A SYSTEMATIC REVIEW ON MACHINE LEARNING INSIDER THREAT DETECTION MODELS, DATASETS AND EVALUATION METRICS

Everleen Nekesa Wanyonyi[1,2], Silvance Abeka[2] and Newton Masinde[2]

[1]Department of Computer Science, Murang'a University of Technology, Murang'a, Kenya
[2]Department of Computer Science, Jaramogi Oginga Odinga University of Science and Technology, Bondo, Kenya

ABSTRACT

Computers are crucial instruments providing a competitive edge to organizations that have adopted them. Their pervasive presence has presented a novel challenge to information security, specifically threats emanating from privileged employees. Various solutions have been tried to address the vice, but no exhaustive solution has been found. Due to their elusive nature, proactive strategies have been proposed of which detection using Machine Learning models has been favoured. The choice of algorithm, datasets and metrics are cornerstones of model performance and hence, need to be addressed. Although multiple studies on ML for insider threat detection have been done, none has provided a comprehensive analysis of algorithms, datasets and metrics for development of Insider Threat Detection models. This study conducts a comprehensive systematic literature review using reputable databases to answer the research questions posed. Search strings, inclusion and exclusion criteria were set for eligibility of articles published in the last decade.

KEYWORDS

Insider threat, mitigation, Deep Learning, Machine Learning and Confusion Matrix

1. INTRODUCTION

The use of computers and the Internet has played a vital role in communication getting more and more ingrained in people's lives worldwide. The global economy now generates billions of dollars a year using the Internet's massive network. Currently, the majority of international economic, commercial, cultural, social, and governmental interactions and activities including those of individuals, non-governmental organisations and governmental institutions are conducted online [1]. The overreliance of organisational functions on cyberspace has made it a crucial component of the global social, political, and economic power hence, susceptible to disruption and manipulation [2]. This therefore makes cyberspace security to be one of the biggest threats to both public and national security since it compromises citizen safety and security and disturbs social and political order [1].

The concept of cyber-security encompasses the vulnerabilities and risks that arise from the new digital landscape as well as the strategies and protocols employed to establish a progressively safe environment. In the last decade, the issue of cybersecurity has emerged as a prominent concern. The surge in information usage has had implications on various aspects, including the protection of trade secrets, privacy considerations, and security concerns [3]. Incidences of

DOI: 10.5121/ijnsa.2023.15603

criminal activities are steadily on the increase with cybercrimes exerting a significant impact on individuals, enterprises, and even nations. A significant rise of unauthorized access of highly sensitive data by hackers, particularly within government and industry groups has led to various negative consequences such as fraud, espionage, and blackmail [4]. [5] cites the virtual conflict that occurred in 2008 between Georgia and Russia on the disputed region of South Ossetia and the cyberattack against Estonia in 2007 as significant cybercrimes that had profound societal consequences.

Threats to information security may originate from inside or outside of an organization. While outsider attacks emanate from external sources due to system vulnerabilities, insiders encompass employees, vendors, or other stakeholders with authorized access [3]. The prevailing cybersecurity challenges of the present era have shifted away from external threats and instead stem predominantly from trustworthy individuals within an organization. In the last three years, insider attacks have escalated to 68% of the total cybercrimes. The statistics are projected to be higher because most commercial entities choose to keep silent and endure the repercussions in order to safeguard their reputation and retain their valuable customers [6]. [7] has projected the rise of the vice to 77% of cybercrimes projecting devastating damages compared to external actors. Consequently, it is stated that the financial impact of insider attacks has risen to a staggering $11.45 million! In developing economies, over 90% of the countries acknowledge having suffered from insider attacks.

Insiders exhibit certain user traits that confer upon them a greater status compared to outsiders. [8] identifies trust, authorized access, system expertise and familiarity of the organization systems as key factors. This set of attributes, when combined with motive, elevates insiders to carry out significant attacks. For example, it is more probable for a system administrator to disclose confidential information since they have both trust and privileged access to systems. [9] define information security as a strategic method employed to protect digital information assets with the aim of attaining the fundamental security objectives, namely Confidentiality, Integrity, and Availability (CIA). Confidentiality regulates access to and disclosure of information, integrity prevents unauthorized modification or destruction of information while availability pertains to the assurance of timely and reliable access to and utilization of information [10].

To address insider threats, [9] propose a comprehensive strategy that incorporates technical and non-technical measures. The technical measures include Intrusion Detection System (IDS), Security Incident and Event Management (SIEM), Access Control Systems (ACS), Honey tokens, and Data Loss Protection (DLP) while non-technical approaches include psychology prediction models, Security Education Training and Awareness and information security policies. Despite the propositions, it has been widely acknowledged that a comprehensive solution to totally eradicate insider threats remains elusive [11]. The need for a proactive solution to insider threats is recommended. As argued by Alsowail and [12], proactive measures permits early detection of insider threats and hence become more effective. Machine Learning (ML) based insider threat detection solutions have been used successfully as proactive solutions [13]. Using user network behavioural characteristics data such as file transfer, browsing and logon, ML algorithms can be effectively trained to detect and classify anomalies in real-time. This means the choice of a dataset for a ML model is key to model performance.

While the choice of a data/dataset has been over emphasized in ML model development, the choice of a ML algorithm is also significant. [14] established the relationship between the dataset and ML algorithm. There exists several algorithms with different characteristics prompting [16] to highlight the need to be keen when selecting a ML algorithm. Consequently, the model's performance in the real world depends heavily on the types of evaluation metrics used to validate and demonstrate the model's robustness [15]. In view of these three factors, this research sought to reveal the most preferred ML algorithms, datasets and evaluation metrics used in the development of ITD models. This knowledge will be of great assistance to ML engineers especially when developing new or improving existing ML models for Insider Threat Detection (ITD).

2. METHODOLOGY

This study employed a Systematic Literature Review (SLR) approach done according to the research questions. Developing effective ITD models is an important step of solving the insider threat menace. To accomplish this, literature has revealed the importance of dataset, algorithm and evaluation metrics selection. This review was therefore done following the questions posed. Focusing on the last decade, the review begun with a general overview of insiders, insider threats and their mitigation strategies followed by looking at the evaluation metrics and datasets used for model development. This process was summarized into three primary stages; planning, implementation, and reporting, as depicted in Figure 1.

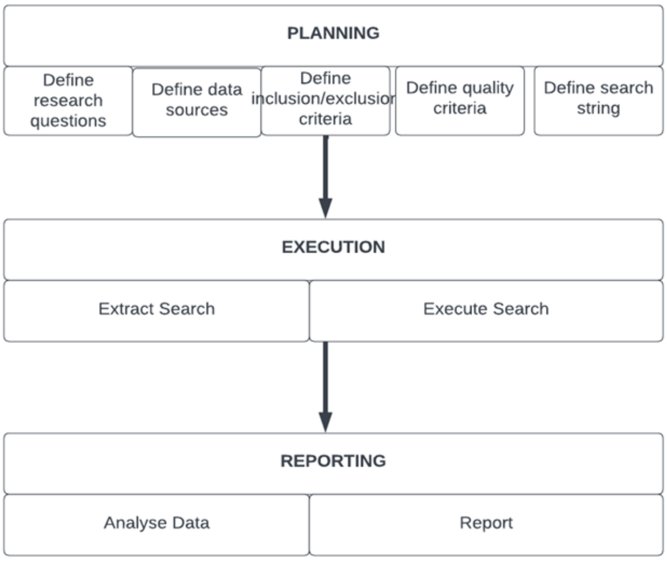

Figure 1. Systematic Literature Review Process [17]

2.1. Planning

There are five steps in the planning phase; definition of search questions, definition of data sources, specification of inclusion and exclusion criteria, definition of quality criteria and definition of search strings.

2.1.1. Definition of Search Questions

The objective of this review was to ascertain the widely used ML and DL based ITD algorithms, datasets and evaluation metrics. The following research questions guided the study.

i. What ML and DL algorithms have been widely used for ITD for the last decade?
ii. Which evaluation metrics have been widely used to validate the ITD models?
iii. What datasets have been preferred for training and evaluation of ITD models?

2.1.2. Definition of Data Sources

This study examined articles on insider threat detection models obtained from online databases, including ACM Digital Library, IEEE Xplore, ScienceDirect, Springer, and Wiley Online Library.

2.1.3. Inclusion/Exclusion Criteria Definition

Articles considered for the study were written in English, either proposed ML or DL techniques for ITD and were published within the past decade. Excluded articles did not meet the specified factors and in addition, they did not specify the datasets and metrics that were used during model development, they did not include empirical results of were a copy or earlier version of the one selected.

2.1.4. Definition of Quality Criteria

The abstracts of the studies were carefully reviewed and selected taking into consideration factors such as language, relevance, and model effectiveness. Furthermore, the inclusion characteristics that were previously stated played a role in facilitating the assessment of quality.

2.1.5. Definition of Search Strings

The study employed the key search terms "insider threat detection," "insider threat mitigation", and "insider threat detection techniques."

2.2. Execution

Search execution and data extraction were the two phases in this stage

2.2.1. Execute Search

The three search strings were employed to ascertain the pertinent scholarly articles from the five designated databases. A total of one thousand publications were obtained by considering the initial two hundred papers from each of the databases. Exclusion and inclusion criteria were applied to the one thousand publications reducing them to 535 articles. Eighty five publications were then selected for comprehensive assessment following the exclusion of research papers that were deemed irrelevant, duplicates, and studies with unavailable full texts. In addition, the review process involved the exclusion of selected papers that lacked explicit datasets and evaluation metrics. As a result, the total number of publications considered for analysis was reduced to 22, as depicted in Figure 2.

2.2.2. Data Extraction

The twenty two publications selected for the SLR are as summarized in Table 1.

2.3. Reporting

In Section 4.1 and 4.2, an examination was conducted on several ML and DL models that have been used in the context of ITD. A total of twenty-two studies were assessed and synthesized with respect to the prevailing algorithm(s), datasets and evaluation metrics. The findings pertaining to the study research questions are presented in the discussion section.

3. TAXONOMY OF INSIDERS AND INSIDER THREATS

3.1. Taxonomy of Insiders

[18] categorizes insiders into six types; the careless insider: a negligent employee who causes a breach of confidentiality unintentionally with no incentive to violate internal information security rules; the naive insider: an employee who is susceptible to the manipulative tactics employed by social engineers and other individuals with malevolent intent; the saboteur: an employee who engages in actions intended to do harm to the organization as a result of personal dissatisfaction or discontentment; the disloyal insider: malicious insiders who plan to leave the organization without informing fellow workers; the moonlighter: those who engage in the unauthorized acquisition of information with the intention of transmitting it to their clients while making efforts to conceal their illicit activities while moles: are highly covert operative strategically embedded within an organization, tasked with surreptitiously acquiring confidential information for the benefit of the sender.

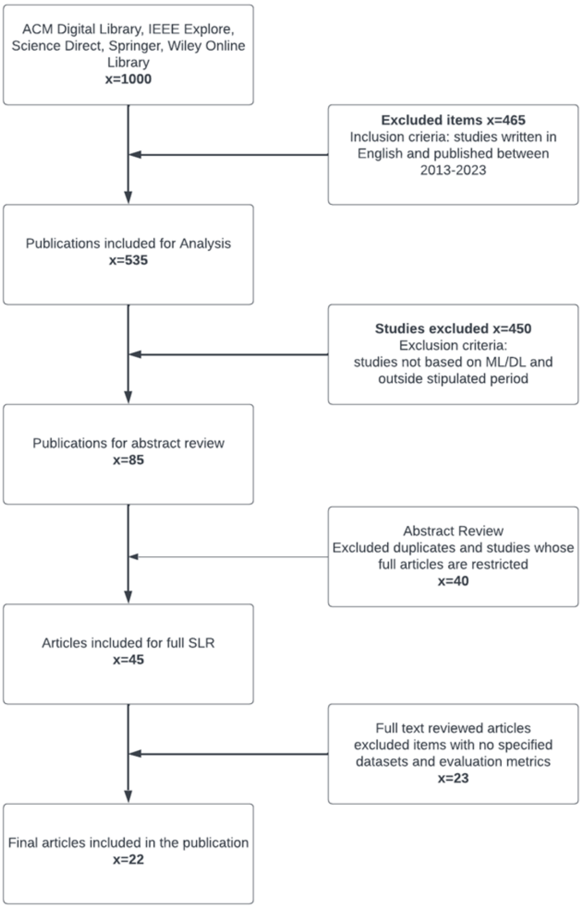
ACM Digital Library, IEEE Explore, Science Direct, Springer, Wiley Online Library
x=1000
Excluded items x=465
Inclusion crieria: studies written in English and published between 2013-2023
Publications included for Analysis
x=535
Studies excluded x=450
Exclusion criteria:
studies not based on ML/DL and outside stipulated period
Publications for abstract review
x=85
Abstract Review
Excluded duplicates and studies whose full articles are restricted
x=40
Articles included for full SLR
x=45
Full text reviewed articles excluded items with no specified datasets and evaluation metrics
x=23
Final articles included in the publication
x=22

Figure 2. SLR Search Execution algorithm adopted from [17]

Table 1. Selected ML and DL articles for SLR

Article Reference Number	**Author(s)**	**ML/DL**	**Year of publicati on**
20	Almehmadi, A., & El-Khatib	ML	2014
43	Al-Mhiqani et al.	DL	2020
12	Alsowail, R. A., & Al-Shehari, T.	ML	2021
42	Chattopadhyay, P., Wang, L., & Tan, Y.-P.	DL	2018
35	Janjua, F., Masood, A., Abbas, H., & Rashid, I.	ML	2020
33	Kandias, M., Stavrou, V., Bozovic, N., &Gritzalis, D.	ML	2013
34	Kim, J., Park, M., Kim, H., Cho, S., & Kang, P.	ML	2019
44	Koutsouvelis, V., Shiaeles, S., Ghita, B., &Bendiab, G.	DL	2020
6	Lu, J., & Wong, R.	DL	2019
40	Moradpoor, N., Brown, M., & Russell, G.	DL	2017
37	Padmavathi, G., Shanmugapriya, D., &Asha, S.	ML	2022
45	Paul, S., & Mishra, S.	DL	2020
39	Peccatiello, R. B., Gondim, J. J. C., & Garcia, L. P. F.	ML	2023
38	Ganapathi, 2023	ML	2023
36	Sav, U., &Magar, G.	ML	2021
46	Sharma, B., Pokharel, P., & Joshi, B.	DL	2020
47	Singh, M., Mehtre, B. M., Sangeetha, S., &Govindaraju, V.	DL	2023
41	Tuor, A., Kaplan, S., Hutchinson, B., Nichols, N., & Robinson, S.	DL	2017
11	Wei, Y., Chow, K.-P., &Yiu, S.-M.	DL	2021
16	Zala, M.	ML	2023
32	Zhang, D., Zheng, Y., Wen, Y., Xu, Y., Wang, J., Yu, Y., &Meng, D.	DL	2018
14	Zheng, P., Yuan, S., & Wu, X.	ML	2022

[19] summarizes insiders into two major groups; intentional and unintentional. Based on the dangers they pose to systems, they are also classified as malicious, careless and moles.[20], [21] group them into pawns: individuals who are deceived into engaging in malevolent actions; goofs: inept or arrogant staff who think they are exempt from security regulations; collaborators: engage in cooperative efforts to commit unlawful acts with other entities, such as competing enterprises

or foreign governmental bodies, and wolfs: who autonomously and maliciously without outside assistance or manipulation breach security policies for financial gain.

3.2. Classification of Insider Threats

[22] define an insider threat as the deliberate misuse of privileges and the violation of an organization's information security policy by individuals with authorized access. The identification of these risks poses a significant challenge due to their diverse nature and their resemblance to the benign activities [7]. Five categories of insider threats are discussed.

3.2.1. Fraud

[7] identifies fraud as a common insider threat that holds the highest prevalence rate, accounting for 61% of occurrences. Fraud encompasses a spectrum of illicit actions, ranging from basic misappropriation of organizational cash to intricate schemes involving the illicit exchange of organizational data for personal benefit [23].

3.2.2. Intellectual Property (IP) Theft

Malicious insiders engage in the unauthorized acquisition of important firm data, including trade secrets, programming code, and customer information, for a diverse array of purposes. This phenomenon is prevalent among individuals who have access to the aforementioned information [23]. The vice specifically focuses on source codes, product information, and proprietary software [24].

3.2.3. Information Technology (IT) Espionage

Referred to as cyber espionage or cyber spying, this phenomenon encompasses the illicit acquisition of personal, sensitive, or proprietary information from persons without their awareness or explicit authorization [25]. This is a prevalent risk that can be perpetrated by any employee [26].

3.2.4. Information Technology sabotage

A highly sophisticated threat majorly committed by insiders with sophisticated IT skills. The occurrence of this threat is predominantly attributed to insiders that possess advanced IT skills [7]. The act of IT sabotage necessitates the possession of privileged access to systems or networks, as well as a comprehensive understanding of their configuration. The attack encompasses a variety of malicious activities, including the introduction of malware, worms, or Trojans, as well as the tampering and interruption of information resources [23].

3.2.5. Unintentional/Accidental Threats

Committed by insiders with authorized access to an organization’s network, system, or data and who act without malicious intent and unwittingly causes harm or substantially increases the probability of future breach against CIA of the organization’s information system resources [27].

4. OVERVIEW OF MACHINE LEARNING AND DEEP LEARNING

In the realm of computing and data analysis, ML, a branch of artificial intelligence (AI), has undergone substantial advancements in recent years. This has facilitated the development of intelligent programs capable of performing tasks with enhanced efficiency and effectiveness. The ML technology is currently regarded as the widely embraced innovation in the context of the fourth industrial revolution (4IR). This is because it enables systems to autonomously acquire knowledge and enhance their performance through experiential learning, without the need for explicit programming [28]. The objective is to provide the machine with a substantial number of instances and allow it to learn autonomously [29]. For example, by providing several methods of representing the numeral 4, the machine will gradually acquire a high level of proficiency in recognizing it.

There are four primary classifications of ML, namely supervised, unsupervised, semi-supervised and reinforcement learning. Supervised learning, sometimes referred to as predictive learning, is the utilization of a labelled dataset to acquire knowledge of a mapping function from input variables to corresponding output variables. Supervised learning is employed for tasks involving classification and regression while unsupervised learning, also known as descriptive learning, aims at discerning patterns within unlabelled data. Reinforcement learning acquires knowledge by actively engaging with the environment and assimilating information from the outcomes of the agent's activities [30].

Deep Learning (DL), a subfield of Machine Learning (ML), places emphasis on Neural Networks (NN) that consist of multiple layers, in contrast to ML which does not incorporate such layers. The design and functionality of these Deep Neural Networks (DNN) emulate the structure and functioning of the human brain, hence facilitating the computational processing and analysis of enormous quantities of complex and unorganized data [29]. ML typically relies on relatively limited datasets with a well-defined structure, whereas DL utilizes enormous quantities of unstructured data to perform intricate computations. According to [28], DL models exhibit superior overall performance compared to ML models after undergoing training.

In order to safeguard the intricate networks and important data of enterprises against internal threats, researchers have adopted ML and DL techniques [16]. According to [30], ML solutions have a higher probability of obtaining favourable outcomes compared to legacy systems. This is because of the capacity to analyze data from multiple perspectives, optimize resource allocation, and automate repetitive tasks with a higher level of accuracy compared to their counterparts [31]. Conversely, DL techniques have experienced a surge in popularity as a result of their capacity to handle complex non-linear problems and also effectively leverage vast amounts of data, particularly in the current era of the Internet of Things (IoT) [32].

The subsequent section employs the timeline spanning from 2013 to 2023 to delineate the utilization of ML and DL algorithms in the context of ITD. Additionally, a discussion of various datasets employed for training and testing the aforementioned models is done. Evaluation metrics employed for analysing the efficacy of such models are also discussed.

4.1. Machine Learning Models for ITD

A model that combines supervised ML techniques, namely Naïve Bayes Multinomial (NBM), Support Vector Machines (SVM), and Logistic Regression (LR), to assess individuals' attitudes towards authorities and law enforcement using comments posted on the social media platform, YouTube is proposed. According to [33], social media offer the capability to monitor user behavior and collect their digital presence, and interpret their attitudes as expressed through their videos, comments, and likes. This information is likely to show the probability of users to attack their organization. This model uses real datasets which were collected using the REST-based API provided by YouTube. For seven years, the data collected comprised of 12,964 individuals, 207,377 videos, and 2,043,362 comments. Precision, Recall, F-Score, and Accuracy were used as evaluation metrics.

[20] propose a model that makes use of supervised Nearest Neighbor (NN) and Functional Tree (FT) classifiers for the purpose of detecting insider threats. This models uses physiological signal monitoring data in contrast to user behavioural data on the premise that, in contrast to, bio-signals they possess an inherent resistance to imitation or alteration. The model, known as the Physiological Signals Monitoring (PSM) is capable of detecting incidents within seconds before to their occurrence. PSM uses a typical fluctuation rate of skin temperature, Galvanic Skin Response (GSR), and electrocardiogram (ECG) amplitude that manifests moments prior to the execution of an event. The dataset for this experiment was collected from 15 people, encompassing both male and female individuals aged between 18 and 35 years and Accuracy was the main evaluation metric.

An "Insider Threat Detection Model Based on User Behavior Modelling and Anomaly Detection Algorithms" was proposed [34]. In this work, Principal Component Analysis (PCA), K-means Clustering (KMC), Gaussian Density Estimation (Gauss), and Parzen Window Density Estimation (Parzen) were integrated as one-class classification methods for the purpose of ITD. E-mails from the CERT r6.2 dataset were used for training and validation. An analysis was conducted on the content of emails and the network communications associated with emails in order to identify any irregularities. The primary assessment metric employed in this investigation was the detection rate.

A study that proposes the use of a supervised ML predictive technique that employs language analysis to assess the risk level of employees based on their email communication was done [35]. The study used TWOs dataset for training and testing. A comparative analysis was conducted using Adaboost, Naive Bayes (NB), Logistic Regression (LR), KNN, Linear Regression (LR), and Support Vector Machine (SVM) algorithms. The results indicated that Adaboost exhibited superior performance in terms of detection accuracy and AUC when compared to the other algorithms.

To detect insider threats based on the analysis of unusual behavior exhibited by individuals within an organization, a model that uses K-Nearest Neighbors (KNN), Histogram-based Outlier Score (HBOS), Local Outlier Factor (LOF), and Principal Component Analysis (PCA) was proposed. The study utilized temporal logon/off and device usage data for individuals affiliated with the CERT r4.2 dataset. The objective was to determine anomalies in user behavior by analysing the mean and mode of login occurrences. This process was facilitated by utilizing the pyOD library [36]. Accuracy of anomaly detection was the only evaluation metric.

The “Insider Data Leakage Detection Using One-Hot Encoding, Synthetic Minority Oversampling, and Machine Learning Techniques” is proposed. The model incorporates a combination of ML methods, including Logistic Regression (LR), Decision Trees (DT), Random Forests (RF), Naive Bayes (NB), k-Nearest Neighbors (k-NN), and Kernel Support Vector Machines (KSVM). The model also used Synthetic Minority Oversampling Technique (SMOTE) to address data imbalance in the dataset. The training and validation process utilized the CERT r4.2 dataset [12]. The primary evaluation metric employed in this study was the AUC-ROC.

A model that utilizes supervised learning approaches to detect instances of hostile insider threat is proposed. This model uses One Class Support Vector Machine (OCSVM) technique for classification. The CMU CERT r4.2 dataset was utilized for model training and testing. The detection rate of the model was used as the main evaluation metric. Though the proposed model achieved a commendable detection rate, DL techniques were proposed for future ITD model development [37].

A real-time ITD model utilizing Dirichlet Marked Hawkes Processes (DMHP), a non-parametric Bayesian, is proposed. The approach is specifically designed to address the challenge of imbalanced datasets in insider threat detection. The Dirichlet process possesses the ability to detect an infinite number of patterns within an infinite set of user activities, whereas the Marked Hawkes process aids in the modelling of user activities based on both temporal and activity type factors. Activities with a high probability indicate innocuous usage, whereas activities with a low probability suggest malicious attacks. The study utilized the CERT r4.2 dataset and AUC was used as the evaluation metric [14].

[38] presents a novel double-layer design for the identification of hostile insider threats. The proposed double-layer architecture aims to mitigate the potential issue of class imbalance during the detection of insider threats by employing sampling techniques. The study uses SVM as a classifier on top of Isolation Forest (IF), Random Forest (RC), Local Outlier Factor (LOF) and K-Nearest Neighbour (K-NN) to identify malicious behavior. The model was trained using the CMU-CERT r3.2 dataset, and its performance was evaluated using recall, f_1 score, and Accuracy metrics.

A comprehensive model for insider threat detection incorporating a range of data science approaches is proposed. This model utilizes supervised and semi-supervised ML methods, as well as data stream analysis and routine retraining approaches. The Elliptic Envelope, LOF, and IF are used. The CERT r4.2 provided by Carnegie Mellon University was the main dataset and evaluation included the assessment of precision, recall, and f_1 Score metrics [39].

In the study conducted by [16], a proposed architectural framework is presented for insider threat detection within the confines of an organization's or corporate network. This research presents a comprehensive strategy for implementing ML approaches in the ITD problem. Supervised ML models, such as logistic regression (LR), NN, RF, and extreme gradient boosting (XG) are employed for the purpose of IT threat detection. The model was trained on the CERT r4.2 insider threat dataset while evaluation metrics included detection rate, recall, precision, and false alarm rate.

4.2. Deep Learning Models for ITD

[40] employed Self-Organizing Maps (SOM) and PCA to construct a model aimed at identifying insider threats in organizational settings. The main dataset utilized for the training and evaluation of this model was obtained from ZoneFox. This is a .csv file containing real data with a total of 2643 records, each consisting of 8 distinct features. Clustering accuracy was the main evaluation metric for this study.

A DL model designed for the purpose of unsupervised ITD in structured cybersecurity data streams was proposed [41]. The researchers employed a hybrid approach, integrating (DNN) and Long Short-Term Memory (LSTM) models, to effectively identify anomalies in real-time. PCA was also used for feature extraction which consolidated the many actions into a unified vector. These vectors were then fed into the DNN and RNN for analysis. The training and evaluation of the model involved the utilization of the CERT r6.2 dataset, with the main evaluation metric being Cumulative Recall.

A classifier created by [32] utilizes a NN model using LSTM to categorize data according to roles. This is achieved by representing user logs as sequences of spoken language. The LSTM model learns the user activity patterns through automated feature extraction, enabling it to detect anomalies when log patterns deviate from the learned model. Precision, Recall, and Accuracy were the main evaluation metrics and the CMU CERT r3.2 was the dataset.

[42] present a methodology called "Scenario-Based Insider Threat Detection from Cyber Activities" that specifically emphasizes the use of behavioural activities. The CERT r4.2 dataset is utilized to derive a time series feature vector based on the actions and behaviours of individual users over a single day. The time-series feature set is classified after balancing the feature set by randomly undersampling the non-malicious class samples. In this study, a Deep Autoencoder Neural Network is employed and its performance is compared with that of the Random Forest and the Multi-Layer Perceptron (MLP) algorithms. Recall and precision served as performance metrics.

The "Insider Catcher", a model that utilizes a DNN and LSTM for the purpose of detecting insider threats was proposed. The model operates based on a log-based anomaly detection technique, wherein it captures logs of user activity considered typical. These logs are subsequently compared with everyday operations in order to detect deviation. LSTM has gained popularity due to its ability to retain information on long-term dependencies within data, hence facilitating the analysis of interrelationships among different data points [6]. In order to assess and verify the model's performance, the researchers employed the CERT r3.2 Insider Threat Dataset, which consisted of 4000 users. Accuracy was the evaluation metric.

A model that aims at detecting insider threats by analysing user behavior is proposed. This study employed GRU (a variant of LSTM) for anomaly detection. The CERT r4.2 dataset, which encompasses a greater number of insider attack incidents compared to its preceding iterations, was used. This study conducted by [43] examined various log files, including logon, file, device, HTTP, email, psychometric, and LDAP logs. The evaluation metric utilized to assess the performance of these log files was accuracy.

The development of an ITD model using Convolutional Neural Network (CNN) method is proposed by [44]. This technique was employed to train the model in the identification of insider threats, utilizing a dataset consisting of photographs. The implementation of this model was carried out using the CERT r4.2 insider threat dataset. The data pertaining to user logins, emails, web browsing history, device data, and user role data were transformed into visual representations. The primary evaluation criterion employed in the study was Accuracy.

A model that uses LSTMAutoencoder to replicate the behaviours of individual employees by analysing their day-to-day activities through time-stamped sequences is proposed. The model, referred to as LAC (LSTM AUTOENCODER with Community), consists of two distinct phases: the community detection phase and the LSTM AUTOENCODER RNN model phase. The LAC model was trained and tested using the CERT r6.2 dataset, which is an open source dataset given by Carnegie Mellon University. The evaluation of the developed model was solely based on the metric of Accuracy [45].

An LSTM-based Autoencoder model was employed for modelling user behavior and session activities, with the objective of identifying data points that deviate significantly from the norm for anomaly detection. The dataset utilized for training and validation purposes is the CERT r4.2 dataset and Accuracy, Recall, and False Positive rate were the evaluation metrics [46].

A novel prediction model that utilizes an unsupervised anomaly detection approach incorporating Cascaded Auto-Encoders (CAEs), Bidirectional Long Short-Term Memory (BiLSTM), and joint optimization network is proposed [11]. This model operates in a proactive manner and provides real-time predictions. The application of feature extraction and density estimation network is employed for the purpose of data purification and optimization, with the aim of mitigating sub-optimal issues. The model underwent testing and validation using the CERT insider threat dataset r6.2 and Recall, Precision and f1 score were the main evaluation metrics.

[47] proposed a model known as "User Behaviour based Insider Threat Detection using a Hybrid Learning Approach”. The proposed model uses a Feed-Forward Artificial Neural Network (FF-ANN), incorporating distance measurements to perform feature selection. Additionally, a Bi-LSTM is employed for anomaly detection. Finally, a SVM is used to classify users into either the benign or malicious categories. Model training and validation was performed using the CERT r4.2 insider threat dataset with Accuracy, Precision, F-measure, and AUC-ROC as evaluation metrics.

5. DISCUSSION

The resolution of the ITD problem has undergone a gradual transformation over its history. Despite the advancements made, empirical research indicates that the issue remains unresolved [24]. The conventional approaches, such as employee screening, security-in-depth, deterrent, and risk management, have proven ineffective in managing the issue at hand, mostly due to the elusive nature of the problem. Furthermore, the implementation of defence-in-depth measures is deemed too costly for the majority of small and medium-sized enterprises [4].

In contrast to earlier studies that employed reactive solutions, recent years have seen the emergence of proactive techniques. Among the proposed remedies, detective tactics have been favoured due to their effectiveness in both prevention and real-time mitigation. In contrast to approaches that rely on analysing digital footprints, proactive strategies are effective because of their preventive nature. The utilization of Artificial Intelligence (AI), from which ML and DL derive, has brought significant transformations in the field of ITD. Currently, both technologies are being widely employed in addressing the vice. Based on the aforementioned review, a number of conclusions can be inferred, which will be further expounded upon in the subsequent sections.

5.1. ML Insider Threat Detection Models

The Fourth Industrial Revolution (4IR or Industry 4.0) has ushered in an era where the digital realm is teeming with a vast array of data. This data encompasses various sources such as the Internet of Things (IoT), cybersecurity, mobile devices, businesses, social media, and health. ML models have played a significant role in the process of inferring meaning from data [28]. Literature demonstrates that since 2013, ML models have been increasingly popular for their exceptional performance in detecting insider threats, surpassing their legacy counterparts. This phenomenon can also be ascribed to the observation that ML models, in comparison to previous solutions, have the capability to handle slightly larger volumes of data and eliminate the requirement for manual intervention in order to detect anomalies and determine their origins [24]. Each ML category for ITD consists of various subtypes. Supervised learning encompasses a range of algorithms, including classification and regression techniques. Classification models categorizes data to ascertain the appropriate group. Example is the classification of emails as Spam or not Spam. The techniques employed in this context encompass Logistic Regression (LR), K-Nearest Neighbours (k-NN), Support Vector Machines (SVM), Naïve Bayes (NB), and Decision Trees. In contrast, regression pertains to the task of making predictions. One illustrative instance involves the anticipation of the expenses associated with constructing a building, which is contingent upon factors such as the geographical placement and the quality of the surrounding infrastructure. The techniques employed in this study encompass Linear Regression (LR), Decision Tree (DT), Support Vector Regression (SVR), Lasso Regression, and Random Forest. There is a greater utilization of supervised ML techniques in the field of ITD. Furthermore, the application of classification has been more prevalent in relation to this subject matter. The rationale behind utilizing online behavioural characteristics of users is to conduct a rigorous study of data in order to determine the potential risk level associated with a certain action. From the examined models, SVM, Logistic Regression (LR), and NB algorithms have frequently been utilized in the development of ITD models. [28] observed that these models, in contrast to traditional applications, exhibit automation, proactivity, and improved performance with larger datasets, making them preferable. Table 2 provides a summary of the ML algorithms widely used for ITD model development for the last decade for the eleven reviewed articles. It can be seen that KNN and K-Means appeared most (frequency of 6) which means that of the ML ITD articles reviewed, these two were used by most researchers.

Ensemble modelling is one of the techniques used to enhance ML model performance and has been used in many of the reviewed articles. This is because by using the two algorithms, the disadvantage of one is neutralized by another's advantages. Additionally, ensembling allows for more features to be added in a model. For example, the use of PCA gives the model the ability to reduce dimensionality while SMOTE helps to correct data imbalance.

5.2. Deep Learning for ITD

DL models have primarily been utilized in the domains of Natural Language Processing (NLP) and computer vision. The current body of research on the use of DL in anomaly detection is limited [30]. This phenomenon arises due to the presence of limited numbers of anomalies within the harvested user data resulting in heavily imbalanced datasets. This is the rationale behind the inclusion of SMOTE within most of the proposed DL insider threat detection models.

Table 2. Common ML algorithms used for ITD between the period 2013 and 2023

Model	**Function**	**Frequency**
KNN	Regression/classification	6
K-Means	Clustering	6
Naïve Bayes	Classification	4
Logistic Regression	Prediction/regression	4
SVM and varieties	Regression/classification	3
PCA	Dimensionality reduction correction	2
Random Forest and Varieties	Regression/classification	5
Local Outlier Factor	Anomaly detection	3
SMOTE	Data imbalance correction	1

LSTM, DNN, Deep Autoencoders, and Self-Organizing Maps (SOM) have frequently been used for ITD. These models possess the capability of effectively processing substantial amounts of data. In addition, they have the capacity to address complex non-linear problems which is one of the main characteristics of DL models. Survival analysis and Few-Shot self-supervised models have gained prominence in the field of ITD due to their efficacy in producing favourable outcomes with imbalanced datasets. The utilization of Convolutional Neural Networks (CNNs) for the purpose of detecting insider threats is a very infrequent phenomenon, mostly due to its reliance on image-based data. The utilization of a purely visual approach in addressing the issue of insider threat detection has demonstrated its efficacy, as evidenced by the findings of [44]. Using user behaviours such as device use, browsing history and logon activities, data was collected and transformed into images to be fed into CNN algorithm for anomaly detection. This was a rare occasion in ITD studies. Table 3 presents a summary of the prevalent DL models employed for the task of ITD for the period 2013-2023.

Table 3. Common DL algorithms for ITD for the period 2013-2023

DL Model	**Function**	**Frequency**
LSTM	Classification	7
DNN	Prediction	3
Autoencoder	Noise removal in data	3
GRU	Classification	1
CNN	Classification	1
Self-Organizing Maps (SOM)	Dimensionality reduction	1
PCA	Dimensionality reduction	2
SMOTE	Data imbalance correction	1

Ensembling techniques have been dominant in DL models too.For instance, the study conducted by [40] integrates SOMs with PCA. Similarly, [11] integrate BiLSTM with CAEs. Furthermore, [47] employ an ensemble approach, combining FF-ANN with LSTM. Thesecombinations result in an enhancedmodel that has improved performance because of the additional features. Table 4 gives a summary of ML and DL characteristics that can be used as a guide to amateur model developers on what to use given a problem.

Table 4.ML and DL models characteristics

Comparison Feature	**Machine Learning**	**Deep Learning**
Definition	ML covers the area where computer science and statistics combine and uses algorithms to perform certain tasks without the need for explicit programming; instead, they identify patterns in the data and project future events based on the introduction of new data.	DL is considered to be an advanced and technically challenging development of ML algorithms. These are algorithms that examine data and make judgments based on a logical framework akin to that of a human
Algorithms	Linear Regression, Logistic Regression, k-NN, SVM, Naive Bayes, Lasso regression and Random Forest.	LSTM, CNN, SOM, DNN, MLP, Deep Belief Networks (DBN), Generative Adversarial Network (GAN), Auto-encoders and GRU
Relationship with AI	A subset of AI	A subset of ML
Computer requirements	Trains on Central Processing Unit (GPU)	Needs Graphic Processing Units (GPU)
Architecture	Does not have layers but uses explicit programming to solve challenges.	Uses layers known as the Artificial Neural Network (ANN)
Volume of Data	Requires smaller datasets	Uses voluminous data
Human intervention	Relies on human intervention/correction	Learns on their own/uses repetition to learns/remembers past mistakes and corrects itself
Accuracy	Low accuracy	High accuracy

Training time	Less training times	Much more training times
Working	Receives input, analyzes, establishes patterns, predicts than given an output	Takes data input through hidden layers (black box) and generates output
Limitations of DL and ML	Overfitting and underfitting, ethical considerations and bias judgement for both ML and DL. Costly to work with in terms of computer resources and data, black box problem and overfitting for DL.	

5.3. Datasets for ITD Models

Out of the twenty two (22) reviewed models in this study, three datasets which include CERT revisions, TWOs and Real dataset have featured predominantly as illustrated in Figure 3.

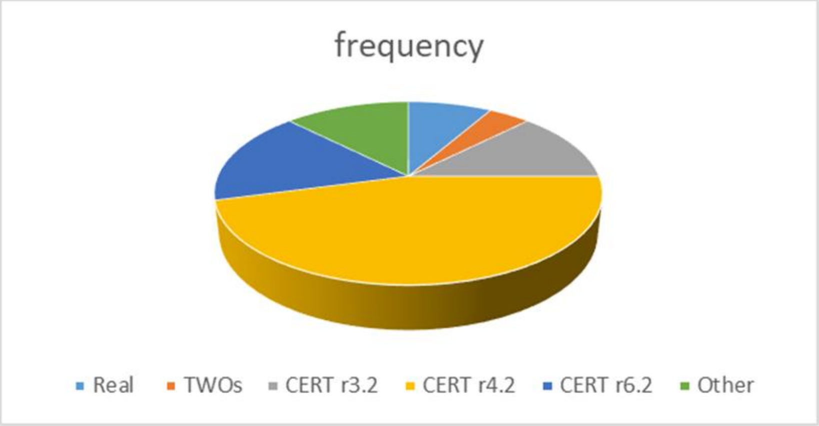

Figure 3. Preferred Insider threat detection datasets

The three revisions of the Community Emergency Response Team (CERT) from Carnegie Mellon University (CMU) feature predominantly. In the graph, real data, TWOs and ‘other’ which featured Wikipedia and Zonefox datasets were used. Standing out is the CERT r4.2 dataset, commonly known as the "deep needle" dataset due to the higher number of "injected" anomalies [7]. The dataset comprises six .csv files encompassing logon, file, device, http, LDAP, and psychometric data. Although it exhibits a significant imbalance, as evidenced by the presence of a mere 70 anomalies within a vast collection of 32,770,227 user behaviours, it is still considered better than other revisions, hence, becoming very popular. The utilization of real data (RD) was avoided among the majority of researchers due to the substantial data requirements associated with ML and DL model development. The acquisition of large quantities of data necessitates significant investments of time and resources which is wasteful. Another key observation is the emergence use of social networks for data harvesting. The traffic towards social media is enormous and any organization in need of voluminous data can easily collect given adherence to the laid out guidelines. This can be illustrated by the use of YouTube data for ITD.

5.4. Evaluation Metrics Used for ITD Models

The quantitative assessment of a ML model's performance and effectiveness is conducted using evaluation metrics. These metrics facilitate the comparison of different models and provide insights into the model's performance. The evaluation of predictive ability, generalization capability, and overall model quality holds significant importance. The selection of appropriate metrics is contingent upon various factors, including the specific issue domain (e.g., regression, clustering, or classification), the type of dataset employed (e.g., images, text, or time series), and the desired outcomes.

Evaluation metrics are utilized at two distinct stages in the process of model building, namely the training stage and the validation stage. In the initial phase, metrics serve as a means of differentiation to identify the most favourable solution. Subsequently, they function as evaluators to assess the efficacy of the model [14]. These assessments aid developers in evaluating the performance of the model in real-world scenarios. The problem of detecting insider threats can be framed as a classification problem, where evaluation criteria are employed based on the confusion matrix.

Classification can be performed using two distinct methods: binary classification and multi-threshold classification. Binary classification metrics encompass various evaluation measures, such as sensitivity (also known as Recall), specificity, positive predictive value (Precision), negative predictive value, Accuracy, and F1 score (also referred to as F Measure). On the other hand, multi-threshold classification involves additional assessment techniques, including the Area Under Curve (AUC) of the Receiver Operating Characteristic (ROC) curve, commonly denoted as AUC-ROC, as well as the Precision-Recall (PR) curve. Figure 4 shows the widely applied evaluation metrics used within the twenty two reviewed articles for ITD models within the past decade.

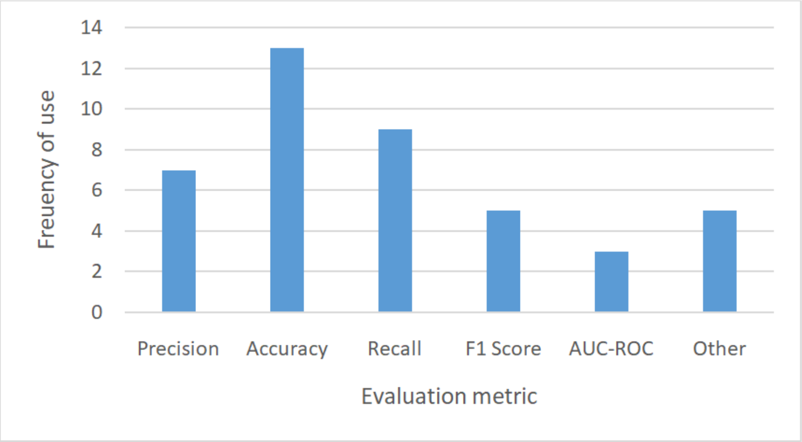

Figure 4. Popular evaluation metrics for ITD models

The most common evaluation metric is used within the reviewed articles is accuracy. Accuracy refers to the proportion of samples that are properly identified. The aforementioned metric is widely favoured; yet, its efficacy diminishes when confronted with imbalanced data because it tends to exhibit bias towards the majority class. Other supplementary evaluation measures include AUC-ROC, Precision, f1 score, and Recall. Integration of several evaluation metrics aims to provide a comprehensive evaluation [12].

Recall is calculated by dividing the total number of relevant samples by the number of accurate positive outcomes. Precision, on the other hand, refers to the proportion of accurately positive results in relation to the classifier's predictions of positive results, while f1 score is the mean between precision and Recall. [48] assert that the utilization of the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is prevalent in the context of multi-threshold classification, hence not used for binary classification. Another metric used among the models reviewed is false positive rate, which quantifies the proportion of negative data points incorrectly classified as positive among all negative data. Using several evaluation metrics enhances model performance in real life because of the all-round tests it has been subjected to.

6. Conclusion

This study examined a decade-long (2013-2023) investigation on ITD models. The study incorporated publications from high quality reputable journals.

Detection of insider threats is a significant field of study that has been extensively explored but never exhausted. The demand for enhanced ITD models has displaced conventional methodologies, as developers increasingly use more contemporary alternatives. The widespread presence of computers and Internet connectivity has resulted in extensive data accumulation that poses challenges for conventional IT infrastructure architectures. Various ML approaches have been widely accepted and continue to be utilized in contemporary applications. The use of ML has been hindered by the advent of Big Data, as it necessitates human intervention for the purposes of detection and interpretation. This implies that a majority of the fundamental ML models are being substituted by DL algorithms. Model selection in conjunction with data pre-processing techniques play a crucial role in enhancing DL model performance. In addition, ensembling has emerged as a significant factor in model performance, however, it is important to exercise caution when implementing this technique to avoid potential increases in processing times. The CERT r4.2 dataset has been widely favoured among researchers for ITD because of its notable feature known as the "deep needle”. Despite the preference, owing to the very low proportion between benign and malicious data, it is imperative to address the issue of data imbalance and dimensionality reduction before feeding it to models for training. In conclusion, Although Accuracy has been popularly used to evaluate ITD models, researchers should understand that it possesses a limitation of ignoring assessment of the model's resilience, hence, the need to incorporate additional evaluation metrics.

References

[1] Li, Y., & Liu, Q. (2021). A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments. Energy Reports, 7, 8176–8186. https://doi.org/10.1016/j.egyr.2021.08.126

[2] Al-Mahrouqi, A., Cianain, C., &Kechadi, T. (2015). Cyberspace Challenges and Law Limitations. International Journal of Advanced Computer Science and Applications, 6, 279–289. https://doi.org/10.14569/IJACSA.2015.060837#sthash.8a23qQ95.dpuf

[3] Viraja, K., &Purandare, P. (2020). A Qualitative Research on the Impact and Challenges of Cybercrimes. Journal of Physics: Conference Series.

[4] Kalakuntla, R., Vanamala, A., &Kolipyaka, R. (2019). Cyber Security. Holistica, 10, 115–128. https://doi.org/10.2478/hjbpa-2019-0020

[5] Adams, J., &Albakajai, M. (2016). Cyberspace: A New Threat to the Sovereignty of the State. https://core.ac.uk/reader/146502700

[6] Lu, J., & Wong, R. (2019). Insider Threat Detection with Long Short-Term Memory. In ACSW 2019: Proceedings of the Australasian Computer Science Week Multiconference (p. 10). https://doi.org/10.1145/3290688.3290692

[7] Saxena, N., Hayes, E., Bertino, E., Ojo, P., Choo, K.-K. R., &Burnap, P. (2020). Impact and Key Challenges of Insider Threats on Organizations and Critical Businesses. Electronics, 9, 1460. https://doi.org/10.3390/electronics9091460

[8] Franqueira, V., Cleeff, A. van, Eck, P. van, &Wieringa, R. (2015). External Insider Threat: A Real Security Challenge in Enterprise Value Webs.

[9] Elmrabit, N., Yang, S.-H., & Yang, L. (2015). Insider threats in information security categories and approaches. https://doi.org/10.1109/IConAC.2015.7313979

[10] NIST. (2020). Data Integrity: Detecting and Responding to Ransomware and Other Destructive Events. https://www.nccoe.nist.gov/publication/1800-26/VolA/index.html

[11] Wei, Y., Chow, K.-P., &Yiu, S.-M. (2021). Insider threat prediction based on unsupervised anomaly detection scheme for proactive forensic investigation. Forensic Science International: Digital Investigation, 38, 301126. https://doi.org/10.1016/j.fsidi.2021.301126

[12] Alsowail, R. A., & Al-Shehari, T. (2022). Techniques and countermeasures for preventing insider threats. PeerJ Computer Science, 8, e938. https://doi.org/10.7717/peerj-cs.938

[13] Jiang, Y., Gu, S., Murphy, K., & Finn, C. (2019). Language as an Abstraction for Hierarchical Deep Reinforcement Learning (arXiv:1906.07343). arXiv. https://doi.org/10.48550/arXiv.1906.07343

[14] Zheng, M., Wang, F., Hu, X., Miao, Y., Cao, H., & Tang, M. (2022). A Method for Analyzing the Performance Impact of Imbalanced Binary Data on Machine Learning Models. MDPI

[15] Mishra, A. (2020, May 28). Metrics to Evaluate your Machine Learning Algorithm. Medium. https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234

[16] Zala, M. (2023). Detection of Insider Threat Forming Using Machine Learning. 11(3).

[17] Kuppusamy, P., GanthanSamy, Maarop, N., Magalingam, P., Kamaruddin, N., Shanmugam, B., &Perumal, S. (2020). Systematic Literature Review of Information Security Compliance Behaviour Theories. Journal of Physics.

[18] Karspersky. (2022). Recognizing different types of insiders. https://encyclopedia.kaspersky.com/knowledge/recognizing-different-types-of-insiders/

[19] Apps, S. C. (2022, June 21). Insider Threat: Definition, Types, Indicators & More. Spanning. https://spanning.com/blog/insider-threats/

[20] Almehmadi, A., & El-Khatib, K. (2014). On the Possibility of Insider Threat Detection Using Physiological Signal Monitoring. Proceedings of the 7th International Conference on Security of Information and Networks, 223–230. https://doi.org/10.1145/2659651.2659654

[21] IBM. (2023). https://www.ibm.com/reports/threat-intelligence

[22] Ophoff, J., Jensen, A., Sanderson-Smith, J., Porter, M., & Johnston, K. (2014). A Descriptive Literature Review and Classification of Insider Threat Research (p. 223). https://doi.org/10.28945/2010

[23] Nurse, J. R. C., Williams, N., Collins, E., Panteli, N., Blythe, J., &Koppelman, B. (2021). Remote Working Pre- and Post-COVID-19: An Analysis of New Threats and Risks to Security and Privacy. In C. Stephanidis, M. Antona, & S. Ntoa (Eds.), HCI International 2021—Posters (Vol. 1421, pp. 583–590). Springer International Publishing. https://doi.org/10.1007/978-3-030-78645-8_74

[24] CISA. (2020). Insider Threat Mitigation Guide.

[25] Freet, D., &Agrawal, R. (2017). Cyber Espionage. https://doi.org/10.1007/978-3-319-32001-4_51-1

[26] Homoliak, I., Toffalini, F., Guarnizo, J., Elovici, Y., & Ochoa, M. (2018). Insight into Insiders and IT: A Survey of Insider Threat Taxonomies, Analysis, Modeling, and Countermeasures. In arXiv e-prints. https://doi.org/10.48550/arXiv.1805.01612

[27] Khan, N., J. Houghton, R., &Sharples, S. (2022). Understanding factors that influence unintentional insider threat: A framework to counteract unintentional risks. Cognition, Technology & Work, 24(3), 393–421. https://doi.org/10.1007/s10111-021-00690-z

[28] Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science, 2(3), 160. https://doi.org/10.1007/s42979-021-00592-x

[29] Raroque, C. (2023). Key Differences: Machine Learning, AI, and Deep Learning. https://aloa.co/blog/differences-between-machine-learning-artificial-intelligence-and-deep-learning, https://aloa.co/blog/differences-between-machine-learning-artificial-intelligence-and-deep-learning

[30] Beigy, H. (2023). Machine learning—Introduction.

[31] Oladimeji, O., Oladimeji, A., &Oladimeji, O. (2021). Machine Learning Models for Diagnostic Classification of Hepatitis C Tests. Frontiers in Health Informatics, 10, 70. https://doi.org/10.30699/fhi.v10i1.274

[32] Zhang, D., Zheng, Y., Wen, Y., Xu, Y., Wang, J., Yu, Y., &Meng, D. (2018). Role-based Log Analysis Applying Deep Learning for Insider Threat Detection. Proceedings of the 1st Workshop on Security-Oriented Designs of Computer Architectures and Processors, 18–20. https://doi.org/10.1145/3267494.3267495

[33] Kandias, M., Stavrou, V., Bozovic, N., &Gritzalis, D. (2013). Proactive insider threat detection through social media: The YouTube case. Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society, 261–266. https://doi.org/10.1145/2517840.2517865

[34] Kim, J., Park, M., Kim, H., Cho, S., & Kang, P. (2019). Insider Threat Detection Based on User Behavior Modeling and Anomaly Detection Algorithms. Applied Sciences, 9(19), Article 19. https://doi.org/10.3390/app9194018

[35] Janjua, F., Masood, A., Abbas, H., & Rashid, I. (2020). Handling Insider Threat Through Supervised Machine Learning Techniques. Procedia Computer Science, 177, 64–71. https://doi.org/10.1016/j.procs.2020.10.012

[36] Sav, U., &Magar, G. (2021). Insider Threat Detection Based on Anomalous Behavior of User for Cybersecurity (pp. 17–28). https://doi.org/10.1007/978-981-15-5309-7_3

[37] Padmavathi, G., Shanmugapriya, D., &Asha, S. (2022). A Framework to Detect the Malicious Insider Threat in Cloud Environment using Supervised Learning Methods. 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom), 354–358. https://doi.org/10.23919/INDIACom54597.2022.9763205

[38] Ganapathi, P. (2023). Malicious insider threat detection using variation of sampling methods for anomaly detection in cloud environment. Computers & Electrical Engineering, 105, 108519. https://doi.org/10.1016/j.compeleceng.2022.108519

[39] Peccatiello, R. B., Gondim, J. J. C., & Garcia, L. P. F. (2023). Applying One-Class Algorithms for Data Stream-Based Insider Threat Detection. IEEE Access, 11, 70560–70573. https://doi.org/10.1109/ACCESS.2023.3293825

[40] Moradpoor, N., Brown, M., & Russell, G. (2017). Insider threat detection using principal component analysis and self-organising map. Proceedings of the 10th International Conference on Security of Information and Networks, 274–279. https://doi.org/10.1145/3136825.3136859

[41] Tuor, A., Kaplan, S., Hutchinson, B., Nichols, N., & Robinson, S. (2017). Deep Learning for Unsupervised Insider Threat Detection in Structured Cybersecurity Data Streams (arXiv:1710.00811). arXiv. http://arxiv.org/abs/1710.00811

[42] Chattopadhyay, P., Wang, L., & Tan, Y.-P. (2018). Scenario-Based Insider Threat Detection From Cyber Activities. IEEE Transactions on Computational Social Systems, 5(3), 660–675. https://doi.org/10.1109/TCSS.2018.2857473

[43] Al-Mhiqani, M. N., Ahmad, R., ZainalAbidin, Z., Yassin, W., Hassan, A., Abdulkareem, K. H., Ali, N. S., &Yunos, Z. (2020). A Review of Insider Threat Detection: Classification, Machine Learning Techniques, Datasets, Open Challenges, and Recommendations. Applied Sciences, 10(15), Article 15. https://doi.org/10.3390/app10155208

[44] Koutsouvelis, V., Shiaeles, S., Ghita, B., &Bendiab, G. (2020). Detection of Insider Threats using Artificial Intelligence and Visualisation. 2020 6th IEEE Conference on Network Softwarization (NetSoft), 437–443. https://doi.org/10.1109/NetSoft48620.2020.9165337

[45] Paul, S., & Mishra, S. (2020). LAC: LSTM AUTOENCODER with Community for Insider Threat Detection. 2020 the 4th International Conference on Big Data Research (ICBDR'20), 71–77. https://doi.org/10.1145/3445945.3445958

[46] Sharma, B., Pokharel, P., & Joshi, B. (2020). User Behavior Analytics for Anomaly Detection Using LSTM Autoencoder—Insider Threat Detection. Proceedings of the 11th International Conference on Advances in Information Technology, 1–9. https://doi.org/10.1145/3406601.3406610

[47] Singh, M., Mehtre, B. M., Sangeetha, S., &Govindaraju, V. (2023). User Behaviour based Insider Threat Detection using a Hybrid Learning Approach. Journal of Ambient Intelligence and Humanized Computing, 14(4), 4573–4593. https://doi.org/10.1007/s12652-023-04581-1

[48] Varoquaux, G., &Colliot, O. (2020). Evaluating machine learning models and their diagnostic value. Machine Learning for Brain Disorders.

AUTHORS

Everleen Nekesa Wanyonyi is a Tutorial Fellow in the School of Computing and Information Technology (SCIT), Department of Computer Science of Murang'a University of Technology, Kenya. She is a PHD student at Jaramogi Oginga Odinga University of Science and Technology with special interest in Information Technology security and audit.

Dr. Newton Masinde is a lecturer at the School of Informatics and Innovative Systems (SIIS) of Jaramogi Oginga Odinga of Science and Technology (JOOUST) specializing in Distributed Computing with research interests in P2P and online social networks. He also serves as the Associate Director of the ICT Directorate.

Prof. Silvance Abeka is an Associate Professor at the School of Informatics and Innovative Systems (SIIS) of Jaramogi Oginga Odinga of Science and Technology (JOOUST). He is researcher and consultant in Technology Enabled Learning, ICT4D, Information Systems, Transformative Leadership and PhD Supervision. He also doubles up as a Director, eLearning at the institution.