Cybersecurity Advances in SCADA Systems

Machine Learning-based Insider Threat Detection and Future Directions

Bakil Al-Muntaser[1], Mohamad Afendee Mohamed[2], Ammar Yaseen Tuama[3], Imran Ahmad Rana[4]
Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Kuala Terengganu, Malaysia[1, 2]
College of Computer Science and Information Technology, University of Kirkuk, Iraq[3]
Superior University, Lahore, Pakistan[4]

***Abstract*—The management of critical infrastructure heavily relies on Supervisory Control and Data Acquisition [SCADA] systems, but as they become more connected, insider attacks become a greater concern. Insider threat detection systems [IDS] powered by machine learning have emerged as a potential answer to this problem. In order to identify and neutralize insider threats, this review paper examines the most recent developments in machine learning algorithms for insider IDS in SCADA security systems. A thorough analysis of research articles published in 2019 and later, focussed on variety of machine learning methods, have been adopted in this review study to better highlight difficulties and challenges being faced by professionals, and how the study will contribute to overcome them. The results show that, in addition to conventional methods, machine-learning based intrusion detection techniques offer important advantages in identifying complex and covert insider attacks. Finding pertinent insider threat data for model training and guaranteeing data privacy and security are still difficult to address. Ensemble techniques and hybrid strategies show potential for improving detection resiliency. In conclusion, machine learning-based insider IDS has the potential to protect critical infrastructures by strengthening SCADA systems against insider attacks. The similarities and differences between cyber physical systems and SCADA systems, emphasizing security challenges and the potential for mutual improvement were also reviewed in this study. In order to be as effective as possible, future research should concentrate on addressing issues with data collecting and privacy, investigating the latest developments in technology, and creating hybrid models. SCADA systems can accomplish proactive and effective defence against insider attacks by integrating machine learning advancements, maintaining their dependability and security in the face of emerging threats.**

***Keywords*—Threat detection; SCADA security; machine learning-based intrusion detection; cyber-physical systems security; insider attack prevention**

I. Introduction

SCADA systems are widely used in areas such as telecommunications, water management, electricity [1], [2]. These systems employ specialized control units like Master Terminal Units (MTUs) and Remote Terminal Units [RTUs] to automate and manage industrial operations [3]. With the advent of Industry 4.0, the integration of Internet of Things [IoT] devices into Cyber-Physical Systems (CPS) has given rise to Industrial IoT. This technology enables real-time monitoring of machine status, providing operators with immediate feedback and facilitating faster operations in industries that heavily rely on electrical machinery, particularly induction motors [1], [4].

Modernized SCADA systems have evolved into highly advanced and intricate technological systems. The adoption of open standard protocols has significantly enhanced their productivity and profitability. SCADA architecture offers numerous benefits, including improved data access, cost-effectiveness, flexibility, configurability, accessibility, and scalability [5], [7]. However, these advancements have also introduced new threats and vulnerabilities [6]. Similar concerns have been raised by other researchers [2], who emphasized the increasing cyber-attacks on SCADA systems due to their rapid evolution, automation, real-time operation, and decentralized, multi-component design. A few researchers [6],[4],[8] highlighted the complexity and severity of the situation, attributing it to the utilization of the Internet for communications within SCADA systems.

Scope of the Review: The study concentrated on investigation and analysis of recent developments in Insider Intrusion Detection Systems (IDS) in SCADA systems. A variety of machine learning methods, deep learning models, advanced algorithms, and other new techniques utilized for insider IDS in SCADA contexts were covered (Tables I to III). The studies have also examined the difficulties encountered when putting these ideas into practice and assess how well they work at identifying insider threats [9]. In this study the developments over the past five years have been discussed.

Research Problem: As the review study focuses on addressing key challenges by exploring the advancements and issues in insider threat detection systems (IDS) for SCADA environments. The research problem lies here is enhancing the detection and prevention of insider threats within SCADA systems using machine learning-based approaches. By examining the latest developments in this field, we aim to provide a comprehensive understanding of how machine learning techniques can be harnessed to fortify SCADA systems against insider attacks.

The significance of this study is twofold: Firstly, it offers insights into the application and efficiency of Insider IDS in SCADA systems, highlighting their capabilities to identify complex and covert insider attacks. Secondly, it delves into the limitations and obstacles faced while implementing these systems, guiding practitioners and researchers towards more effective and secure solutions. By bridging the gap between the evolving threat landscape and SCADA security, this review contributes to the advancement of cyber-physical system protection and ensures the continuous operation of critical infrastructures [8].

A. Research Objectives: The Key Objectives of this Study are;

1) to give a thorough review of insider IDS's application and efficiency in SCADA systems.

2) to recognize the main issues and new developments in insider threat detection and prevention in SCADA environments.

3) to assess the effectiveness and constraints of the current insider IDS systems and suggest additional areas for development.

B. Research Questions: Based on above Objectives the Study Addressed following Questions;

1) What are the prevalent methods and strategies employed in insider IDS for SCADA systems, and how have they evolved to address the growing challenges posed by insider attacks?

2) What are the key issues and challenges in insider threat detection and prevention in SCADA environments, and what are the recent advancements in machine learning-based intrusion detection techniques to overcome these challenges effectively?

3) How effective are the current insider IDS systems in SCADA environments, and what are their strengths and weaknesses? Additionally, what are the potential areas for further development to enhance the capabilities of insider IDS systems in SCADA?

This paper is structured as follows: In the Review of Literature section, we delve into the comprehensive body of existing research to establish the context and identify the gaps that motivate our study. The Review Methodology section outlines our approach to analyzing recent advancements and challenges in Insider IDS for SCADA systems. We have then presented our findings in the Discussion and Analysis section, where we explored the strengths, limitations, and potential solutions of different techniques. Then, at challenges and recommendations section, we addressed the identified gaps and provide insights into enhancing the effectiveness of Insider IDS. The Comparison of IDS SCADA and Cyber Physical section highlighted the distinctive features of both realms and their interaction [10]. Finally, we concluded by summarizing our key findings, followed by a discussion of limitations and future research directions.

By following this structure, we aim to offer a holistic understanding of the advancements and challenges in Insider IDS for SCADA systems, providing valuable insights to both researchers and practitioners in the field (Table IV).

II. Review of Literature

A. Machine Learning Techniques for Insider IDS in SCADA Systems

SCADA systems are vital in maintaining and controlling different infrastructures, such as power grids, water treatment facilities, and transportation networks. These systems come with higher risk of insider attacks as they become more networked and accessible [10], [11]. The security and dependability of SCADA systems are seriously threatened by insider assaults carried out by anyone with authorized access to the system.

Traditional security solutions frequently fall short in addressing these internal risks. Machine Learning [ML] based Insider Intrusion Detection Systems [IDS] have become a potential method for improving insider attack detection and prevention in SCADA systems. The machine learning is commonly known as a branch of artificial intelligence [12]. ML has the capacity to examine massive amounts of data, spot trends, and spot unusual behaviour that might be indicative of hostile intent.

For insider IDS in SCADA systems, machine learning techniques comprise on developing models from past data in order to understand typical system behaviour and spot abnormalities that can point to insider assaults. These methods benefit from flexibility and the capacity to learn from fresh facts, enabling them to advance alongside new dangers [13].

The use of machine learning in insider IDS enables the detection of intricate and covert attacks that rule-based or signature-based methods can miss. Machine learning algorithms may recognize small anomalies and identify aberrant actions that depart from established standards by learning patterns and behaviours from data.

It is not without issues, yet, to implement machine learning methods for insider IDS in SCADA systems [1], [14]. As recognized insider threat data is often difficult to come by due to the frequency of such instances, data collection, pre-processing, and classification can be challenging jobs. Furthermore, it is essential to protect the confidentiality and security of critical SCADA system data while developing machine learning models [14].

The goal of the current study is to examine the most recent developments in insider IDS techniques for SCADA systems. It has explored the many methodologies, approaches, and algorithms applied in this field (Tables I to II). The review has also gone through the advantages, difficulties, and potential future research paths of using insider IDS to improve the security and resilience of SCADA systems [15].

Insider IDS in SCADA systems may strengthen the ability to identify and respond against insider attacks by utilizing machine learning techniques. The subsections below will provide in-depth discussion on particular machine learning techniques and methodology used in insider IDS, as well as the difficulties and opportunities presented by their application in SCADA systems [16].

TABLE I. IDS SOLUTIONS COMPARISONS

IDS Solution	Algorithm Description	Scalability	Limitations	Advantages	Conditions for Algorithm Use
Snort	Rule-based Systems	Snort is suitable for small to medium-sized networks, easily scaled up by adding more hardware or utilizing distributed deployments [10].	Limited in detecting novel or zero-day attacks, requires regular updates to stay up-to-date with emerging threats [10].	Easy customization and quick deployment, widely used and well-established in various network environments.	Rule-based systems are the primary approach in Snort. Machine learning and statistical analysis are used to a limited extent for rule generation and identifying abnormal behavior. [10],[14]
Suricata	Rule-based Systems	Suricata is highly scalable and can handle large network traffic volumes, making it suitable for enterprise-level deployments.	Deployment complexity may require more expertise, continuous tuning for minimizing false positives [15].	Accurate detection of known attack patterns, data fusion capability enhances detection of complex attack scenarios.	Rule-based systems are the primary approach in Suricata. Data fusion is used to correlate data from multiple sources and enhance detection capabilities [15]. [16].
Bro/Zeek	Rule-based Systems, Statistical Analysis	Bro/Zeek is highly scalable and can handle large network traffic volumes, making it suitable for enterprise-level deployments.	Requires more computational resources due to the comprehensive analysis it performs, potential performance impact in high traffic scenarios.	Comprehensive network monitoring capabilities, rule-based approach with added statistical analysis for anomaly detection.	Rule-based systems are the primary approach in Bro/Zeek. Statistical analysis is used to identify anomalies and detect patterns within network traffic [19].20].
McAfee	Rule-based Systems, Machine Learning	McAfee IDS solutions are designed for scalability and can be deployed in various network environments, including small to large enterprise networks.	Complexity in managing machine learning models, potential false positives/negatives depending on training data quality.	Combination of rule-based systems with machine learning techniques for enhanced threat detection, suitable for different network environments [23].	Rule-based systems are the primary approach in McAfee IDS solutions. Machine learning is used to detect unknown or evolving threats in some versions of the solution [20],[25].
Cisco Firepower IPS	Rule-based, Statistical Analysis, Graph-based, Machine / Deep Learning, Data Fusion, NLP, Hybrid Approaches	Cisco Firepower IPS is designed to scale for large enterprise networks and can handle high network traffic volumes.	Complex deployment and management, potential resource-intensive processing for advanced techniques, dependence on quality training data.	Comprehensive set of detection techniques, ability to handle high network traffic volumes, advanced approaches like machine learning and graph-based analysis [14], 16].	Multiple algorithms are used based on the specific requirements and features. The solution employs rule-based systems, statistical analysis, machine learning, deep learning, graph-based approaches, and data fusion [23].

In order to solve the particular difficulties, problems, and traits of insider IDS [Intrusion Detection Systems] in SCADA systems, machine learning approaches have been developed and customized [17]. These modifications take into account the unique demands and limitations of SCADA systems, such as the necessity for rapid and precise anomaly detection and real-time monitoring and large-scale data processing. We have gone through some of the most significant modifications made to machine learning methods for insider IDS in SCADA systems in this part.

- Feature Engineering: In SCADA systems, feature engineering is key to machine learning for insider IDS. Domain-specific features must be identified and designed due to the nature of SCADA data, which consists of time-series measurements, sensor readings, and control orders [18]. These characteristics record crucial facets of system activity and give machine learning models useful input. Statistical measurements, signal processing methods, and frequency domain analysis are a few examples of features frequently found in SCADA systems.
- Imbalanced Data: Compared to typical system behaviour, insider attacks are frequently infrequent occurrences. As a result, datasets become unbalanced, with a vastly greater number of regular instances than attacks [1]. Machine learning algorithms may have trouble effectively identifying the minority class of attacks. The performance of machine learning algorithms can be strengthened by rebalancing the dataset using a variety of techniques, such as oversampling, under sampling, or the use of ensemble methods like SMOTE (Synthetic Minority Over-Sampling Technique).
- Real-time Processing: SCADA systems work in real-time settings where it's crucial to promptly detect and respond to insider threats. For SCADA systems to handle the high-speed data streams they produce, machine learning algorithms must be modified. Models can update and react in real-time thanks to methods like

online learning or streaming algorithms, ensuring the prompt detection of insider threats [19].

- Model Interpretability: In the context of SCADA systems, machine learning model interpretability is crucial. The reasoning behind the judgments made by the models must be understood by system administrators and security staff. Rule extraction, feature importance analysis, or the application of explainable AI [XAI] methodologies are a few techniques that can give insights into model behaviour and improve user confidence in and comprehension of the applied machine learning models [9].

- Resource Constraints: SCADA systems frequently use constrained computational resources, necessitating the development of effective and portable machine learning models. Reduce the computing requirements of the models without sacrificing performance by using methods like model compression, model pruning, or the usage of simplified architectures [10].

- Transfer Learning and Domain Adaptation: Transfer Learning and Domain Adaptation methods are useful because classified insider threat data is hard to get in SCADA systems [10]. Retrained models from related domains can be used as a starting point or to bootstrap the training process, which may then be fine-tuned using the specific data from the SCADA system. By using this strategy, the machine learning models perform better and the problem of data scarcity is lessened.

The improvements and adjustments of machine learning approaches for insider threats in SCADA systems take into account the particular difficulties and traits of these systems. They make it possible to create intrusion detection systems that are effective and efficient, capable of handling unbalanced data, providing interpretability, and operating within the resource limitations of SCADA environments [12]. In order to strengthen the security of SCADA systems against insider assaults, current research in this area intends to further improve and refine these adaptations along with investigating fresh approaches.

Insider IDS [Intrusion Detection Systems] for SCADA [Supervisory Control and Data Acquisition] systems provide compelling potential for fresh methodologies and applications. These methods improve insider threat detection and prevention by utilizing the capabilities of machine learning algorithms. We'll talk about a few cutting-edge methods and uses of machine learning in this part as they relate to insider IDS for SCADA systems [5], [7].

- Behaviour Analysis: Using machine learning algorithms for behavioural analysis in SCADA systems is one cutting-edge strategy. Machine learning models can identify variations that can be signs of insider assaults by learning the typical patterns of user behaviour and system operations [10]. To respond to changing system behaviour, these models can be continuously updated and trained using past data.

- Anomaly Detection: Machine learning approaches are excellent at recognizing anomalies, which is essential for identifying insider threats in SCADA systems [9]. Models are able to spot alterations from the norm that could be indicative of malicious activity by learning from the typical system behaviour. Auto encoders, Gaussian mixture models, or one-class SVMs are examples of unsupervised learning techniques that can be used to quickly identify anomalies and flag questionable behaviour [10].

- Ensemble methods: To increase detection accuracy and robustness, ensemble approaches mix various machine learning models. Ensembles of models can be built using methods like bagging, boosting, or stacking that were trained on various subsets of the data or with various techniques. This ensemble-based strategy reduces false positives and improves insider IDS's overall performance in SCADA systems [12], [16].

B. Deep Learning Models

For insider IDS in SCADA systems, deep learning models like deep neural networks, have demonstrated potential in a number of fields. These models are capable of discovering complex features and patterns from unprocessed data, which makes it possible to identify sophisticated insider attacks [13]. For evaluating network traffic and time-series data, Convolutional Neural Networks [CNNs] and Recurrent Neural Networks [RNNs] are frequently used in deep learning-based IDS.

- Graph-Based Approaches: SCADA systems frequently have a complicated network structure, where components and dependencies can be represented as graphs. The relationships between system items can be captured and potential insider threats can be identified using graph-based machine learning techniques (Tables II and III). In SCADA systems, suspicious patterns, attack pathways, and alert priority can all be found using methods like graph neural networks and graph clustering techniques [17].

- Hybrid Approaches: To take use of their complimentary qualities, hybrid approaches incorporate various machine learning techniques such as rule-based systems, statistical analysis, and machine learning algorithms. These methods make it possible to combine several detection techniques, and they improve insider IDS in SCADA systems' precision and robustness [12].

- Natural Language Processing [NLP]: Textual information from logs, configuration files, and system messages can offer crucial information for spotting insider threats in SCADA systems according to Natural Language Processing [NLP]. The NLP techniques can be used to process and analyse this textual data, extract pertinent data, and spot potential attack indications.

A thorough insider IDS architecture can benefit from the use of machine learning models that can be trained to categorize and analyze text input [18].

These revolutionary methods and insider IDS for SCADA systems uses of machine learning techniques offer enormous potential to improve security and safeguard crucial infrastructures. Continuous research and development in these field aims to improve the effectiveness of these methods, handle the changing problems brought on by insider threats in SCADA systems, and further these approaches [6].

In addition to the machine learning techniques, deep learning models, and advanced algorithms mentioned earlier, there are other tools and approaches that can be explored in the context of insider IDS in SCADA systems (Table IV). Here are a few examples:

- Statistical Approaches: Data from SCADA systems can be examined for patterns, trends, and abnormalities using statistical approaches. Techniques like time series analysis, statistical process control, and multivariate analysis can shed light on unusual behaviour or departures from the way a system is supposed to work [7].
- Rule-based Systems: To identify potential insider threats, rule-based systems use a set of established rules or criteria. These guidelines may be drawn from professional judgment or widely accepted commercial norms. To improve detection accuracy, rule-based systems are frequently employed in conjunction with other methods [16].
- Data Fusion: To increase the precision of insider threat detection, data fusion merges data from many sources, including sensor data, network logs, and user behaviour. Complex patterns and correlations that might not be obvious when evaluating individual data sources might be found by integrating various data streams [20].
- Evolutionary Algorithms: To maximize a solution, evolutionary algorithms imitate the processes of natural selection and genetic evolution. They can be used to improve insider IDS parameters, feature choices, or model architectures in SCADA systems. Examples of evolutionary algorithms include differential evolution, genetic algorithms, and particle swarm optimization [4].
- Reinforcement learning includes teaching an agent how to make decisions sequentially in a setting to maximize a reward signal. This method can be used to develop IDS systems that are self-adaptive and self-learn in order to dynamically respond to insider threats in SCADA systems [19].
- Graphical probabilistic models called Bayesian networks are used to describe ambiguous relationships between variables. They can be used to simulate causal chains and interdependencies within a SCADA system, making it easier to spot irregularities and possible insider threats.
- Fuzzy Logic is a mathematical framework for handling deliberation and decision-making in the face of uncertainty. In insider IDS for SCADA systems, it can be used to describe imprecise or uncertain knowledge, enabling more adaptable and reliable detection procedures [13], [19].

This broadens the selection of tools available for identifying insider threats in SCADA systems by offering alternatives to conventional machine learning and deep learning techniques. When choosing and implementing these approaches, it's crucial to thoroughly evaluate their applicability for the specific SCADA environment and take into account their advantages, disadvantages and performance traits.

C. *Cyber Physical Systems*

Cyber-Physical Systems (CPS) are networked systems that combine computing and communication skills with physical aspects [10]. CPS denotes to the integration of SCADA systems with the latest information and communication technologies in the context of this review study. CPS is essential for controlling critical infrastructure, but it also creates new security risks, especially with regard to insider attacks [21],[22]. To increase the security of CPS, the study intends to investigate the most recent advancements and breakthroughs in Insider Intrusion Detection Systems (IDS) based on machine learning. The study seeks to pinpoint major issues and offer suggestions for proactive security against insider assaults in CPS contexts by examining the use and efficacy of various machine learning approaches (Table III).

III. IDS SOLUTIONS

Machine learning is a subfield of artificial intelligence (AI) which entails training computers to learn from data and make judgments or predictions without being explicitly programmed. ML approaches can be applied to Intrusion Detection Systems (IDS) to improve the solutions' capacity for identifying potential hacking attempts or anomalies [5].

The major components of the IDS solutions like Snort, Suricata, Bro/Zeek, McAfee IPS, and Cisco Firepower IPS—use established rules or patterns to identify known threats. These IDS systems can, however, be enhanced using ML algorithms to make them more intelligent and flexible [15],[23].

An outline of how ML relates to IDS solutions is provided below: Data is necessary for ML algorithms to learn from and generate predictions with. In the case of IDS, ML models can be trained using historical network traffic data [24].

Learning Patterns: Machine learning algorithms examine the training data and discover patterns or traits that distinguish between legitimate and harmful activity. For instance, they can spot specific network traffic patterns linked to particular kinds of attacks [13].

Making Predictions: Following training, ML models can be used to generate predictions about incoming network traffic that hasn't yet been seen. The models look for any signs of an ongoing assault or unusual activity by comparing the observed traffic patterns with what they have learnt from the training data.

Adaptability and Anomaly Detection: ML algorithms are also capable of identifying anomalies, which are atypical patterns or behaviours that don't correspond to well-known assault patterns. Due to their versatility, ML-based IDS solutions can recognize new or undiscovered assaults [12], [13].

Continuous Improvement: As they are exposed to more data over time, ML models can continuously train and get better. They can refresh their knowledge to recognize new risks and respond to changing attack methodologies [12].

In summary, ML algorithms are employed in IDS solutions to learn from historical network traffic data, discover patterns linked to legitimate and nefarious conduct, and generate forecasts about impending attacks or anomalies. This makes IDS solutions more proficient in identifying and stopping network intrusions.

It's important to note that the conditions for algorithm use as displayed in Table II are general guidelines which may vary based on the specific configurations, versions, and deployment environments of each IDS solution. The performance metrics for each technique, such as detection accuracy, false positive rate, and response time, are compared thoroughly in the Table II. These were selected based on studies [26], 27].

TABLE II. PERFORMANCE METRICS REVIEW FOR INSIDER IDS SCADA SYSTEMS TECHNIQUES

Technique	Detection Accuracy	False Positive Rate	Response Time	Applicability / Usage Condition	Limitation	Advantages	Other Information
Behavioural Analysis	High	Low	Fast	Effective for identifying insider threats	May not capture all insider attack patterns	Captures unusual patterns of behavior that may indicate insider attacks [1],[3]	Measures the accuracy of detecting insider threats based on behavior
Anomaly Detection	High	Low	Fast	Effective for identifying unknown threats	May result in false positives for certain data distributions	Can detect previously unseen insider attack patterns	Measures the accuracy of detecting anomalies in network traffic [3]
Ensemble Methods	High	Low	Fast	Effective for improving overall accuracy	Complexity may lead to higher resource requirements [9]	Combines multiple models to reduce false positives and increase accuracy [14]	Measures the overall detection accuracy of the ensemble
Deep Learning Models	High	Low	Fast	Effective for complex pattern recognition	Requires large amounts of labelled data	Can identify intricate patterns and detect novel insider threats	Measures the accuracy of detecting threats based on deep learning models
Graph-Based Approaches	High	Low	Fast	Effective for capturing network relationships	May face challenges in large-scale networks	Can detect insider threats by analyzing complex relationships in SCADA	Measures the accuracy of detecting threats based on graph analysis [15]
Hybrid Approaches	High	Low	Fast	Effective for improving overall accuracy	Requires careful integration of different techniques	Combines multiple methods to enhance detection capabilities	Measures the overall detection accuracy [14],[15]
Natural Language Processing [NLP]	High	Low	Fast	Effective for analysing textual data	Requires pre-processing of unstructured data	Can identify indicators of potential insider attacks from text data [21]	Measures the accuracy of detecting insider threats based on NLP

IV. REVIEW METHODOLOGY

A structured approach has been used in the review study to collect, evaluate, and synthesize relevant research on insider IDS in SCADA systems. The methodology is comprised on following steps:

Literature Search: To find studies, conferences, and journal publications from the previous five years [2019 to the present] that are relevant to the objectives of the research, a thorough search was carried out in research databases including IEEE Xplore, WOS, ScienceDirect, and Google Scholar.

The selection of relevant research is based on previously established inclusion and exclusion criteria. This study has taken into account works that concentrate on machine learning methods, deep learning models, advanced algorithms, and new insider IDS strategies in SCADA systems. Papers that don't fit the criteria for scope or quality were not being considered.

Data Extraction: Important details from the chosen articles, including the title, authors, publication year, methodology, algorithms employed, performance measures, and conclusions were taken out. In order to facilitate comparison and analysis, such information was displayed in a tabular manner.

The search was conducted in June 2023, and the years 2019 through 2023 were taken into consideration. 183 papers were obtained as a result, including: 59 papers from Science Direct, 77 papers from WOS, Google scholar, and 47 papers from IEEE Xplore. The final collection contained 94 papers after manual inspection and the elimination of the replicated publications.

Additionally, we chose potential articles based on the titles and abstracts, paying particular attention to those that offered novel suggestions for NIDS-specific to SCADA [5],[14]. There were 51 papers in the remaining collection.

After reviewing the complete candidate papers, the goal was to collect a set of original and similar solutions. As a result, we disregarded publications that were similar in content but had different authors or described the outcomes of the same initiatives, as well as papers that lacked IDS evaluation findings. 27 papers were selected in the end.

Data Analysis: To find patterns, trends, and insights regarding the most recent developments in insider IDS for SCADA systems, the retrieved data was evaluated and synthesized. A comparative analysis was done to assess how well various strategies and algorithms perform [Reference Table IV].

Development of a Conceptual Framework: Based on the research and analysis, conceptual framework was constructed to classify and comprehend the numerous machine learning methods, deep learning models, and advanced algorithms utilized in insider IDS for SCADA systems (Table III).

TABLE III. CPS AND SCADA SECURITY LANDSCAPE AND LEARNINGS REVIEW

Comparison Aspect	Cyber-Physical Systems [CPS]	SCADA Systems	Security Landscape Comparison	Advantages for Learning
Integration of Technologies	CPS integrates cyber and physical components to create interconnected systems [6].	SCADA focuses on integrating sensors, actuators, and control systems in industrial environments.	Both CPS and SCADA require secure integration of diverse technologies to prevent cyber-physical attacks and ensure data integrity.	CPS can learn from SCADA's focus on industrial protocols and network segmentation to enhance security. SCADA can learn from CPS's advanced encryption and authentication techniques for securing data flow in integrated systems [17].
Data Collection and Analysis	CPS relies on data from sensors and other sources for real-time monitoring and control.	SCADA systems gather data from sensors and devices to monitor industrial processes.	Both CPS and SCADA must secure data collection and analysis to prevent unauthorized access or tampering.[20]	CPS can learn from SCADA's data filtering techniques for efficiently handling large data streams. SCADA can learn from CPS's data analytics capabilities to improve predictive maintenance and anomaly detection [21].
Real-time Monitoring and Control	CPS provides real-time feedback and control in various domains.	SCADA allows operators to monitor and control industrial processes in real-time.	Both CPS and SCADA face real-time security challenges, requiring robust authentication and authorization mechanisms.	CPS can learn from SCADA's focus on redundancy and fail-safe mechanisms for continuous real-time control. SCADA can learn from CPS's distributed control architecture for improved system resiliency [22].
Connectivity and Communication	CPS uses communication networks to facilitate data exchange between cyber and physical components.	SCADA relies on communication networks for data transmission between central control and field devices.	Both CPS and SCADA must ensure secure communication channels to prevent unauthorized access and data interception.	CPS can learn from SCADA's strict access control policies and encryption techniques for secure communication. SCADA can learn from CPS's adaptive communication protocols for handling dynamic network conditions [22].
Security Challenges	CPS and SCADA face security challenges due to their interconnected nature.	Ensuring data and communication security is crucial to prevent cyber-attacks and disruptions.	Both CPS and SCADA must address security challenges related to insider threats, remote access, and supply chain vulnerabilities [21].	CPS can learn from SCADA's robust anomaly detection mechanisms for detecting suspicious activities. SCADA can learn from CPS's threat intelligence integration for proactive identification of potential cyber threats.
Industrial Applications	CPS finds applications in manufacturing, energy, healthcare, transportation, etc.	SCADA is commonly used in industrial sectors for process control and automation.	Both CPS and SCADA play critical roles in enhancing efficiency, automation, and optimization of processes in their respective domains [22].	CPS can learn from SCADA's domain-specific protocols and standards for seamless integration into industrial applications. SCADA can learn from CPS's adaptability to diverse industrial settings for improved flexibility and scalability.

TABLE IV. IDS ALGORITHMS REVIEW

Algorithm	Usage Conditions	Core Strengths	Core Weaknesses	Opportunities	Core Threats	Supporting Technologies & Industries
Ensemble Methods	Diverse dataset, multiple models	Improved prediction accuracy, model robustness	Increased complexity and computational resources [4]	Ensemble learning techniques, model combination	Overfitting, model selection, ensemble diversity	Various machine learning frameworks, successful in various industries such as finance, healthcare, and retail
Evolutionary Algorithms	Complex optimization problems	Effective for global optimization, handle constraints	Computationally expensive, slow convergence	Optimization of insider IDS parameters and features [10], [11]	Premature convergence, parameter tuning	Genetic algorithms, particle swarm optimization [7]
Hybrid Models	Diverse techniques, flexible	Combines strengths of different approaches	Complexity in model integration, interpretability trade-off [10],[13]	Improved detection accuracy, adaptable systems	Increased complexity, model integration challenges [20]	Integration of machine learning and rule-based systems
Anomaly Detection	Unusual patterns, outlier detection	Identifies unknown and rare insider threats	Difficulty in defining normal behaviour, high false positives	Uncovering novel insider threats, pattern recognition	Sensitivity to data quality, evolving threats	Statistical analysis, unsupervised learning, successful in various industries such as cybersecurity and fraud detection [12], [13]
Graph-based Methods	Network or relationship data	Captures complex relationships, detects structural anomalies	High computational cost for large graphs, graph construction	Identifying suspicious connections, network analysis	Scalability, graph sparsity, noise in data	Network analysis tools, successful in social networks, cybersecurity, and transportation systems [7],[20]
Reinforcement Learning	Sequential decision-making tasks	Adapts to dynamic environments, learns from interactions	High sample complexity, sensitivity to reward design	Adaptive and self-learning IDS systems	Exploration-exploitation trade-off, reward design	Q-learning, deep reinforcement learning frameworks
Bayesian Networks	Uncertain relationships, probabilistic reasoning	Captures causal dependencies, handles uncertainty	Requires prior knowledge, limited scalability	Modelling complex relationships, uncertainty handling	Learning structure from data, complexity in learning	Probabilistic programming, successful in medical diagnosis, risk assessment, and fault diagnosis [28], [29]
Fuzzy Logic	Handling imprecise knowledge, reasoning under uncertainty	Captures vague and uncertain information	Interpretability, intuitive reasoning	Modelling linguistic variables, fuzzy rule-based systems	Knowledge representation, fuzzy rule tuning	Fuzzy logic controllers, successful in industrial automation, decision support systems, and robotics [23]

V. CHALLENGES AND RECOMMENDATIONS

A. Challenges

Supervisory Control and Data Acquisition [SCADA] systems must be protected from internal threats by insider intrusion detection systems [IDS]. Insider attacks are more likely as SCADA systems grow more digitalized and networked, therefore strong and flexible security measures are required. This article examines the most recent difficulties encountered by Insider IDS in SCADA systems and makes suggestions to improve their efficiency and resiliency.

Data overload: Sensors, control systems, and communication networks all contribute to the massive volumes of data that SCADA systems produce. Insider IDS faces substantial challenges in processing and evaluating this data in real-time because doing so calls for strong computational capabilities and effective data handling techniques [1].

Anomaly Detection in Complex Environments: In the extremely dynamic and complex SCADA environment, insider IDS must be able to differentiate between legal deviations and probable insider assaults. Finding the right balance between detecting real anomalies and setting out false alarms is difficult, and SCADA operations vary widely [30].

Zero-Day assaults: Conventional signature-based detection techniques may have difficulty spotting new and unidentified assaults, such as zero-day threats. Insiders can take advantage of previously unknown flaws, making more sophisticated detection methods that go beyond specified rules necessary [8],[9].

Unbalanced Data Distribution: In SCADA systems, legitimate activity outweighs criminal activity by a large margin. The Insider IDS algorithms may be biased toward usual patterns as a result of this imbalance, which could affect how accurately they detect threats [1].

Limited Access to Training Data: Due to the sensitive nature of SCADA systems, obtaining labeled training data for Insider IDS can be difficult. Accurate machine learning models could be difficult to construct because of data access limits and privacy issues.

Adaptability to Evolving Attacks: Insider IDS must change in order to detect new and sophisticated threats as insider attack strategies change. To keep the system resilient to new attack vectors, regular upgrades and ongoing learning are essential [18].

SCADA systems need real-time monitoring and reaction capabilities in order to quickly minimize insider threats. The integrity and safety of critical infrastructure can be severely compromised by any lag in detection and reaction.

B. Recommendations to Address Challenges:

Machine Learning-Based Detection: To improve Insider IDS capabilities, use machine learning techniques like deep learning models and anomaly detection algorithms. Even in complicated SCADA setups, these techniques may learn from prior data and spot patterns that can point to insider assaults [30].

Effective data preparation and feature engineering are essential to overcoming the problems brought on by data overload and unbalanced distributions. Insider IDS's accuracy can be improved by selectively choosing and extracting the most pertinent features [7].

Implement mechanisms for machine learning models that allow for continuous training and updating. The system's ability to respond to changing insider threats is enhanced by routinely providing it with fresh data.

Encourage cooperation between SCADA operators, vendors, and cybersecurity professionals so they can share threat intelligence and experiences. Collaboration can result in the discovery of fresh attack pathways and the creation of more potent detection methods.

Adopt hybrid IDS strategies that incorporate rule-based systems, statistical analysis, and machine learning techniques. Utilizing the advantages of several detection techniques can improve detection precision and lower false positives [1],[26].

User Monitoring and Behavioral Profiling: Use behavioral profiling of users and administrators to find alterations in normal patterns of activity. User monitoring can offer insightful information about shady behavior and possible insider threats [17].

Real-Time Incident Response: Create real-time response plans to quickly stop insider attacks. SCADA systems can avoid future harm with automated reactions like isolating hacked devices or halting unauthorized activity.

VI. Comparison of IDS SCADA and Cyber Physical Systems (Table III)

Cyber-Physical Systems [CPS] and SCADA systems are similar and dissimilar in many ways such as the integration of technologies, data collecting and analysis, real-time monitoring and control, connectivity and communication, security issues, and industrial applications. Due to their interconnected nature, CPS and SCADA both need to handle security issues such supply chain vulnerabilities and insider threats [22]. Additionally, for these systems to be protected from cyber-physical threats, unauthorized access, safe integration of various technologies and communication networks is essential.

The strengths of CPS and SCADA can be used to improve each other's security and effectiveness. The emphasis on industrial protocols and network segmentation in SCADA can help CPS, and SCADA can gain knowledge from CPS's cutting-edge encryption and authentication methods for securing data flow in integrated systems. Additionally, SCADA can benefit from CPS's distributed control architecture and adaptive communication protocols for increased resiliency and flexibility in a variety of industrial settings, while CPS can learn from SCADA's data filtering strategies and emphasis on redundancy for improved efficiency and continuous real-time control [20],[21].

CPS and SCADA can improve their capabilities and security by sharing best practices and leveraging one another's strengths, so enhancing the general effectiveness, safety, and dependability of industrial processes and cyber-physical systems [6], [17], [20].

VII. Discussion and Analysis

The study has conducted a comprehensive analysis of developments, obstacles, and efficiency to strengthen SCADA security against insider attacks. Examining prevalent techniques like behavioral analysis and anomaly detection, the research identifies their advantages over others for identifying insider threats (Table II). It also reveals barriers to insider IDS integration, such as the difficulty of interpreting models and the lack of available data. Encouraging advancements like explainable AI and federated learning are discussed that may open up new possibilities for cooperative threat detection [30].

The part on research objectives offers a thorough analysis of the use of insider IDS, new advancements, and efficacy evaluation. For researchers and professionals looking to strengthen vital infrastructures, it delivers insightful information.

The primary goal of this study was to offer a thorough analysis of the use and efficacy of insider IDS in SCADA systems. The study discovered that machine learning-based IDS systems have developed significantly to meet the expanding problems encountered by insider threats through a thorough analysis of the most recent developments [11], [12], [14]. These machine learning methods are good at interpreting massive volumes of data, recognizing patterns, and identifying unusual behavior that can point to malicious intent.

The review covered a wide range of machine learning techniques, such as natural language processing [NLP], graph-based approaches, ensemble methods, deep learning models, anomaly detection, behavioral analysis, and ensemble methods. The benefits and drawbacks of each method for identifying insider attacks on SCADA systems were examined (Tables II and III).

Important findings from studies revealed that behavioral analysis, applying machine learning algorithms, successfully learned typical patterns of user behavior and system operations, detecting variations suggestive of insider threats [10], [12]. Another machine learning-driven strategy called anomaly detection, which takes its nods from typical system activity and looks for abnormalities from established patterns, proved essential in identifying complex attacks.

The second objective of the study was to determine whether current insider threat detection and mitigation [IDS] techniques were appropriate for SCADA systems. The study indicated that insider attacks present certain complications that conventional security measures often are unable to meet. Machine learning-based systems demonstrated their capability for learning from new data and adaptability, continuously improving alongside new threats [5], [11], [29] (Tables I and IV).

The ability to integrate different machine learning models using ensemble approaches like bagging, boosting, and stacking has been shown to improve detection accuracy and boost insider IDS in SCADA systems [2],[6],[8]. The accuracy and robustness of IDS solutions were further enhanced by the hybrid techniques that included various detection techniques.

The third objective of the review study was to find the key challenges and constraints encountered while using insider IDS solutions in SCADA systems. Due to the frequency of such incidents, the study noted the challenge in obtaining recognized insider threat data, making data collection, pre-processing, and classification difficult jobs. The review study findings reveal that it may be addressed by integrating Statistical Analysis techniques with rule based methods in IDS solutions (Tables I and III).

It also emphasized that it's very crucial to protect the security and confidentiality of crucial SCADA system data when creating machine learning models. In order to increase the transparency and interpretability of machine learning-driven IDS solutions and make sure that professionals can understand and trust the judgments made by these models, the review emphasized the necessity to integrate explainable AI. For greater transparency and confidence in the detection process, explainable AI needs to be integrated with IDS solutions in SCADA systems. Potential approaches that can accomplish this goal include ensemble methods, graph-based approaches, and rule-based systems. As they provide concrete concepts for decision-making, rule-based systems are simple to understand [27]. By revealing specific model contributions, ensemble approaches, and explicable AI techniques increase transparency. Understanding complicated relationships and attack pathways is made possible by graph-based techniques and explainable AI [30], [31]. By utilizing these methods, IDS systems in SCADA can deliver precise, comprehensible data, improving security analysts' capacity to recognize and efficiently address insider threats.

VIII. Conclusion, Limitations and Future Directions

In conclusion, the discussion and analysis of this study offered valuable insight into the most recent developments, advantages, and difficulties of insider IDS in SCADA systems. The outcomes highlighted the significance of utilizing machine learning techniques as well as their adaptability and potential to change SCADA security. The review not only advanced academic knowledge in this field but also provided practitioners seeking to strengthen the security of vital infrastructures with practical recommendations (Tables II to IV].

IX. Limitations

Limited Scope: The study excludes other alternative strategies and technologies that can improve insider threat detection in favor of Insider Intrusion Detection Systems [IDS] for SCADA systems.

1) Data availability: An important limitation was the lack of pertinent insider threat data available for model training and evaluation. The findings' applicability to other situations may be restricted by the dearth of data from actual insider attacks.

2) Time restrictions: The study only included research papers published in 2019 and later, which may have left out some pertinent studies conducted earlier.

3) Generalization: Due to differences in system architectures, data formats, and threat settings, the conclusions and suggestions may not be universally applicable to all SCADA environments.

X. Future Directions

Enhanced Data acquiring: To improve the training and assessment of machine learning models, future research might concentrate on acquiring more thorough and varied insider threat data.

Explainable AI Integration: To improve the understand ability and transparency of machine learning-driven IDS solutions, more research can study and apply cutting-edge explainable AI methodologies.

Implementation in the real world: It would be helpful to conduct field tests and case studies to evaluate the usefulness and efficacy of machine learning-based IDS in actual SCADA environments.

Hybrid ways: Researching and creating hybrid ways that combine the benefits of various techniques, such as ensemble methods and rule-based systems, may result in the detection of insider threats being more reliable and accurate.

Exploring federated learning strategies, which enable cooperative model training across several SCADA systems without exchanging sensitive data, may be a promising step toward resolving data privacy issues.

Resistance to Adversarial Attacks: It would be beneficial to conduct research to make machine learning models more resistant to adversarial attacks, which might be used by knowledgeable insiders.

Collaboration between industries: Working together with SCADA system manufacturers and industrial groups could make it easier for machine learning-based IDS solutions to be adopted in practical environments.

The field of insider IDS for SCADA systems can go further by resolving these constraints and pursuing the indicated future directions, protecting critical infrastructures from the growing threat of insider threats.

XI. Conflict of Interest Statement

The authors affirm that they have no known conflicts of interest that would have appeared to have an impact on the research presented in this study.

References

[1] Balla, A., Habaebi, M. H., Elsheikh, E. A., Islam, M. R., & Suliman, F. M. [2023]. The Effect of Dataset Imbalance on the Performance of SCADA Intrusion Detection Systems. Sensors, 23[2], 758.

[2] Salahudin, F., & Setiyono, B. [2019, September]. Design of Remote Terminal Unit [RTU] Panel Supply Monitoring Based on IOT Case Study at PLN. In 2019 6th International Conference on Information Technology, Computer and Electrical Engineering [ICITACEE] [pp. 1-6]. IEEE.

[3] Elhady, A. M., El-Bakry, H. M., & Abou Elfetouh, A. [2019]. Comprehensive risk identification model for SCADA systems. Security and Communication Networks, 2019.

[4] Huang, J. C., Zeng, G. Q., Geng, G. G., Weng, J., Lu, K. D., & Zhang, Y. [2023]. Differential evolution-based convolutional neural networks: An automatic architecture design method for intrusion detection in industrial control systems. Computers & Security, 132, 103310.

[5] Öztürk, T., Turgut, Z., Akgün, G., & Köse, C. [2022]. Machine learning-based intrusion detection for SCADA systems in healthcare. Network Modeling Analysis in Health Informatics and Bioinformatics, 11[1], 47.

[6] Alanazi, M., Mahmood, A., & Chowdhury, M. J. M. [2022]. SCADA vulnerabilities and attacks: A review of the state-of-the-art and open issues. Computers & Security, 103028.

[7] Qian, J., Du, X., Chen, B., Qu, B., Zeng, K., & Liu, J. [2020]. Cyber-physical integrated intrusion detection scheme in SCADA system of process manufacturing industry. IEEE Access, 8, 147471-147481.

[8] Livinus Obiora Nweke, “A Survey of Specification-based Intrusion Detection Techniques for Cyber-Physical Systems” International Journal of Advanced Computer Science and Applications[IJACSA], 12[5], 2021. http://dx.doi.org/10.14569/IJACSA.2021.0120506

[9] Saheed, Y. K., Abdulganiyu, O. H., & Tchakoucht, T. A. [2023]. A Novel Hybrid Ensemble Learning for Anomaly Detection in industrial sensor networks and SCADA systems for smart city infrastructures. Journal of King Saud University-Computer and Information Sciences, 35[5], 101532.

[10] Alem, S., Espes, D., Nana, L., Martin, E., & De Lamotte, F. [2023]. A novel bi-anomaly-based intrusion detection system approach for industry 4.0. Future Generation Computer Systems.

[11] Sahani, N., Zhu, R., Cho, J. H., & Liu, C. C. [2023]. Machine Learning-based Intrusion Detection for Smart Grid Computing: A Survey. ACM Transactions on Cyber-Physical Systems, 7[2], 1-31.

[12] Mohammad Azmi Ridwan, Nurul Asyikin Mohamed Radzi, Kaiyisah Hanis Mohd Azmi, Fairuz Abdullah and Wan Siti Halimatul Munirah Wan Ahmad, “A New Machine Learning-based Hybrid Intrusion Detection System and Intelligent Routing Algorithm for MPLS Network” International Journal of Advanced Computer Science and Applications[IJACSA], 14[4], 2023. http://dx.doi.org/10.14569/IJACSA.2023.0140412

[13] Yang, H., Cheng, L., & Chuah, M. C. [2019, June]. Deep-learning-based network intrusion detection for SCADA systems. In 2019 IEEE Conference on Communications and Network Security [CNS] [pp. 1-7]. IEEE.

[14] Sivakumar, S., Raffik, R., Kumar, K. K., & Hazela, B. [2023, January]. Scada energy management system under the distributed decimal of service attack using verification techniques by IIoT. In 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering [ICECONF] [pp. 1-4]. IEEE.

[15] Khan, I. A., Pi, D., Khan, Z. U., Hussain, Y., & Nawaz, A. [2019]. HML-IDS: A hybrid-multilevel anomaly prediction approach for intrusion detection in SCADA systems. IEEE Access, 7, 89507-89521.

[16] Bugshan, N., Khalil, I., Kalapaaking, A. P., & Atiquzzaman, M. [2023]. Intrusion Detection-Based Ensemble Learning and Microservices for Zero Touch Networks. IEEE Communications Magazine, 61[6], 86-92.

[17] Asiri, M., Saxena, N., Gjomemo, R., & Burnap, P. [2023]. Understanding indicators of compromise against cyber-attacks in industrial control systems: a security perspective. ACM transactions on cyber-physical systems, 7[2], 1-33.

[18] Alimi, O. A., Ouahada, K., Abu-Mahfouz, A. M., Rimer, S., & Alimi, K. O. A. [2021]. A review of research works on supervised learning algorithms for SCADA intrusion detection and classification. Sustainability, 13[17], 9597.

[19] Xingjie Huang, Jing Li, Jinmeng Zhao, Beibei Su, Zixian Dong and Jing Zhang, “Research on Automatic Intrusion Detection Method of Software-Defined Security Services in Cloud Environment” International Journal of Advanced Computer Science and Applications[IJACSA], 14[4], 2023. http://dx.doi.org/10.14569/IJACSA.2023.0140406

[20] Agrawal, N., & Kumar, R. [2022]. Security perspective analysis of industrial cyber physical systems [I-CPS]: A decade-wide survey. ISA transactions, 130, 10-24.

[21] Yu, Z., Gao, H., Cong, X., Wu, N., & Song, H. H. [2023]. A Survey on Cyber-Physical Systems Security. IEEE Internet of Things Journal.

[22] Isern, J., Jimenez-Perera, G., Medina-Valdes, L., Chaves, P., Pampliega, D., Ramos, F., & Barranco, F. [2023]. A Cyber-Physical System for integrated remote control and protection of smart grid critical infrastructures. Journal of Signal Processing Systems, 1-14.

[23] Alsakran, F., Bendiab, G., Shiaeles, S., & Kolokotronis, N. [2019, December]. Intrusion detection systems for smart home iot devices: experimental comparison study. In International Symposium on Security in Computing and Communication [pp. 87-98]. Singapore: Springer Singapore.

[24] S. V. B. Rakas, M. D. Stojanović and J. D. Marković-Petrović, "A Review of Research Work on Network-Based SCADA Intrusion Detection Systems," in IEEE Access, vol. 8, pp. 93083-93108, 2020, doi: 10.1109/ACCESS.2020.2994961.

[25] Sangeetha, K., Shitharth, S., & Mohammed, G. B. [2022]. Enhanced SCADA IDS security by using MSOM hybrid unsupervised algorithm. International Journal of Web-Based Learning and Teaching Technologies [IJWLTT], 17[2], 1-9.

[26] Potnurwar, A. V., Bongirwar, V. K., Ajani, S., Shelke, N., Dhone, M., & Parati, N. (2023). Deep Learning-Based Rule-Based Feature Selection for Intrusion Detection in Industrial Internet of Things Networks. *International Journal of Intelligent Systems and Applications in Engineering*, *11*(10s), 23-35.

[27] Al-Muntaser, B., Mohamed, M. A., & Tuama, A. Y. (2023). Real-Time Intrusion Detection of Insider Threats in Industrial Control System Workstations Through File Integrity Monitoring. *International Journal of Advanced Computer Science and Applications*, *14*(6).

[28] Mendonça, Y. V., Naranjo, P. G. V., & Pinto, D. C. (2022). The Role of Technology in the Learning Process. *Emerging Science Journal*, *6*(Special Issue), 280-295.

[29] Kandel, I., Castelli, M., & Manzoni, L. (2022). Brightness as an augmentation technique for image classification. *Emerging Science Journal*, *6*(4), 881-892.

[30] Chatterjee, J., & Dethlefs, N. [2020, September]. Temporal causal inference in wind turbine scada data using deep learning for explainable AI. In Journal of Physics: Conference Series [Vol. 1618, No. 2, p. 022022]. IOP Publishing.

[31] Nwakanma, C. I., Ahakonye, L. A. C., Njoku, J. N., Odirichukwu, J. C., Okolie, S. A., Uzondu, C., ... & Kim, D. S. [2023]. Explainable artificial intelligence [xai] for intrusion detection and mitigation in intelligent connected vehicles: A review. Applied Sciences, 13[3], 1252.