control for accessing sensitive healthcare information for therapeutic purposes. Much as there is the need to provide tighter security measures in controlling accesses, there is the need to strike a balance in order not to prevent legitimate accesses of healthcare data for therapeutic reasons [3]. In access control management in healthcare, access to personal health data and personal data filing systems for therapeutic purposes must be granted following a specific decision based on "the completed or planned implementation of measures for the medical treatment of the patient" [4]. So, access must only be granted to those with official need [5, 6]. While providing restrictions against unauthorized accesses, there are some provisions to comply with the availability trait of Confidentiality, Integrity and Availability (CIA) during emergency situations. These include the provision for self-authorization. Self-authorization or “Break-the-Glass” is a “technical measure which has been established for health personnel to be able to gain access to personal health data and personal data as and when necessary” [6]. However, self-authorization must be verified for abuse and clear misuse must be followed up as data breach [4, 2].

What is challenging is how to detect misuse in broad range of accesses [3, 6]. Broad range of accesses via self-authorization results in tones of variant data[7], making it complex to manually determine legitimate accesses. But in the light of recent increase in data breaches within healthcare, it has become necessary to adopt to state-of-the-art methods to understand the challenges often face by users in their effort to comply with security requirements. So, in Healthcare Security Practice Analysis, Modeling and Incentivization (HSPAMI) project [8], this data-driven was identified and adopted as one of the approaches. The role of data-driven and Artificial Intelligence (AI) approach in this study is to aid in modeling and analyzing healthcare staff’s security practices in their access control logs [8]. The intention is to understand anomaly practices in healthcare in the scope of data-driven and AI and to determine the security practice challenges often faced by healthcare workers while performing their duties. The results will provide knowledge to guide for finding better approaches to security practice in healthcare. But there are different type of data sources and AI methods which can be used in this approach [8]. A systematic review was therefore adopted to identify into details, the data sources and AI methods which can be adopted in related studies.

According to Verizon, the healthcare sector, globally, witnessed about 503 data breaches, which resulted in the compromise of up to 15 million records, in 2008 [9, 5]. This figure was a triple of that number of data breaches recorded in 2017. Also, the number of records compromised within the healthcare sector in 2019 far exceeds those recorded in 2018[9]. Unfortunately, more than half of these data breaches were perpetuated by insiders [9]. The report opined that[9] about (83%), of the adversaries were motivated by financial gains, (3%) was due to convenience, (3%) was also due to grudges and (2%) was as a result of industrial espionage. The current situations imply that the number of data breaches within the healthcare sector have surpassed that of the financial sector and almost equals those of other public sectors [9].

These situations have raised concerns among relevant stakeholders, and many are wondering the reasons behind the spike in the number of data breaches within the healthcare sector. Some of these reasons can be easily deduced because the healthcare data have economic value and as such a possible target for malicious actors [10, 11]. Also, healthcare data has scientific and societal values which makes them very attractive for cyber criminals. In fact, the report by [12] indicates that patient medical records are sold for about $1000 on the dark web. Another reason that could be attributed to the data breaches within the healthcare is that most healthcare personnel are more interested with their core healthcare duties and are ill-prepared to handle healthcare information security issues. This situation provides cyber criminals with the opportunity to exploit healthcare systems.

Although there have been improvements in technical measures such as firewall, intrusion detection and prevention systems, antiviruses and security governance configurations, the development of “human firewall” has not been considered [13, 14]. The “human firewall” refers to the information security conscious care behavior of the insiders [15]. It has not received equal attention like the technical measures and thus, cyber criminals seek to exploit it for easy access [16]. Healthcare insiders have access privileges which enable them to provide therapeutic care to patients, however, through errors or deliberate actions, they can compromise the confidentiality, integrity and availability (CIA) of healthcare data. Also, it is possible for an attacker to masquerade as insiders to compromise healthcare data through social engineering and other methods [17, 18].

It is usually the case that access control mechanisms within the healthcare sector are designed with a degree of flexibility to facilitate efficient patient management [19]. Even though such design considerations are very important and can meet the availability attribute of the CIA, they make the healthcare systems vulnerable. This is because the flexibility can be abused by the insiders

[19]. Also, an attacker who could obtain an insiders' access privilege can exploit this flexibility to have larger access. A successful data breach could have many consequences like a denial of service for timely medical services, corrode trust between the patient and healthcare providers, breaches to an individuals' privacy [20] and huge fines to healthcare providers by national and international regulatory bodies. The general objective of this study is to review for appropriate data sources and their features in addition to AI methods that can be used to determine irregularities in security practice among healthcare workers. The anomalies can be further processed to determine the dilemmas in security practice of healthcare staffs. The purpose is towards understanding the challenges often faced by healthcare staff in pursuance of their duties. The knowledge can be used to improve upon the security measures and develop effective incentives towards enhancing security in the healthcare sector.

2. Background

Security practice of healthcare staffs includes how healthcare professionals respond to the security controls and measures towards achieving the CIA goals of the healthcare organizations[3, 5, 7]. Healthcare professionals are required to conduct their work activities in a security conscious manner to maintain the CIA of healthcare environment [4]. For instance, borrowing of access credentials could jeopardize the purpose of access control for authorized users and legitimate accesses. Additionally, the inability to understand social engineering scammers' behavior can lead to healthcare data breaches [8].

Various ways can be adopted to observe, model and analyze healthcare professionals' security practices [8]. Perception and social and cultural theories can be adopted to analyze the healthcare staffs' security perception, social, cultural and socio-demographic characteristics against their required security practices [8, 16]. Also, Attack-Defense simulation can be used to measure how healthcare staffs understand social engineering related tricks [8, 16]. Furthermore, data-driven approach with artificial intelligence (AI) methods could be adopted to understand the security risk of each healthcare professionals [8, 16]. The findings can then help decision makers to introduce appropriate incentive methods and solve issues which are hindering sound information security practice towards enhancing conscious care behavior.

Advances in computational and data sciences along with engineering innovations in medical devices have prompted the need for the application of AI in the healthcare sector [21, 22] [21, 22]. This has the potential of improving care delivery and revolutionizing the healthcare industry. AI can be referred to as the use of complex algorithms and software to imitate human cognitive functions [23, 24]. It involves the application of computer algorithms in the process of extracting meaning from complicated data and to make intelligent decisions without direct human input [23, 22]. AI is increasingly impacting every aspects of our lives and the healthcare sector is not an exception. In recent years, the healthcare sector is experiencing massive deployments of AI in the bid to improve the overall healthcare delivery. However, we rely on the classification of the application of AI in healthcare described in [25] to briefly discuss deployment of AI in healthcare.

The deployment of AI in healthcare sector has been classified in [25] to include: expert systems, machine learning, natural language processing, automated planning and scheduling, and image and signal processing. Expert systems are AI programs that have been trained with real cases to execute complicated tasks [26]. Machine learning employs algorithms to identify patterns in data and learn from them and its applications can be grouped into three, namely: supervised learning, unsupervised learning, and reinforcement learning [25, 22]. Natural language processing facilitates the use of AI to determine the meaning of a text by using algorithm to identify key words and phrases in natural language [25]. For automated planning and scheduling, it is an emerging field in the use of AI in healthcare that is concerned with the organization and prioritization of the necessary activities in order to obtain desired aim [25] . And image and signal processing involve the use of AI to train information extracted from a physical occurrence (images and signals) [25] .

The common characteristics of all these applications is the utilization of massive data that is being generated in the healthcare sector to make better informed decisions. For instance, the collection of healthcare staffs' generated data, has been used for disease surveillance, decision support systems, detecting fraud and enhancing privacy and security [10]. In fact, the code of conduct for healthcare sector of Norway require the appropriate storage and protection of access logs of healthcare information systems for security reasons [4]. The healthcare staffs' accesses within the network or electronic health records (EHR), leaves traces of their activities which can be logged and reconstructed to form their unique profiles [4]. The healthcare staffs' accesses within the network or electronic health records (EHR), leaves traces of their activities which can be logged and reconstructed to form their unique profiles [5]. So, the appropriate AI methods can then be used to mine in such logs to determine the unique security practices of the healthcare staffs. Such findings can support management to adopt to the suitable incentivization methods towards improving on the security conscious care behavior in healthcare. Therefore, this study aims to explore for the appropriate AI methods and data sources that can be used to observe, model and analyzed the security practices of healthcare staffs.

Healthcare Security Practice Analysis, Modelling and Incentivization (HSPAMI), is an ongoing research project in which an aspect involves modelling and analyzing data with AI methods to determine the security practices of healthcare staffs, towards improving their security conscious care behavior. In analyzing healthcare related data, there is the need to consider details of the methods and data sources in view of the unique and critical nature of the sector. In a related study, Walker-Roberts et al., conducted a systematic review of "the availability and efficacy of countermeasures to internal threats in healthcare critical infrastructure" [27]. Among various teams few machine learning methods were identified to be used for intrusion detections and preventions. The methods that were identified are Petri net, Fuzzy logic, K-NN, Decision tree (RADISH system)[27, 28, 29] and inductive machine learning methods [27, 28, 30]. In a similar way, Islam et al conducted a systematic review on data mining for healthcare analytics [31]. Categories such as healthcare sub-areas, data mining techniques, type of analytics, data and data sources were considered in the study. Most of the data analysis were for clinical and administrative decision making. The data sources were mostly human generated from electronic health records. Other studies which explored for related methods includes [32] and [33].

Even though, the studies [27, 31] were in healthcare context, details of the algorithms and data sources were not considered. For instance, the features of the data sources and algorithm performance methods, were not deeply assessed in their studies. Additionally, the studies of [32] and [33] were general and not healthcare specific. So unique challenges within healthcare environment were not considered in their study. To this end, this study explored in detail, AI methods and data sources in healthcare that can be efficiently used for modeling and analyzing healthcare professionals' behavior. Healthcare professionals and healthcare staffs were used interchangeably in this study to include but not limited to nurses, physicians, laboratory staff and pharmacies who access patients records for therapeutic reasons.

See the front matter of this document for examples. You are recommended to conform your choice to the journal you are submitting to.

2.1. Scope, Problem Specification and Contribution

Following the recent increase in data breaches in healthcare, a study called Healthcare Security Practice Analysis, Modeling and Incentivization (HSPAMI) project has been initiated to understand the gaps in security practices of the healthcare professionals [8, 16]. The results will help to provide better ways of incorporating conscious care behavior among healthcare staffs. Various approaches were identified in HSPAMI study to include psycho-socio-cultural context [8, 16], Attack and defense simulations on social engineering context [8] and data driven using AI approaches [8].

The general objective of this study is to review for appropriate data sources and their features in addition to AI methods that can be used to determine irregularities in security practice among healthcare workers. The specific objectives of this study focused on reviewing the data-driven and AI approaches to identify, assess and analyze the state-of-the-art data-driven and artificial intelligence (AI) algorithms along with their design strategies, and challenges. The findings are towards analyzing healthcare professionals' security practices in the scope of data-driven and human generated data in Healthcare Security Practice Analysis, Modeling and Incentivization (HSPAMI) project. Based on the findings, a framework was then presented for future empirical studies in healthcare which involve AI, data-driven and security practice. So, psycho-socio-cultural context and attack-defense simulations are beyond the scope of this paper.

Having analyzed related studies [27, 28, 29, 30, 31], some details of data sources and AI methods that can be used in this study were not provided. For instance, amidst various data sources which are generated by healthcare staffs, which of them is most appropriate to be used in analyzing the security practice? Which AI methods have been pinpointed to be used for modeling and analyzing healthcare security practice? What evaluation techniques are most appropriate in this context and how were these methods adjusted to curtail biases in the midst of various access pointed such as self-authorization during emergency care scenarios and busy schedules of healthcare staff? The general objective of this study is to conduct a systematic review towards modeling and analyzing healthcare staff's generated access logs to enhance security practice. The specific objectives include identifying, analyzing, and assessing of data input sources and AI algorithms and related evaluation techniques. The remaining sections of this study has the method section which describes the approach in this review. This is followed with the result section which presents the findings in this study and the framework for future empirical studies. The findings and framework were then discussed in the discussion section. The entire study was then summarized in the conclusion section.

3. Methods

A literature search was conducted between June 2019 and December 2019 through Google Scholar Science Direct and Elsevier, IEEE Explore, ACM Digital. Different key words such as “Healthcare”, “staff”, “employee”, “Information security”, “behavior”, “Practice”, “Threat”, “Anomaly detection”, “Intrusion detection”, “Artificial Intelligence” and “Machine Learning “, were used. For a good quality searching approach, the key words were combined using Boolean functions of ‘AND’, ‘OR’ and ‘NOT’. For instance, the search string which was generated in PubMed is as follows;

((Intrusion[All Fields] AND Detection[All Fields]) OR (Anomaly[All Fields] AND Detection[All Fields])) AND (”health”[MeSH Terms] OR ”health”[All Fields]) AND ((”artificial intelligence”[MeSH Terms] OR (”artificial”[All Fields] AND ”intelligence”[All Fields]) OR ”artificial intelligence”[All Fields]) OR (”machine learning”[MeSH Terms] OR (”machine”[All Fields] AND ”learning”[All Fields]) OR ”machine learning”[All Fields])) AND (”information”[All Fields] AND Security[All Fields]) AND ((”behaviour”[All Fields] OR ”behavior”[MeSH Terms] OR ”behavior”[All Fields]) OR ”practice”[All Fields]) Peer reviewed journals and articles were considered. The inclusions and exclusions criteria were developed based on the objective of the study and through rigorous discussions among the authors.

Basic selection was done by initially skimming through the titles, abstracts, and keywords to retrieve records which were in line with the inclusion and exclusion criteria. Duplicates were filtered out and articles, which seems relevant, based on the inclusion and exclusion criteria, were fully read, and evaluated. Each of the authors independently read and assessed all the selected articles and judged to either be included or exclude. Using the inclusion and exclusion criteria as a guideline, discrepancies were discussed and resolved among the authors. Other appropriate articles were also retrieved using the reference list of accepted literature. Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram was used to report the article selection

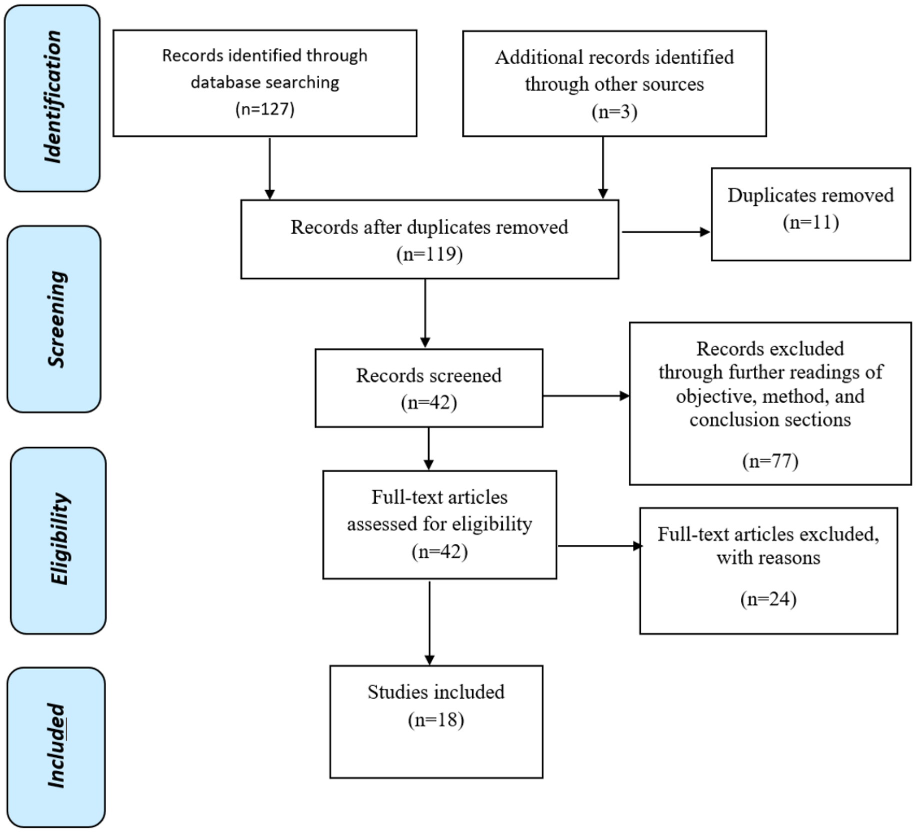

Identification
Screening
Eligibility
Included
Records identified through database searching (n=127)
Additional records identified through other sources (n=3)
Records after duplicates removed (n=119)
Duplicates removed (n=11)
Records screened (n=42)
Records excluded through further readings of objective, method, and conclusion sections (n=77)
Full-text articles assessed for eligibility (n=42)
Full-text articles excluded, with reasons (n=24)
Studies included (n=18)

Figure 1: Flowchart of the systematic review process[34].

and screening [34] as shown in Fig. 1.

3.1. Inclusion and Exclusion Criteria

For an article to be included in the review, the articles relating to anomaly detection or intrusion detection in healthcare using artificial intelligence methods in healthcare professionals' generated access logs data or patterns were included. Any other article outside the above stated scope (such as articles in medical cyber-physical devices, body area networks etc.) including literature in other languages, except English, were excluded.

3.2. Data Collection and Categorization

3.3. Literature Evaluation and Analysis

The selected articles were assessed, analyzed and evaluated, based on the above defined categories. The analysis was performed on each of the categories (Type of AI method, type of input, input source, prepossessing, learning techniques, performance methods etc.) to evaluate the state-of-the-art approaches. Percentages of the attributes of the categories were calculated based on the total number of counts (n) of each type of the attribute. Some studies used multiple categories, therefore, the number of counts of these categories exceeded the total number of articles of these systems presented in the study.

4. Results

After searching in the various online databases, a total of 130 records were initially identified by following the guidelines of the inclusion and exclusion criteria in the reading of titles, abstracts, and keywords. A further assessment of these articles through skimming of the objective, method and conclusion sections led to an exclusion of 77 articles which did not meet the defined inclusion criteria. After removing duplicates, 42 articles were fully read and judged. After the full text reading, a total of 18 articles were included in the study and analysis as shown in the Fig. 1. As shown in the Figure 2 & 3, the topic of data-driven and AI for analyzing healthcare security practice has seen consistent interest.

As shown in Fig. 2, most of the literature were identified in google scholar and followed by IEEE Explore and ACM Digital Library. The articles were published between 2010 and 2019 as shown in Fig. 3

4.1. Evaluation and Analysis

4.2. Algorithms

The algorithms which were found in the review are as shown in the table in figure 4. KNN method was mostly used (17%), followed by Bayesian Network (14%) and C4.5 decision tree (10%).

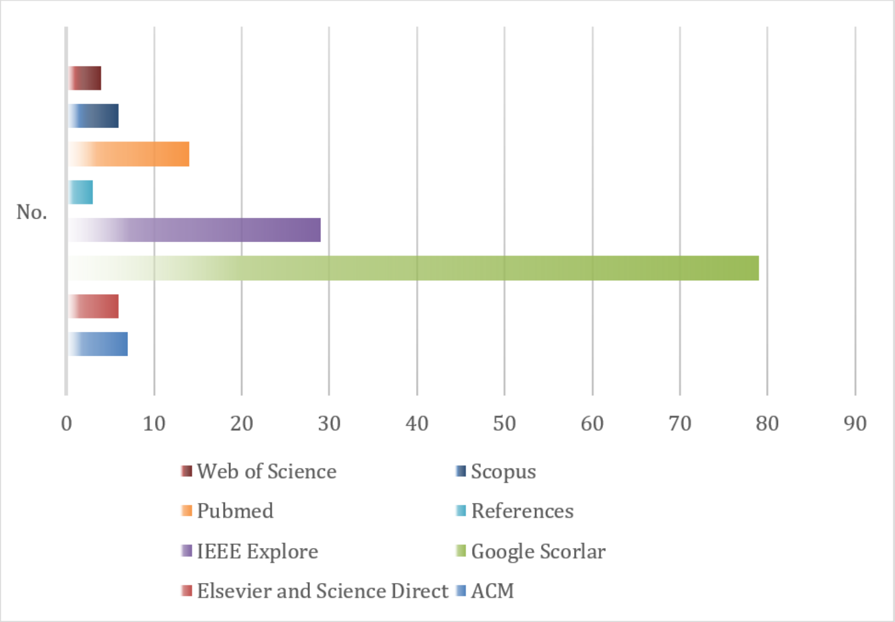


Figure 2: Literature Sources.

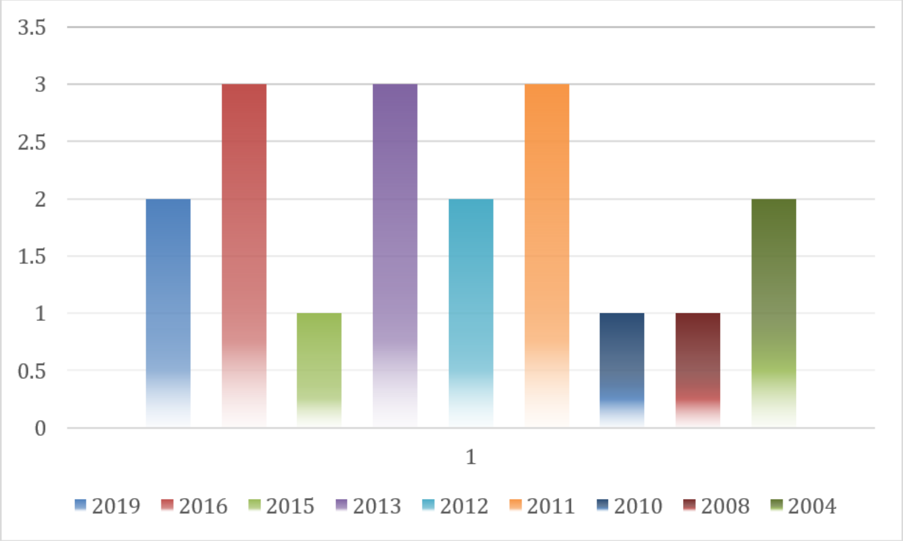
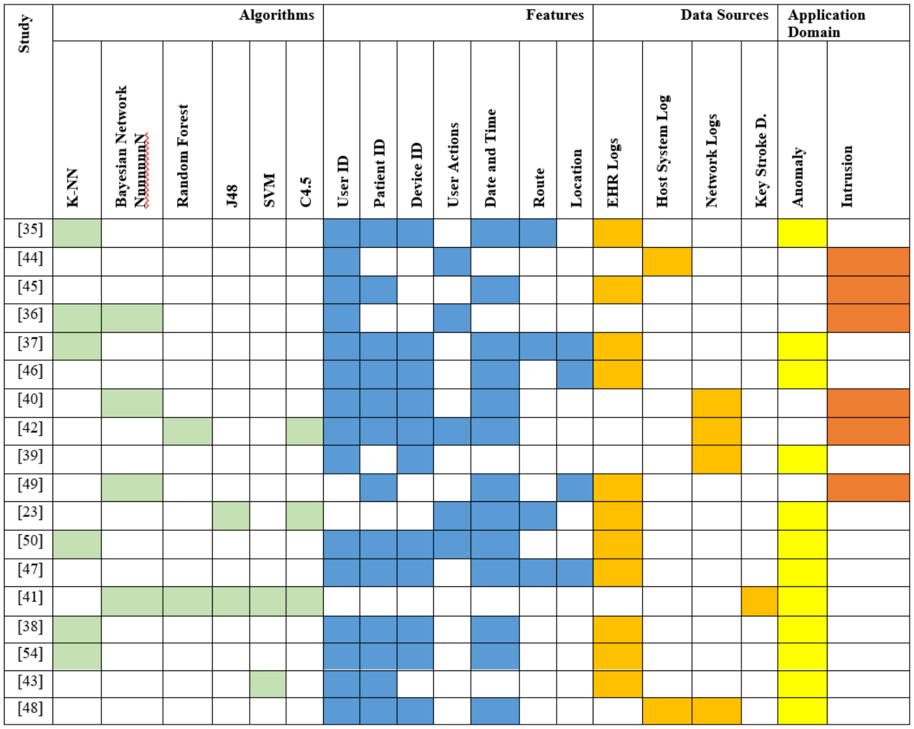


Figure 3: Yearly Distribution.



Study	Algorithms						Features							Data Sources				Application Domain	
	K-NN	Bayesian Network NnnnnnnN	Random Forest	J48	SVM	C4.5	User ID	Patient ID	Device ID	User Actions	Date and Time	Route	Location	EHR Logs	Host System Log	Network Logs	Key Stroke D.	Anomaly	Intrusion
[35]	■						■	■	■		■	■		■				■	
[44]							■			■					■				■
[45]							■	■			■			■					■
[36]	■	■					■			■									■
[37]	■						■	■	■		■	■	■	■				■	
[46]							■	■	■		■		■	■				■	
[40]		■					■	■	■		■					■			■
[42]			■			■	■	■	■	■	■					■			■
[39]							■		■							■		■	
[49]		■						■			■		■	■					■
[23]				■		■				■	■	■		■				■	
[50]	■						■	■	■	■	■			■				■	
[47]							■	■	■		■	■	■	■				■	
[41]		■	■	■	■	■											■	■	
[38]	■						■	■	■		■			■				■	
[54]	■						■	■	■		■			■				■	
[43]					■		■	■						■				■	
[48]							■	■	■		■				■	■		■	

Figure 4: Algorithms, Features, their related Data Sources and application domain

4.3. Performance Methods

Regarding performance methods as shown in Table 5, ROC (22%) Receiver Operating Characteristic ROC Curve (16%) and Recall (16%).

other findings in the review were the application scenarios of the AI methods, the data format of the input data and the ground truth which refers to the expected security practice. Other dimensions of the results are privacy preserving data mining approach and the nature of the data sources.

On the application scenario, the studies in the review were mostly applied for anomaly detection (60%) and Intrusion detection (40%).

Regarding file format, Comma separated values (CSV) was commonly used as the file format [39, 44]. Some studies also used SQL file format[45, 49].

In the review, the ground truth was being established with similarity measures [39, 36, 50], observed practices[44, 46] and historical data of staffs' practices[45]. Privacy preserving methods which were adopted in study are tokenization [39], de-identification [37] and removal of medical information [23]. The nature of the data sources which were used in the studies were mostly Real data (80%) and synthetic data (20%).

4.4. Framework Towards Implementation

Based on the review, a concept was depicted (as shown in figure 5) on how data-driven and artificial intelligence (AI) could be adopted to analyze logs of electronic health records (EHR) in security practice.

The concept primarily consists of data sources such as network, EHR or workstation logs as shown in figure 5. These logs are generated based on healthcare staff activities in accessing resources such as patients, printers, medical devices and physical security systems. The logs go through prepossessing phase [22] such as cleaning. The essential features are then selected with the appropriate methods including filter methods, wrapper methods or the combined filter and wrapper approach [22]. Having obtained the appropriate features, the ground truth can also be established. A machine learning method can then be created, trained and used to detect patterns of unusual security practices. The

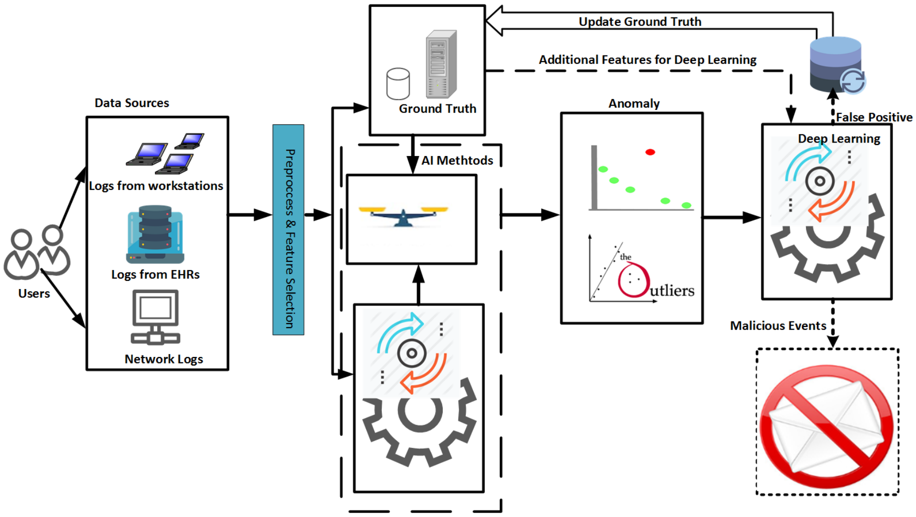


Figure 5: Concept of Data-Driven and AI for EHR log Analysis

outliers can feature be processed to determine if they were for malicious. The flow diagram of this concept is presented as the framework as shown in figure 6.

Figure 6 is a flowchart representation of the framework. The healthcare staffs' generated logs serve as input data sources to the data processing stage. At the data processing stage, data prepossessing such as data construction and cleaning and feature selection activities are performed [22]. The typical security practices of users are further established from the processed data in the "Build Normal Pattern" process. In addition, the ground-truth can also be established from the processed data. An anomaly detection model (which is incorporated with the suitable machine learning methods) would then be trained and used to detect anomalies from the test set labeled as "Get Current Events" in figure 6. Since abnormal security practices does not necessary concludes malicious acts, a further extermination would be conducted on detected anomalies. From Figure 6, the malicious detection aspect of the framework assesses the anomalies into details to conclude on whether the detected atypical security practices were malicious or not. Non-malicious events are deemed false alarms and the ground

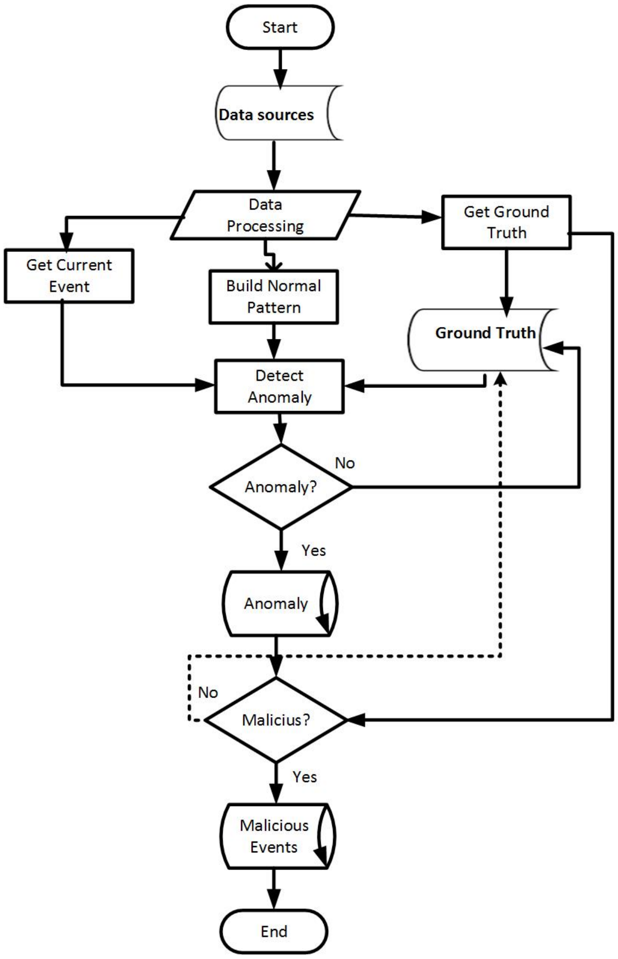


Figure 6: Flowchart framework for analyzing healthcare security practice in data-driven context

truth is updated.

5. Discussion

The main purpose of this systematic review was to find details of Artificial Intelligence (AI) methods and suitable healthcare staffs' generated security practice data, that can be efficiently mined to determine the status of healthcare security practices with respect to required security practices. The main findings in the study are as shown in Table 6. With reference to Figure 1, 2 and 3 and Table 1, there were 15 studies which met the inclusion and exclusion criteria. Recently, a related systematic review for countermeasures against internal threats in healthcare also found about 5 machine learning methods,[27] which were fit for such measures. This suggests that the adoption of AI methods for modeling and analyzing healthcare professionals' generated security practice data, is still an emerging topic of academic interest.

5.1. AI Methods

As shown in Table 2 and Table 6, various algorithms were identified in the study, but the most used methods were KNN and BN algorithms. K-Nearest Neighbors (kNN) is a supervised learning -based classification algorithm [46] which gets its intelligence from labeled data. The KNN then tries to classify unlabeled data item based on the category of the majority of most similar training data items known as K. The similarity between two data items in KNN, can be determined with the Euclidean distance of the various respective feature vectors of the data item. Another method which was mostly used is Bayesian Network (BN). BN is a probabilistic classifier algorithm, based on the assumption that, related pair of features used for determining an outcome are independent of each other and equal [46]. There are two commonly used methods of BN for classifying text, thus the multi-variant Bernoulli and multinomial models. KNN and BN algorithms were mostly used based on their comparatively higher detection accuracy. For instance, in an experimental assessment of KNN and

BNN for security countermeasures of internal threats in healthcare, both KNN and BN had over 90% accuracy. BN performed better (94%) than the KNN (93%). In a related study [27], the KNN method was found to have higher detection rate with high true positive rates and low false positive rate. The major issue with KNN in the context of healthcare staff security generated data is the lack of appropriate labeled data [51, 23, 35, 39]. Within the healthcare setting, emergencies often dictate needs. In such situations, broader accesses for resources are normally allowed, making it challenging for reliable labeled data [51, 23, 35, 39]. Therefore, in adopting KNN for empirical studies, the availability of appropriate labeled data should be considered but, in the absence of labeled data, unsupervised clustering methods such as K means clustering could also be considered[25].

5.2. input data

The input data which was mostly used include EHR logs and Network data. A study which was conducted by Yeng et al., for observational measures towards profiling healthcare staffs' security practices, identified various sources including EHR logs, browser history, network logs, and patterns of keystroke dynamics [8]. Most EHR systems uses an emergency access control mechanism, known as "break the glass "or self-authorization" [**?**]. This enables healthcare staffs, to access patients' medical records during emergency situations without passing through conventional procedures for access authorization. A study into access control methods in Norway [**?**] revealed that about 50% of 100,000 patients records were accessed by 12,298 healthcare staffs (representing about 45% of the users) through self-authorization. In such a scenario, EHR remains a vital source for analyzing for deviations of required healthcare security practices.

Regarding Ground Truth, it refers to the baseline, often used for training algorithms [52]. The detection efficiency of the algorithms can be negatively impacted if the accuracy of the ground-truth is low. As shown in Table 6, various methods such as similarity measures, observed data and historical methods were used. Similarity measure compares security practices with other healthcare professionals who have similar security practices. Observed measure is a control approach of obtaining the ground truth whereby some users were observed to conduct their security practices under a supervised, required security practices [41]. But the historical data basically relied on past records with a trust that, the data is reliable enough for training set. These methods can be assessed for adoption in related studies.

5.3. Features and data format

EHR contains most of the features which were identified in this review as shown in Table 6. Features such as patients ID, Actions, and User ID are primary features in EHR logs. The actions of the users such as deletion, inserting, updating and various routes such as diagnosis, prescriptions, and drugs dispensing can be tracked in EHR logs [**?**].

5.4. Application Scenario and Privacy preserving log analysis

The application of AI methods to analyze data-driven, generated by healthcare professional security practice, is a reactive approach. With such approaches, the primary aim is to determine deviations or outliers in healthcare security practices and further process these anomalies for possible malicious activities. As most of the algorithms were applied for anomaly detection (60%), such methods can be used to initially detect outliers. Deep leaning methods such as BN can then be used to further analyze the outliers for possible intrusions. This would help in privacy preserving at the same time while saving resources. Privacy preserving in data mining provides method to efficiently analyze data while shielding the identifications of the data subjects in a way to respect their right to privacy. For instance, limited number of less sensitive features can be used with KNN-based algorithms and if there exist outliers, BN methods can then be applied on only large number of the outliers to further assess these anomalies. In the review, deidentification, tokenization and sensitive data removals were some of the methods adopted to preserve privacy.

5.5. Framework for healthcare security practice analysis in data-driven approach

There are various challenges in analyzing healthcare staffs' security practice in the context of data-driven. Most prominent among them includes how to detect malicious activities with static rules. This is due to the provision of broad range of accesses of resources to healthcare staffs based on the critical nature of the sector. Imagine a patient record becomes inaccessible to a healthcare provider during an emergency. This will contravene the availability aspect of the confidentiality, integrity and availability (CIA) trait. While providing such broad accesses to healthcare personnel, there is therefore the need to be able to determine illegitimate accesses in modeling and analyzing security practices [4, 8, 16]. To mitigate false alarms, this framework initially determines for anomalous patterns of users. The anomalies include strange security practices of healthcare staff. The staff and their related anomalous security practices are further accessed with deep learning methods to determine if the anomalies were malicious or not as shown in Figure 5 and Figure 6.

5.6. Conclusion

Based on the galloping rate of data breaches in healthcare, Healthcare Security Practice Analysis, Modeling and Incentivization (HSPAMI) project was initiated to observe, model and analyze healthcare staffs' security practices. One of the approaches in the HSPAMI project is the adoption of AI methods for modeling and analyzing healthcare staffs' generated security practice data [8, 53, 54]. This systematic review was then conducted to identify, assess and analyze the appropriate AI methods and data sources. Out of about 130 articles which were initially identified in the context of human-generated healthcare data for security measures in healthcare, 15 articles were found to meet this inclusion and exclusion criteria. After the assessment and analysis, various methods such as KNN, Bayesian Network and Decision Trees (C4.5) algorithms were mostly applied on Electronic Health Records (EHR) Logs and Network logs with varying input features of healthcare staffs' security practices. A framework was also developed based on the review towards analyzing healthcare security practice in data-driven approach.

With these algorithms, security practice of healthcare staffs, can then be studied. Deviations of security practices from required healthcare staffs' security behavior can be examined to define appropriated incentives towards improving conscious care security practice. Analyzing healthcare staff security practice with AI seems to be a new research focus area and this resulted into the inclusion of only 15 articles in this study. Among these included articles, there were no adequate recorded performance scores. As a result, the study could not adequately perform a comparative assessment of the performance of the identified algorithms. Further work would include the detailed assessment of these algorithms towards practical analysis of real healthcare staffs' generated logs.

5.7. Abbreviations

AI: Artificial Intelligence AUC: Area Under the Curve BN: Bayesian Network CIA: Confidentiality, Integrity and Availability CSV: Comma Separated Values EHR: Electronic Health Records KNN K-Nearness Neighbor ROC: Receiver Operative Curve HSPAMI: Healthcare Security Practice Analysis Modeling and Incentivization SVM: Support Vector Machine

References

[1] M. E. Whitman, H. J. Mattord, Principles of information security, 6th Edition, Cengage Learning, 2017.

[2] L. O. Nweke, P. Yeng, S. Wolthusen, B. Yang, Understanding attribute-based access control for modelling and analysing healthcare professionals' security practices, International Journal of Advanced Computer Science and Applications 11 (2) (2020) 683–690. doi:10.14569/IJACSA.2020.0110286.

[3] L. Røstad, O. Edsberg, A study of access control requirements for healthcare systems based on audit trails from access logs, in: 22nd Annual Computer Security Applications Conference (ACSAC 2006), 11-15 December 2006, Miami Beach, Florida, USA, IEEE Computer Society, 2006, pp. 175–186. doi:10.1109/ACSAC.2006.8.

[4] e-helse, Direktoratet for, Code of conduct for information security and data protection in the healthcare and care services sector (2018).

[5] P. Yeng, A. Szekeres, B. Yang, E. A. Snekkenes, Framework for healthcare staffs' information security practice analysis: Psycho-socio-cultural context (preprint). doi:10.2196/preprints.17604.

[6] C. A. Ardagna, S. D. C. di Vimercati, S. Foresti, T. Grandison, S. Jajodia, P. Samarati, Access control for smarter healthcare using policy spaces, Comput. Secur. 29 (8) (2010) 848–858. doi:10.1016/j.cose.2010.07.001.

[7] E. Baro, S. Degoul, R. Beuscart, E. Chazard, Toward a literature-driven definition of big data in healthcare, BioMed Research International 2015 (2015). doi:10.1155/2015/639021.
URL http://dx.doi.org/10.1155/2015/639021

[8] P. K. Yeng, B. Yang, E. A. Snekkenes, Framework for healthcare security practice analysis, modeling and incentivization, in: 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, December 9-12, 2019, IEEE, 2019, pp. 3242–3251. doi:10.1109/BigData47090.2019.9006529.

[9] Verison, 2019 data breaches investigation report (2019).

[10] S. Chandra, S. Ray, R. T. Goswami, Big data security in healthcare: Survey on frameworks and algorithms (2017). doi:10.1109/iacc.2017.0033.

[11] C. Humer, J. Finkle, Your medical record is worth more to hackers than your credit card (2014).

[12] M. Garrity, Patient medical records sell for $1k on dark web (2019). URL https://www.beckershospitalreview.com/cybersecurity/patient-medical-records-sell-for-1k-on-dark-web.htmlhttps://www.beckershospitalreview.com/cybersecurity/patient-medical-records-sell-for-1k-on-dark-web.html

[13] S. D. Cannoy, A. F. Salam, A framework for health care information assurance policy and compliance, Commun. ACM 53 (3) (2010) 126–131. doi:10.1145/1666420.1666453.

[14] N. S. Safa, M. Sookhak, R. von Solms, S. Furnell, N. A. Ghani, T. Herawan, Information security conscious care behaviour formation in organizations, Comput. Secur. 53 (2015) 65–78. doi:10.1016/j.cose.2015.05.012.

[15] M. E. Whitman, P. Fendler, J. Caylor, D. Baker, Rebuilding the human firewall (2005). doi:10.1145/1107622.1107646.

[16] P. K. Yeng, B. Yang, E. A. Snekkenes, Healthcare staffs' information security practices towards mitigating data breaches: A literature survey, in: B. Blobel, M. Giacomini (Eds.), pHealth 2019 - Proceedings of the 16th International Conference on Wearable Micro and Nano Technologies for Personalized Health - 10-12 June 2019, Genoa, Italy, Vol. 261 of Studies in Health Technology and Informatics, IOS Press, 2019, pp. 239–245. doi:10.3233/978-1-61499-975-1-239.

[17] Network firewalls: Perimeter defense - dummies (2018).

[18] J. B. Predd, S. L. Pfleeger, J. Hunker, C. Bulford, Insiders behaving badly, IEEE Secur. Priv. 6 (4) (2008) 66–70. doi:10.1109/MSP.2008.87.

[19] A. McLeod, D. Dolezel, Cyber-analytics: Modeling factors associated with healthcare data breaches, Decision Support Systems 108 (2018) 57–68.

[20] J. Kwon, M. E. Johnson, The market effect of healthcare security: Do patients care about data breaches?, in: 14th Annual Workshop on the Economics of Information Security, WEIS 2015, Delft, The Netherlands, 22-23

June, 2015, 2015.
URL http://www.econinfosec.org/archive/weis2015/papers/WEIS_2015_kwon.pdf

[21] A. Shaban-Nejad, M. Michalowski, D. L. Buckeridge, Health intelligence: how artificial intelligence transforms population and personalized health, Nature Medicine 1 (2018). doi:10.1038/s41746-018-0058-9.

[22] I. Kononenko, M. Kukar, MACHINE LEARNING AND DATA MINING: Introduction to Principles and Algorithms, United Kingdom: Horwood Publishing Limited, 2007.

[23] T. Ziemniak, Use of machine learning classification techniques to detect atypical behavior in medical applications, in: H. Morgenstern, R. Ehlert, S. Frings, O. Göbel, D. Günther, S. Kiltz, J. Nedon, D. Schadt (Eds.), Sixth International Conference on IT Security Incident Management and IT Forensics, IMF 2011, Stuttgart, Germany, May 10-12, 2011, IEEE Computer Society, 2011, pp. 149–162. doi:10.1109/IMF.2011.20.

[24] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, Y. Wang, Artificial intelligence in healthcare: past, present and future, BMJ (2017).

[25] B. Wahl, A. Cossy-Gantner, S. Germann, N. R. Schwalbe, Artificial intelligence (ai) and global health: how can ai contribute to health in resource-poor settings?, BMJ Global Health 3 (2018) e000798. doi:10.1136/bmjgh-2018-000798.

[26] M. Vihinen, C. Samarghitean, Medical expert systems, Current Bioinformatics 3 (1) (2008) 56–65.

[27] S. Walker-Roberts, M. Hammoudeh, A. Dehghantanha, A systematic review of the availability and efficacy of countermeasures to internal threats in healthcare critical infrastructure, IEEE Access 6 (2018) 25167–25177. doi:10.1109/ACCESS.2018.2817560.

[28] B. Bose, B. Avasarala, S. Tirthapura, Y. Chung, D. Steiner, Detecting insider threats using RADISH: A system for real-time anomaly detection in heterogeneous data streams, IEEE Systems Journal 11 (2) (2017) 471–482. doi:10.1109/JSYST.2016.2558507.

[29] M. Gafny, A. Shabtai, L. Rokach, Y. Elovici, Detecting data misuse by applying context-based data linkage (2010). doi:10.1145/1866886.1866890.

[30] Y. Chen, S. Nyemba, W. Zhang, B. A. Malin, Specializing network analysis to detect anomalous insider actions, Security Informatics 1 (1) (2012) 5. doi:10.1186/2190-8532-1-5.

[31] M. Islam, M. Hasan, X. Wang, H. Germack, M. Noor-E-Alam, A systematic review on healthcare analytics: Application and theoretical perspective of data mining, Healthcare 6 (2018) 54. doi:10.3390/healthcare6020054.

[32] I. A. Gheyas, A. E. Abdallah, Detection and prediction of insider threats to cyber security: a systematic literature review and meta-analysis, Big Data Analytics 1 (2016). doi:10.1186/s41044-016-0006-0.

[33] I. Ghafir, M. Husák, V. Prenosil, A survey on intrusion detection and prevention systems (2014).

[34] PRISMA, Prisma (2018).
URL http://www.prisma-statement.org/

[35] A. Boddy, W. Hurst, M. Mackay, A. E. Rhalibi, Density-based outlier detection for safeguarding electronic patient record systems, IEEE Access 7 (2019) 40285–40294. doi:10.1109/ACCESS.2019.2906503.

[36] J. J. G. Adeva, J. M. P. Atxa, Intrusion detection in web applications using text mining, Eng. Appl. Artif. Intell. 20 (4) (2007) 555–566. doi:10.1016/j.engappai.2006.09.001.

[37] S. Gupta, C. Hanson, C. A. Gunter, M. Frank, D. M. Liebovitz, B. A. Malin, Modeling and detecting anomalous topic access, in: K. Glass,

R. Colbaugh, A. Sanfilippo, A. Kao, M. Gabbay, C. D. Corley, J. Li, L. Khan, A. Wynne, L. Coote, W. Mao, D. Zeng, A. Yaghoobi (Eds.), 2013 IEEE International Conference on Intelligence and Security Informatics, Seattle, WA, USA, June 4-7, 2013, IEEE, 2013, pp. 100–105. doi:10.1109/ISI.2013.6578795.

[38] Y. Chen, B. A. Malin, Detection of anomalous insiders in collaborative environments via relational analysis of access logs, in: R. S. Sandhu, E. Bertino (Eds.), First ACM Conference on Data and Application Security and Privacy, CODASPY 2011, San Antonio, TX, USA, February 21-23, 2011, Proceedings, ACM, 2011, pp. 63–74. doi:10.1145/1943513.1943524.

[39] A. Boddy, W. Hurst, M. Mackay, A. E. Rhalibi, A hybrid density-based outlier detection model for privacy in electronic patient record system (2019). doi:10.1109/infoman.2019.8714701.

[40] N. Amálio, G. Spanoudakis, From monitoring templates to security monitoring and threat detection, in: A. Cotton, O. Dini, A. F. Gómez-Skarmeta, M. Ion, M. Popescu, M. Takesue (Eds.), Proceedings of the Second International Conference on Emerging Security Information, Systems and Technologies, SECURWARE 2008, August 25-31, 2008, Cap Esterel, France, IEEE Computer Society, 2008, pp. 185–192. doi:10.1109/SECURWARE.2008.58.

[41] T. E. Wesolowski, P. Porwik, R. Doroz, Electronic health record security based on ensemble classification of keystroke dynamics, Applied Artificial Intelligence 30 (6) (2016) 521–540. doi:10.1080/08839514.2016.1193715.

[42] D. Pierrot, N. Harbi, J. Darmont, Hybrid intrusion detection in information systems, in 2016 International Conference on Information Science and Security (ICISS) (2016). doi:10.1109/icissec.2016.7885857.

[43] A. K. Menon, X. Jiang, J. Kim, J. Vaidya, L. Ohno-Machado, Detecting in-

appropriate access to electronic health records using collaborative filtering, Mach. Learn. 95 (1) (2014) 87–101. doi:10.1007/s10994-013-5376-1.

[44] T. A. Tchakoucht, M. Ezziyyani, M. Jbilou, M. Salaün, Behavioral apraoch for intrusion detection, in: 12th IEEE/ACS International Conference of Computer Systems and Applications, AICCSA 2015, Marrakech, Morocco, November 17-20, 2015, IEEE Computer Society, 2015, pp. 1–5. doi:10.1109/AICCSA.2015.7507118.

[45] E. Costante, D. Fauri, S. Etalle, J. den Hartog, N. Zannone, A hybrid framework for data loss prevention and detection, in: 2016 IEEE Security and Privacy Workshops, SP Workshops 2016, San Jose, CA, USA, May 22-26, 2016, IEEE Computer Society, 2016, pp. 324–333. doi:10.1109/SPW.2016.24.

[46] X. Li, Y. Xue, B. A. Malin, Detecting anomalous user behaviors in workflow-driven web applications, in: IEEE 31st Symposium on Reliable Distributed Systems, SRDS 2012, Irvine, CA, USA, October 8-11, 2012, IEEE Computer Society, 2012, pp. 1–10. doi:10.1109/SRDS.2012.19.

[47] H. Zhang, S. Mehrotra, D. M. Liebovitz, C. A. Gunter, B. A. Malin, Mining deviations from patient care pathways via electronic medical record system audits, ACM Trans. Management Inf. Syst. 4 (4) (2013) 17:1–17:20. doi:10.1145/2544102.

[48] A. Siraj, R. B. Vaughn, S. M. Bridges, Decision making for network health assessment in an intelligent intrusion detection system architecture, Int. J. Inf. Technol. Decis. Mak. 3 (2) (2004) 281–306. doi:10.1142/S0219622004001057.

[49] B. Asfaw, D. Bekele, B. Eshete, A. Villafiorita, K. Weldemariam, Host-based anomaly detection for pervasive medical systems, in: CRiSIS 2010, Proceedings of the Fifth International Conference on Risks and Security of Internet and Systems, Montreal, QC, Canada, October 10-13, 2010, IEEE Computer Society, 2010, pp. 1–8. doi:10.1109/CRISIS.2010.5764923.

[50] Y. Chen, S. Nyemba, B. A. Malin, Detecting anomalous insiders in collaborative information systems, IEEE Trans. Dependable Secur. Comput. 9 (3) (2012) 332–344. `doi:10.1109/TDSC.2012.11`.

[51] C. S. Gates, N. Li, Z. Xu, S. N. Chari, I. Molloy, Y. Park, Detecting insider information theft using features from file access logs, in: M. Kutylowski, J. Vaidya (Eds.), Computer Security - ESORICS 2014 - 19th European Symposium on Research in Computer Security, Wroclaw, Poland, September 7-11, 2014. Proceedings, Part II, Vol. 8713 of Lecture Notes in Computer Science, Springer, 2014, pp. 383–400. `doi:10.1007/978-3-319-11212-1\_22`.
URL `https://doi.org/10.1007/978-3-319-11212-1_22`

[52] P. Smyth, U. M. Fayyad, M. C. Burl, P. Perona, P. Baldi, Inferring ground truth from subjective labelling of venus images, in: G. Tesauro, D. S. Touretzky, T. K. Leen (Eds.), Advances in Neural Information Processing Systems 7, [NIPS Conference, Denver, Colorado, USA, 1994], MIT Press, 1994, pp. 1085–1092.

[53] P. K. Yeng, B. Yang, E. Snekkenes, Observational measures for effective profiling of healthcare staffs' security practices, in: V. Getov, J. Gaudiot, N. Yamai, S. Cimato, J. M. Chang, Y. Teranishi, J. Yang, H. V. Leong, H. Shahriar, M. Takemoto, D. Towey, H. Takakura, A. Elçi, S. Takeuchi, S. Puri (Eds.), 43rd IEEE Annual Computer Software and Applications Conference, COMPSAC 2019, Milwaukee, WI, USA, July 15-19, 2019, Volume 2, IEEE, 2019, pp. 397–404. `doi:10.1109/COMPSAC.2019.10239`.

[54] J. Kim, J. M. Grillo, A. A. Boxwala, X. Jiang, R. B. Mandelbaum, B. A. Patel, D. Mikels, S. A. Vinterbo, L. Ohno-Machado, Anomaly and signature filtering improve classifier performance for detection of suspicious access to ehrs, in: AMIA Annual Symposium Proceedings, Vol. 2011, American Medical Informatics Association, 2011, p. 723.

Table 1: Data categorization

No.	**Categorization**	**Definition**
1	Type of AI method	This category includes explicit machine learning methods such as, Support Vector Machine (SVM), Bayesian network, etc.
2	Type of Input	This category includes the features which were used by the algorithm. This could include access location, time, log in failed attempts etc.
3	Input Sources	This attribute refers to the kind of access logs data, which was used in the study. Such sources include browser history, network logs, host-based activity logs and electronic health records logs
4	Data Format, Type, Size, and Data Source	This category include file format such as XML, CSV
5	Input Prepossessing	Defines how the data was prepossessed from unstructured to structured, and how missing and corrupted input data were handled.
6	Application Scenario	This category defines the context of which the algorithm was implemented such as intrusion or anomaly detection.
7	Ground Truth	Refers to the kind of training set used in training the model.
8	Privacy approach	This defines the privacy method used to safeguard the privacy right of individuals who contributed to the data source.
9	Performance Metrics or Evaluation Criteria	This includes the measures used to assess the accuracy of the study. It includes metrics such as specificity, sensitivity, receiver operating characteristic (ROC) curves, and others
10	Nature of Data Sources	This category specifies if the data used was synthetic or real data.

Table 2: Algorithms and their respective proportions

Algorithm.	**Count**	%
K-Nearest Neighbors (KNN)[35, 36, 37, 30, 38]	5	17
Bayesian Network (BN)[39, 36, 40, 41]	4	14
C4.5[23, 42, 41]	3	10
Random Forest[42, 41]	2	7
J48[23, 41]	2	7
Principal Component Analysais(PCA) [38]	2	7
Decision Tree[38]	1	3
SVM[39, 43]	1	3
k-Means[44]	2	7
Spectral Project Method	2	7
Ensemble averaging and a human-in-the-loop model [39]	1	3
Partitioning Around Medoids with k estimation (PAMK) [42]	1	3
White-box anomaly detection system[45]	1	3
C5.0 [23, 42, 41]	1	3
Hidden Markov Model (HMM)[46]	1	3
Graph-Based[47]	1	3
Logistic Regression[43]	1	3
Linear Regression[43]	1	3
Fuzzy Cognitive Maps[48]	1	3

Table 3: Performance Methods

Performance Method.	**Count**	**%**
Receiver Operating Characteristic (ROC) Curve	7	26
Area Under ROC (AUC) curve	3	11
Recall (Sensitivity)	5	19
Precision	4	15
Accuracy	2	7
True Negative Rate/Specificity	3	11
F-Score	2	7
Root Mean Square Error(RMSE)	1	4

Table 4: Principal findings

Category.	**Most Used**
Algorithms	KNN and Bayesian Networks
Features	User IDS, Patient IDs, Device ID, Date and Time, Location, Route and Actions
Data sources	Electronic health Records (EHR) logs and Network logs
Application Domain	Anomaly Detection
Performance Methods	True Positive, False Positive, False Negative, ROC curve, AUC
Data Format	CSV
Nature of Data Sources	Real Data logs
Ground Truth	Similarity measures and observed data
Privacy preserving approaches	Tokenization and deidentification