The Application of Artificial Intelligence in the Detection of Malicious Insider Threats: A Review

Lead Author

Oguntunde, B.A

Affiliation:

Department of Computer Science, Redeemer's University Ede, Osun State

Abstract

Insider threats are a growing threat to organizations' security, resulting in a significant increase in cyberattacks. As organizations continue to rely on digital systems and data, the potential for malicious insider threats has heightened the need for advanced detection methods using Artificial Intelligence (AI) technology. A malicious insider is an individual granted legitimate access to an organization and exploits this privilege for personal or other reasons to compromise information assets' confidentiality, integrity, or availability. A simple review of forty-seven (47) articles identified from various academic databases was conducted. In this review paper, we explore the current state of research on the application of AI techniques for the detection of malicious insider threats in the cybersecurity space by examining the different AI-based approaches and techniques that have been employed for the detection of malicious insider threats, types of data source and how effective the AI models are through the evaluation metrics utilized. The academic literature reveals a wide range of advancements in artificial intelligence related to the detection of insider threats. The Computer Emergency Response Team (CERT) dataset has the highest usage of 68%, while accuracy and precision have the highest usage of 26% and 21%, respectively, in terms of performance metrics, with Machine learning as the most used AI technique compared to others. Additionally, the paper outlines future research directions. It serves as a starting point for young researchers and a yardstick for experienced researchers in proposing new methodologies to enhance the effectiveness of insider threat detection.

Co-Authors: Ogunde A.O. Odim, M.O, Kayode, A.A. Department of Computer Science, Redeemer's University, Ede, Osun state. Nigeria.

Keywords: Artificial intelligence, Cybersecurity, Deep Learning, insider threat detection, Machine learning, Malicious insider threat

Introduction

The application of AI in detecting malicious insider threats has become increasingly important in cybersecurity. As organizations continue to rely on digital systems and data, the potential for malicious insider threats has heightened the need for advanced detection methods using AI technology. According to the Cybersecurity Insiders 2024 report, there has been an exponential increase in insider attacks. Between 2019 and 2024, insider attacks increased from 66% to 76%, stressing the urgent need for enhanced detection and mitigation strategies, including continuous monitoring and proactive defenses (Schulze, 2024)

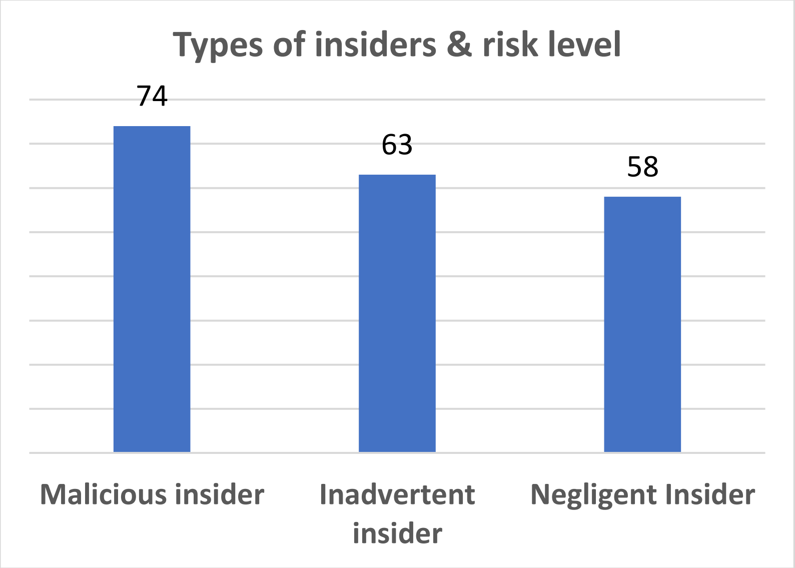
A malicious insider is an individual granted legitimate access to an organization and exploits this privilege for personal or other reasons to compromise the organization's confidentiality, integrity, or availability of information assets (Al-shehari & Alsowail, 2021). As these people are often culprits of the greatest losses endured in organizations, unwanted insider access is labeled an "insider threat." Unfortunately, characterizing abnormal or unwanted activity by trusted people is difficult as it often resembles normal activity. Security mechanisms to protect significant data often rely on perimeters that block access from outside the organizations. This can be a false sense of security. Although an organization may be secure over its perimeter, the critical data are often exposed to potential attackers that have legitimate access within the organizations. Protection mechanisms only focus on potential threats outside an organization, disregarding the amplified risks when granting legitimate access to significant data. Insiders can make more sophisticated threats through privileged access (Abiodun et al., 2023).

In the literature of insider threat detection, there are several definitions for insider threats. According to the CERT Insider Threat Center, an "insider threat" can be defined as "a current or former employee, contractor, or business partner who has or had authorized access to an organization's network, system, or data and intentionally exceeded or misused that access in a manner that negatively affected the confidentiality, integrity, or availability of the organization's information or information systems (Capelli, Moore, &

Trzeciak, 2012)." Other definitions provided by the Computer Security Resource Center and the National Institute of Standards and Technology (NIST) are "an employee, contractor or business partner who has inside information concerning an organization's security practices, data or other assets" and "a user of a computer system who exploits their authorization to compromise the confidentiality, integrity or availability of the system," respectively. Those definitions consider that an insider threat is an activity that harms the organization and is committed by a trusted user (Al-Shehari & Alsowail, 2021).

Insider threats are classified into three types, as shown in Figure 1: malicious, inadvertent, and credential misuse. Malicious threats are defined as events in which current or former employees or contractors commit acts against an organization with the intention to harm it.

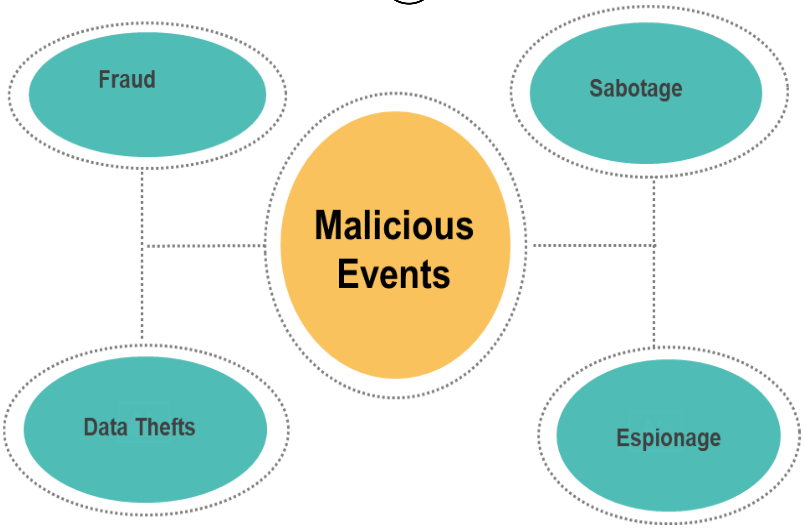
Figure 1: Types of insider threats



Source: Cybersecurity Insiders 2024 report

Malicious events can be further classified as fraud, data theft, sabotage, or espionage, as shown in Figure 2. For instance, ex-employees can steal or leak sensitive data to competitors,

Figure 2: Malicious Events



Source: Alsowail & Al-Shehari, 2022

tamper or destroy data, damage the organization's reputation or use privileged information for personal gains. Inadvertent threats are events in which employees or contractors unintentionally execute actions that harm the organization. Inadvertent events can be further classified as errors or negligence. For instance, negligence can be an employee who fails to follow IT policy and exposes sensitive data. Credential misuse threats are defined as events in which an employee or contractor's account is used by an unauthorized user, regardless of whether such action has malicious or inadvertent intentions (Alsowail & Al-Shehari, 2022). This review will explore the current state of AI techniques and their effectiveness in identifying and preventing malicious insider threats in organizations by exploring the current methodologies and technological frameworks for detecting malicious insider threats. It identifies and synthesizes relevant research from various databases, subsequently providing a comparative analysis. Furthermore, it examines the categories of data utilized, the machine/deep learning algorithms implemented to identify these threats, and the performance metrics employed in evaluating the developed models.

Challenges of Traditional Detection Methods

To secure data against various compromises in the past, conventional methods such as intrusion detection, firewalls, and data encryption were employed. However, these techniques fail to protect sensitive information from individuals possessing authorized access to such data(Arif et al., 2023; Al-shehari & Alsowail, 2021). Insiders might misinterpret sensitive data for several motives that endanger organizations and national security, such as curiosity, personal grievance, or financial advices as selling intellectual property. Moreover, the inside activity may not always be malicious. Still, inappropriate access may cause weakening of the protected data, such as authentication secret data utilized for several services such as biometric securities (Alsowail & Al-Shehari, 2022).

Related works

In this section, previous review papers on insider threat detection are reviewed. In literature,Yuan & Wu, (2021)paper systematically reviews existing literature on deep learning applications for insider threat detection, examining key algorithms, datasets, and evaluation metrics. It identifies a critical research gap in the lack of standardized datasets for evaluating these models and the limited studies on their real-world deployment and scalability. The authors also address the challenges in implementing deep learning solutions, such as data privacy concerns and the need for more robust models to withstand adaptive attacks. The authors underscore the growing importance of deep learning in cybersecurity while pointing out the need for further research to overcome the current limitations and fully harness its potential in insider threat detection. In their paper, Anju et al., (2022)identify key research gaps, such as the challenge of differentiating behavioral patterns between insiders and regular users and the lack of clearly defined insider risks that hampers effective detection. Despite these challenges, machine learning techniques offer promising opportunities for improving detection methods. The authors assess various methodologies and algorithms, contributing valuable insights into how machine learning can enhance the identification of insider threats in the cyber world. Through their comprehensive analysis, the paper provides a deeper understanding of the difficulties in detecting insider threats and highlights the potential for machine learning to address these issues effectively. Prajitno, Hadiyanto, & Rochim (2023) paper provides a thorough exploration of research opportunities in insider threat detection using machine learning (ML) methods. The authors classify existing detection methods into three categories: combination, selection, and singular focus. Through this review, they identify significant research gaps and propose directions for future work to enhance insider threat detection solutions. The paper is focused specifically on studies that utilize ML algorithms. This classification helps in identifying research gaps and suggesting new avenues for enhancing insider threat detection. Despite its thorough approach, the study's scope is limited to research utilizing machine learning algorithms and is constrained by the methodology employed.

Wanyonyi, Abeka, & Masinde (2023) provided an in-depth analysis of machine learning (ML) models for insider threat detection, focusing on algorithms, datasets, and evaluation metrics. The paper underscores the crucial role of these elements in developing effective insider threat detection systems. It highlights that while various datasets are essential for training and testing ML models, there is a need for a comprehensive understanding of their application to enhance model performance. Similarly, the choice of evaluation metrics is critical for assessing the effectiveness of these models, yet specific metrics are not detailed in the provided context.

Alzaabi & Mehmood (2024) highlighted the critical role of machine learning, particularly NLP and time-series analysis, in enhancing insider threat detection by identifying communication patterns and temporal behaviors. While these advancements offer significant benefits, the authors underscore ongoing challenges such as data scarcity, model interpretability, adversarial attacks, and more scalable, real-time, and adaptable systems. They advocate for integrating explainable AI and hybrid models to improve detection robustness and suggest collaborative intrusion detection to enhance accuracy and scalability, addressing limitations like data integration issues and human factor resistance. Atadoga et al., (2024) review emphasizes the fundamental role of machine learning (ML) in enhancing network security by improving incident response, detecting evolving threats in real-time, and reducing false positives. While ML proves effective in areas like intrusion and malware detection, the review highlights challenges such as data privacy, model interpretability, adversarial attacks, and the scalability of ML-based security solutions. The authors also note gaps in addressing network security across specific industries and the impact of emerging technologies. The authors advocate for future research on privacy-preserving algorithms, context-aware security, and explainable AI to build trust and transparency in ML models, while stressing the need for collaboration, training, and adherence to best practices in advancing ML for network security.

In the paper of Ismaila & Adeleke (2023), the authors presented an analysis of insider threat detection mechanisms, highlighting the evolving nature of these threats and the need for innovative approaches to address them. The study identifies key challenges, including the rapid technological advancements and the limitations of existing detection methods, which contribute to the rise in insider threats across organizations. The findings suggest that ongoing research is necessary to develop more robust and adaptable detection systems that can effectively identify and mitigate insider threats in diverse environments**.**

Naseer (2024) thoroughly examined the benefits of combining ML with cyber threat intelligence (CTI), noting that this integration facilitates more comprehensive data analysis and bolsters threat detection capabilities. However, it also acknowledges persistent challenges such as risk assessment, data accuracy, and sophisticated tactics employed by cybercriminals, which complicate the identification of attackers. The paper also touches on the contributions of supervised and unsupervised learning in cybersecurity and the importance of content aggregation and threat intelligence management. Despite these advancements, the study points out that difficulties remain in addressing the evolving nature of cyber threats and enhancing implementation strategies. Ugochukwu et al., (2024) investigate both the contributions and limitations of ML in cybersecurity. It notes that ML enhances threat detection and automates decision-making processes for quicker responses. However, challenges such as adversarial attacks, skewed datasets, and the interpretability of ML models are significant concerns. The authors stress the need for a holistic approach that integrates technological solutions with ethical considerations to address these challenges effectively. The paper emphasizes combining human expertise with machine intelligence to build robust defenses against evolving cyber threats by showcasing successful applications and addressing the complexities involved.

Yilmaz & Can (2024) posited that conventional methods struggle with detecting threats and face hurdles such as limited insider threat datasets and the dynamic behavior of employees. The study proposes leveraging AI techniques to improve detection capabilities, offering insights into how these technologies can strengthen organizational defenses. Additionally, the paper outlines future research directions, including integrating multimodal data analysis, human-centric approaches, privacy-preserving techniques, and explainable AI, to address existing challenges and enhance the effectiveness of insider threat detection.

Research Methodology

This section presents the literature review protocol followed in renewing existing studies on insider threat detection using artificial intelligence.

The search strategy of the study and selection criteria

A simple literature review was conducted for works done between 2014 to 2024 (a decade). Several databases were queried based on the search keywords with appropriate search strings to gather the required papers. Table 1 shows the academic database searched and the equivalent papers identified:

Table 1: Analysis of sourced database and corresponding no of articles

S/N	Database Source	No of Articles
	ACM Digital Library	**9**
	ScienceDirect	**4**
	IEEE Explore	**12**
	Wiley	**5**
	Researchgate	**13**
	Research4life	**4**
	Total	**47**

Source: Authors

Search Keywords: The search keywords are carefully selected based on the aim of the review, which is ("insider threat" OR "malicious insider") AND ("artificial intelligence" OR "machine learning" OR "deep learning" OR "NLP" OR "AI").

For the final selection of the identified paper, inclusion and exclusion criteria were applied as follows:

Inclusion: Peer-reviewed journal articles, conference proceedings and book chapter

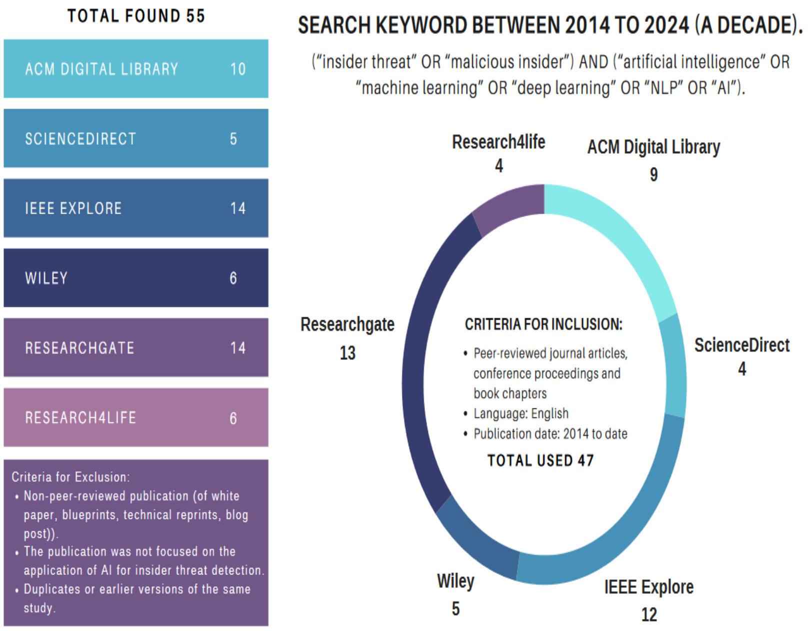
Language: English

Publication date: 2014 to date
Exclusion: Non-peer-reviewed publication (of white paper, blueprints, technical reprints, blog post). Publication not focused on the application of AI for insider threat detection.

Duplicates or earlier versions of the same study.

The initial search on the academic databases identified a total of 55 articles. However, after applying the inclusion extinct criteria, the total number left is 47. Figure 4 depicts the full selection process.

Figure 4: Selection Process Diagram



("insider threat" OR "malicious insider") AND ("artificial intelligence" OR "machine learning" OR "deep learning" OR "NLP" OR "AI").

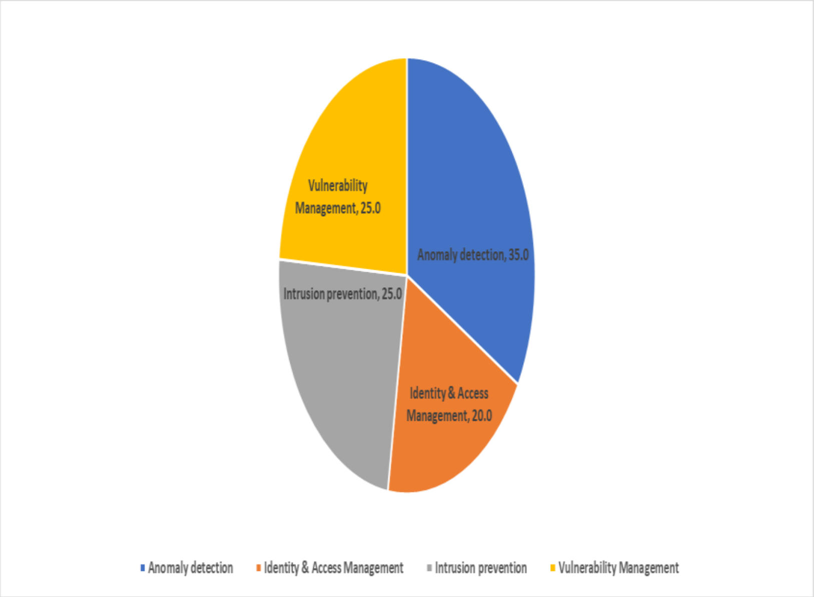
Criteria for Exclusion:

- Non-peer-reviewed publication (of white paper, blueprints, technical reprints, blog post)).
- The publication was not focused on the application of AI for insider threat detection.
- Duplicates or earlier versions of the same study.

CRITERIA FOR INCLUSION:

- Peer-reviewed journal articles, conference proceedings and book chapters
- Language: English
- Publication date: 2014 to date

Figure 5: AI usage in security automation

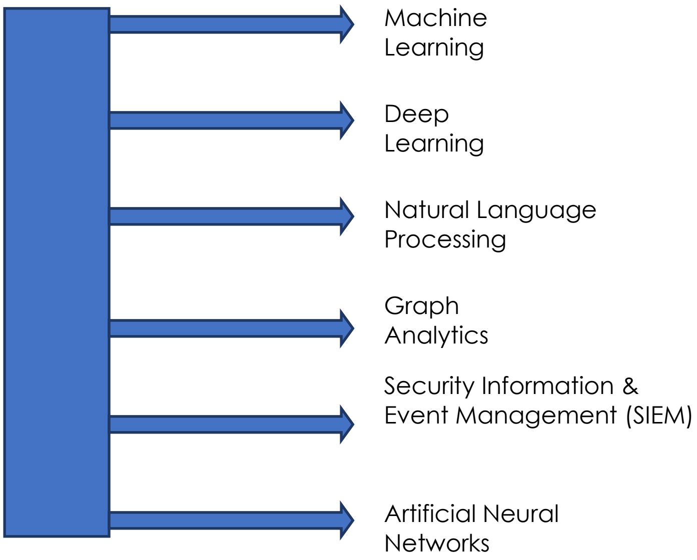


Source: Manoharan & Sarker, 2024

Detection of malicious insider threats using Artificial Intelligence (AI)-based techniques has been a growing research area, building its foundation upon decades of postures towards dealing with insider threats.

Artificial intelligence (AI) plays a significant role in insider threat detection. It improves the detection of malicious activities and enhances cybersecurity for organizations across the globe. Machine learning, a branch of AI, helps analyze user behavioral patterns, identify anomalies, and classify users as malicious insiders or benign insiders based on various features such as user activity logs, resources accessed, and the time of access, among others. Deep learning, a subset of machine learning, has become one of the most influential AI technologies over the last decade. Deep learning often demands high-performance computing and massive amounts of data for building intelligent applications. Natural Language Processing, or NLP, is a subfield of artificial intelligence that deals with the interaction between computers and humans through natural language. The ultimate objective of NLP is to read, decipher, understand, and make sense of human languages in a valuable way. With the rise of users with access to relevant information, insights begin to emerge only through the relationship between those accesses. A graph analytic approach allows one to aggregate many observations into one global picture and to select the most suspicious users. This makes graph analytics the most relevant technique to use alongside anomaly detection-based methods. Figure 6displays some of the AI tools used in insider threat detection.

Figure 6: Some AI tools used in insider threat detection


Machine Learning
Deep Learning
Natural Language Processing
Graph Analytics
Security Information & Event Management (SIEM)
Artificial Neural Networks

Source: Authors

Recent Insider Threat Detection Techniques

This section presents various AI techniques that could be used to detect malicious insider threat incidents in the cybersecurity environment, either to improve security in organization networks or to safeguard sensitive data/ intellectual property.

Gavai et al., (2015) worked on detecting insider threats from enterprise social and online activity data by identifying abnormal behavior in employees' enterprise social and online activity data. The author processes and extracts relevant features indicative of insider threat behavior. This includes features extracted from social data, including email communication patterns and content, and online activity data, such as web browsing patterns, email frequency, and file and machine access patterns. The study employed an anomaly detection method called isolation forest to identify anomalous events. The result shows an ROC score of 0.77, demonstrating that the proposed approach fairly successfully identifies insider threat events.

Tuor et al., (2017) worked on deep Learning for unsupervised insider threat detection in structured cybersecurity data streams. The authors pointed out that analyzing an organization's computer network activity is a key component of early detection and mitigation of insider threats. The authors also posited that insider threat is a growing concern for many organizations. The study developed a model that decomposes anomaly scores into the contributions of individual user behavior features using deep and recurrent neural networks based on CERT insider threat Dataset v6.2 and threat detection recall as their performance metrics. The result shows that long short-term memory performed equivalently to the deep neural network.

Goldberg et al., (2017) developed a prototype system (PRODIGAL) for insider threat detection as a test-bed for exploring various detection and analysis methods. The data and test environment, system components, and the core method of unsupervised detection of insider threat leads were presented to document the study. The authors also discuss a core set of experiments valuating the prototype's ability to detect both known and unknown malicious insider behaviuors. The experimental results showed the ability to detect a large variety of insider threat scenario instances imbedded in real data with no prior knowledge of what scenarios were present or when they occurred.

Chattopadhyay et al., (2018) conducted some study on scenario-based insider threat detection from cyber activities by presenting a method for detecting insider threats using time-series categorization of user activity. Initially, the user activity logs were used to construct a collection of single-day characteristics. The statistics of each single-day feature over a period of time were then used to create a time-series feature vector. Malicious and non-malicious threats were classified. Cost-sensitive data adjustment techniques were used randomly to sample the non-malicious class instances. Two-layered deep auto encoder neural networks were used as a classifier and its performance was compared with other popularly used classifiers: random forest and multilayer perceptron. The study revealed that both deep auto encoder and random forest classifiers classified the data-adjusted time-series feature set with high precision, recall, and f-score which are 0.9868, 0.9870 and 0.9742, respectively. Although multilayer perceptron had a high recall, it suffered from a lower precision and f-score compared to the other two classifiers.

Yuan et al., (2018) presented a study on insider threat detection with deep neural network by framing insider threat detection as an anomaly detection task and use anomalous behavior of a user as indicative of insider threat. The Long Short-Term Memory (LSTM) extracts user behavior features from sequences of user actions and generates fixed-size feature matrices. The Convolutional Neural Network (CNN) classifies fixed-size feature matrices as normal or anomaly. The proposed method was evaluated using the CERT Insider Threat dataset V4.2. The result shows that the method can successfully detect insider threat with AUC = 0.9449.

Hall et al., (2019)presented some study on predicting malicious insider threat scenarios using organizational data and a heterogeneous stack-classifier using the CERT dataset r4.2 along in a series of ML classifiers by aggregating the algorithm into a meta classifier. The result showed that the meta classifier had an accuracy of 96.2% and an area under the ROC curve of 0.988. However, the model developed was a generalized classifier model which did not give room for testing the instance of data tailored to each scenario which creates more performance classifiers than the generalized classifiers.

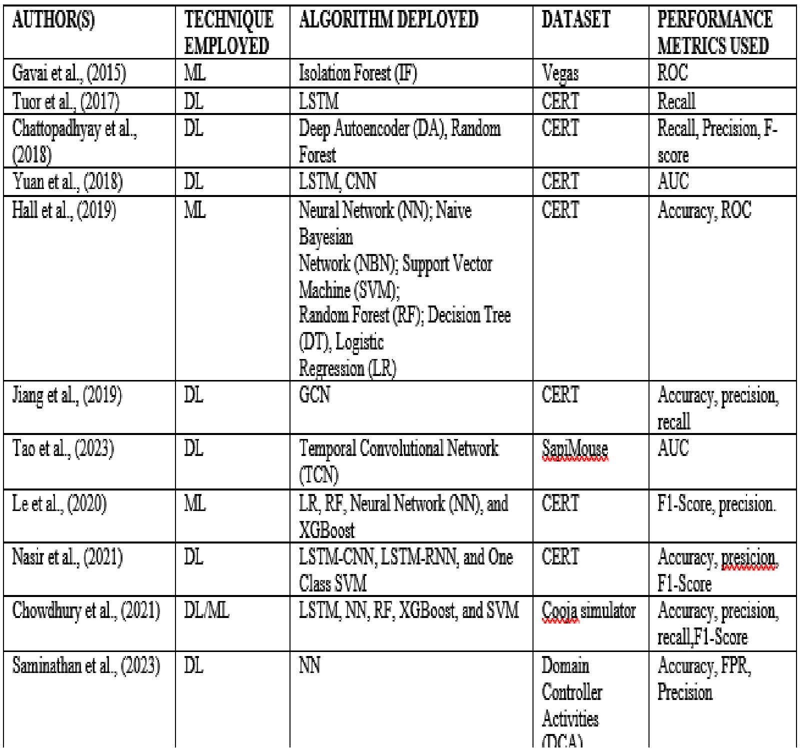
Jiang et al., (2019) conducted a study on anomaly detection with graph convolutional networks (GCN) for insider threat and fraud detection. The study pointed out the importance of connections or relationships between entities in the detecting of anomalous behavior and associated threat groups by describing a Graph Convolutional network-based anomaly detection model. The proposed model was evaluated using real-life datasets by insider threat and was compared with some widely used algorithms. The outcome showed that it delivered the highest detection accuracy of 93%.

Tao et al., (2023) proposes the Efficient Channel Attention mechanism (ECA-TCN) method to enhance user authentication using mouse dynamics data, addressing insider threat detection challenges in information security management. The method focuses on extracting personalized features from mouse dynamics, significantly improving authentication accuracy and time efficiency. The results, evaluated on the Sapi Mouse dataset from Sapientia Hungarian University of Transylvania (2020), demonstrate that the ECA-TCN method achieves higher AUC value of 96% compared to other models, making it a more effective solution for real-time user authentication.

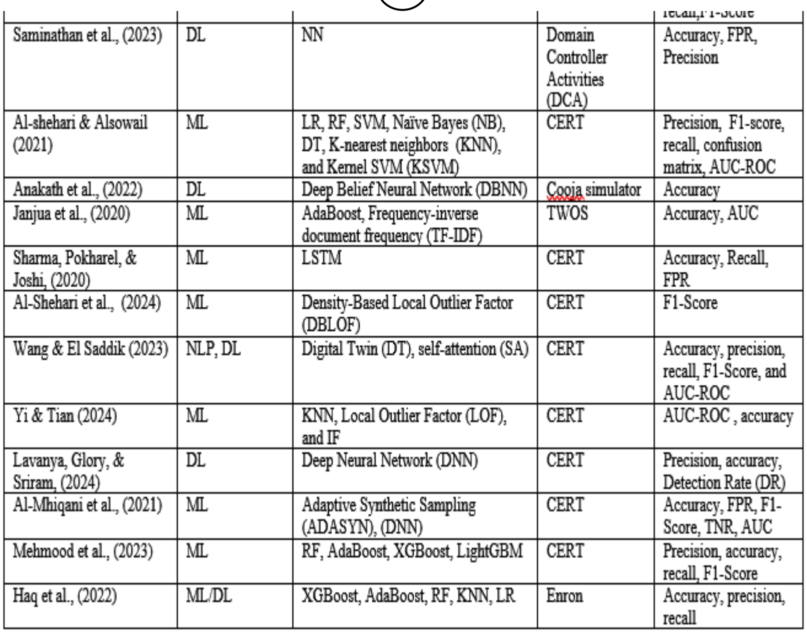
Le et al., (2020) The study develops a machine learning (ML) system for detecting insider threats in corporate networks, achieving high accuracy and low false positive rates despite limited ground truth data. The research employs algorithms such as Logistic Regression, Random Forest (RF), Neural Network (NN), and XGBoost, with RF outperforming others in detection performance, F1-score, and precision. User-session data proved effective for detecting malicious insiders with minimal delay, while user-day data excelled in intellectual property theft scenarios. The system's efficacy was tested on the CERT insider threat dataset, with future work focusing on incorporating temporal information in user actions and exploring non-Markovian decision-making in models.

A brief summary of the recent work on insider threat detection using artificial intelligence techniques is presented in table 2. We classify them on different techniques employed: machine leaning (ML), Deep Learning (DL), Natural Language Processing (NLP), and Others. We also include the algorithms, dataset and the performance metrics utilized.

Table 2: Brief summary of some reviewed papers showing the trends of detection methods



AUTHOR(S)	TECHNIQUE EMPLOYED	ALGORITHM DEPLOYED	DATASET	PERFORMANCE METRICS USED
Gavai et al., (2015)	ML	Isolation Forest (IF)	Vegas	ROC
Tuor et al., (2017)	DL	LSTM	CERT	Recall
Chattopadhyay et al., (2018)	DL	Deep Autoencoder (DA), Random Forest	CERT	Recall, Precision, F-score
Yuan et al., (2018)	DL	LSTM, CNN	CERT	AUC
Hall et al., (2019)	ML	Neural Network (NN); Naive Bayesian Network (NBN); Support Vector Machine (SVM); Random Forest (RF); Decision Tree (DT), Logistic Regression (LR)	CERT	Accuracy, ROC
Jiang et al., (2019)	DL	GCN	CERT	Accuracy, precision, recall
Tao et al., (2023)	DL	Temporal Convolutional Network (TCN)	SapiMouse	AUC
Le et al., (2020)	ML	LR, RF, Neural Network (NN), and XGBoost	CERT	F1-Score, precision.
Nasir et al., (2021)	DL	LSTM-CNN, LSTM-RNN, and One Class SVM	CERT	Accuracy, presicion, F1-Score
Chowdhury et al., (2021)	DL/ML	LSTM, NN, RF, XGBoost, and SVM	Cooja simulator	Accuracy, precision, recall,F1-Score
Saminathan et al., (2023)	DL	NN	Domain Controller Activities (DCA)	Accuracy, FPR, Precision



				recall,F1-Score
Saminathan et al., (2023)	DL	NN	Domain Controller Activities (DCA)	Accuracy, FPR, Precision
Al-shehari & Alsowail (2021)	ML	LR, RF, SVM, Naïve Bayes (NB), DT, K-nearest neighbors (KNN), and Kernel SVM (KSVM)	CERT	Precision, F1-score, recall, confusion matrix, AUC-ROC
Anakath et al., (2022)	DL	Deep Belief Neural Network (DBNN)	Cooja simulator	Accuracy
Janjua et al., (2020)	ML	AdaBoost, Frequency-inverse document frequency (TF-IDF)	TWOS	Accuracy, AUC
Sharma, Pokharel, & Joshi, (2020)	ML	LSTM	CERT	Accuracy, Recall, FPR
Al-Shehari et al., (2024)	ML	Density-Based Local Outlier Factor (DBLOF)	CERT	F1-Score
Wang & El Saddik (2023)	NLP, DL	Digital Twin (DT), self-attention (SA)	CERT	Accuracy, precision, recall, F1-Score, and AUC-ROC
Yi & Tian (2024)	ML	KNN, Local Outlier Factor (LOF), and IF	CERT	AUC-ROC , accuracy
Lavanya, Glory, & Sriram, (2024)	DL	Deep Neural Network (DNN)	CERT	Precision, accuracy, Detection Rate (DR)
Al-Mhiqani et al., (2021)	ML	Adaptive Synthetic Sampling (ADASYN), (DNN)	CERT	Accuracy, FPR, F1-Score, TNR, AUC
Mehmood et al., (2023)	ML	RF, AdaBoost, XGBoost, LightGBM	CERT	Precision, accuracy, recall, F1-Score
Haq et al., (2022)	ML/DL	XGBoost, AdaBoost, RF, KNN, LR	Enron	Accuracy, precision, recall

Source: Authors

Peccatiello, Gondim, & Garcia (2023) paper investigated insider threat detection using AI and machine learning (ML) algorithms, focusing on real-life scenarios through a data stream approach and anomaly detection. The study employs Isolation Forest (ISOF), Elliptic Envelope (EV), and Local Outlier Factor (LOF) algorithms, combined with semi-supervised and supervised learning, and retraining procedures to enhance detection. ISOF achieved the best recall for both benign and malign classes, especially effective with a two-month retraining interval, while EV was the fastest. The research highlights the limitations of batch learning for real-world scenarios and the practicality of semi-supervised approaches. The Insider Threat Dataset (ITD) from Carnegie Mellon University, containing LDAP, device, email, HTTP, logon, and file logs, was utilized for evaluation.

Nasir et al. (2021). The study presented a deep learning model for insider threat detection through behavioral analysis, utilizing the CMU CERT synthetic insider threat dataset r4.2, which includes data from 1,000 users across logon, device, HTTP, email, file, psychometric, and LDAP logs. The proposed model, which combines Structural Anomaly Detection and Psychological Profiling, achieves a high accuracy of 90.60%, precision of 97%, and an F1 Score of 94%, outperforming other techniques like LSTM-CNN, LSTM-RNN, and One Class SVM. Despite its success, the study highlights the need for more diverse insider threat scenarios and real-life evaluations to further validate the model's robustness.

Hong et al. (2023). The paper introduced the ResHybnet model for insider threat detection, utilizing a combination of manual and automated feature engineering, including an LSTM auto-encoder for extracting features from sequential user activities. The model integrates GNN and CNN components and focuses on detecting threats based on daily behaviors. Tested on the CERT4.2 dataset (LDAP, Device, Email, File, HTTP activities) and the SEA command history dataset, ResHybnet outperformed other models by 1.97% and improved the F1 score by 0.56%. Despite its effectiveness, challenges such as imbalanced classification, limited datasets, and privacy concerns remain, with future work aiming to incorporate more data, online learning, and privacy solutions.

Chowdhury et al. (2021). The paper proposed a novel insider attack in IoT systems that exploits RPL vulnerabilities, simulated using the Cooja simulator in Contiki IoT. A machine learning framework, employing models such as LSTM, neural networks, Random Forest, XGBoost, and SVM, is developed to detect these attacks with high accuracy. XGBoost achieved the highest classification accuracy of 93.8%, followed closely by Random Forest at 93.7%. The proposed features, including source and destination, were key in detecting the attacks. The study highlights the ineffectiveness of existing security mechanisms against IoT-specific attacks and suggests future work to design an adaptive security framework. The evaluation used two datasets: one with a short duration of 100 minutes and another collected over a 24-hour period to capture learning patterns.

Saminathan et al. (2023). The paper proposed a deep learning model utilizing an ANN-based autoencoder for detecting insider cyber threats, achieving a high accuracy of 94.3%, a false positive rate of 11.1%, and a precision of 89.1%. The model focuses on anomaly detection through user and entity behavior analysis, using the Rectified Linear Activation Unit (ReLU) function in each layer. It outperforms existing methods in terms of precision, recall, and accuracy, particularly addressing the challenges posed by remote work in insider threat detection. The study employs large public datasets with diverse features, as well as locally generated data on mouse, keyboard, CPU, and memory usage. However, there is a noted lack of focus on real-time detection, scalability, and adaptability of the proposed model.

Al-shehari & Alsowail (2021) The study presented a machine learning model for detecting insider data leakage, achieving a high AUC-ROC value of 0.99. The model effectively addresses bias and class imbalance issues through data preprocessing techniques such as one-hot encoding, feature scaling, and the SMOTE technique, enhancing the precision, recall, and F-measure of various ML algorithms like Logistic Regression (LR), Decision Trees (DT), Random Forest (RF), Naive Bayes (NB), and K-Nearest Neighbors (KNN). The model's robustness was tested using the CERT dataset from Carnegie Mellon University (version R4.2.tar.bz), which contains activity logs of 1,000 insiders. Despite its success, the study highlights the need for more focused scenarios on insider data leakage detection.

Fei & Zhou (2024). The study introduced the ITI (Insider Threat Investigation) method, which enhances insider threat detection by leveraging causal graph analysis. ITI effectively reduces the scale of causal graphs and ranks true malicious employees higher than existing methods like AIRTAG, showing superior performance even in scenarios with varying data sizes. The method uses alarm features, entities, and rarity, combined with a sequence-based approach for pattern extraction and a graph convolution neural network (GCN) for analysis. Evaluated using CERT dataset versions r4.2, r5.2, and r6.2, the ITI method demonstrates significant improvements in detecting insider threats, although it lacks comparison with more diverse investigation methods and has limited scalability discussions for large organizations.

Haq et al. (2022). The study evaluated insider threat detection using machine learning (ML) and deep learning (DL) models, finding that ML models, such as XGBoost, Ada-Boost, Random Forest (RF), K-Nearest Neighbors (KNN), and Logistic Regression (LR), consistently outperform DL models in terms of accuracy, precision, and recall. The research also employs Word2Vec and GLoVe NLP models for transfer learning, demonstrating that pre-trained models offer higher accuracy compared to those built from scratch. Despite the higher performance of ML models, the study highlights the ongoing challenges of low accuracy and high false alarm rates in current detection approaches. Using real datasets, including the Enron dataset, the study emphasizes the need for real-time detection and the integration of non-technical aspects, while noting the lack of a standard framework for evaluating insider threat detection systems.

Anakath et al. (2022). The study explored insider attack detection in cloud computing networks using a Deep Belief Neural Network (DBN) model, which leverages user behavior patterns such as mouse movements and keystrokes to identify internal threats. The DBN model outperforms existing models like SVM and LSTM, achieving 99% accuracy in detecting insider attacks. By monitoring user behavior for feature extraction and anomaly detection, the model addresses the challenges of identifying malicious insider activities in cloud environments. The study emphasizes the effectiveness of DBN in enhancing security compared to traditional methods but notes a lack of ensemble models for more comprehensive detection. Open-source datasets were used for the simulation evaluation.

Janjua et al. (2020). The study focuses on detecting insider threats through linguistic analysis of emails, using supervised machine learning techniques. The AdaBoost algorithm, in combination with TF-IDF for text preprocessing, achieved the highest performance, with 98.3% accuracy and an AUC of 0.983 in classifying malicious emails. The research highlights the effectiveness of this approach in correctly identifying 98% of both malicious and normal emails. The study addresses issues related to limited data and overfitting by utilizing simpler models. Future work suggests exploring deep learning classifiers to enhance model performance further. The research utilized the TWOS dataset and analyzed the CERT dataset with a Hidden Markov method, but there is limited discussion on unsupervised learning techniques and comparisons with deep learning models.

Al-Shehari et al. (2024). The study introduced the Density-Based Local Outlier Factor (DBLOF) algorithm to enhance insider threat detection in imbalanced cybersecurity environments, particularly addressing challenges with skewed datasets like CERT r4.2. The DBLOF model achieved a 98% F-score and a 98.9 F1 score at a 0.02 contamination rate, demonstrating its effectiveness in identifying malicious insider activities as outliers. The research highlights the model's strength in detecting rare but dangerous insider threats, though it notes limitations in interpretability and scalability for real-time, high-dimensional data. Future work suggests focusing on advanced optimization techniques and cost-benefit analysis to improve the model's generalizability and reduce the risk of overfitting. The dataset used includes merged and cleaned insider activity logs from the CERT r4.2 dataset.

Wang & El Saddik (2023). The study introduced the DTITD (Distilled Transformer for Insider Threat Detection) framework, which employs deep learning and NLP techniques, notably using BERT and GPT-2 for data augmentation, to tackle insider threat detection. The DistilledTrans model, a simplified transformer architecture, outperforms existing models in accuracy, precision, and AUC, effectively addressing the challenges of data imbalance and unknown threats. While demonstrating high performance on CERT datasets, the paper suggests future work in sentiment analysis and user profile integration but lacks detailed explanations on model adjustments and result interpretation. Results: recall 84.62%; F1 90.53%, AUC 97.37%, an accuracy of 93.53%, and a precision of 98.13%.

Prasad, Nayak, & Krishna (2024). The study focused on improving insider threat detection in cybersecurity by addressing dataset imbalances through oversampling and under sampling techniques. It applies five machine learning (ML) algorithms—Logistic Regression, Decision Tree, Random Forest, Adaboost, and Naive Bayes—to balanced datasets, with ensemble learning and Principal Component Analysis (PCA) further enhancing model performance. The Random Forest, Adaboost, and Decision Tree algorithms achieved particularly strong results, with precision, recall, and accuracy being notably high. The study reports an F-score of 98%, indicating significant improvement over existing models. It highlights the challenges of dataset imbalance and the benefits of using advanced ML techniques, such as the DBLOF algorithm, for insider threat detection. However, concerns about model generalizability, overfitting, and scalability in real-time scenarios are acknowledged, suggesting the need for further optimization and testing. The research uses the CERT r4.2 insider threat dataset, emphasizing the importance of data preprocessing and cleaning for effective analysis.

Yi & Tian (2024). The proposed hybrid model enhanced insider threat detection by combining unsupervised and supervised learning techniques, integrating outlier scores to boost the predictive power of supervised classifiers. This approach achieves 86.12% accuracy while using only 20% of the computing budget, significantly outperforming other anomaly detection methods by up to 12.5%. The model effectively captures temporal information, improving early detection of insider threats. The study focuses on addressing challenges like data imbalances and the complexities of detecting threats due to authorized access, using the CERT r4.2 dataset for validation.

Lu & Wong (2019). The Insider Catcher system leveraged LSTM (Long Short-Term Memory) models to effectively detect malicious insider threats by analyzing user behavior patterns. Tested on the CERT Insider Threat Dataset V6.2, which includes real enterprise data, the system outperformed existing log-based anomaly detection strategies, proving its superior capability in distinguishing normal behavior from malicious activities. Recommendations for future work include testing other RNN algorithms and enhancing text analytics to improve detection accuracy further. Despite its success, the study notes a lack of comparison with other RNN models and highlights the need for enhanced content analysis. Recall 90%, Precision 72%, and F-measure 80%.

Wang, Sun, & Zhou (2023). The proposed insider threat detection method utilized deep clustering of multi-source behavioral events, employing an end-to-end deep neural network to learn user behavior features and improve detection accuracy. This method outperforms existing techniques like BAIT, Isolation Forest, and Random Forest, achieving AUC values exceeding 90 for all abnormal behavior types, with a 98 AUC and 99.8 recall for identifying malicious insiders. While effective in diverse enterprise environments, the study lacks a comparison with other models and discussions on scalability. The approach was tested using the CMU-CERT r4.2 dataset, capturing behaviors of 1000 employees.

Al-Shehari et al. (2023). The proposed insider threat detection model leveraged the anomaly-based Isolation Forest algorithm to address the class imbalance problem, achieving a 98% accuracy and an f-score of 99% on the CERT r4.2 dataset, which includes data from 1000 users and 7 malicious insiders. The model incorporates the Synthetic Minority Oversampling Technique (SMOTE) to improve detection performance in imbalanced datasets, outperforming previous studies. The approach is particularly effective in identifying insider data leakage attacks. Future work includes hyper-parameter tuning, feature selection, and exploring deep learning techniques, although the lack of real-world datasets remains a challenge in this research area.

Liu et al. (2019). A novel approach for insider threat detection utilizes the Word2vec model to transform security logs into texts, enabling effective identification of malicious insiders. The method involves components like Log2text and text2corpus for organizing data, improving detection accuracy by analyzing word similarities within security logs. The approach has demonstrated scalability and effectiveness in practical applications, with performance metrics such as True Positive Rate (TPR) and False Positive Rate (FPR) evaluated under various parameters. While the method shows promise, future work aims to enhance it by integrating more raw security log information and exploring more efficient models like Doc2vec. The approach has been tested on the CMU CERT v6.2 Programs insider threat database, showcasing its capability in real-world scenarios.

Sheykhkanloo & Hall (2020) examined the detection of insider threats by utilizing a spread subsample as their balancing technique. The authors reported that while dataset balancing techniques did not significantly improve performance metrics, they did reduce model-building time. The impact of adjusting classifier parameters, such as those in algorithms like J48, SVM, Naive Bayes, and Random Forest, was more pronounced on imbalanced data. The study highlights the importance of classifier tuning over dataset balancing in improving detection accuracy. Future work suggests exploring additional balancing techniques to enhance the detection of insider threats in imbalanced datasets.

Khan et al. (2020). The study focuses on detecting insider attacks in IoT environments using a lightweight AI-based approach. Utilizing the Levenshtein (LV) distance measurement technique, the proposed algorithm effectively addresses security challenges posed by internal threats in IoT devices. The approach improves accuracy while reducing false positives and computational overhead, making it a cost-effective solution. Implemented in R Studio, this AI-based detection method is compared with state-of-the-art techniques, demonstrating superior performance in identifying malicious insider activities in IoT systems.

Nicolaou, Shiaeles, & Savage, (2020) presented a bio-inspired model utilizing swarm intelligence algorithms and the EvoloPy-FS framework was developed to mitigate insider threats by improving the accuracy and speed of detecting malicious behavior in large datasets. This model employs bio-inspired computing for automating feature selection in machine learning models, focusing on optimizing feature subsets for anomaly detection. The methodology involved using unsupervised learning algorithms on synthetic datasets to detect outliers effectively. Results demonstrated that the bio-inspired model maintained near-optimal performance, similar to using the original features, thus enhancing model performance and aiding in the prevention of insider threats. The study highlights the growing security concerns posed by insider threats and the effectiveness of bio-inspired models in addressing these challenges through improved feature selection optimization.

Lavanya, Glory, & Sriram (2024) developed a novel insider threat detection model, combining Enhanced Bidirectional Generative Adversarial Networks (EBiGAN) and Deep Neural Network with Predictive Inference (DNN-PI), utilizes improved Principal Component Analysis (PCA), Bidirectional GAN, and Bayesian Optimization. This methodology addresses data imbalance issues in IoT-enabled institutions, achieving high detection rates with minimal false alarms. Improved PCA enhances user functionality samples and outlier estimations, while the Bidirectional GAN includes an additional discriminator for quality assurance of samples. Bayesian Optimization with the PI acquisition function fine-tunes the DNN model's hyperparameters, improving detection rates. Comparisons with SMOTE demonstrate EBiGAN's superior performance in classifier models, balancing data retention and computational efficiency. The proposed model effectively enhances security by addressing challenges in generalizability and interpretability of existing insider threat detection techniques. Datasets utilized are CERT v6.1 and v6.2

Al-Mhiqani et al. (2021) proposed AD-DNN model, which enhanced insider threat detection by integrating Adaptive Synthetic Sampling (ADASYN) and Deep Neural Networks (DNN), addressing the challenges of imbalanced data in traditional machine learning techniques. Implemented using Python with TensorFlow in an Ubuntu 18.04.5 LTS environment, the AD-DNN model was evaluated using the CERT r4.2 dataset, which includes various user activities such as logon/logoff, device usage, email, HTTP, and file activities. Evaluation metrics such as accuracy, false positive rate, F-Score, and true-negative rate showed that the AD-DNN model achieved a high AUC of 95, outperforming other classifiers like SVM, DNN, and LSTM. The integration of ADASYN effectively addressed data imbalance, significantly improving detection accuracy and overall performance, making it superior to current insider threat detection systems.

Mehmood et al. (2023) in their study proposed ML-based system detects insider threats using ensemble learning techniques, leveraging Random Forest (RF), AdaBoost, XGBoost, and LightGBM algorithms. This methodology employs bagging and boosting techniques for enhanced performance in insider threat detection and classification, utilizing data aggregation and normalization during preprocessing. Experiments conducted on the CERT dataset showed that the LightGBM algorithm outperformed others, achieving the highest accuracy of 97%, while RF, AdaBoost, and XGBoost achieved accuracies of 86%, 88%, and 88.27%, respectively. The study highlights the effectiveness of ensemble learning in identifying insider attacks, particularly privilege escalation attacks, and suggests future research to expand dataset sizes for further model enhancement. Utilized CERT r4.2

Sallam & Bertino (2019) proposed a system that detects insider threats in databases using advanced anomaly detection techniques, which monitor data access rates to identify suspicious user behavior. It employs components like Profiler, Mediator, and A-Detector for query inspection, leveraging syntactic, data-centric, and temporal features to represent user activity. The methodology includes both preliminary and deep inspections of query rates and tuple retrievals. Evaluation of real database query logs and T-DBMS logs during the training phase demonstrated that these techniques achieve high anomaly detection accuracy with low false alarm rates, effectively identifying data aggregation and tracking updates. Despite challenges in tracking data access frequencies and detecting sophisticated data misuse scenarios, the proposed methods show low error rates with sufficient data availability. However, the study lacks comparisons with existing anomaly detection solutions and discusses scalability and real-time implementation challenges only briefly.

Williams et al. (2022) examined the effectiveness of artificial neural networks (ANNs) in detecting insider threats at a nuclear facility, specifically the Nuclear Engineering Teaching Laboratory (NETL). The research aims to identify deviations that could indicate malicious insider activities by analyzing operational patterns. The ANN-based approach, tested using the ReconaSense AI Platform, successfully detected off-normal behaviors and enhanced insider threat detection and mitigation (ITDM). The study highlights the importance of collective behavior analysis in improving security measures. While the results show promise, further research is needed to address technical limitations and the impact of human policies, suggesting the integration of additional sensing data for future evaluations.

Lo et al. (2018) presented a study that focused on insider threat detection using distance measurement techniques and Hidden Markov Models (HMM) in cybersecurity. Techniques like Damerau-Levenshtein, Cosine, and Jaccard distances are analyzed for detecting behavior changes, with HMM showing the highest overall detection rate of 0.69. While distance methods detect unique malicious users faster, combining these techniques with HMM can detect up to 80% of insiders. The research highlights the difficulty in detecting insider threats with traditional signature detection methods and emphasizes the effectiveness of combining distance measurement techniques. Evaluations using the CERT r4.2 dataset indicate that these methods, despite their varying rates, enhance detection accuracy and computational speed. However, the study lacks detailed exploration of threshold techniques for distance measurements and comparisons of various distance algorithms.

Hu et al. (2019) proposed a method that utilized mouse dynamics and deep learning for insider threat detection, achieving continuous user authentication every 7 seconds with a low false acceptance rate of 2.94%. The study involved experimenting with data from ten users, demonstrating the method's effectiveness. Mouse actions were mapped to images to preserve features, and data augmentation techniques such as flipping and rotating were applied. Three experiments confirmed the method's effectiveness: Experiment A validated the CNN network for identity authentication, Experiment B demonstrated quick and continuous user verification, and Experiment C tested practical application using 'test-files' data. The method outperforms existing techniques that require minutes for data collection, making it a promising approach for addressing insider threats in intranet security. The experiments used the Balabit Mouse Dynamics Challenge dataset, illustrating the method's high accuracy and low error rates despite the small experimental data size and simulated malicious data usage.

Aljably, Tian, & Al-Rodhaan, (2020) study explores a model for privacy preservation in multimedia social networks using a combination of supervised and unsupervised machine learning techniques integrated with access control models. It achieved over 95% accuracy in detecting anomalous behavior using a Bayesian classifier, outperforming other methods such as SVM, Isolation Forest, PCA, and the Kolmogorov-Smirnov test. The model dynamically adjusts permission assignments based on detected anomalies, enhancing both privacy and security. While it successfully safeguards user data and prevents unauthorized access, the study highlights a need for further comparison with emerging anomaly detection techniques and a deeper exploration of its impact on different types of multimedia data.

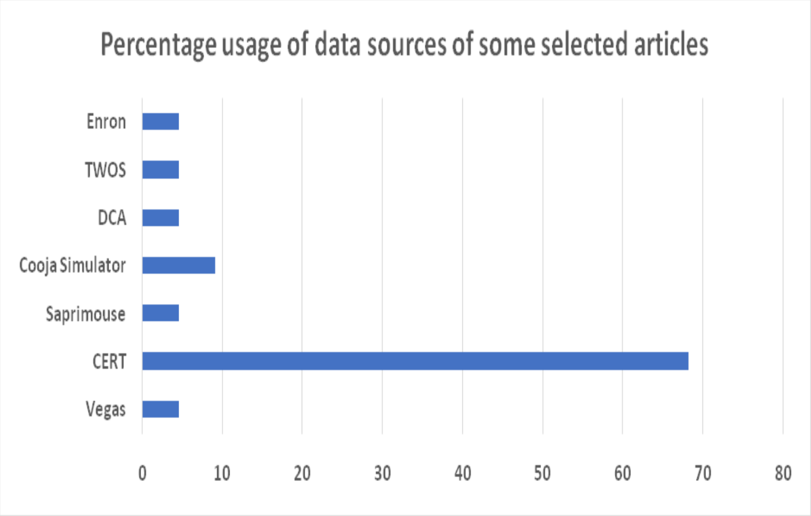
Li et al. (2021) introduced an Insider Threat Detection (IGT) method that leverages image-based feature representation and geometric transformations to enhance anomaly detection. The IGT method converts unsupervised anomaly detection into a supervised image classification task, using computer vision techniques to improve precision and reduce computational complexity. It outperforms traditional unsupervised approaches, including autoencoder-based methods, with improvements in AUROC by 4% and 2%. The proposed method effectively addresses challenges in insider threat detection, offering superior performance on the Carnegie Mellon University CERT insider threat dataset.

Zhang et al., (2021) proposed a method for insider threat detection, which employed ensemble learning combined with self-supervised learning, achieving high AUCs of 99.2% and 95.3% on the CERT4.2 and CERT6.2 datasets, respectively. This method addresses the extreme imbalance between legitimate and malicious user behavior data using an over-bootstrap sampling strategy to balance the training data. Entity representation based on TF-IDF improves detection effectiveness. Despite challenges like overfitting due to the high proportion of one data type, the proposed method surpasses competing methods in detection metrics, including DR, F1, and AUC values. The approach leverages ensemble learning and self-supervised learning to effectively detect insider threats, significantly outperforming other methods in terms of AUC values. The findings are supported by evaluations on the CERT4.2 and CERT6.2 datasets, demonstrating the method's superior performance in mitigating the harmful effects of insider threats compared to existing detection techniques.

Sharma, Pokharel, & Joshi (2020) study examined using an LSTM Autoencoder for anomaly detection in user behavior analytics, specifically targeting insider threats. By modeling user activities, the approach achieved a high accuracy of 90.17%, with a True Positive Rate of 91.03% and a False Positive Rate of 9.84%. The model effectively detects anomalies by analyzing reconstruction errors from user sessions. While focusing on improving true positive rates, the study also addresses challenges such as limited anomalous data and difficulties in detecting unknown threat patterns. The research utilizes the CERT insider threat dataset, highlighting the effectiveness of LSTM Autoencoders in enhancing cybersecurity through accurate anomaly detection.

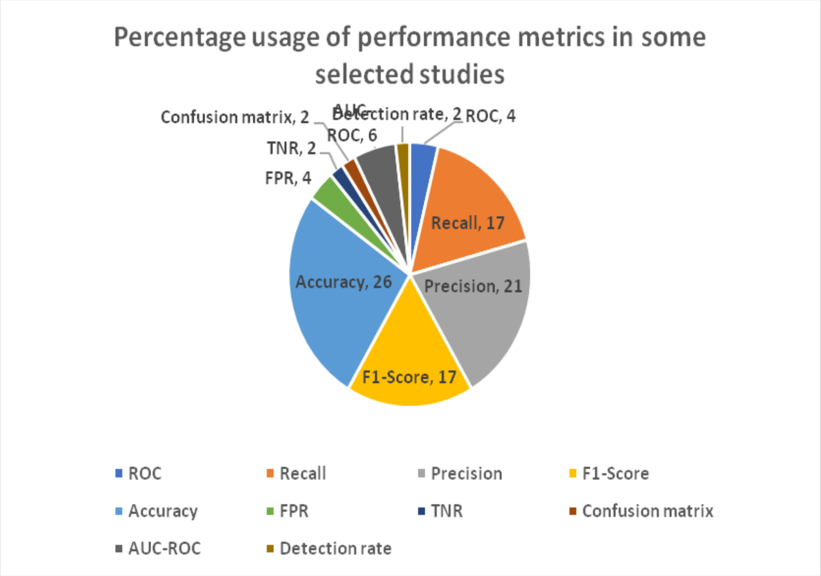
Insider threat detection is challenging due to the unassuming nature of the malicious activities. With the abundant use of information technology, extensive data is generated concerning users' and systems' behaviors, including text records such as logs, audits, and emails. Moreover, most insider activities are benign, making it uneconomical to label labeled data for further classification. Nevertheless, the proposed algorithms achieve satisfactory performance in the datasets analyzed and show promising potential for automatic detection concerning insider activities. Figure 7 and Figure 8 display the percentage of data sources employed by some of the reviewed articles and the performance metrics used, respectively.

Figure 7: Percentage usage of data sources of some selected articles



Source: Authors

Figure 8: Percentage usage of performance metrics in some of the selected studies



Source: Authors

Future Trends and Directions

In 2022, organizations around the globe witnessed a 31% increase in insider threats, taking the number of cases up to 2000. This necessitated the improvement of existing capabilities and the enhancement of operational security. Artificial Intelligence (AI) has emerged as one of the most useful tools for industries to improve efficacy and efficiency significantly (Alsowail & Al-Shehari, 2022). Applying artificial intelligence (AI) techniques to detect insider threats caused by malicious actors is beneficial to organizations in preventing misuse (Arif et al., 2023).

The academic literature reveals a wide range of advancements in artificial intelligence related to the detection of insider threats. Initial endeavors focused on developing machine learning-based techniques and detection of insider threats within companies. However, several AI techniques have emerged and been applied in different sectors and, as such, should be tapped into for the timely detection of malicious insider threats.

Conclusion and Recommendations

In this paper, we explore the current state of AI techniques and their effectiveness in identifying and preventing malicious insider threats in organizations. We explore the current methodologies and technological frameworks for detecting malicious insider threats and examine the types of data employed along with the evaluation metrics utilized.

Preserving organizational data privacy has emerged as a paramount concern for data-intensive companies. Unfortunately, employees with access to confidential data often pose the most significant threat to such information. Various factors are behind these attacks, such as feelings of negligence, grievances against the company, or merely their monetary value. As verified by leading security analysis firms, yearly financial losses are experienced due to insider threats and data breaches. It has been difficult to combat this issue due to the attackers' degree of freedom and trusted access to confidential data. Focusing solely on preventing data leaks to outside attackers is inadequate to protect potentially exposed data. Methods like encryption, firewalls, and intrusion detection systems are ineffective in controlling insider misconduct, as they often disregard activities performed with legitimate access. Therefore, organizations must adequately monitor their data and system usage patterns, behavior, and occurrences and understand their truthfulness.

Artificial intelligence, with its powerful tools and methods, can explore this terrain to design effective strategies for combating insider threats. Equipped with machine learning and deep learning techniques, adversarial models of inference and big data analysis can be trained accordingly to evaluate the need for improper data accesses efficiently. After sufficient training with legitimate data use patterns or occurrences, these models can be incorporated into the monitoring systems to detect suspicious activities on the data by actively observing the system in use.

References:

Abiodun, M. K., Adeniyi, A. E., Victor, A. O., Awotunde,J. B., Atanda, O. G., and Adeniyi, J. K. (2023). Detection and prevention of data leakage in transit using lstm recurrent neural network with encryption algorithm. In the 2023 International Conference on Science, Engineering and Business for Sustainable Development Goals (SEB-SDG), volume 1, pages 01–09.

Agrawal, J., Kalra, S. S., & Gidwani, H. (2023). AI in cyber security. *International Journal of Communication and Information Technology*, *4*(1), 46–53. https://doi.org/10.33545/2707661x.2023.v4.i1a.59

Akoh Atadoga, Enoch Oluwademilade Sodiya, Uchenna Joseph Umoga, & Olukunle Oladipupo Amoo. (2024). A comprehensive review of machine learning's role in enhancing network security and threat detection. *World Journal of Advanced Research and Reviews*, *21*(2), 877–886. https://doi.org/10.30574/wjarr.2024.21.2.0501

Al-Mhiqani, M. N., Isnin, S. N., Ahmed, R., & Abidi, Z. Z. (2021). An Integrated Imbalanced Learning and Deep Neural Network Model for Insider Threat Detection. *International Journal of Advanced Computer Science and Applications*, *12*(1), 1–5.

Al-Shehari, T. A., Rosaci, D., Al-Razgan, M., Alfakih, T., Kadrie, M., Afzal, H., & Nawaz, R. (2024). Enhancing Insider Threat Detection in Imbalanced Cybersecurity Settings Using the Density-Based Local Outlier Factor Algorithm. *IEEE Access*, *12*(January), 34820–34834. https://doi.org/10.1109/ACCESS.2024.3373694

Al-Shehari, T., Al-Razgan, M., Alfakih, T., Alsowail, R. A., & Pandiaraj, S. (2023). Insider Threat Detection Model Using Anomaly-Based Isolation Forest Algorithm. *IEEE Access*, *11*(October), 118170–118185. https://doi.org/10.1109/ACCESS.2023.3326750

Al-shehari, T., & Alsowail, R. A. (2021). An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques. *Entropy*, *23*(10). https://doi.org/10.3390/e23101258

Aljably, R., Tian, Y., & Al-Rodhaan, M. (2020). Preserving Privacy in Multimedia Social Networks Using Machine Learning Anomaly Detection. *Security and Communication Networks*, *2020*. https://doi.org/10.1155/2020/5874935

Alsowail, R. A., & Al-Shehari, T. (2022). Techniques and countermeasures for preventing insider threats. *PeerJ Computer Science*, *8*. https://doi.org/10.7717/PEERJ-CS.938

Alzaabi, F. R., & Mehmood, A. (2024). A Review of Recent Advances, Challenges, and Opportunities in Malicious Insider Threat Detection Using Machine Learning Methods. *IEEE Access*, *12*, 30907–30927. https://doi.org/10.1109/ACCESS.2024.3369906

Anakath, A. S., Kannadasan, R., Joseph, N. P., Boominathan, P., & Sreekanth, G. R. (2022). Insider attack detection using deep belief neural network in cloud computing. *Computer Systems Science and Engineering*, *41*(2), 479–492. https://doi.org/10.32604/csse.2022.019940

Anju, A., Krishnamurthy, M., Nithakalyani, M., Shalini, K., & Haritha, R. (2022). A Review to Analyze Insider Threats Using Machine Learning Techniques. In . *In International Conference on Information and Communication Technology for Competitive Strategies*. https://doi.org/10.1007/978-981-19-9638-2_55

Arif, H., Kumar, A., Fahad, M., & Hussain, H. K. (2023). *Future Horizons: AI-Enhanced Threat Detection in Cloud Environments: Unveiling Opportunities for Research*. *2*(2), 242–251.

Capelli, D., Moore, A., & Trzeciak, R. (2012). *The CERT guide to insider threats*. USA. Retrieved from https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=The+CERT+Guide+to+Insider+Threats+%5B2012%5D&oq=the+

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

Chowdhury, M., Ray, B., Chowdhury, S., & Rajasegarar, S. (2021). A Novel Insider Attack and Machine Learning Based Detection for the Internet of Things. *ACM Transactions on Internet of Things*, 2(4), 1–23. https://doi.org/10.1145/3466721

Fei, K., & Zhou, J. (2024). An Insider Threat Investigation Method by Graph Analysis with Log Texts. *ACM International Conference Proceeding Series*, 19–23. https://doi.org/10.1145/3672121.3672126

Gavai, G., Sricharan, K., Gunning, D., Hanley, J., Singhal, M., & Rolleston, R. (2015). Detecting insider threat from enterprise social and online activity data. *MIST 2015 - Proceedings of the 7th ACM CCS International Workshop on Managing Insider Security Threats, Co-Located with CCS 2015*, 13–20. https://doi.org/10.1145/2808783.2808784

Goldberg, H. G., Young, W. T., Reardon, M. G., Phillips, B. J., & Senator, T. E. (2017). Insider threat detection in PRODIGAL. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2017-Janua*, 2648–2657. https://doi.org/10.24251/hicss.2017.320

Hall, A. J., Pitropakis, N., Buchanan, W. J., & Moradpoor, N. (2018). Predicting Malicious Insider Threat Scenarios Using Organizational Data and a Heterogeneous Stack-Classifier. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 5034–5039. https://doi.org/10.1109/BigData.2018.8621922

Haq, M. A., Khan, M. A. R., & Alshehri, M. (2022). Insider Threat Detection Based on NLP Word Embedding and Machine Learning. *Intelligent Automation and Soft Computing*, *33*(1), 619–635. https://doi.org/10.32604/iasc.2022.021430

Hong, W., Yin, J., You, M., Wang, H., Cao, J., Li, J., … Man, C. (2023). A graph empowered insider threat detection framework based on daily activities. *ISA Transactions*, *141*, 84–92. https://doi.org/10.1016/j.isatra.2023.06.030

Hu, T., Niu, W., Zhang, X., Liu, X., Lu, J., & Liu, Y. (2019). An Insider Threat Detection Approach Based on Mouse Dynamics and Deep Learning. *Security and Communication Networks*, *2019*. https://doi.org/10.1155/2019/3898951

Ismaila, I., & Adeleke, N. D. (2023). Systematic Literature Review and Metadata Analysis of Insider Threat Detection Mechanism.

International Journal of Computer Science and Mobile Computing, *12*(4), 60–88. https://doi.org/10.47760/ijcsmc.2023.v12i04.007

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

Jiang, J., Chen, J., Gu, T., Choo, K. K. R., Liu, C., Yu, M., … Mohapatra, P. (2019). Anomaly Detection with Graph Convolutional Networks for Insider Threat and Fraud Detection. *Proceedings - IEEE Military Communications Conference MILCOM*, *2019-Novem*, 109–114. https://doi.org/10.1109/MILCOM47813.2019.9020760

Khan, A. Y., Latif, R., Latif, S., Tahir, S., Batool, G., & Saba, T. (2020). Malicious Insider Attack Detection in IoTs Using Data Analytics. *IEEE Access*, *8*, 11743–11753. https://doi.org/10.1109/ACCESS.2019.2959047

Lavanya, P., Glory, H. A., & Sriram, V. S. (2024). Mitigating Insider Threat: A Neural Network Approach for Enhanced Security. *IEEE Access*, *12*(June), 73752–73768. https://doi.org/10.1109/ACCESS.2024.3404814

Le, D. C., Zincir-Heywood, N., & Heywood, M. I. (2020). Analyzing Data Granularity Levels for Insider Threat Detection Using Machine Learning. *IEEE Transactions on Network and Service Management*, *17*(1), 30–44. https://doi.org/10.1109/TNSM.2020.2967721

Li, D., Yang, L., Zhang, H., Wang, X., Ma, L., & Xiao, J. (2021). Image-Based Insider Threat Detection via Geometric Transformation. *Security and Communication Networks*, *2021*. https://doi.org/10.1155/2021/1777536

Liu, L., Chen, C., Zhang, J., De Vel, O., & Xiang, Y. (2019). Insider Threat Identification Using the Simultaneous Neural Learning of Multi-Source Logs. *IEEE Access*, *7*, 183162–183176. https://doi.org/10.1109/ACCESS.2019.2957055

Lo, O., Buchanan, W. J., Griffiths, P., & Macfarlane, R. (2018). Distance measurement methods for improved insider threat detection. *Security and Communication Networks*, *2018*. https://doi.org/10.1155/2018/5906368

Lu, J., & Wong, R. K. (2019). Insider Threat Detection with Long Short-Term Memory. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3290688.3290692

Manoharan, A., & Sarker, M. (2024). Revolutionizing Cybersecurity: Unleashing the Power of Artificial Intelligence and Machine Learning for Next-Generation Threat Detection. *International Research Journal of Modernization in Engineering Technology and Science*, (12), 2151–2164. https://doi.org/10.56726/irjmets32644

Mehmood, M., Amin, R., Muslam, M. M. A., Xie, J., & Aldabbas, H. (2023). Privilege Escalation Attack Detection and Mitigation in Cloud Using Machine Learning. *IEEE Access*, *11*(April), 46561–46576. https://doi.org/10.1109/ACCESS.2023.3273895

Moekthi Prajitno, N. T., Hadiyanto, H., & Rochim, A. F. (2023). Research Opportunity of Insider Threat Detection based on Machine Learning Methods. *5th International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2023*, 292–296. https://doi.org/10.1109/ICAIIC57133.2023.10067010

Naseer, I. (2024). Machine Learning Applications in Cyber Threat Intelligence: A Comprehensive Review. *The Asian Bulletin of Big Data Management*, *3*(2), 190–200. https://doi.org/10.62019/abbdm.v3i2.85

Nasir, R., Afzal, M., Latif, R., & Iqbal, W. (2021). Behavioral Based Insider Threat Detection Using Deep Learning. *IEEE Access*, *9*, 143266–143274. https://doi.org/10.1109/ACCESS.2021.3118297

Nicolaou, A., Shiaeles, S., & Savage, N. (2020). Mitigating insider threats using bio-inspired models. *Applied Sciences (Switzerland)*, *10*(15). https://doi.org/10.3390/app10155046

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*(July), 70560–70573. https://doi.org/10.1109/ACCESS.2023.3293825

Prasad, P. S. S., Nayak, S. K., & Krishna, M. V. (2024). Enhanced Insider Threat Detection Through Machine Learning Approach With Imbalanced Data Resolution. *Journal of Theoretical and Applied Information Technology*, *102*(3), 914–926.

Sallam, A., & Bertino, E. (2019). Result-based detection of insider threats to relational databases. *CODASPY 2019 - Proceedings of the*

9th ACM Conference on Data and Application Security and Privacy, 133–143. https://doi.org/10.1145/3292006.3300039

Saminathan, K., Mulka, S. T. R., Damodharan, S., Maheswar, R., & Lorincz, J. (2023). An Artificial Neural Network Autoencoder for Insider Cyber Security Threat Detection. *Future Internet*, *15*(12). https://doi.org/10.3390/fi15120373

Schulze, H. (2024). 2024 INSIDER THREAT REPORT: CYBERSECURITY INSIDERS. In *Vormetric*.

Sharma, B., Pokharel, P., & Joshi, B. (2020). User Behavior Analytics for Anomaly Detection Using LSTM Autoencoder-Insider Threat Detection. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3406601.3406610

Sheykhkanloo, N. M., & Hall, A. (2020). Insider threat detection using supervised machine learning algorithms on an extremely imbalanced dataset. *International Journal of Cyber Warfare and Terrorism*, *10*(2), 1–26. https://doi.org/10.4018/IJCWT.2020040101

Tao, X., Yu, Y., Fu, L., Liu, J., & Zhang, Y. (2023). An insider user authentication method based on improved temporal convolutional network. *High-Confidence Computing*, *3*(4), 100169. https://doi.org/10.1016/j.hcc.2023.100169

Tuor, A., Kaplan, S., Hutchinson, B., Nichols, N., & Robinson, S. (2017). Deep learning for unsupervised insider threat detection in structured cybersecurity data streams. *AAAI Workshop - Technical Report*, *WS-17-01*-(2012), 224–234.

Ugochukwu Ikechukwu Okoli, Ogugua Chimezie Obi, Adebunmi Okechukwu Adewusi, & Temitayo Oluwaseun Abrahams. (2024). Machine learning in cybersecurity: A review of threat detection and defense mechanisms. *World Journal of Advanced Research and Reviews*, *21*(1), 2286–2295. https://doi.org/10.30574/wjarr.2024.21.1.0315

Wang, J., Sun, Q., & Zhou, C. (2023). Insider Threat Detection Based on Deep Clustering of Multi-Source Behavioral Events. *Applied Sciences (Switzerland)*, *13*(24). https://doi.org/10.3390/app132413021

Wang, Z. Q., & El Saddik, A. (2023). DTITD: An Intelligent Insider Threat Detection Framework Based on Digital Twin and Self-Attention Based

Deep Learning Models. *IEEE Access*, *11*(September), 114013–114030. https://doi.org/10.1109/ACCESS.2023.3324371

Wanyonyi, E. N., Abeka, S., & Masinde, N. (2023). A Systematic Review on Machine Learning Insider Threat Detection Models, Datasets and Evaluation Metrics. *International Journal of Network Security & Its Applications*, *15*(6), 37–56. https://doi.org/10.5121/ijnsa.2023.15603

Williams, A. D., Abbott, S. N., Shoman, N., & Charlton, W. S. (2022). Results from Invoking Artificial Neural Networks to Measure Insider Threat Detection & Mitigation. *Digital Threats: Research and Practice*, *3*(1). https://doi.org/10.1145/3457909

Yi, J., & Tian, Y. (2024). Insider Threat Detection Model Enhancement Using Hybrid Algorithms between Unsupervised and Supervised Learning. *Electronics (Switzerland)*, *13*(5). https://doi.org/10.3390/electronics13050973

Yilmaz, E., & Can, O. (2024). Unveiling Shadows: Harnessing Artificial Intelligence for Insider Threat Detection. *Engineering, Technology and Applied Science Research*, *14*(2), 13341–13346. https://doi.org/10.48084/etasr.6911

Yuan, F., Cao, Y., Shang, Y., Liu, Y., Tan, J., & Fang, B. (2018). Insider threat detection with deep neural network. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer International Publishing. https://doi.org/10.1007/978-3-319-93698-7_4

Yuan, S., & Wu, X. (2021). Deep learning for insider threat detection: Review, challenges and opportunities. *Computers and Security*, *104*, 102221. https://doi.org/10.1016/j.cose.2021.102221

Zhang, C., Wang, S., Zhan, D., Yu, T., Wang, T., & Yin, M. (2021). Detecting Insider Threat from Behavioral Logs Based on Ensemble and Self-Supervised Learning. *Security and Communication Networks*, *2021*. https://doi.org/10.1155/2021/4148441