IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 January 2024, accepted 3 February 2024, date of publication 26 February 2024, date of current version 1 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3369906

SURVEY

A Review of Recent Advances, Challenges, and Opportunities in Malicious Insider Threat Detection Using Machine Learning Methods

FATIMA RASHED ALZAABI AND ABID MEHMOOD, (Member, IEEE)
College of Engineering, Abu Dhabi University, Abu Dhabi, United Arab Emirates

Corresponding author: Abid Mehmood (abid.mehmood@adu.ac.ae)

This work was supported by the Abu Dhabi University's Office of Research and Grant Programs. Abu Dhabi University's office of sponsored programs in the United Emirates funded this endeavor under Grant 19300752.

ABSTRACT Insider threat detection has become a paramount concern in modern times where organizations strive to safeguard their sensitive information and critical assets from malicious actions by individuals with privileged access. This survey paper provides a comprehensive overview of insider threat detection, highlighting its significance in the current landscape of cybersecurity. The review encompasses a broad spectrum of methodologies and techniques, with a particular focus on classical machine-learning approaches and their limitations in effectively addressing the intricacies of insider threats. Furthermore, the survey explores the utilization of modern deep learning and natural language processing (NLP) based methods as promising alternatives, shedding light on their advantages over traditional methods. The comprehensive analysis of results from experiments utilizing NLP and large language models to detect malicious insider threats on the CMU CERT dataset reveals promising insights. Studies surveyed in this paper indicate that these advanced techniques demonstrate notable efficacy in identifying suspicious activities and anomalous behaviors associated with insider threats within organizational systems. Additionally, the survey underscores the potential of NLP and large language model-based approaches, which can enhance threat detection by deciphering textual and contextual information. In the conclusion section, the paper offers valuable insights into the future directions of insider threat detection. It advocates for the integration of more sophisticated time-series-based techniques, recognizing the importance of temporal patterns in insider threat behaviors. These recommendations reflect the evolving nature of insider threats and emphasize the need for proactive, data-driven strategies to safeguard organizations against internal security breaches. In conclusion, this survey not only underscores the urgency of addressing insider threats but also provides a roadmap for the adoption of advanced methodologies to enhance detection and mitigation capabilities in contemporary cybersecurity paradigms.

INDEX TERMS Insider threat detection, privilege escalation, anomaly detection, user action graph, cyber security, user behavior, temporal information, pre-trained language models, word embedding, CERT dataset.

I. INTRODUCTION

Insider threat detection is a critical component of cybersecurity, primarily concerned with identifying and mitigating security risks originating from individuals with authorized access to an organization's systems, data, or facilities. This domain has gained increasing significance due to the rising frequency and severity of insider threats, encompassing activities such as data theft, fraud, sabotage, and espionage by trusted insiders. Statistical evidence underscores the urgency of addressing this issue, with insider threats responsible for 35% data breaches in a 2023 Verizon Data Breach

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Huang.

Investigations Report [1]. Malicious insider threats, in particular, inflict damage 20 times greater than external actors and incur an average cost of $4.5 million per breach.

This survey delves into the multifaceted field of insider threat detection, elucidating the approaches and technologies employed while considering behavioral, technical, and organizational aspects. It also explores the evolving landscape of insider threats in the context of technological advancements, remote work paradigms, and the integration of artificial intelligence and machine learning-driven solutions. In sectors like banking, healthcare, government, technology, and workplace safety, identifying and classifying malicious insiders are essential for safeguarding sensitive data, maintaining privacy, ensuring compliance with regulations, preventing financial losses, and addressing the evolving nature of insider threats in a technologically advancing world.

The relevance of insider threat detection and classification is further emphasized by the evolving methods and motivations of malicious insiders, the increasing volume of sensitive data, stringent regulatory requirements, potential financial repercussions, and the enhanced capabilities provided by AI and machine learning for detection and classification in modern organizations [2].

II. BACKGROUND AND MOTIVATION

The historical development of insider threat detection has evolved significantly over time, reflecting the changing landscape of security challenges. Initially, insider threat detection relied on simplistic methods, primarily rooted in interpersonal trust and rudimentary surveillance practices [3], [4]. During the pre-digital era, the primary focus was on physical security, with security personnel playing a pivotal role in identifying suspicious behavior among employees. However, as technology advanced, so did the nature of insider threats.

With the advent of computer technology and the digitalization of business processes, insider threats transformed in the 1980s and 1990s, taking the form of unauthorized access to computer systems and data theft [5]. To counter these emerging threats, the field of insider threat detection underwent adaptations. Early strategies included the development of access control mechanisms and user authentication systems, such as password-based systems and user access logs. While effective at identifying unauthorized access, these methods had limitations when it came to detecting subtler insider threats, like privileged users abusing their access rights [3], [6].

As technology continued to advance, insider threat detection methods evolved accordingly. In the 2000s, behavioral analytics emerged as a promising approach. This method involved monitoring and analyzing user behavior to identify deviations from established patterns. By creating user profiles and scrutinizing changes in behavior, organizations could better detect anomalies indicative of insider threats. Concurrently, data loss prevention (DLP) solutions gained prominence. DLP systems focused on monitoring and preventing the unauthorized transfer of sensitive data, a critical aspect of combating insider threats related to data exfiltration [7]. Furthermore, the integration of machine learning and artificial intelligence (AI) into insider threat detection marked a significant turning point. AI-driven solutions could analyze vast datasets in real-time, enabling organizations to detect subtle patterns and anomalies that traditional methods might overlook. Machine learning models excelled at identifying changes in user behavior, promptly flagging potentially malicious activities [8].

In recent years, the convergence of technologies like big data analytics, cloud computing, and the Internet of Things (IoT) has further expanded the scope of insider threat detection. These technologies generate massive amounts of data that can be harnessed to enhance detection capabilities. Additionally, the shift toward remote and hybrid work environments, accelerated by the COVID-19 pandemic, has introduced new complexities to insider threat detection. Employees now access organizational resources from various locations and devices, demanding adaptability in detection strategies [9].

III. LITERATURE REVIEW

The literature survey begins by delineating four fundamental facets within the domain of insider threats: types of insider threats, insider threat detection approaches, data sources and features, and machine learning and data mining techniques. These constituent elements represent critical dimensions of research and inquiry surrounding the complex issue of insider threats to organizational security. The following sections will provide an in-depth exploration of each of these components, explaining the existing body of knowledge, and discerning prevalent trends and developments therein. Through this survey, the research aims to offer a comprehensive overview of the extant literature in the field, thereby contributing to the broader understanding and effective mitigation of insider threats.

A. MAJOR CATEGORIES OF INSIDER THREATS

Insider threats can be categorized based on various criteria, each shedding light on different aspects of these security challenges. One fundamental categorization of insider threats revolves around the individual's intent and motivations. This criterion differentiates between malicious intent, accidental actions, and fraudulent activities. One such classification is shown in Figure 1.

1) MALICIOUS INSIDER THREATS

Malicious insider threats encompass actions taken by individuals with the deliberate intent to harm their organization. These individuals may engage in activities such as data theft, sabotage, or espionage, driven by a desire for revenge, ideological beliefs, or personal gain [10]. Their actions are characterized by a conscious and purposeful effort to compromise security. Perhaps one of the most famous cases of a malicious insider threat, Edward Snowden, a former

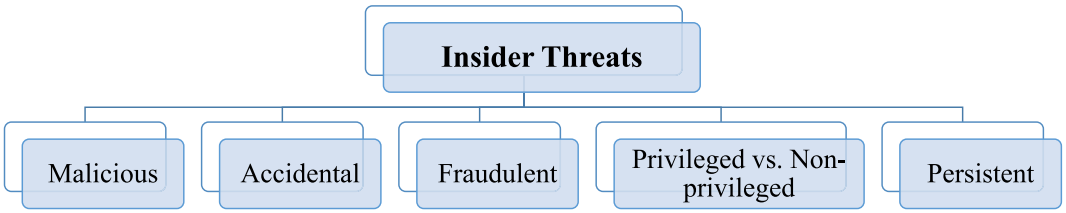

FIGURE 1. Major categories of insider threats.

contractor for the U.S. National Security Agency (NSA), intentionally leaked classified information in 2013 [11]. Snowden's actions were driven by his belief that the NSA's mass surveillance programs violated privacy rights. He exposed extensive details about these programs, leading to significant repercussions and debates on government surveillance. Bradley Manning, a U.S. Army intelligence analyst, intentionally leaked classified military documents and diplomatic cables in 2010 to WikiLeaks. Manning's actions were motivated by a desire to expose what he perceived as government wrongdoing and human rights abuses. This massive data breach had wide-ranging diplomatic and security implications [12].

2) ACCIDENTAL INSIDER THREATS

Conversely, accidental insider threats involve actions taken by individuals without any malicious intent. These may include inadvertent data breaches, misconfigurations, or errors in judgment that inadvertently compromise security. Accidental insider threats often stem from a lack of awareness or training regarding security protocols and best practices. The Facebook Data Leak - In 2019, Facebook unintentionally exposed hundreds of millions of user records due to a misconfigured server [13]. This accidental insider threat occurred when sensitive data, including user passwords, was stored in an unsecured format. It was a result of an oversight in Facebook's data handling practices rather than malicious intent. In 2017, a bug in the Dropbox file-sharing service allowed users to log in to accounts without the correct password. This accidental insider threat stemmed from a software flaw, enabling unauthorized access to user accounts and data [14]. Dropbox addressed the issue promptly once it was identified.

3) FRAUDULENT INSIDER THREATS

Fraudulent insider threats pertain to actions taken by individuals with the intent to deceive or commit fraud within the organization. Such threats may involve embezzlement, financial manipulation, or other deceptive practices aimed at personal financial gain. Fraudulent insiders manipulate their position within the organization to exploit vulnerabilities for their illicit objectives. The Enron scandal [15] of the early 2000s involved several top executives engaging in fraudulent insider threats. Key individuals, including the CEO and CFO, manipulated financial statements and concealed massive debts to inflate the company's stock value. The fraud ultimately led to the bankruptcy of Enron and the loss of billions of dollars for investors and employees. In 2016, Wells Fargo faced a scandal when it was revealed that employees had fraudulently opened millions of unauthorized bank accounts and credit cards in customers' names. This fraudulent insider threat aimed at meeting sales targets and earning incentives, resulting in financial harm to affected customers and damage to the bank's reputation [16].

4) PRIVILEGED VS. NON-PRIVILEGED INSIDER THREATS

This categorization distinguishes between privileged insiders, who typically possess elevated access within the organization, and non-privileged insiders, who have limited influence and access. A notable case illustrating the significance of privileged insiders involves the 2013 cyberattack on Target's payment system, where cybercriminals compromised credit card information for millions of customers. This breach was facilitated by an HVAC contractor with privileged access to Target's network for maintenance purposes [17]. In contrast, the 2017 Equifax data breach, affecting 147 million individuals, was not caused by a privileged insider but resulted from a vulnerability in Equifax's website software, which was exploited by cybercriminals to gain unauthorized access to sensitive customer data [18].

5) DURATION OF INSIDER THREATS

Additionally, insider threats can be categorized based on their duration or persistence. Some insider threats may manifest as isolated incidents or short-term actions, while others may involve prolonged, continuous, or recurring activities. Understanding the duration of insider threats is crucial for devising effective detection and mitigation strategies. Edward Lee Howard, a former CIA officer, posed a long-term insider threat by defecting to the Soviet Union in 1985. He had access to classified intelligence for an extended period [19], which he potentially shared with Soviet authorities over several years, compromising U.S. national security.

B. INSIDER THREAT DETECTION APPROACHES

Insider threat detection is a critical area of focus in the field of cybersecurity, necessitating a variety of approaches and techniques for effective mitigation. The presented summary in Table 1 provides a comparison between the diverse ML and DL techniques employed in insider threat detection.

TABLE 1. A comprehensive overview of the diverse ML and DL approaches employed in detecting insider threats.

Paper Ref.	Technique			Approach		
	ML	DL	Hybrid	Behavior-Based	Network-Based	Other
[20]	×	×	✓	✓	×	×
[21]	✓	×	×	×	×	✓
[22]	✓	×	×	×	✓	×
[23]	✓	×	×	×	×	Behavioral + Rule based
[24]	✓	×	×	✓	×	×
[25]	✓	×	×	✓	×	×
[26]	×	✓	×	✓	×	×
[27]	×	×	✓	✓	×	×
[28]	×	×	✓	×	×	Behavioral + Network Traffic Analysis based
[29]	✓	×	×	×	×	Behavioral + Network Traffic Analysis based
[30]	×	✓	×	✓	×	×
[31]	✓	×	×	✓	×	×
[32]	✓	×	×	✓	×	×
[33]	✓	×	×	×	✓	×
[34]	×	×	✓	×	×	Behavioral + Network Traffic Analysis based
[35]	×	×	✓	✓	×	×
[36]	×	✓	×	✓	×	×
[37]	×	×	✓	✓	×	×
[38]	×	✓	×	×	×	Behavioral + Content based
[39]	✓	×	×	✓	×	×
[40]	✓	×	×	✓	×	×
[41]	×	✓	×	✓	×	×
[42]	×	×	✓	✓	×	×
[43]	×	×	✓	✓	×	×
[44]	×	×	✓	✓	×	×
[45]	×	✓	×	✓	×	×
[46]	✓	×	×	✓	×	×
[47]	×	✓	×	✓	×	×
[48]	×	✓	×	✓	×	×
[49]	×	×	✓	✓	×	×
[50]	✓	×	×	✓	×	×
[51]	✓	×	×	✓	×	×
[52]	×	×	✓	✓	×	×
[53]	✓	×	×	✓	×	×
[54]	✓	×	×	✓	×	×
[55]	✓	×	×	✓	×	×
[56]	✓	×	×	×	×	Behavioral + Rule based
[57]	✓	×	×	✓	×	×
[58]	×	×	✓	✓	×	×
[59]	✓	×	×	✓	×	×

These approaches and techniques can be broadly categorized into several key strategies. These various approaches for insider threat detection collectively contribute to enhancing an organization's capability to identify and respond to the diverse and evolving nature of insider threats. The malicious insider threat methods can be classified into different categories based on the underlying technique used.

Some frequently employed methodologies used in the malicious threat detection literature are shown in Figure 2 and their details are as follows:

1) BEHAVIOR-BASED DETECTION METHODS

Behavior-based detection methods involve the analysis of user actions and activities, employing algorithms and techniques such as user behavior analytics (UBA), time-series analysis, clustering, and sequence pattern mining to create models of normal user behavior [60], [61]. Deviations from these models can trigger alerts for further investigation, making them suitable for identifying novel insider threats and subtle anomalies that rule-based approaches might miss. While behavior-based methods provide context-rich insights into potential insider threats, they can produce false positives when legitimate user behaviors change and may demand substantial computational resources for continuous monitoring. Additionally, they may struggle to detect sophisticated insider threats that imitate normal user behavior. The authors of [62] and [63] show that the effectiveness of these methods hinges on the quality of the baseline behavior models.

2) RULE-BASED DETECTION METHODS

Rule-based detection methods [64] rely on predefined rules or policies to identify suspicious activities. These methods often use regular expressions to search for specific patterns in log files or network traffic and signature-based rules in intrusion detection systems (IDS) to detect known malicious activities. Network-based rule engines, such as Snort, employ

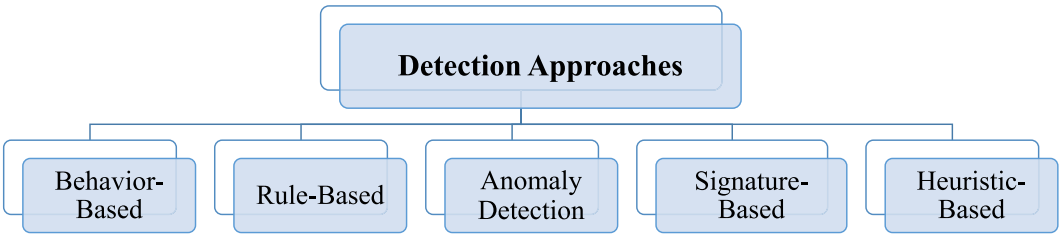

FIGURE 2. Insider threat detection approaches.

preconfigured rules to identify network traffic patterns associated with insider threats [65]. Rule-based detection is effective in identifying known attack patterns or policy violations, offering control over the detection process for customization to specific security requirements [64], [66]. However, these methods may struggle to adapt to emerging or novel insider threats that do not match predefined rules. They heavily depend on expert knowledge to define and update rules, potentially leading to delays in addressing rapidly evolving threats and the possibility of false negatives when new tactics are employed by insider threats.

3) ANOMALY DETECTION METHODS

Insider threats pose a significant challenge to the security of organizational systems, as malicious activities initiated by individuals with privileged access can lead to severe consequences. Addressing this concern requires robust and sophisticated anomaly detection methods that can discern abnormal behavior from the vast volumes of legitimate user activities. In recent years, the integration of machine learning (ML) and deep learning (DL) techniques has emerged as a promising avenue for enhancing the accuracy and efficiency of insider threat detection.

Machine learning techniques, such as clustering, classification, and various ensemble methods, have shown considerable success in identifying anomalous patterns in user behavior [40], [44], [67], [68], [69]. Unsupervised learning approaches, including clustering algorithms like K-Means and hierarchical clustering, are adept at grouping similar behavioral patterns and flagging outliers [61], [70], [71], [72]. Moreover, supervised learning methods, such as support vector machines (SVM) and random forests, leverage labeled datasets to train models capable of discriminating between normal and malicious activities [73], [74], [75]. Ensemble methods, such as boosting and bagging, amalgamate multiple base models to enhance overall predictive performance, demonstrating efficacy in detecting subtle deviations indicative of insider threats [48], [50], [51], [76],

Several studies [45], [49], [55], [57], [62] have demonstrated that deep learning, characterized by the use of neural networks with multiple layers, has remarkable capabilities in capturing intricate patterns within complex datasets. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, designed to handle sequential data, excel in capturing temporal dependencies in user behavior, making them well-suited for insider threat detection [63], [77]. Additionally, autoencoders, a form of unsupervised deep learning, have shown promise in reconstructing normal patterns and identifying anomalies [78]. The ability of deep learning models to automatically extract hierarchical representations from raw data contributes to their effectiveness in discerning subtle anomalies that may elude traditional methods.

Despite the progress made in leveraging ML and DL for insider threat detection, several challenges persist. Issues related to data scarcity, model interpretability, and adversarial attacks necessitate continued research efforts. Future directions should explore the integration of explainable AI techniques, the development of hybrid models combining the strengths of traditional methods and deep learning, and the enhancement of models' robustness against adversarial manipulations. Moreover, collaborative research efforts should focus on the creation of benchmark datasets that mirror real-world insider threat scenarios to facilitate fair comparisons and advancements in the field.

4) SIGNATURE-BASED DETECTION METHODS

Signature-based insider threat detection relies on predefined patterns or signatures to identify known malicious activities perpetrated by insiders. In this approach, specific behavioral patterns or indicators of compromise associated with insider threats are cataloged in a signature database [79]. These signatures encompass a range of anomalous actions, including unauthorized access to sensitive data, abnormal data exfiltration, or suspicious system interactions. The detection mechanism operates by comparing the observed behavior within an organization's network or systems to these predefined signatures. When a match is identified, the system triggers an alert or takes predefined actions, such as blocking access or notifying security personnel. The authors of [80] and [81]. suggests that while signature-based detection is effective in recognizing well-established and documented insider threats, its limitation lies in its inability to identify novel or previously unseen malicious activities, as it heavily relies on historical data and known attack patterns.

5) HEURISTIC-BASED DETECTION METHODS

Heuristic-based insider threat detection approaches in [82] and [83] present proactive ways to identifying potential

malicious activities by establishing a set of rules, guidelines, or heuristics that define normal behavior within an organization's network or information systems. Unlike signature-based detection, which relies on predefined patterns, heuristic-based methods focus on recognizing deviations from established norms. The heuristics are typically derived from a combination of industry best practices, security policies, and domain-specific knowledge. By defining what is considered typical behavior for users or system processes, these heuristics enable the detection of anomalies that may indicate insider threats.

One key advantage of heuristic-based detection, as suggested by the authors of [84], lies in its adaptability to evolving threat landscapes and its capability to identify previously unseen malicious activities. Rather than relying on historical data and predefined signatures, heuristic-based approaches can dynamically adjust to changes in user behavior or system interactions. For instance, heuristics may include rules that flag abnormal data access patterns, unauthorized privilege escalations, or unusual network traffic. This flexibility allows organizations to respond to emerging threats and tailor their detection mechanisms to specific operational contexts.

However, the effectiveness of heuristic-based insider threat detection relies heavily on the careful design and fine-tuning of the heuristics themselves as demonstrated in [10], [42], and [61]. Striking a balance between sensitivity to detect genuine threats and avoiding false positives is crucial. Additionally, continuous monitoring and refinement of heuristics based on feedback from real-world incidents contribute to the ongoing improvement of detection accuracy. Integrating heuristic-based methods with other detection approaches, such as machine learning or anomaly detection, can further enhance the overall resilience of insider threat detection systems, providing a comprehensive defense against both known and novel insider threats.

C. DATA SOURCES AND FEATURES

The field of malicious insider threat detection relies on diverse and purpose-driven data sources and datasets to unveil covert activities within organizations. Network logs, including firewall and intrusion detection system records, offer real-time insights into network traffic and suspicious communication patterns. User behavior data, encompassing login histories and file access logs, provide critical information for profiling normal and aberrant behavior. Textual sources like emails, chat messages, and incident reports offer contextual information and communication patterns. These multifaceted data sources empower researchers to develop robust detection methodologies to counter malicious insider threats.

The TWOS dataset, generated through a gamified competition [85], provides authentic insider threat instances. This publicly available dataset includes data on masquerader and traitor behaviors, defensive and offensive strategies, making it a valuable resource for malicious threat detection research.

The Sentiment140 dataset [86], while not designed for malicious insider threat detection, can be potentially useful by applying sentiment analysis techniques to identify negative sentiments on social media platforms. It offers additional context and early warning signals when used in conjunction with internal data sources.

The CERT dataset, created by the CERT Insider Threat Center at Carnegie Mellon University, is a comprehensive collection of logs and data related to insider threat incidents [87]. It falls into the category of labeled, real-world insider threat datasets and is highly valuable for research in this field. It enables the development and testing of sophisticated detection algorithms and models.

Text-based features within the CERT dataset provides critical insights into malicious insider activities Natural language processing techniques facilitate the analysis of communication patterns, sentiment shifts, and linguistic anomalies. Machine learning models leverage textual data to discern abnormal behavior, identify coordination among malicious insiders, and reveal emergent topics of interest. The integration of machine learning and artificial intelligence techniques enhances the precision of malicious insider threat detection when combined with complementary information sources. Table 2 shows a few activity types included in CERT dataset along with their attributes.

The RUU (Are You You?) Dataset, introduced by Salem and Stolfo [88], comprises host-based events from 34 regular users and includes masquerade sessions conducted by 14 individuals. This dataset serves as a valuable resource for researching masquerader detection and insider threat identification, offering real-world insights into deceptive activities within computer systems.

D. LEARNING-BASED INSIDER DETECTION TECHNIQUES

The field of malicious insider detection has witnessed a notable surge in interest in recent years, owing to the increasing recognition of the grave security risks posed by individuals within an organization who abuse their privileged access to sensitive information for malicious purposes. To combat this threat, various machine learning and data mining techniques have been employed, offering promising avenues for the detection of insider threats. This academic discourse aims to provide an overview of the prevalent techniques and algorithms utilized for this purpose, while also setting forth noticeable trends, and highlighting their respective advantages and disadvantages. Table 3 presents a comparative analysis of machine learning and deep learning approaches as observed in recent studies within the malicious insider detection domain. The table also emphasizes the performance metrics and datasets employed in these studies

A dominant trend in the domain of malicious insider detection involves the utilization of supervised learning techniques. These methods rely on labeled datasets comprising historical insider threat incidents, allowing algorithms to learn patterns indicative of malicious behavior. Commonly

TABLE 2. Various types of employee activities containing textual content.

Activity Type	User	Activity Attributes		Activity Content (Keywords)
Email	HMC1847	To	Noah.Jeremy.Vasquez@dtaa.com; Germane.Fredericka.Velez@dtaa.com	come publicized neglected swimmers trail regularly biking native earthfill km last colder 18th protection southwards under their higher concentrations system than nine flat mixed dog season 3 national still corps heavy base suffered swimmers have o number cut divides complex count coliforms upper occurred today could waves danger 149087 drastic back 1973 detected wet there 55 evidence year bells trail included poor
		From	Hannah.Mollie.Callahan@dtaa.com	
HTML	HRE1950	URL	`http://td.com/History_of_the_National_Hockey_League_19421967/mikita/2051596542.aspx`	distance english rescue earlier holy 50 affairs written policy 12 test 12 single onto claptons nearly stories premiere feel shock guarded hurt precious comment words reviews filler fear acceptable
File	TSG0262	Filename	`R:\79L99n6\H7RHJS5J.zip`	50-4B-03-04-14 moved imaging underwent key late appearance span ontario due compiled month 07 sedins final leaders ability doug another presidents improving donation by joseph quadruple 104 agreed 16 brian upon built all to handsome searching track wounded mike march one developer owned 5000 stepping lists orange metacritic second more supervisor currently initial

employed supervised learning algorithms include logistic regression, decision trees, random forests, and support vector machines. Their advantage lies in their ability to provide interpretable results and establish a direct connection between input features and threat prediction. Nevertheless, the efficacy of supervised techniques is contingent upon the availability of comprehensive and accurate training data, which can be challenging to obtain due to the low occurrence rate of insider threats.

Unsupervised learning techniques have also gained traction in the context of malicious insider detection. Anomalous behavior is detected by identifying deviations from established norms without relying on labeled data. Clustering and anomaly detection algorithms, such as k-means, DBSCAN, and isolation forests, have been employed for this purpose. The principal advantage of unsupervised learning is its ability to uncover novel and previously unknown threats. However, these techniques often produce higher false-positive rates, necessitating further investigation and verification of flagged anomalies.

In recent years, deep learning algorithms, particularly recurrent neural networks (RNNs), convolutional neural networks (CNNs) and Graph neural Networks (GNNs) have gained prominence for their ability to analyze complex, high-dimensional data such as user activity logs and network traffic. Deep learning models excel at capturing intricate patterns and dependencies in data, rendering them well-suited for insider threat detection in intricate and dynamic environments. Nonetheless, the inherent black-box nature of deep learning models poses challenges in terms of interpretability, raising concerns regarding their explainability which is an important issue in the domain of cyber security.

Another discernible trend involves the fusion of multiple data sources, including user behavior analytics, system logs, and network traffic data, to enhance the accuracy of insider threat detection. The amalgamation of diverse data types allows for a more comprehensive understanding of user activities, thereby reducing false positives and improving detection rates. However, this approach introduces new challenges related to data pre-processing, feature engineering, model selection and computational complexity.

Recent work [20] explores the application of one-class algorithms for data stream-based insider threat detection, using a hybrid analysis approach that combines offline batch training and online stream retraining. The approach shows promise in handling highly imbalanced datasets and adapting to concept changes in the data stream. However, it requires labeled normal samples, may lead to false positives, and needs further validation in real-world scenarios to assess its practical effectiveness.

In [21], the authors have proposed a machine learning-based system for detecting and classifying insider threats in cloud computing. The approach uses an ensemble approach with four machine learning algorithms, namely Random Forest, AdaBoost, XGBoost, and LightGBM, applied to a customized dataset from the CERT Insider Threat Tools dataset. The advantages of the approach used in this paper include higher accuracy in classification compared to previous studies. However, a potential disadvantage of the technique is the requirement for a large amount of data to train ML models effectively.

The authors of [22] propose a novel approach for detecting insider threats based on constructing a time-series user action graph (UAG) and using an ensemble autoencoder model for anomaly detection. The approach achieves outstanding performance in detecting anomalies and offers a more accurate and comprehensive understanding of user behavior. Nevertheless, the approach's limitations include a reliance on

TABLE 3. A Comparative analysis of machine and deep learning approaches with performance metrics and dataset overview.

Paper Ref.	Algorithm(s)	Performance Metrics					Dataset(s) Used		
		Acc.	Prec.	F1-Score	AUC	Other(s)	CERT r4.2	CERT r6.2	Other(s)
[5]	SVM, MLP & RF	×	1.0	0.81	×	Gmean = 0.76 Kappa = 0.01 Recall = 0.68	×	×	ITD
[6]	LightGBM	.97	.97	.88	×	Recall = 0.83	×	×	Customized CERT Dataset
[7]	Ensemble Autoencoder	×	×	×	0.85-0.9	DR = 0.87 FPR = 0.14	✓	×	×
[8]	Elliptic Envelope	×	×	0.4	×	×	×	✓	×
[9]	DBSCAN & K-Means	×	0.5	0.62	×	Recall = 0.83	×	×	Civil Aviation Flight University of China Web Cluster System logs
[10]	LDA & Spider Monkey Optimization (SMO)	0.98	0.99	0.98	×	Recall = 0.97	×	×	CERT r3.1, CERT r4.1
[11]	Deep Attention Neural network	0.93	0.94	0.97	0.99	Recall = 0.97	×	✓	×
[12]	SMOTE & LSTM	0.99	0.99	0.99	0.99	Recall = 0.99	✓	×	×
[13]	Dual-Domain GCN	0.98	0.96	0.98	×	Recall = 0.98	✓	✓	×
[14]	Random Forest	×	0.98	0.98	×	Recall = 0.98	×	✓	×
[15]	Log2Graph	×	×	×	0.99	×	×	✓	×
[16]	XGBoost	×	×	0.92	×	DR = 0.95	×	×	CERT r5.2
[17]	Convolutional Neural Network	.98	.98	.98	×	Recall = 0.98	×	×	TWOS
[18]	Random Forest	0.99	0.99	0.99	×	Recall = 0.99	×	✓	×
[19]	Multi-Edge Weight Relational GCN	0.99	0.98	0.99	0.99	Recall = 0.99	×	✓	×
[20]	Random Forest	×	0.98	0.98	0.99	Recall = 0.97	×	×	Simulated environment with synthetic data
[21]	Day-level SeqA-ITD	×	0.95	0.95	0.99	Recall = 0.95	✓	×	×
[22]	SMOTE & SVM	0.99	0.98	0.99	0.99	Recall = 0.99	×	×	Data generated from CWGAN-GP
[23]	ResNet18 Based Siamese Architecture	×	0.98	×	0.97	Recall = 0.96 FPR = 0.02	×	×	CERT r5.2
[24]	XGBoost	0.99	0.99	0.99	×	Recall = 0.99	×	×	CERT Insider Threat Tools dataset
[25]	XGBoost	0.98	0.98	0.98	×	Recall = 0.98	✓	×	
[26]	LSTM & CNN	0.99	0.99	0.99	0.99	Recall = 0.99	×	×	DARPA OpTC & LANL
[27]	CGANs, SVMs and RF	×	×	0.85-0.99	0.94-1.0		×	×	Synthetic data from CyberSim Tool
[29]	Tree Based User Behavior Modelling	0.98.	0.98	0.98	×	Recall = 0.98	×	×	User data in the IaaS cloud environment
[30]	One-class SVM & Autoencoder Based Behavior Sequence Network	×	×	×	0.95	Recall = 0.94 FPR = 0.12	×	✓	×
[31]	BERT & Bi-LSTM	×	×	0.94	×	×	✓	×	×
[32]	RF with Randomized Weighted Fuzzy Feature Set (RF-RWFF)	0.98	0.99	×	×	Recall = 0.97	×	✓	×
[33]	LSTM & CNN	0.99	0.94	0.95	0.99	Recall = 0.96	×	✓	×
[34]	VAE	0.96	0.92	0.94	×	Recall = 0.96	✓	×	×
[35]	Bi-LSTM & SVM	0.98	0.98	0.98	0.99	Recall = 0.98	✓	×	×
[36]	Label Propagation (LP), Label Spreading (LS) & Self Training (ST) Based SSL	×	×	0.92	0.98	×	×	×	✓
[37]	COPOD	×	0.98	0.98	0.98	Recall = 0.98	✓	×	×
[38]	RNN & LSTM	×	0.94	0.94	0.94	Recall = 0.94	×	✓	×
[40]	Ensemble of Local Outlier Factor, Isolation Forest & One-class SVM	×	×	×	0.98	TPR = 1.0 FPR = 0.001	✓	✓	×
[41]	LSTM	0.93	0.93	0.93	×	Recall = 0.93	×	×	CERT r5.2
[42]	Doc2vec	×	×	×	×	TPR = 0.95 FPR = 0.05	×	✓	×
[44]	Entity Portrait, Rule Matching & Iterative Attention	0.97	0.98	0.97	×	Recall = 0.97	×	×	CERT (version not mentioned)
[45]	Autoencoder	×	0.99	0.98	0.99	Recall = 0.98	×	✓	×
[46]	Random Forest	×	×	0.87	0.94	×	×	×	TWOS & Enron
[47]	Random Forest	×	×	0.98	0.99	×	×	✓	×

a substantial amount of training data and the possibility of false positives.

Recent work [23] explores the application of unsupervised machine learning for insider threat detection, using a hybrid solution that leverages both misuse and anomaly detection. The approach based on Isolation Forest REF successfully detects malicious insider activity during the training phase but shows negligible success during the testing phase. The pros of this technique include its simplicity and the ability to detect unknown threats. At the same time, the shortcomings revolve around the lack of information on known malicious insiders and their actions. Another study utilizing a semi-supervised anomaly detection model for insider threat detection in website cluster systems is proposed in [24]. The model combines agglomerative hierarchical clustering and heuristic-based log parsing techniques and utilizes user behavior, system, and time-based features for classification. The model achieves high accuracy in detecting unknown attacks but has the potential for model overfitting and difficulty in processing audit log data.

This research work in [25] proposes a methodology for insider threat detection using a combination of Linear Discriminant Analysis (LDA) REF and Spider Monkey Optimization (SMO) REF algorithms. The approach uses psychological sentiment analysis based on email and network browsing data to detect malicious insiders with negative emotions. The advantages of this approach include its ability to use advanced machine learning techniques and detect subtle and adaptive insider attacks. However, the approach also has limitations such as the lack of labeled data and the subjective nature of psychological sentiment analysis.

Recent work [26] achieves the development of a Continuous Audit System (CAS) for insider threat detection using a deep attention neural network REF. A deep attention neural network for insider threat detection operates by leveraging hierarchical attention mechanisms to focus on specific features or behaviors within user activities, enabling refined analysis for potential security risks. The model excels in capturing subtle anomalies and patterns, enhancing its ability to detect insider threats effectively. The advantages include improved accuracy, adaptability to evolving attack methods, and the ability to handle complex, multi-modal data sources. However, challenges lie in the interpretability of attention weights, potential biases in training data, and the computational demands of deep architectures. Additionally, fine-tuning and maintaining such models require significant resources, and their success is contingent on the continuous evolution of threat landscapes and the adaptation of detection strategies

Inspired by the previous achievements of recurrent models for the insider threat detection task, [27] achieves the detection of insider threats through the proposed S-LSTM model, which combines the SMOTE technique for handling unbalanced data and the LSTM algorithm for sequential activity analysis. A SMOTE-based Long Short-Term Memory (LSTM) model for insider threat detection operates by utilizing the Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance in sequential data. This approach involves generating synthetic instances of the minority class, enhancing the LSTM's ability to learn from imbalanced datasets containing rare instances of insider threats. The benefits include improved sensitivity to potential threats, enhanced generalization, and better overall model performance. However, challenges include the risk of overfitting due to synthetic data generation, the need for careful hyperparameter tuning for both SMOTE and LSTM components, and potential difficulties in interpreting the model's decisions. Another recent study [40] has developed an ML-based model that uses SMOTE for resampling the malicious users class and XGBoost algorithm with hyperparameter tuning for detecting malicious insider threats. The advantages of this technique include its robustness and ability to handle structured data effectively. However, a potential shortcoming lies in the requirement for a substantial amount of training data, which may limit its effectiveness in detecting unknown or zero-day attacks.

Recent work [28] proposes a novel approach, the Dual-Domain Graph Convolutional Network, for insider threat detection that combines user behavior and structural information to achieve improved accuracy and reduced false positives. Dual-domain GCN uses a hybrid technique of machine learning and graph convolutional networks to extract embeddings from both the original topology structure and the fused structure. It works by leveraging graph-based representations of relationships between entities in both user behavior and contextual information domains. The advantages of this approach include the ability to capture different scale information, while the disadvantage is the need for a large amount of data for training.

Reference [29] proposes a semi-supervised approach for insider threat detection using user and entity behavior analytics (UEBA) and a combination of behavioral and network traffic analysis. The proposed framework utilizes a combination of behavior-based and rule-based detection methods to classify user behavior as normal or anomalous. The strengths of this approach include its ability to detect zero-day and spear phishing attacks efficiently, while its weaknesses include the need for a large amount of data for training and the possibility of false positives.

A recent study on insider threat classification [30] introduces Log2Graph, an insider threat detection method based on graph convolutional neural networks (GCN) that combines supervised and unsupervised methods. It operates by transforming log data into a graph representation, where nodes represent entities (such as users or devices) and edges denote relationships and interactions. This graph-based approach allows for the extraction of meaningful patterns and anomalies, providing a comprehensive view of user behavior. The benefits include improved interpretability, enhanced detection of complex attack patterns, and the ability to capture temporal dependencies in user activities. However, challenges involve the need for effective graph

construction and potential scalability issues with large datasets. Additionally, the success of Log2Graph relies on the quality of feature engineering, the choice of relevant graph algorithms, and the adaptability to evolving threat scenarios. In another study employing graph neural network [34], the authors achieve robust anomaly-based insider threat detection using a Multi-Edge Weight Relational Graph Neural Network (MEWRGNN) approach. The technique combines GCN with anomaly detection to capture the contextual relationship of user behaviors over time, enabling accurate identification of insider threats. The MEWRGNN approach demonstrates the ability to learn from limited sample data sets and achieve quick and accurate insider threat detection. It provides a certain degree of interpretability through ranking the contribution of different edge-representation features. However, the approach requires a large amount of expert knowledge and prior knowledge to provide decision support for detection, which could be a potential limitation in practical applications.

The research study in [31] experimentally evaluates insider threat detection methods based on temporal representation using a combination of supervised learning anomaly detection methods. It employs Logical Regression, multi-layer perceptron, Random Forest, and Extreme Gradient Boosting, to model various threats and anomalies, aiming to obtain the baseline of normal user behaviors and detect abnormal behaviors. The approach is applied to the synthetic insider threat public dataset from CERT, which contains benign data from normal users and malignant data from malicious internal employees across four threat scenarios. The authors of [32] propose an approach for analyzing insider threat detection using email content analysis, employing a hybrid technique that combines sentiment analysis with machine learning algorithms such as CNN, RNN, Decision Tree, and SVM. The approach achieves high accuracy in real-time detection of insider threats by considering email content and user behavior.

A recent research study [33] improves collaborative intrusion detection networks against insider attacks using a supervised learning technique based on decision tree and random forest algorithms. The technique provides a challenge-based trust mechanism for defending against conventional insider attacks. Collaborative intrusion detection for insider threat detection employing network traffic analysis works by aggregating information from multiple sources within a network. This approach leverages the collective insights of various detection systems to identify anomalous user behaviors indicative of insider threats. The advantages of collaborative intrusion detection include increased accuracy through diverse data sources, improved scalability, and the ability to detect sophisticated attacks. However, potential shortcomings may involve challenges in standardizing data formats across different systems, addressing privacy concerns related to information sharing, and the risk of false positives if there is inconsistency in the interpretation of network traffic patterns.

Another approach to mitigate insider threats through the integration of access control and cyber deception methods is proposed in [35]. It employs a combination of behavior-based and rule-based detection methods along with genetic algorithm (GA) which is a population-based algorithm and produces evolving solutions for real-world problems. Therefore, the proposed approach can detect both known and unknown insider threats, but it has the potential for false positives and requires ongoing maintenance.

SeqA-ITD framework, which aims to improve Insider Threat Detection by capturing malicious user behavior's temporal and sequential patterns through the use of a multi-granular enhanced LSTM is proposed in [36]. This method operates by modeling sequential patterns in user activities to identify potential insider threats. Leveraging temporal dependencies, Seq-ITD analyzes the sequence of events to detect deviations from normal behavior, providing a dynamic approach to threat detection. The advantages include the ability to capture subtle anomalies over time, adaptability to evolving threat scenarios, and improved sensitivity to insider threats with nuanced behavioral changes. However, shortcomings may arise in handling noisy data, defining appropriate time windows for analysis, and the potential for false positives if normal behavioral variations are not adequately accounted for. Successful implementation of Seq-ITD requires careful consideration of threshold parameters, feature engineering, and continuous refinement to balance accurate detection with minimizing false alarms in real-world scenarios.

Recent work [37] proposes an approach for robust insider threat detection using adversarial training and synthetic data generation. The approach uses a hybrid of ML and DL techniques, including multi-class classifiers trained with adversarial training and synthetic data generation using a conditional Wasserstein generative adversarial network with gradient penalty (CWGAN-GP). The method works by utilizing a generative model to create realistic synthetic representations of normal user behavior and then distinguishing these from actual user activities using a discriminator network. By conditioning the GAN on relevant contextual information, it captures the fine details of insider threats within the generated samples. The advantages include the generation of diverse and realistic synthetic data, improved model generalization, and the potential to adapt to evolving threat behaviors. However, challenges involve the risk of overfitting, the need for careful tuning of hyperparameters, and potential biases in the generated data. Additionally, the Conditional Wasserstein GAN with Gradient Penalty may require substantial computational resources, and the interpretability of the generated features poses a challenge in understanding the reasoning behind insider threat predictions.

The authors of [38] utilizes a Siamese architecture and an improved contrastive loss function to address the challenge of detecting insider threats on imbalanced datasets. The proposed Siamese-architecture Insider Threat Detection (SITD) method uses ResNet18 and the contrastive loss

function, to judge whether input sample pairs belong to the same category, thereby avoiding model overfitting and high False Positive Rates. Utilizing a Siamese architecture with an improved contrastive loss function for insider threat detection involves training a neural network to learn a similarity metric between pairs of user activities, emphasizing differences for anomalies. The Siamese network's advantage lies in its ability to discern subtle variations in behavior, especially in the context of insider threats. The improved contrastive loss function further refines the model by encouraging the network to minimize the distance between similar instances while maximizing the separation between dissimilar ones. This approach offers enhanced discriminative power and adaptability to varying threat scenarios. However, potential shortcomings include the need for careful hyperparameter tuning, the challenge of handling imbalanced datasets, and the requirement for representative training samples to capture the diversity of insider threat behaviors.

To detect malicious insider threats in cloud environments, [39] proposes a supervised learning-based anomaly detection framework, employing One-Class Support Vector Machine (OC-SVM). The technique utilizes user log details, including web activity, device connectivity, and login information, to train the activity logs. Utilizing OC-SVM for insider threat detection involves training the model on normal user behavior and identifying deviations as potential threats. The OC-SVM maps the data into a high-dimensional space, aiming to encapsulate the normal behavior within a tight boundary. The advantages include simplicity, effectiveness in handling imbalanced datasets, and the ability to detect outliers without requiring labeled insider threat instances during training. However, shortcomings may arise in situations where normal behavior is diverse and hard to define precisely, leading to false positives.

Recent study [40] has developed an ML-based model that uses SMOTE for resampling the malicious users class and XGBoost algorithm with hyperparameter tuning for detecting malicious insider threatsThe advantages of this technique include its robustness and ability to handle structured data effectively. However, a potential shortcoming lies in the requirement for a substantial amount of training data, which may limit its effectiveness in detecting unknown or zero-day attacks.

The authors of [41] has developed DeepTaskAPT, a task-tree based deep learning method for detecting insider advanced persistent threats (APTs). It utilizes a combination of Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs), to analyze users' log templates and predict malicious activities. The technique leverages behavioral-based anomaly detection to effectively identify malicious traces in various attack scenarios. The advantages of this approach include its ability to handle unseen log models, adapt to new trends, and detect insider APTs with high accuracy. However, it relies on expert input to validate anomalies and may have limitations in handling false positives. The combination of LSTM and CNN techniques is also utilized in [48], which proposes a novel approach for detecting insider threats through behavioral analysis of users. The proposed approach claims to have low processing and memory requirements for rapid malicious insider detection.

Conditional Generative Adversarial Networks (CGANs) has been applied in [42] to address the scarcity of real-world data and skewed class distribution. The proposed technique combines behavioral profiling features with a hybrid approach that integrates machine learning and deep learning methods. By generating synthetic data using CGANs and employing a combination of algorithms such as Support Vector Machines (SVMs) and Random Forest (RF) classifiers, the paper demonstrates the effectiveness of multi-class anomaly detection in insider activity analysis. The advantages of this approach lie in its ability to mitigate data scarcity and class imbalances, leading to improved detection performance. However, a potential limitation is the reliance on a substantial amount of labeled data for training, which may pose challenges in real-world deployment scenarios.

The study in [43] introduces BTDetect, an insider threats detection approach based on behavior traceability for IaaS environments. The approach utilizes a tree-based modeling technique to construct a normal behavior tree that describes legal operations of cloud users and employs virtualization behavior keyword matching technology to identify malicious internal threats through tree-based integrity analysis. BTDetect achieves high recognition rates and the ability to identify unknown threats. However, it relies on a known set of normal behaviors, which may limit its effectiveness in detecting novel or evolving threats. Additionally, there is a potential for false positives due to the reliance on predefined behavior patterns. approach based on behavior traceability.

Another recent research paper [44] proposes Behavior Sequence Network (BS-Net), a semi-supervised approach for insider threat detection using a combination of behavior-based and anomaly detection methods. The approach involves dividing the original data flow into short sequences and extracting features from raw log data to create user behavior portraits. Utilizing One-Class Support Vector Machine (OCSVM) in conjunction with an autoencoder involves training the SVM on normal behavior and employing the autoencoder to reconstruct features, emphasizing deviations as potential threats. The OCSVM helps identify outliers, while the autoencoder captures complex patterns and anomalies in the data. The advantages include adaptability to imbalanced datasets and robustness against noise. However, potential shortcomings may include challenges in setting optimal hyperparameters, sensitivity to variations in data distribution, and difficulties in interpreting the combined model. In this context, the effectiveness of the Variational Autoencoder(VAE) is demonstrated in [49] in identifying internal threats with high accuracy. The advantages of this technique lie in its ability to detect complex and unknown insider threats without human intervention.

However, the approach may be prone to false positives and requires a substantial amount of data for training, which could be considered as its shortcomings.

Another approach to insider threat detection using deep learning techniques is proposed in [45], which make use of pre-trained language models and attention-based Bi-LSTM. The proposed framework for Insider Threat Detection is based on Bidirectional Encoder Representations from Transformers BERT. It leverages temporal-semantic representation to capture the sequential relationship among user behavior and achieves high precision and recall in day-level insider threat detection. The advantages of the proposed approach include its ability to capture the sequential relationship among user behavior and its outperformance of baselines. However, the approach relies on rare labeled data for behavior-level detection, which is a limitation.

The authors of [46] have identified the need of insider threat detection in the domain of lockchain technology, and proposes an architecture for detecting and mitigating malicious activities of compromised insider nodes in blockchain networks. The approach continuously monitors individual node behavior using a combination of server-side authentication and blockchain technology. By leveraging behavioral analysis and real-time verification of transactions, the architecture aims to identify and isolate insider attacks in real-time. The advantages of this approach include its potential to detect insider threats and its compatibility with heterogeneous IDS nodes. However, the technique may be susceptible to false positives and requires a trusted third party to manage the server-side authentication, which could be considered a limitation.

Randomized Weighted Fuzzy Feature Set (RF-RWFF) algorithm based on random forest is proposed in [47], for detecting insider cyber-attacks. The technique leverages machine learning and anomaly detection to effectively identify potential insider threats within organizations. This algorithm operates by employing randomized feature selection and fuzzy logic to enhance the interpretability and robustness of the model. The algorithm randomly selects features and assigns weights based on their significance, creating a diverse set of feature subsets. Fuzzy logic is then applied to evaluate the degree of membership of each feature in the subset. The advantages of RF-RWFF include increased interpretability, adaptability to various data types, and the ability to handle imbalanced datasets effectively.

The research work in [50] achieves an enhanced insider threat detection method by utilizing a bi-LSTM for feature extraction and SVM for classification in insider threat detection involves using bi-LSTM to capture temporal dependencies and extract relevant features from sequential data. The bidirectional nature of the LSTM allows it to consider past and future contexts, enhancing its ability to understand complex patterns in user behavior. The extracted features are then fed into an SVM for classification, leveraging the SVM's robustness in handling high-dimensional data and its capacity for effective binary classification. The advantages of this hybrid approach include improved representation learning, adaptability to sequential data, and the ability to handle non-linear relationships in the feature space. However, potential shortcomings may include increased computational complexity, the need for careful parameter tuning, and potential challenges in interpreting the combined model. A recent paper [53] proposes a novel framework, DeepMIT, for detecting malicious insider threats using a combination of RNN and LSTM. The approach is based on anomaly detection and models user behaviors as time sequences. The advantages of the approach include its ability to handle multi-source data. However, the approach may require a large amount of training data and may not be suitable for real-time detection. Recent research work in [55] proposes another approach based on Long Short-Term Memory (LSTM). The approach uses system logs as a natural language sequence and extracts patterns from these sequences to create workflows of sequences of actions that follow a natural language logic and control flow. However, the approach requires a good dataset for monitoring tasks and successful analysis, which can be a limitation.

The authors of [51] achieve insider threat detection by employing label propagation (LP), label spreading (LS), and self-training (ST) based semi-supervised learning, utilizing a combination of labeled and unlabeled data. The techniques used primarily involve semi-supervised machine learning algorithms, such as self-training and co-training, to leverage the limited labeled data available for detecting insider threats. The advantages of this approach include its potential to detect zero-day attacks and its ability to work with a small amount of labeled data. However, the technique's effectiveness may be limited by the quality of the labeled data and the need for skilled cybersecurity analysts to interpret the results and make informed decisions based on the detected anomalies.

The researchers in [52] have proposed the Copula Based Outlier Detection (COPOD) method in combination with the tree structure method for representing user behavior. The unsupervised learning technique used to analyze multi-dimensional feature data without the limitation of the number of dimensions. The advantages of this approach include its ability to process large-scale complex and heterogeneous data, detect abnormal users effectively, and outperform common unsupervised learning algorithms. However, a potential shortcoming is the unpredictability of insider personnel behavior, which may require a large amount of data for training and may not capture all possible insider threat scenarios.

The authors of [54] propose an unsupervised learning-based approach for detecting insider threats using four popular unsupervised machine learning methods, including Local Outlier Factor (LOF), Isolation Forest (IF), One-Class Support Vector Machine (OCSVM), and Principal Component Analysis (PCA). The research investigates various methods employing different data representations with temporal information, such as concatenation, percentile, and

mean or median difference. The objective is to interpret changes in user activities that could signify the detection of anomalous behaviors. The findings of the study indicate that an autoencoder using percentile representation proves to be the most effective combination for anomaly detection. Temporal data represented in percentile format exhibits substantial improvements over the original extracted data, enabling efficient insider threat detection even with minimal investigation budgets and demonstrating generalizability to new data.

Doc2vec-based behavioral analysis of multi-source security logs is proposed in [56]. The technique leverages unsupervised machine learning to identify anomalous behaviors by comparing them with the neighbours' behaviors, enabling spatio-temporal anomaly detection. The advantages of this approach include its ability to handle texts of any length and its flexibility in employing various corpora and metrics for posterior probability. However, the technique heavily depends on the quality and quantity of training data, and may not be robust against well-defined insider attack scenarios. Additionally, the performance of certain spatial metrics may require careful tuning or combination with other metrics in practical applications. The authors of [57] explore the utilization of process mining for insider threat detection from audit logs. It achieves this by employing a behavioral-based approach, utilizing process mining techniques such as visual analysis, conformance checking, and declarative conformance checking. The paper demonstrates the potential of process mining in detecting complex insider threats and providing specialized hints for analysts. However, the approach requires expert knowledge for effective implementation and may be susceptible to false positives, highlighting the need for further refinement and evaluation of the technique.

Recent research work in [58] proposes a hybrid intelligent system for insider threat detection that combines entity portrait, rule matching, and iterative attention techniques. The approach is supervised in nature and uses a combination of behavior-based, rule-based, and anomaly detection methods. In this study's experimental phase, a Convolutional Neural Network (CNN) model was trained on images generated by transforming the CERT dataset into a visual feature set to detect potential threats. By posing ITD as an image classification problem, the study aims to address the question of whether each user's activity could be classified as ''malicious'' or not within the Information System.

E. EVALUATION METRICS

The evaluation of malicious insider threat detection techniques within the cybersecurity domain encompasses the utilization of various performance metrics, each possessing distinct advantages and limitations that necessitate consideration in addressing the specific challenges associated with insider threats. Table 4 shows the performance of these evaluation metrics in the context of malicious insider detection.

Two fundamental metrics in the assessment of insider threat detection are True Positive (TP) and True Negative (TN) counts. TP represents the correct identification of genuine insider threats, while TN corresponds to accurately classifying benign activities as non-threats [89]. These metrics are appreciated for their simplicity and ease of interpretation. Nevertheless, they are criticized for providing an incomplete assessment, as they do not incorporate False Positives (FP) and False Negatives (FN), rendering them insensitive to class imbalances found in insider threat datasets, particularly when insider threats are rare.

False Positives (FP) and False Negatives (FN) metrics are vital for understanding the errors made by insider threat detection models [90]. FP identifies benign activities incorrectly flagged as threats, whereas FN identifies missed actual threats. However, they are deemed insufficient as standalone metrics due to their isolation from TP and TN, potentially leading to one-sided evaluations.

Precision, Recall, and the F1 Score offer a balanced assessment of detection techniques by addressing some of the limitations of TP, TN, FP, and FN. Precision assesses the accuracy of positive predictions, while Recall measures how many of the actual threats have been identified. The F1 Score combines these metrics to provide a holistic view of a detection technique's performance. However, they do not directly consider true negatives, potentially limiting their informativeness in non-threat-dominant datasets and may not be ideal in cases of significant class imbalance.

Specificity (True Negative Rate) and False Positive Rate (FPR) become prominent in settings focused on non-threat classification. Specificity quantifies the correct identification of true negatives among non-threats, while FPR measures the ratio of false positives among actual non-threats. These metrics are valuable for minimizing false alarms but may emphasize non-threat classification over threat detection performance.

The Area Under the ROC Curve (AUC-ROC) and Area Under the Precision-Recall Curve (AUC-PR) are comprehensive metrics that consider the entire range of detection thresholds. AUC-ROC evaluates the trade-off between true positive rate and false positive rate, offering an overall assessment of detection performance. AUC-PR focuses on the precision-recall trade-off, which is especially relevant for imbalanced datasets. These metrics provide flexibility in threshold selection but introduce subjectivity into the evaluation process.

IV. CHALLENGES AND OPEN PROBLEMS

Detecting malicious insider threats is a challenging and evolving task that requires a combination of technical, behavioral, and organizational measures. Malicious insiders are individuals within an organization who misuse their access and privileges to intentionally harm the organization's data, systems, or operations. Addressing this issue involves

TABLE 4. Commonly used evaluation metrics in malicious insider threat detection literature.

Performance Metric	Description	Advantages	Shortcomings
FPR	Rate of normal activities flagged as threats	Mitigates alert fatigue due to imbalance	May miss true insider threats if minimized too much
TPR	Rate of actual threats correctly identified	Captures genuine threats effectively	Can lead to increased false alarms
Precision	Ratio of true threats to total identified threats	Minimizes false alarms	Can result in a reduced TPR due to class imbalance
F1-score	Harmonic mean of precision and recall	Balances precision and recall	Does not reveal the precision-recall trade-off i.e. May not fully address class imbalance issues
ROC Curve	Trade-off plot between TPR and FPR	Assesses discriminative power	Does not directly account for class imbalance
AUC	Area under the ROC curve	Summarizes overall discriminative power	Sensitivity to Imbalanced Data and lack of selectivity

FPR: False Positive Rate, TPR: True Positive Rate, ROC Curve: Receiver Operating Characteristic Curve

dealing with various challenges, including technical, human, and organizational aspects.

A. USER BEHAVIOR ANALYSIS

Using User Behavior Analysis (UBA) for detecting malicious insider threats presents several technical challenges that require careful consideration. Firstly, managing the sheer volume and variety of data involved in UBA can be overwhelming, as it encompasses logs, network traffic, and user activity records. The integration and processing of diverse data sources pose significant hurdles, particularly in large organizations with complex IT infrastructures. Secondly, distinguishing between normal and suspicious behavior is crucial for effective threat detection but can be challenging due to the inherent variability in user activities. This noise reduction task necessitates accurate baseline establishment for each user, demanding thorough analysis and understanding of typical behavior patterns across different roles and departments. Furthermore, contextual analysis plays a pivotal role in UBA, as the legitimacy of user actions often depends on factors such as time, location, device, and user role. Incorporating contextual information into threat detection algorithms requires sophisticated modeling techniques to accurately assess the risk associated with user behaviors.

Another critical concern is the prevalence of false positives, where benign activities are mistakenly flagged as malicious, leading to alert fatigue and reduced system efficacy. Balancing detection accuracy with minimizing false positives is essential for the successful implementation of UBA systems. Additionally, ensuring compliance with privacy regulations such as GDPR and HIPAA while analyzing user behavior presents a significant technical and ethical challenge [21]. UBA systems must handle sensitive user data securely and anonymize it where necessary to uphold privacy requirements.

Real-time detection adds another layer of complexity, demanding efficient processing and response mechanisms to identify and mitigate threats promptly. Finally, addressing adversarial behavior, where malicious insiders intentionally attempt to evade detection, requires the development of robust algorithms capable of detecting and mitigating such threats effectively.

B. DATA ENCRYPTION AND EXFILTRATION

Detecting attempts to exfiltrate sensitive data through encrypted channels within the realm of malicious insider threat detection presents several intricate technical challenges that demand precise attention. Primarily, the encryption of communications, such as SSL/TLS, poses a formidable barrier to conventional inspection methods. Encryption obscures the payload of network traffic, rendering traditional deep packet inspection (DPI) techniques ineffective without access to decryption keys or compromising user privacy [24]. Consequently, this technical hurdle impedes the real-time monitoring and analysis of encrypted traffic for indicators of malicious insider activity.

Privacy concerns further complicate the decryption and inspection of encrypted traffic, as it entails potential violations of user privacy and regulatory compliance. Decrypting communications for inspection raises ethical and legal considerations, necessitating the development of privacy-preserving techniques that enable threat detection without compromising user confidentiality. Moreover, the detection of data exfiltration attempts through encrypted channels is compounded by the challenge of false positives. Analyzing encrypted traffic without considering the context may result in the misidentification of benign communications as malicious, leading to alert fatigue and diminishing the efficacy of threat detection mechanisms. Overcoming this challenge requires the implementation of advanced anomaly detection algorithms capable of discerning between normal encrypted traffic patterns and potential security threats based on contextual cues and behavioral analysis.

Furthermore, malicious insiders may exploit encrypted communication channels between internal endpoints to

exfiltrate data covertly, evading detection by traditional monitoring solutions. This underscores the importance of endpoint visibility and analysis capabilities to identify anomalous patterns indicative of insider threats, even within encrypted communications. In addition, the limitations of conventional Data Loss Prevention (DLP) solutions in inspecting encrypted traffic exacerbate the challenge of detecting insider threats. Traditional DLP techniques, reliant on pattern matching and content inspection, are ill-equipped to handle encrypted data effectively. Developing advanced DLP solutions that leverage techniques such as traffic flow analysis and metadata inspection to analyze encrypted traffic without compromising security or privacy is imperative to addressing this challenge effectively [42].

Lastly, malicious insiders may employ sophisticated obfuscation techniques, such as encryption within encryption or steganography, to conceal data exfiltration attempts within encrypted traffic [26]. Detecting such covert channels necessitates the deployment of advanced traffic analysis algorithms capable of identifying subtle anomalies indicative of malicious behavior, such as unusual traffic patterns or data transfer volumes.

C. INSIDER COLLABORATION

Detecting malicious insider threats becomes notably challenging when insiders collaborate with external threat actors or other insiders due to several technical hurdles. Traditional anomaly detection techniques may struggle to identify suspicious behavior, especially when insiders coordinate their actions to avoid detection. Experimental findings [27], [38] indicate that existing anomaly detection methods often fail to distinguish between coordinated insider threats and normal activities. Moreover, integrating and correlating data from multiple sources, such as network logs and user activity logs, introduces challenges in data fusion and correlation. Studies have shown that existing correlation methods face limitations in accurately identifying patterns indicative of collaborative insider threats, particularly in the presence of noise or false positives in the data.

Furthermore, malicious insiders frequently exploit weaknesses in identity and access management systems to collaborate with external threat actors or share credentials with other insiders, bypassing security controls [45]. Encrypted communication channels and evasion techniques further complicate detection efforts, making it challenging to identify malicious intent hidden within encrypted traffic. To mitigate these challenges, advanced techniques such as traffic analysis and behavioral monitoring are necessary as demonstrated in [57]. Additionally, contextual understanding of user behavior and system interactions is crucial for distinguishing between legitimate collaborations and malicious activities. Incorporating contextual information, such as user roles and historical behavior, can significantly enhance the accuracy of insider threat detection systems, particularly in identifying collaborative insider threats. Addressing these technical challenges requires interdisciplinary research efforts that combine expertise in cybersecurity, machine learning, data analytics, and human factors.

D. CREDENTIAL MISUSE

Detecting malicious insider threats poses significant technical challenges, particularly when insiders exploit their own or others' credentials to gain unauthorized access. One of the primary obstacles lies in distinguishing between legitimate and malicious activities within a sea of seemingly normal user behaviors. Traditional security measures often rely on rule-based systems or anomaly detection techniques, which may struggle to accurately identify malicious actions perpetrated by insiders who possess valid credentials. Various research studies [36], [56] highlight that malicious insiders often blend their activities with routine tasks, making it difficult for conventional systems to identify malicious intent solely based on behavioral deviations.

In conclusion, detecting malicious insider threats presents significant technical challenges, particularly when insiders exploit their own or others' credentials to gain unauthorized access. These challenges stem from the difficulty in distinguishing between legitimate and malicious activities, the obfuscation tactics employed by insiders, and the sheer volume and complexity of data to be analyzed. Addressing these challenges necessitates the development of advanced detection mechanisms capable of differentiating between benign and malicious behaviors while scaling to handle the intricacies of modern IT environments.

E. BEHAVIORAL CHANGES

Detecting malicious insider threats becomes particularly challenging when there is a sudden change in employee behavior indicating malicious intent. One of the primary obstacles stems from the necessity to differentiate between genuine shifts in behavior due to legitimate reasons and those indicative of malicious activities. Traditional anomaly detection techniques may struggle to accurately discern between benign deviations and malicious actions, especially when the sudden change is subtle or context-dependent. Research in [35] highlights the difficulty in establishing a baseline for normal behavior and determining the significance of deviations, particularly in dynamic environments where job roles and responsibilities evolve over time.

Moreover, the challenge is compounded by the need to swiftly detect and respond to such behavioral changes to mitigate potential damage. Time is of the essence in identifying insider threats, as delays in detection can lead to prolonged exposure and increased risk to organizational assets. Experimental findings in [37] underscore the importance of timely detection in mitigating the impact of insider threats, emphasizing the critical need for detection mechanisms capable of quickly identifying suspicious behavioral patterns amidst the noise of everyday activities.

Additionally, the issue of false positives poses a significant challenge in the context of sudden changes in employee behavior. Alert fatigue resulting from a high volume of

false alarms can desensitize security personnel, leading to genuine threats being overlooked or disregarded. A study by Kim, et al. [52] demonstrates the adverse effects of false positives on the efficacy of insider threat detection systems, highlighting the importance of refining detection algorithms to minimize false alarms while maintaining high detection rates.

Furthermore, the interpretation of behavioral changes requires context-aware analysis to distinguish between legitimate reasons for deviations and those indicative of malicious intent. Contextual information such as job role, access privileges, and recent events can significantly influence the assessment of behavioral anomalies. Researchers in [43] emphasize the importance of incorporating contextual cues into insider threat detection models to enhance their accuracy and reduce false positives. However, contextual analysis introduces additional complexity, requiring sophisticated algorithms capable of dynamically adjusting detection thresholds based on contextual factors

F. PRIVACY CONCERNS

Detecting malicious insider threats while preserving employee privacy necessitates sophisticated technical solutions that balance the intricacies of threat detection with stringent privacy requirements. One substantial technical challenge lies in effectively differentiating between normal user behavior and malicious actions without compromising individual privacy. Traditional anomaly detection methods, such as rule-based systems or statistical modeling, often yield high false positive rates due to their inability to accurately discern between benign and malicious activities. To address this, advanced machine learning techniques like deep learning algorithms coupled with anomaly detection frameworks, such as Isolation Forests or One-Class SVMs, can be employed [38], [39]. These models can learn complex patterns in user behavior and identify deviations indicative of malicious intent while minimizing false positives through continuous learning and adaptation.

Furthermore, the need to preserve privacy imposes constraints on the types of data that can be collected and analyzed. Techniques like data anonymization and pseudonymization can help mitigate these constraints by masking personally identifiable information (PII) while still allowing for effective threat detection [52]. Additionally, leveraging federated learning frameworks enables collaborative model training across distributed data sources while preserving data privacy at each source. Tools like PySyft and TensorFlow Privacy provide implementations of federated learning and privacy-preserving machine learning techniques, allowing organizations to build robust insider threat detection systems without compromising employee privacy.

The decentralized nature of insider threats poses another technical hurdle, as malicious activities may occur across disparate systems and endpoints. Endpoint detection and response (EDR) solutions offer a promising approach to address this challenge by continuously monitoring endpoint activities and identifying suspicious behavior indicative of insider threats. Tools such as Carbon Black and CrowdStrike Falcon utilize behavioral analysis and machine learning algorithms to detect anomalies and potential indicators of insider threats while minimizing the collection of personally identifiable information [35].

Moreover, the adoption of privacy-enhancing technologies such as differential privacy and homomorphic encryption can facilitate secure data analysis without compromising individual privacy [33]. Differential privacy techniques add noise to query responses to prevent the reconstruction of individual-level information, thereby safeguarding employee privacy while still allowing for meaningful analysis of aggregated data. Similarly, homomorphic encryption enables computations on encrypted data, allowing organizations to perform analytics on sensitive information without exposing plaintext data to unauthorized parties. Libraries like Microsoft SEAL and PySyft offer implementations of homomorphic encryption and differential privacy techniques, enabling organizations to conduct privacy-preserving data analysis for insider threat detection purposes.

G. INTEGRATION OF SECURITY SYSTEMS

Detecting malicious insider threats while ensuring seamless integration between various security systems presents several technical challenges that require careful consideration. One significant challenge is the heterogeneity of data sources and formats across different security systems. In an enterprise environment, security data often resides in disparate systems such as network logs, endpoint detection and response (EDR) solutions, user behavior analytics (UBA) platforms, and identity and access management (IAM) systems [91]. Integrating these diverse data sources poses challenges in terms of data normalization, correlation, and analysis.

Experimental findings have shown that inconsistencies in data formats and semantics hinder the effectiveness of threat detection algorithms. For instance, a study [51] found that integrating data from multiple security systems without proper normalization led to decreased accuracy in detecting insider threats. Misaligned timestamps, inconsistent field naming conventions, and varying levels of granularity across data sources impede the correlation of events and patterns indicative of malicious behavior. Consequently, security teams encounter difficulties in identifying relevant signals amidst the noise and false positives generated by disparate data sources.

Furthermore, another technical challenge lies in the real-time processing and analysis of large volumes of heterogeneous security data. Traditional SIEM (Security Information and Event Management) systems often struggle to keep pace with the velocity and volume of data generated by modern enterprise environments. Experimental studies conducted in [59] demonstrated that SIEM solutions without adequate scalability and computational resources experience

latency issues, leading to delays in threat detection and response.

To address these challenges, researchers have proposed the adoption of advanced analytics techniques such as machine learning and artificial intelligence (AI) for insider threat detection. However, integrating machine learning models across different security systems requires careful orchestration and model management to ensure seamless operation. Experimental evaluations conducted by the authors in [37] emphasized the importance of model interpretability and explainability in the context of integrated security systems. Black-box machine learning models may hinder the trust and adoption of integrated solutions, especially in regulated industries where explainability is crucial for compliance.

Moreover, privacy concerns and regulatory requirements add another layer of complexity to the integration of security systems for insider threat detection. Another study [53] highlighted the challenges of preserving data privacy while aggregating and correlating sensitive information across multiple systems. Compliance with regulations such as GDPR (General Data Protection Regulation) and HIPAA (Health Insurance Portability and Accountability Act) necessitates the implementation of privacy-preserving techniques such as differential privacy or federated learning, which can further complicate the integration process.

Deploying effective malicious insider threat detection systems in real-world security settings introduces a set of challenges that extend beyond the technical realm, encompassing issues related to scalability, adaptability, and compliance. One significant challenge involves the scalability and performance of detection systems, as organizations with large user bases and complex infrastructures require algorithms and models capable of efficiently handling vast datasets and scaling horizontally to accommodate growth [8], [20]. Real-time detection and response pose another challenge, requiring systems to reduce latency in analysis and decision-making. Achieving prompt responses to insider threats demands the optimization of algorithms for real-time processing and the implementation of automated response mechanisms that don't disrupt regular operations [36].

The adaptability of insider threat detection systems to organizational dynamics is crucial. Organizations often experience shifts in structures, roles, and responsibilities, impacting the efficacy of detection mechanisms. Addressing this challenge involves designing systems that can dynamically adapt to changes, automate role-based access adjustments, and minimize false positives during structural shifts [30], [48]. Data integration and correlation challenges arise due to the need to merge diverse data sources from different security systems for comprehensive threat analysis. Standardized protocols for data exchange, connectors for seamless integration with existing security tools, and correlation mechanisms are essential for providing a holistic view of potential threats.

Addressing these challenges requires a collaborative effort involving cybersecurity researchers, software developers, organizational leaders, and legal experts. The deployment of insider threat detection systems in real security environments demands a nuanced approach that considers not only technical aspects but also navigates the intricacies of organizational dynamics and regulatory landscapes [34], [39]. In conclusion, addressing malicious insider threats involves a multifaceted approach that combines technical solutions, employee training, and organizational policies. As the threat landscape evolves, ongoing research and development efforts are necessary to stay ahead of new challenges and emerging attack vectors.

V. FUTURE TRENDS

The ongoing evolution of information technology has brought about an accompanying rise in the sophistication of cyber threats, particularly those originating from within an organization—the malicious insider threat. As enterprises struggle with the growing challenges posed by insider threats, the field of malicious insider threat detection is poised to witness significant advancements in the coming years. This academic discussion delves into potential future trends, drawing upon current technological landscapes and emerging cybersecurity imperatives.

A. ENHANCED BEHAVIORAL ANALYTICS AND MACHINE LEARNING

Current efforts in malicious insider threat detection heavily rely on behavioral analytics and machine learning algorithms. These systems scrutinize user activities, identifying aberrations from established norms. Future trends indicate a progression towards enhanced behavioral analytics, leveraging more extensive datasets and refined machine learning models. For instance, advanced algorithms could incorporate contextual information such as time-of-day, location, and user roles to improve the accuracy of threat detection [3]. The shift towards predictive modeling and anomaly detection based on deep learning techniques is anticipated to fortify organizations against sophisticated insider threats.

B. USER ENTITY BEHAVIOR ANALYTICS (UEBA) ADVANCEMENTS

User Entity Behavior Analytics (UEBA) has emerged as a pivotal technology for profiling and analyzing user behavior to discern potential threats. The future trajectory of UEBA involves heightened sophistication, incorporating diverse contextual factors [3], [6], [60]. The integration of graph analytics within UEBA systems is projected to unveil intricate relationships between users and entities, thereby providing a nuanced understanding of potential threats. This expansion of UEBA capabilities aims to detect subtle patterns and anomalies that may signify an insider threat.

C. ENDPOINT DETECTION AND RESPONSE (EDR) TAILORED FOR INSIDERS

Endpoint Detection and Response (EDR) solutions, designed to identify and mitigate threats at the endpoint level, are

expected to undergo a transformation with a dedicated focus on malicious insider threats [92]. EDR systems will likely integrate features that closely monitor file access patterns, application usage, and user authentication behavior to discern anomalous activities indicative of insider threats [93].

In sum, the future trends in malicious insider threat detection are poised to witness a convergence of advanced technologies, including enhanced behavioral analytics, machine learning algorithms, UEBA sophistication, and tailored EDR solutions. These developments are crucial for organizations seeking to fortify their cybersecurity postures in an era where insider threats pose increasingly intricate challenges. The academic community and cybersecurity practitioners must collaboratively engage in research and development to stay ahead of the evolving threat landscape and safeguard organizational assets.

D. FUSING MULTIMODAL THREAT DETECTION

The rapid rise of malicious insider threats demands sophisticated detection methods that surpass the limitations of analyzing isolated data streams. Multimodal threat detection emerges as a powerful paradigm, fusing information from diverse sources like user behavior logs, network traffic, and content analysis to paint a comprehensive picture of user intent and uncover hidden malicious activities. This multifaceted approach offers several compelling advantages. Firstly, each data modality provides unique insights. Network traffic might unveil anomalous data transfers, while emails could expose suspicious communication, and documents might reveal unauthorized access attempts. By stitching these perspectives together, a holistic understanding of user behavior emerges, revealing the full context of potential threats as demonstrated in [30]. Secondly, multimodal fusion mitigates the issue of false positives prevalent in single-modal analysis. Anomalies in one data stream might be legitimate in another, acting as cross-validation and reducing the burden of false alarms while streamlining investigations.

Promising fusion techniques exist to harness this multimodal power. Early fusion directly combines raw data from various sources before feature extraction, leveraging sensor fusion algorithms or probabilistic models. Mid-level fusion, on the other hand, fuses features extracted from individual modalities, potentially employing feature selection and dimensionality reduction for efficient analysis. Finally, late fusion makes decisions based on independent analysis results from each modality, using methods like weighted voting or rule-based approaches [32].

However, challenges remain. Data integration and preprocessing pose hurdles due to the heterogeneity of data formats and varying levels of granularity across modalities. Balancing effective threat detection with user privacy necessitates careful consideration of data anonymization and access control mechanisms. Additionally, ensuring explainability and interpretability of how different modalities contribute to threat detection decisions is crucial for building trust and enabling human oversight [20], [50]. By addressing these challenges and continuously exploring advanced fusion techniques, multimodal threat detection holds immense potential and promises to revolutionize the detection and classification of malicious insider threats.

E. CONTINUOUS LEARNING AND ADAPTATION

Insider threats, characterized by their nuanced behavioral deviations and exploitation of novel attack vectors, necessitate dynamic and adaptable security solutions. Continuous learning and adaptation emerge as potent weapons in this battle, offering the ability to proactively identify, understand, and mitigate insider threats in real time. This paradigm leverages the power of machine learning (ML) algorithms. Supervised learning empowers systems to classify known insider attack patterns, while unsupervised learning uncovers anomalous behavior deviations. Adaptive anomaly detection, employing techniques like k-Nearest Neighbors and Isolation Forests, dynamically adjusts anomaly thresholds based on evolving user baselines and emerging threats [38]. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, excel at analyzing sequential data, allowing them to learn complex temporal patterns indicative of insider activity [20]. For privacy-sensitive scenarios, federated learning facilitates collaborative learning across decentralized datasets without compromising individual data privacy [4], [8].

However, there exist some challenges, for instance, data availability and quality pose hurdles due to privacy concerns and limited labeled data. Explainability and interpretability of ML models are crucial for building trust and enabling human oversight, requiring techniques like LIME and SHAP [31]. Additionally, computational resources for training and running complex ML models can be a barrier for smaller organizations.

Despite these challenges, the potential for continuous learning and adaptation is undeniable. By embracing this paradigm and addressing its hurdles, adaptive and dynamic detection methods can be developed that can learn, evolve, and respond effectively to the ever-changing threat landscape. This ultimately empowers security professionals to stay ahead of malicious actors, safeguarding their digital infrastructure from the evolving tactics of malicious insider threats.

VI. CONCLUSION

In conclusion, this survey paper has provided a comprehensive overview of the state-of-the-art in malicious insider threat detection using learning-based methods. The ever-evolving landscape of insider threats necessitates continuous innovation in detection techniques, and this survey has highlighted the pivotal role that learning-based methods play in addressing these challenges. Notably, the integration of Natural Language Processing (NLP), advances in time-series analysis, and deep learning has emerged as a promising frontier in this domain.

The incorporation of NLP techniques enables the analysis of unstructured textual data, offering deeper insights into employee communications and intentions. Meanwhile, advances in time-series analysis and deep learning provide the capacity to model and detect subtle temporal patterns and anomalies within user activities. These advancements collectively enhance the detection performance, enabling organizations to identify and respond to malicious insider threats more effectively. As the threat landscape continues to evolve, ongoing research and innovation in learning-based methods, particularly those leveraging NLP and advanced temporal analysis, will remain critical in safeguarding organizational security against the growing menace of insider threats.

In today's context, insider threat detection is of paramount importance due to the increasing digitalization of business operations and remote work arrangements. Malicious insiders, with authorized access, pose a significant risk to data security, financial stability, and reputation. Compliance regulations further mandate the need for robust detection methods to protect sensitive information and mitigate legal liabilities. Effective insider threat detection, including advanced techniques like deep learning, NLP, and temporal analysis, is essential to safeguard organizations in an evolving and interconnected landscape.

REFERENCES

[1] P. Langlois, A. Pinto, D. Hylender, and S. Widup, "DBIR 2023 data breach investigations report 10K 20K 30K about the cover," Tech. Rep., Jun. 2023.

[2] X. Kan, Y. Fan, J. Zheng, C.-H. Chi, W. Song, and A. Kudreyko, "Data adjusting strategy and optimized XGBoost algorithm for novel insider threat detection model," *J. Franklin Inst.*, vol. 360, no. 16, pp. 11414–11443, Nov. 2023.

[3] S. Asha, D. Shanmugapriya, and G. Padmavathi, "Malicious insider threat detection using variation of sampling methods for anomaly detection in cloud environment," *Comput. Electr. Eng.*, vol. 105, Jan. 2023, Art. no. 108519.

[4] P. Zheng, S. Yuan, and X. Wu, "Using Dirichlet marked Hawkes processes for insider threat detection," *Digit. Threats, Res. Pract.*, vol. 3, no. 1, pp. 1–19, Oct. 2021.

[5] M. AlSlaiman, M. I. Salman, M. M. Saleh, and B. Wang, "Enhancing false negative and positive rates for efficient insider threat detection," *Comput. Secur.*, vol. 126, Mar. 2023, Art. no. 103066.

[6] D. Li, L. Yang, H. Zhang, X. Wang, L. Ma, and J. Xiao, "Image-based insider threat detection via geometric transformation," *Secur. Commun. Netw.*, vol. 2021, pp. 1–18, Sep. 2021.

[7] I. H. Montano, J. J. García Aranda, J. R. Diaz, S. M. Cardín, I. D. L. T. Díez, and J. J. P. C. Rodrigues, "Survey of techniques on data leakage protection and methods to address the insider threat," *Cluster Comput.*, vol. 25, no. 6, pp. 4289–4302, Dec. 2022.

[8] B. Bin Sarhan and N. Altwaijry, "Insider threat detection using machine learning approach," *Appl. Sci.*, vol. 13, no. 1, p. 259, Dec. 2022.

[9] A. M. Rahmani, S. Bayramov, and B. Kiani Kalejahi, "Internet of Things applications: Opportunities and threats," *Wireless Pers. Commun.*, vol. 122, no. 1, pp. 451–476, Jan. 2022.

[10] Y. Wei, K.-P. Chow, and S.-M. Yiu, "Insider threat prediction based on unsupervised anomaly detection scheme for proactive forensic investigation," *Forensic Sci. Int., Digit. Invest.*, vol. 38, Oct. 2021, Art. no. 301126.

[11] H. Coyne, "The untold story of Edward Snowden's impact on the GDPR," *Cyber Defense Rev.*, vol. 4, no. 2, pp. 65–80, 2019.

[12] P. L. Bellia, "Wikileaks and the institutional framework for national security disclosures," *Yale Law J.*, vol. 121, no. 6, pp. 1448–1526, 2012.

[13] M. Fuller, "Big data and the Facebook scandal: Issues and responses," *Theology*, vol. 122, no. 1, pp. 14–21, Jan. 2019.

[14] P. Devine, Y. S. Koh, and K. Blincoe, "Evaluating software user feedback classifier performance on unseen apps, datasets, and metadata," *Empirical Softw. Eng.*, vol. 28, no. 2, p. 26, Mar. 2023.

[15] F. Gannon, "Learning from ENRON," *EMBO Rep.*, vol. 3, no. 9, p. 805, Sep. 2002.

[16] D. Shichor and J. W. Heeren, "Reflecting on corporate crime and control: The Wells Fargo banking saga," *J. White Collar Corporate Crime*, vol. 2, no. 2, pp. 97–108, Jun. 2021.

[17] X. Shu, K. Tian, A. Ciambrone, and D. Yao, "Breaking the target: An analysis of target data breach and lessons learned," Jan. 2017, *arXiv:1701.04940*.

[18] J. Thomas, "A case study analysis of the Equifax data breach 1 A case study analysis of the Equifax data breach," Tech. Rep., Dec. 2019.

[19] D. Wise, *The Spy Who Got Away: The Inside Story of Edward Lee Howard, the CIA Agent Who Betrayed His Country's Secrets and Escaped to Moscow*, 1st ed. New York, NY, USA: Random House, 1988.

[20] R. B. Peccatiello, J. J. C. Gondim, and L. P. F. Garcia, "Applying one-class algorithms for data stream-based insider threat detection," *IEEE Access*, vol. 11, pp. 70560–70573, 2023.

[21] M. Mehmood, R. Amin, M. M. A. Muslam, J. Xie, and H. Aldabbas, "Privilege escalation attack detection and mitigation in cloud using machine learning," *IEEE Access*, vol. 11, pp. 46561–46576, 2023.

[22] X. Wang, J. Jiang, Y. Wang, Q. Lv, and L. Wang, "UAG: User action graph based on system logs for insider threat detection," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jul. 2023, pp. 1027–1032.

[23] R. Yousef, M. Jazzar, A. Eleyan, and T. Bejaoui, "A machine learning framework & development for insider cyber-crime threats detection," in *Proc. Int. Conf. Smart Appl., Commun. Netw. (SmartNets)*, Jul. 2023, pp. 1–6.

[24] Y. Li and Y. Su, "The insider threat detection method of university website clusters based on machine learning," in *Proc. 6th Int. Conf. Artif. Intell. Big Data (ICAIBD)*, May 2023, pp. 560–565.

[25] A. Mittal and U. Garg, "Prediction and detection of insider threat detection using emails: A comparision," in *Proc. 2nd Int. Conf. Electr., Electron., Inf. Commun. Technol. (ICEEICT)*, Apr. 2023, pp. 01–06.

[26] S. K. Jah Rizvi, K. F. Javed, and M. Moazam, "CAS—Attention based ISO/IEC 15408–2 compliant continuous audit system for insider threat detection," in *Proc. 3rd Int. Conf. Artif. Intell. (ICAI)*, Feb. 2023, pp. 153–157.

[27] S. Besnaci, M. Hafidi, and M. Lamia, "Dealing with extremly unbalanced data and detecting insider threats with deep neural networks," in *Proc. Int. Conf. Adv. Electron., Control Commun. Syst. (ICAECCS)*, Mar. 2023, pp. 1–6.

[28] N. T. Moekthi Prajitno, H. Hadiyanto, and A. F. Rochim, "Research opportunity of insider threat detection based on machine learning methods," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, Feb. 2023, pp. 292–296.

[29] X. Li, X. Li, J. Jia, L. Li, J. Yuan, Y. Gao, and S. Yu, "A high accuracy and adaptive anomaly detection model with dual-domain graph convolutional network for insider threat detection," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 1638–1652, 2023.

[30] M. Z. A. Khan, M. M. Khan, and J. Arshad, "Anomaly detection and enterprise security using user and entity behavior analytics (UEBA)," in *Proc. 3rd Int. Conf. Innov. Comput. Sci. Softw. Eng. (ICONICS)*, 2022, pp. 1–9.

[31] K. Fei, J. Zhou, L. Su, W. Wang, Y. Chen, and F. Zhang, "A graph convolution neural network based method for insider threat detection," in *Proc. IEEE Int. Conf. Parallel Distrib. Process. Appl., Big Data Cloud Comput., Sustain. Comput. Commun., Social Comput. Netw. (ISPA/BDCloud/SocialCom/SustainCom)*, Dec. 2022, pp. 66–73.

[32] G. Lu, H. Zhang, T. Liu, K. Liao, and C. Feng, "Experimental evaluation of insider threat detection methods based on temporal representation," in *Proc. IEEE 10th Int. Conf. Inf., Commun. Netw. (ICICN)*, Aug. 2022, pp. 682–688.

[33] A. Mittal and U. Garg, "A proposed approach to analyze insider threat detection using emails," in *Proc. 3rd Int. Conf. Comput., Autom. Knowl. Manage. (ICCAKM)*, Nov. 2022, pp. 1–6.

[34] R. Singh and R. Deorari, "Enhancing collaborative intrusion detection networks against insider attack using supervised learning technique," in *Proc. IEEE 2nd Mysore Sub Sect. Int. Conf. (MysuruCon)*, Oct. 2022, pp. 1–6.

[35] J. Xiao, L. Yang, F. Zhong, X. Wang, H. Chen, and D. Li, "Robust anomaly-based insider threat detection using graph neural network," *IEEE Trans. Netw. Service Manag.*, vol. 20, no. 3, pp. 3717–3733, Sep. 2023.

[36] M. Alohaly, O. Balogun, and D. Takabi, "Integrating cyber deception into attribute-based access control (ABAC) for insider threat detection," *IEEE Access*, vol. 10, pp. 108965–108978, 2022.

[37] F. Zhang, X. Ma, and W. Huang, "SeqA-ITD: User behavior sequence augmentation for insider threat detection at multiple time granularities," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2022, pp. 1–7.

[38] R. G. Gayathri, A. Sajjanhar, and Y. Xiang, "Adversarial training for robust insider threat detection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2022, pp. 1–8.

[39] S. Zhou, L. Wang, J. Yang, and P. Zhan, "SITD: Insider threat detection using Siamese architecture on imbalanced data," in *Proc. IEEE 25th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, May 2022, pp. 245–250.

[40] G. Padmavathi, D. Shanmugapriya, and S. Asha, "A framework to detect the malicious insider threat in cloud environment using supervised learning methods," in *Proc. 9th Int. Conf. Comput. for Sustain. Global Develop. (INDIACom)*, Mar. 2022, pp. 354–358.

[41] S. K. Mamidanna, C. R. K. Reddy, and A. Gujju, "Detecting an insider threat and analysis of XGBoost using hyperparameter tuning," in *Proc. Int. Conf. Adv. Comput., Commun. Appl. Informat. (ACCAI)*, Jan. 2022, pp. 1–10.

[42] M. Mamun and K. Shi, "DeepTaskAPT: Insider APT detection using task-tree based deep learning," in *Proc. IEEE 20th Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, Oct. 2021, pp. 693–700.

[43] R. G. Gayathri, A. Sajjanhar, Y. Xiang, and X. Ma, "Anomaly detection for scenario-based insider activities using CGAN augmented data," in *Proc. IEEE 20th Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, Oct. 2021, pp. 718–725.

[44] L. Lin, S. Li, X. Lv, and B. Li, "BTDetect: An insider threats detection approach based on behavior traceability for IaaS environments," in *Proc. IEEE Int. Conf. Parallel Distrib. Process. Appl., Big Data Cloud Comput., Sustain. Comput. Commun., Social Comput. Netw. (ISPA/BDCloud/SocialCom/SustainCom)*, Sep. 2021, pp. 344–351.

[45] D. Zhu, H. Sun, N. Li, B. Mi, and T. Xi, "BS-net: A behavior sequence network for insider threat detection," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Sep. 2021, pp. 1–6.

[46] W. Huang, H. Zhu, C. Li, Q. Lv, Y. Wang, and H. Yang, "ITDBERT: Temporal-semantic representation for insider threat detection," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Sep. 2021, pp. 1–7.

[47] O. Ajayi and T. Saadawi, "Detecting insider attacks in blockchain networks," in *Proc. Int. Symp. Netw., Comput. Commun. (ISNCC)*, Oct. 2021, pp. 1–7.

[48] P. Varsha Suresh and M. Lalitha Madhavu, "Insider attack: Internal cyber attack detection using machine learning," in *Proc. 12th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, Jul. 2021, pp. 1–7.

[49] R. Nasir, M. Afzal, R. Latif, and W. Iqbal, "Behavioral based insider threat detection using deep learning," *IEEE Access*, vol. 9, pp. 143266–143274, 2021.

[50] E. Pantelidis, G. Bendiab, S. Shiaeles, and N. Kolokotronis, "Insider threat detection using deep autoencoder and variational autoencoder neural networks," in *Proc. IEEE Int. Conf. Cyber Secur. Resilience (CSR)*, Jul. 2021, pp. 129–134.

[51] M. Singh, B. Mehtre, and S. Sangeetha, "User behaviour based insider threat detection in critical infrastructures," in *Proc. 2nd Int. Conf. Secure Cyber Comput. Commun. (ICSCCC)*, May 2021, pp. 489–494.

[52] D. C. Le, N. Zincir-Heywood, and M. Heywood, "Training regime influences to semi-supervised learning for insider threat detection," in *Proc. IEEE Secur. Privacy Workshops (SPW)*, May 2021, pp. 13–18.

[53] X. Sun, Y. Wang, and Z. Shi, "Insider threat detection using an unsupervised learning method: COPOD," in *Proc. Int. Conf. Commun., Inf. Syst. Comput. Eng. (CISCE)*, May 2021, pp. 749–754.

[54] D. Sun, M. Liu, M. Li, Z. Shi, P. Liu, and X. Wang, "DeepMIT: A novel malicious insider threat detection framework based on recurrent neural network," in *Proc. IEEE 24th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, May 2021, pp. 335–341.

[55] D. C. Le and N. Zincir-Heywood, "Anomaly detection for insider threats using unsupervised ensembles," *IEEE Trans. Netw. Service Manag.*, vol. 18, no. 2, pp. 1152–1164, Jun. 2021.

[56] Q. Ma and N. Rastogi, "DANTE: Predicting insider threat using LSTM on system logs," in *Proc. IEEE 19th Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, Dec. 2020, pp. 1151–1156.

[57] L. Liu, C. Chen, J. Zhang, O. De Vel, and Y. Xiang, "Doc2vec-based insider threat detection through behaviour analysis of multi-source security logs," in *Proc. IEEE 19th Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, Dec. 2020, pp. 301–309.

[58] M. Macak, I. Vanát, M. Merjavý, T. Jevocin, and B. Buhnova, "Towards process mining utilization in insider threat detection from audit logs," in *Proc. 7th Int. Conf. Social Netw. Anal., Manag. Secur. (SNAMS)*, Dec. 2020, pp. 1–6.

[59] V. Koutsouvelis, S. Shiaeles, B. Ghita, and G. Bendiab, "Detection of insider threats using artificial intelligence and visualisation," in *Proc. 6th IEEE Conf. Netw. Softwarization (NetSoft)*, Jun. 2020, pp. 437–443.

[60] K. Randive, R. Mohan, and A. M. Sivakrishna, "An efficient pattern-based approach for insider threat classification using the image-based feature representation," *J. Inf. Secur. Appl.*, vol. 73, Mar. 2023, Art. no. 103434.

[61] X. Ye and M.-M. Han, "An improved feature extraction algorithm for insider threat using hidden Markov model on user behavior detection," *Inf. Comput. Secur.*, vol. 30, no. 1, pp. 19–36, Jan. 2022.

[62] D. C. Le and N. Zincir-Heywood, "Exploring adversarial properties of insider threat detection," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, Jun. 2020, pp. 1–9.

[63] R. A. Alsowail and T. Al-Shehari, "Empirical detection techniques of insider threat incidents," *IEEE Access*, vol. 8, pp. 78385–78402, 2020.

[64] C. Zhang, S. Wang, D. Zhan, T. Yu, T. Wang, and M. Yin, "Detecting insider threat from behavioral logs based on ensemble and self-supervised learning," *Secur. Commun. Netw.*, vol. 2021, pp. 1–11, Nov. 2021.

[65] T. Al-Shehari and R. A. Alsowail, "An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques," *Entropy*, vol. 23, no. 10, p. 1258, Sep. 2021.

[66] N. Elmrabit, S.-H. Yang, L. Yang, and H. Zhou, "Insider threat risk prediction based on Bayesian network," *Comput. Secur.*, vol. 96, Sep. 2020, Art. no. 101908.

[67] C. Soh, S. Yu, A. Narayanan, S. Duraisamy, and L. Chen, "Employee profiling via aspect-based sentiment and network for insider threats detection," *Exp. Syst. Appl.*, vol. 135, pp. 351–361, Nov. 2019.

[68] L. Liu, J. Yang, and W. Meng, "Detecting malicious nodes via gradient descent and support vector machine in Internet of Things," *Comput. Electr. Eng.*, vol. 77, pp. 339–353, Jul. 2019.

[69] C. Cheh, U. Thakore, A. Fawaz, B. Chen, W. G. Temple, and W. H. Sanders, "Data-driven model-based detection of malicious insiders via physical access logs," *ACM Trans. Model. Comput. Simul.*, vol. 29, no. 4, pp. 1–25, Nov. 2019.

[70] F. L. Greitzer, J. Purl, Y. M. Leong, and P. J. Sticha, "Positioning your organization to respond to insider threats," *IEEE Eng. Manag. Rev.*, vol. 47, no. 2, pp. 75–83, 2nd Quart., 2019.

[71] D. Noever, "Classifier suites for insider threat detection," 2019, *arXiv:1901.10948*.

[72] R. G. Gayathri, A. Sajjanhar, and Y. Xiang, "Image-based feature representation for insider threat classification," *Appl. Sci.*, vol. 10, no. 14, p. 4945, Jul. 2020.

[73] M. Singh, B. M. Mehtre, and S. Sangeetha, "User behavior profiling using ensemble approach for insider threat detection," in *Proc. IEEE 5th Int. Conf. Identity, Secur., Behav. Anal. (ISBA)*, Jan. 2019, pp. 1–8.

[74] J. Wang, L. Cai, A. Yu, and D. Meng, "Embedding learning with heterogeneous event sequence for insider threat detection," in *Proc. IEEE 31st Int. Conf. Tools Artif. Intell. (ICTAI)*, Nov. 2019, pp. 947–954.

[75] J. Jiang, J. Chen, T. Gu, K.-K. R. Choo, C. Liu, M. Yu, W. Huang, and P. Mohapatra, "Warder: Online insider threat detection system using multi-feature modeling and graph-based correlation," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Nov. 2019, pp. 1–6.

[76] S. Yuan, P. Zheng, X. Wu, and H. Tong, "Few-shot insider threat detection," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manag.*, Oct. 2020, pp. 2289–2292.

[77] D. C. Le, N. Zincir-Heywood, and M. I. Heywood, "Analyzing data granularity levels for insider threat detection using machine learning," *IEEE Trans. Netw. Service Manage.*, vol. 17, no. 1, pp. 30–44, Mar. 2020.

[78] A. Y. Khan, R. Latif, S. Latif, S. Tahir, G. Batool, and T. Saba, "Malicious insider attack detection in IoTs using data analytics," *IEEE Access*, vol. 8, pp. 11743–11753, 2020.

[79] H.-Y. Kwon, T. Kim, and M.-K. Lee, "Advanced intrusion detection combining signature-based and behavior-based detection methods," *Electronics*, vol. 11, no. 6, p. 867, Mar. 2022.
[80] R. A. Alsowail and T. Al-Shehari, "Techniques and countermeasures for preventing insider threats," *PeerJ Comput. Sci.*, vol. 8, p. e938, Apr. 2022.
[81] M. I. Khan, S. N. Foley, and B. O'Sullivan, "Database intrusion detection systems (DIDs): Insider threat detection via behaviour-based anomaly detection systems—A brief survey of concepts and approaches," in *Proc. Int. Symp. Emerg. Inf. Secur. Appl.* Springer, 2021, pp. 178–197.
[82] V. A. Yerdon, R. W. Wohleber, G. Matthews, and L. E. Reinerman-Jones, "A simulation-based approach to development of a new insider threat detection technique: Active indicators," in *Advances in Human Factors in Cybersecurity* (Advances in Intelligent Systems and Computing), Orlando, FL, USA. Springer, 2019, pp. 3–14.
[83] W. Szczepanik and M. Niemiec, "Heuristic intrusion detection based on traffic flow statistical analysis," *Energies*, vol. 15, no. 11, p. 3951, May 2022.
[84] P. Manoharan, J. Yin, H. Wang, Y. Zhang, and W. Ye, "Insider threat detection using supervised machine learning algorithms," *Telecommun. Syst.*, pp. 1–17, Dec. 2023. [Online]. Available: https://link.springer.com/article/10.1007/s11235-023-01085-3#article-info
[85] A. Harilal, F. Toffalini, J. Castellanos, J. Guarnizo, I. Homoliak, and M. Ochoa, "Twos: A dataset of malicious insider threat behavior based on a gamified competition," in *Proc. MIST*. New York, NY, USA: Association for Computing Machinery, 2017, pp. 45–56.
[86] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," Tech. Rep., 2009.
[87] R. Trzeciak, "The cert insider threat database," Carnegie Mellon Univ., Softw. Eng. Inst. Insights (BLOG), Tech. Rep., Aug. 2011.
[88] M. B. Salem and S. J. Stolfo, "Modeling user search behavior for masquerade detection," in *Recent Advances in Intrusion Detection*, R. Sommer, D. Balzarotti, and G. Maier, Eds. Berlin, Germany: Springer, 2011, pp. 181–200.
[89] F. L. Greitzer and T. A. Ferryman, "Methods and metrics for evaluating analytic insider threat tools," in *Proc. IEEE Secur. Privacy Workshops*, May 2013, pp. 90–97.
[90] E. N. Wanyonyi, S. Abeka, and N. Masinde, "A systematic review on machine learning insider threat detection models, datasets and evaluation metrics," *Int. J. Netw. Secur. Appl.*, vol. 15, no. 6, pp. 37–56, Nov. 2023.
[91] A. Kim, J. Oh, J. Ryu, and K. Lee, "A review of insider threat detection approaches with IoT perspective," *IEEE Access*, vol. 8, pp. 78847–78867, 2020.
[92] S. Chandel, S. Yu, T. Yitian, Z. Zhili, and H. Yusheng, "Endpoint protection: Measuring the effectiveness of remediation technologies and methodologies for insider threat," in *Proc. Int. Conf. Cyber-Enabled Distrib. Comput. Knowl. Discovery (CyberC)*, Oct. 2019, pp. 81–89.
[93] F. G. Siji and O. P. Uche, "An improved model for comparing different endpoint detection and response tools for mitigating insider threat," *Indian J. Eng.*, vol. 20, no. 53, pp. 1–13, Jun. 2023.

FATIMA RASHED ALZAABI received the B.S. degree in electrical and electronics engineering from Khalifa University, United Arab Emirates. She is currently pursuing the M.S. degree with Abu Dhabi University. Her current research interests include cyber security systems and natural language processing techniques.

ABID MEHMOOD (Member, IEEE) received the Ph.D. degree in computer science from Deakin University, Australia. He is currently an Assistant Professor with Abu Dhabi University. He has multiple research publications in well-reputed journals. His research interests include machine learning, privacy, information security, data mining, and cloud computing.