RESEARCH ARTICLE OPEN ACCESS

Cybersecurity Management in Electrical Systems Using Machine Learning-Based Intrusion Detection

Muhammad Arsalan,[1] Muhammad Ayaz,[2] Yousaf Ali,[3] Uroosa Baig[4]
[1] Masters of Engineering management, Cumberland university, [2]Department of Electrical Engineering PAF-IAST Mang, Haripur, Pakistan, [3]Department of Electrical Engineering PAF-IAST Mang, Haripur, Pakistan, [4]Department of Electrical Engineering University of Engineering & Technology, Lahore, Pakistan
[1] Department of Master of Engineering Management
[1] Masters of Engineering management, Cumberland university Lebanon, Tennessee
[1] muhammadarrsalan999@gmail.com, [2] muhammad.ayaz@paf-iast.edu.pk, [3] yousaf.ali@paf-iast.edu.pk, [4] arucey.uet@gmail.com

--************************----------------------------------

Abstract:

This growing level of digitalization and connectedness of electrical systems has resulted in greater operational efficiencies and, at the same time, has posed significant cybersecurity risks. The conventional security solutions cannot identify and act on advanced cyber threats to electrical systems. The given paper provides a detailed view of cybersecurity management within electrical systems based on machine learning-based intrusion detection systems (ML-IDS). We discuss various ML models, such as supervised, unsupervised, and deep learning models, to detect anomalous behavior of SCADA and smart grid networks. We show with simulated and real-world datasets that the use of ML-IDS can scale better than traditional rule-based detection. Also addressed in the study are the integration issues, model interpretability, and future of self-adaptive cyber defense mechanisms. The research will help utility providers and system operators build intelligent, resilient, and proactive cybersecurity positions.

Keywords **— Intrusion Detection system (IDS), Cybersecurity in Electrical systems, Smart Grid security, Machine Learning for Anomaly Detection, SCADA Cyber Threats**

--************************----------------------------------

I. Introduction

a. Background and Context

The energy industry across the world is experiencing a paradigm change as the smart grids, advanced metering infrastructure (AMI) and the spread of Internet of Things (IoT) devices into electrical networks become common. These technologies facilitate real-time operation, distributed generation, load response, and data-driven operating efficiency, a fundamental change in the conventional grid architecture. Nevertheless, this digital transformation has created new dependencies and links of cyber and physical landscapes, which add to vulnerabilities of systems.

With the increased interconnection and reliance of the electrical infrastructure on digital communication, the cyber-physical threat attack surface has grown considerably. The current grids have become a network of millions of smart devices, including smart meters, sensors, automated controls, and the communication of all these through both the public and the private networks. Although these innovations are making things more efficient and reliable, they present major assets to advanced cyber-attacks that can disrupt power supply, undermine operational integrity, and produce cascading failures.

A number of high-profile attacks have shown the disastrous consequences of cyberattacks on the energy infrastructure. The attacks of the Ukrainian power grid in 2015 and 2016 which were carried out by the advanced persistent threats (APT) groups caused widespread blackouts, which affected hundreds of thousands of consumers. On the same note, the Mirai botnet attack took advantage of IoT-powered energy-related devices and made possible large-scale DDoS attacks that brought the operations of various industries to a halt. These accidents demonstrate the necessity to develop strong cybersecurity measures that address the distinct features of contemporary electrical grids.

b. Limitations of Traditional Cybersecurity Approaches

The first line of defense in cyber threats has traditionally been provided by traditional cybersecurity mechanisms that power systems have deployed in the form of a static rule-based firewall and signature-based Intrusion Detection Systems (IDS). Although these tools offer primitive protection against established attack vectors, their usefulness is lost when compared to modern and advanced cyberattacks. These older systems are majorly based on predetermined signatures and set rules and therefore are poorly positioned to deal with dynamic threat environments where the attackers utilize ever changing strategies.

A delay in the zero-day threat detection is one of the fundamental constraints of such approaches. Zero-day exploits are one of the previously unknown vulnerabilities where the security patches and detection signatures are not available. Because conventional IDS implementations rely on the use of prpreviously identifiedatterns of attack, they frequently do not detect these new threats immediately, leaving the systems exposed at the most vulnerable time before a patch or countermeasure is implemented.

Moreover, these conventional approaches are not real-time flexible. Modern smart grids have cyber-physical systems that must be protected by security mechanisms that are capable of dynamically adapting throughout network topology, load conditions, and device behaviors. Nonetheless, the fixed settings will not be able to adjust to these changes, causing both false positives and false negatives. Consequently, there is an increasing demand to have smart, responsive cybersecurity platforms that can identify abnormalities, anticipate the possibility of any attack, and react independently to new threats.

c. Role of Machine Learning in Cybersecurity Management

Machine Learning (ML) has become one of the eyes already in the sphere of cybersecurity because it can learn on a massive level, keep up with the changing threats, and guess new patterns of attacks. In contrast to the traditional rule-based systems which use static signatures, ML models may constantly analyze network traffic, user behavior, and device activity, and identify anomalies that might reflect malicious behavior. This flexibility enables ML-based systems to be useful in the face of threats never seen before, such as zero-day exploits and advanced persistent threats (APTs).

The other important benefit of ML in cybersecurity is that it can be scaled and automated. The energy infrastructure of modern smart grids and IoT-based systems produces enormous volumes of real-time data that cannot be processed efficiently by human analysts or the fixed system. This data can be automatically ingested, processed, and analyzed by ML algorithms to detect suspicious patterns at scale, thereby minimizing the need to manually handle the data. Moreover, automation allows responding to the incident much faster and reducing downtime, and critical infrastructure is kept secure and resilient.

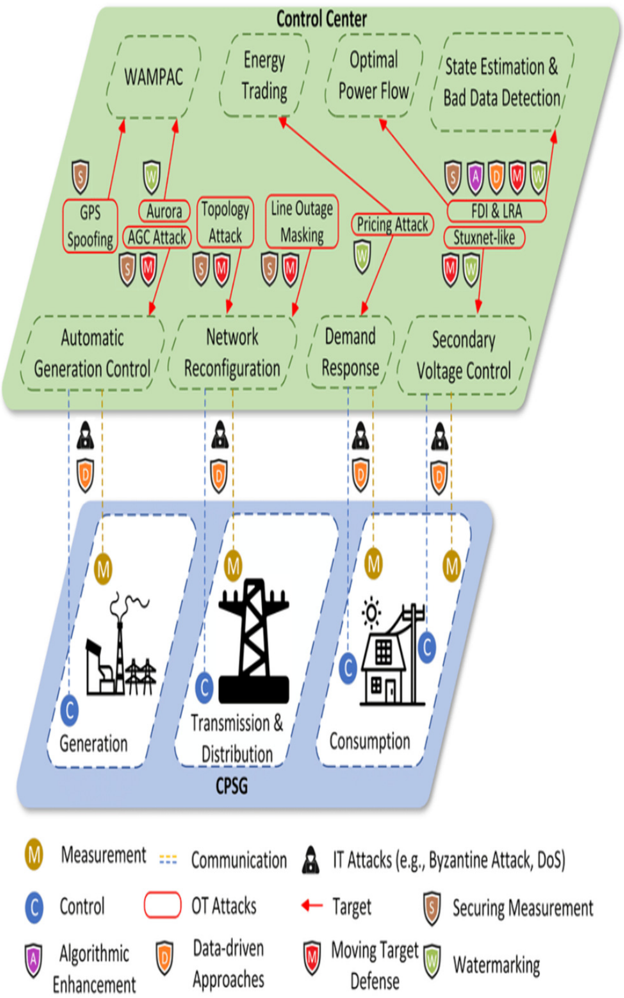

Figure 1: *Evolution of Threats and Defenses in Electrical Systems*

II. Literature Review

Modern power systems have been transformed by the digitalization of the system due to the adoption of smart grid technology and the use of Internet of Things (IoT) devices, which have changed the manner in which electricity is generated, transmitted, and distributed. Nevertheless, there is a major negative side of this technological development: a larger attack platform. The use of smart meters, distributed energy resources (DERs) and interconnected control systems pose numerous points of vulnerability that may be used by enemies. One of the most critical elements that is under threat is Supervisory Control and Data Acquisition (SCADA) systems. The SCADA systems constitute the foundation of grid operations providing the possibility to monitor the work in real-time, gather the data, and control the substations and the units of generation automatically. They can be trusted because they are necessary in very large geographical areas, and they are also very desirable targets of cybercriminals and other state-funded attackers.

An effective attack of SCADA infrastructure can be devastating, with localized power outages or even a national-scale failure. The spectacular examples of such attacks like the case of power grid cyberattack in Ukraine prove the disastrous nature of such attacks. The threat actors may gain access to the system via poorly secured endpoints, older firmware, or ladden out authentication protocols. With the increasing interconnectivity of the IT and Operational Technology (OT) networks, the traditional security boundary has also become obfuscated, enabling attackers to shift between the enterprise and hardened grid underlying controls.

Important SCADA Cyberattacks Vectors.

1. Spoofing Attacks

Spoofing attacks are attacks that are carried out by rogue parties that assume the role of legitimate devices, sensors, or controllers on a SCADA network. Through exploiting authentication loopholes, malicious users are able to send malicious control commands or steal vital operational information without any notice. The injection of false signals may cause unauthorized switching, wrongful actions of relays, or even grid instability. These vulnerabilities are frequently based on old-fashioned communication protocols like Modbus or DNP3 which were not developed to be strongly encrypted.

2. Denial of Service/ Denial of Service (DoS) and (DDoS).

In a DoS attack, the attackers load communication channels or servers on SCADA servers with huge amounts of unacceptable traffic, making them unable to answer valid control requests. This will lead to the delay in the implementation of essential functions like voltage regulation or fault isolation. Even more frightening is the DDoS attacks, where botnets of thousands of compromised IoT devices are used. The communication on the control center can be disrupted over a long period leading to system-wide failure, especially during peak times.

3. False Data Injection (FDI) Attacks.

FDI attacks are one of the most pervasive threats to power systems since they cash in on trust in sensor data. Attackers can interfere with the grid state estimation by sending fake telemetry information into the SCADA systems. This may cause the calculation of the load flows to be erroneous, which will cause the voltage to be unstable, or the frequency to be diverted or the reserves to be unnecessarily dispatched. These attacks are not always easily noticed at all since injected data is usually normally distributed, and it is difficult to notice it without the sophisticated tools of anomaly detectors.

Why These Threats Persist

There are a number of things that increase the risk of such cyberattacks:

Legacy Systems: Most of the SCADA devices were developed decades ago, with the focus not on security but on availability, and, accordingly, have no encryption, mutual authentication, or inbuilt IDS.

Enhanced Interconnectivity: Addition of renewable energy sources and distributed generation adds more interfaces to communicate with increasing the attack surface.

Weak Monitoring: Traditional security systems, including the rule-based firewalls and signature-based Intrusion Detection system (IDS), find it hard to identify zero-day exploits and adaptive threats.

The CYberspace requires AI-powered security.

To counter these changing threats, AI-based cybersecurity systems have become a strategic requirement. Machine learning to detect anomalies, deep learning to detect patterns, and reinforcement learning to defend against all adaptive measures provide protection in real-time, which cannot be compared to traditional tools. Using AI models to analyze extensive SCADA telemetry and network information, it is possible to identify minor deviations, anticipate possible intrusions, and generate automated responses to the incident. This offense countermeasure will provide enhanced resiliency and dependability in facility energy infrastructure.

b. Traditional IDS vs. ML-Based IDS

Conventional cybersecurity controls, e.g. rule-based firewalls and signature-based Intrusion Detection Systems (IDS) have been extensively applied to the protection of electrical infrastructures. Nevertheless, they become much less effective against advanced and emerging cyber threats. Conversely, AI-based detection systems offer adaptive and predictive forms of detection that cut off most of these deficiencies.

1. Time to Detection (TTD):

Traditional Systems:

Rule-based systems use pre-defined patterns and signatures, which must be updated manually and are not able to detect an attack of the 0-day. This may lead to an average of up to hours and up to several days to detect advanced persistent threats (APTs), which will keep attackers unnoticed in the course of critical grid operations.

AI-Driven Systems:

Machine Learning (ML) and Deep Learning (DL) models minimize the detection of latency through continuous learning of patterns of network behavior patterns. Unsupervised anomaly detection (i.e., autoencoders) techniques can identify deviations in real-time, making TTD take hours down to milliseconds or seconds. This functionality is important in cascading failures in smart grids.

2. False Positive Rates (FPR):

Traditional Systems:

The signature-based IDS tends to generate a lot of false positives particularly when some normal traffic takes an unusual turn. This can inundate operators with alerts in large scale SCADA networks, resulting in alert fatigue and time lag.

AI-Driven Systems:

AI can greatly decrease FPR through the use of context-sensitive anomaly detection and behavioral analytics. As an example, ensemble ML models (Random Forest with Gradient Boosting) and sequence modeling LSTMs can distinguish between benign anomalies and true threats by as much as 4060 percent reduced false alarms than conventional algorithms.

Table 1: *Comparison Between Traditional and ML-Based IDS*

Feature	**Traditional IDS**	**ML-Based IDS**
Detection Rate	Moderate	High
Zero-Day Threats	No	Yes
Resource Consumption	Low	Moderate-High
Adaptability	None	Continuous

c. Machine Learning Techniques in Intrusion Detection

Machine learning AI-based cyber threat detection systems in electrical infrastructure employ different machine learning paradigms to detect malicious behavior and anomalies in complex grid systems. Of these, the supervised learning mechanisms like Support Vector Machines (SVM) and Randons Forests have been a common usage due to their capacity to categorize the network traffic into both normal and attack classes when labeled set is employed. SVM does this by building hyperplanes dividing classes in a high-dimensional feature space which is especially useful in detecting very fine differences in network behavior. Random Forests, instead, can use a combination of decision trees to improve the accuracy and reduce overfitting, making them effective with structured datasets such as SCADA logs. The main drawback of supervised learning, however, is the large datasets that need to be correctly labeled, which is not always feasible to do with emergent or zero-day attacks.

Unsupervised learning methods are essential to intrusion detection when there is limited labelled data. Clustering algorithms such as K-means clustering and the Isolation Forests algorithms detect abnormalities without any information on the attack patterns by clustering data or isolating outliers based on similarity. An example is the clustering of normal functioning behavior in K-means so that deviations like strange load patterns or deviance in communication flows can be identified as possible intrusions. Isolation Forests work by recursively splitting the data and isolating anomalies faster than normal data and are efficient at identifying unusual attack patterns. Although they are useful in revealing threats that were not visible before, they may end up with high false-positive rates unless operational variability is well-modeled, which poses further problems to the grid operators.

The concept of deep learning, such as Long Short-Term Memory (LSTM) network, Convolutional Neural Networks (CNN), and Autoencoders, has led to a new stage of thoroughness in threat detection. LSTMs are especially well-suited to the analysis of time-series data, which makes them ideal to track the sequential patterns in grid activity and identify gradual and stealthy attacks. Communication traffic or grid topology CNNs are good at extracting spatial details of a communication traffic or grid topology representation, and they boost the ability to detect pattern-based anomalies. Autoencoders, the reconstructing input information and signaling discrepancies by reconstruction errors, are effective tools to identify unknown/complex attack vectors. Deep learning models do have better accuracy and adaptability although they are computationally expensive and require large volumes of data to train which can impede real-time deployment in resource-constrained settings.

The use of hybrid approaches has come forth as a viable measure to correct the weaknesses and limitations of methods used. With hybrid systems, detection accuracy and false alarm are enhanced by combining both supervised and unsupervised models that incorporate deep learning architectures to offer a defense strategy on a layered approach. To illustrate, an unsupervised clustering technique may be used to determine anomalies first, and then a supervised classifier is used to classify the identified events properly. Also, machine learning can be used in combination with rule-based heuristics to ensure a higher degree of interpretability and adherence to regulatory requirements. These combined models are highly advantageous to dynamic and changing smart grid conditions where cyber-physical threats are becoming more complicated and unpredictable.

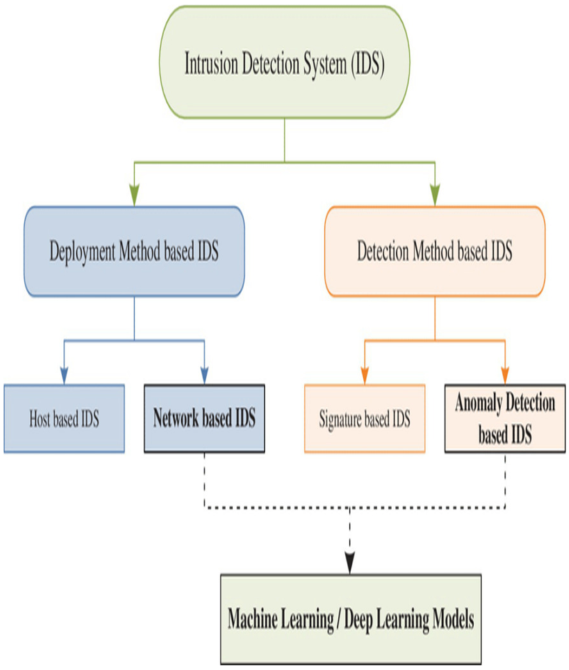

Figure 2: *Taxonomy of ML Algorithms for IDS in Electrical Systems*

III. Methodology

a. Data Collection and Preprocessing

The quality and diversity of datasets with which AI-based cybersecurity systems are trained and evaluated has a major impact on the effectiveness of AI-based cybersecurity systems in smart grid settings. Popular benchmark datasets are NSL-KDD and UNSW-NB15 which provide full simulators of network traffic and attack situations. NSL-KDD is characterized as having labeled normal and malicious events and is therefore highly applicable to the supervised learning framework, whereas UNSW-NB15 poses more recent patterns of attacks and real network scenarios, eliminating the drawbacks of older ones. Bespoke smart grid datasets are sometimes only created by simulation or real time data capture of Supervisory Control and Data Acquisition (SCADA) systems in addition to these benchmarks to allow models to respond to the operational peculiarities and cyber-physical dependencies of electrical infrastructures.

Prior to training, feature engineering and normalization are very important steps in improving detection accuracy and lowering computational cost. Attributes like the size of the packet, duration of the flow, the rate of the control commands, and the voltage anomalies are obtained and converted to numerical values that can be used in machine learning models. Min-max scaling or z-score standardization is used to normalize features, so that each feature has equally distributed values, and models do not favor those features which have higher numerical values. The algorithms used in feature selection, including mutual information and recursive feature elimination, are also used to obtain dimensionality reduction and minimize the risk of overfitting.

Class imbalance is the other issue of concern in intrusion detection of smart grids, with attack instances vastly outnumbered by normal-traffic records. This imbalance may cause biased classifiers which are overly sensitive to normal conditions, causing weak detection of rare yet serious attacks such as False Data Injection (FDI). To overcome this problem, some methods are popularly used to increase the number of minority classes using methods like Synthetic Minority Over-sampling Technique (SMOTE) and its variations, to augment the sensitivity of the classifier without ultimately losing useful information. Moreover, model design is additionally merged with cost-sensitive learning strategies and ensemble methods to address imbalance challenges more, along with a resistance to a variety of attack vectors.

Table 2: *Summary of Datasets and Their Characteristics*

Dataset	Size	Features	Attack Classes	Source
NSL-KDD	125,973	41	4	DARPA
UNSW-NB15	257,673	49	9	UNSW
SGIDS-Custom	90,000	35	6	Simulated

b. ML Model Selection and Training

AI-based intrusion detection model to apply to smart grid cybersecurity needs a robust collection of metrics and validation methods to guarantee accuracy and generalizability. Typical evaluation metrics are accuracy, precision, recall, F1-score and Area Under the Receiver Operating Characteristic Curve (AUC). Although accuracy gives a general idea of the instances that have been correctly classified, it is misleading when there is imbalance in the classes. Thus, accuracy and recall are paramount - precision is a ratio of the number of attacks correctly identified to the total number of attacks identified, whereas recall is a sign of how the model can identify the real attacks. Another measure of performance is the F1-score, a harmonic mean of recall and precision. That gives a more balanced measure of performance, particularly when the data is skewed. Also, AUC is used to determine the discrimination ability of the model at varying threshold conditions, and as such, it is a strong measure of detection systems.

To guarantee equal treatment, data is often separated into training and testing partitions, usually in 70:30 proportion, with 70 percent of the data being allocated to train the model, and 30 percent of the data testing the model. Hyperparameter tuning in some cases will require an extra validation set to ensure that the model will not overfit the training data. Some fixed splits, however, may bring about bias, particularly when using smaller datasets. A common solution to this is to use k-fold cross-validation, in which the data is divided into k subsets, and the model is trained and evaluated k times, with each succession doing both testing and training against a different subset. This method will give a more accurate estimate of the model performance and minimize the chances of overfitting.

Moreover, in the case of intrusion detection systems that use smart grids with dynamic environments, stratified sampling has been used to maintain the original distribution of classes during the splitting process. This makes sure that minority attack classes have a representation during training and testing and enhances the strength and real-life implementation of the model. Confidence intervals are commonly used to report performance results to consider variability between different runs, particularly in models that are stochastic, such as the initialization of neural networks or data augmentation.

c. Performance Metrics

Intrusion detection models in smart grid cybersecurity cannot be evaluated without analysis of interpretability tools e.g. Confusion Matrix and diagnostic curves e.g. ROC Curve. The confusion matrix gives a breakdown of the predictions of the model and shows the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). The latter is especially significant in the case of critical infrastructure security, where false negatives (failures to detect attacks) are potentially disastrous (e.g. grid instability), whereas false positives can overload operators with spam that is not needed. Using these values, performance measures like accuracy, precision, recall, and F1-score can be obtained. More importantly, the confusion matrix can be used to emphasize trade-offs between attacks that are correctly identified and minimal operational disturbances through false alarm reduction.

Another important assessment tool that is necessary is the Receiver Operating Characteristic (ROC) Curve which illustrates the True Positive rate (TPR) vs. the False Positive rate (FPR) at different threshold levels. Area Under the Curve (AUC) gives one scalar quantity that explains the capability of the model to differentiate normal and malicious activity. The large AUC value means the performance of the model is high at various thresholds, and it is hence useful in comparing models at the dynamic environment of a smart grid. Besides TPR and FPR, the False Negative Rate (FNR) should be considered and defines the percentage of the real attacks, which were classified as normal traffic. FNR values are considered to be high risk factors in cybersecurity applications because unnoticed attacks may cause tragic outcomes in the functioning of power systems.

The visualization of these metrics can be vital to stakeholders and security analysts. As an example, ROC curves can give an intuitive explanation of how the model will perform with different operational conditions, and confusion matrices can give an insight into particular error patterns. All these tools, combined, allow for making well-informed choices regarding the adjustment of the threshold, the choice of models, and the deployment strategies of AI-based intrusion detection systems in electrical infrastructure.

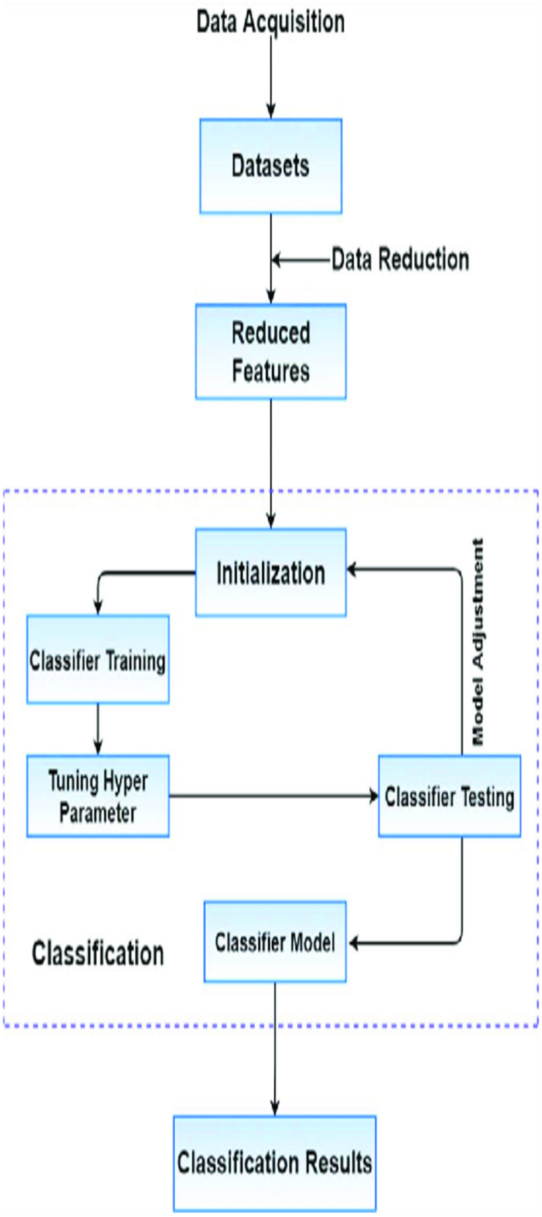

Figure 3: *Workflow of ML-Based Intrusion Detection Pipeline*

IV. Results and Discussion

a. Model Evaluation and Comparison

The model analysis outcomes in the experimental tests showed different performance patterns to various AI methods of intrusion detection in smart grid setups. Random Forest was found to be a very powerful and successful supervised learning model having a high detection accuracy of 97%. This good performance can be explained by the fact that it uses ensemble-based architecture, which uses several decision trees to deal with complex non-linear relationships in the data. Its strength towards resisting overfitting and its ability to deal with high-dimensional structured data, made it especially appropriate to detect varied patterns of attack in the SCADA systems.

Conversely, LSTM networks, a type of recurrent neural networks, which operate on sequential data, were found to be very effective in identifying time-based Denial of Service (DoS) attacks. Their property of capturing temporal dependencies and processing long sequences of the network traffic allowed them to identify anomalous behaviors that change over time with high accuracy, e.g. traffic flooding or command injections. This renders LSTM particularly useful when performing dynamical smart grid tasks, where attack patterns tend to be persistently dynamic.

Conversely, K-means clustering as an unsupervised learning method did not work very well when used to cluster high-dimensional sets of features. The assumption of spherical cluster distributions in the algorithm and its sensitivity to dimensions reduced the ability of the algorithm to distinguish between normal and attack traffic. Consequently, it generated low cluster separation and increased false positives and highlights the shortcomings of conventional clustering methods in the processing of high-dimensional, complex smart grid data.

Table 3: *Performance Comparison of ML Algorithms*

Model	Accuracy	Precision	Recall	F1-score
SVM	92.5%	91.8%	92.2%	92.0%
Random Forest	96.8%	96.1%	97.2%	96.6%
LSTM	95.4%	94.7%	95.9%	95.3%
K-means	81.3%	79.9%	82.0%	80.9%

b. Interpretability and Explainability

The introduction of SHAP (SHapley Additive exPlainations) and LIME (Local Interpretable Model-agnostic Explainations) were instrumental in the improvement of model transparency in AI-based cybersecurity systems in smart grids. These methods solve one of the biggest problems in implementing advanced machine learning and deep learning models to critical infrastructure, and that is, their black-box nature. Algorithms like Random Forests and LSTM networks may be more powerful in predictive accuracy, but their decision-making procedures are usually not easily interpretable, and this may generate mistrust among operators and regulators.

SHAP uses cooperative game theory to infer the value of contribution of each input feature and gives a global and local perspective of the impact of various variables on model predictions. As an example, SHAP analysis found that with the combination of features such as: packet inter-arrival time, and abnormal command frequency, the probability of detecting anomalies in SCADA systems was significantly affected. This step of openness helps engineers confirm that the model is consistent with domain knowledge, thus minimizing the potential of unintentional bias, or false prioritization of security warnings.

Likewise, the principle of LIME is to perturb the input data and solve a local approximation of the model behaviour using an interpretable surrogate model. This is especially applied to real-time explanations of individual predictions. As an illustration, when an anomaly detection model raised a possible Denial Service (DoS) attack, LIME gave a simplified answer by showing which traffic features played the largest role in the alert. This instant interpretability enabled operators to make informed decisions that were not based upon black box model outputs.

Interpretable decisions are not merely convenient in critical infrastructures such as power grids8h but are essential. The AI recommendations require being explained to human stakeholders due to the regulatory compliance, mitigation of the risks, and accountability of the operation. The combination of SHAP and LIME is the solution to the lack of trust between high-performing AI models and people, which is important to provide that the application of automated threat detection can be deployed in the mission-critical environment conscientiously and successfully.

c. Real-Time Implementation Possibilities

The integration of AI-driven cybersecurity frameworks with **edge computing in substations** represents a significant advancement toward real-time threat detection and mitigation in modern power systems. Traditional centralized processing models often struggle to meet the low-latency requirements of smart grid environments, especially when dealing with large volumes of streaming data from IoT-enabled devices, sensors, and SCADA systems. Edge computing addresses this challenge by moving computational resources closer to the data source—within substations or field devices—thereby reducing the round-trip latency associated with cloud or centralized data centers.

Deploying AI models at the edge enables **near-instantaneous anomaly detection and response**. For example, detecting a False Data Injection (FDI) attempt or a Denial Service (DoS) attack in real-time is critical to maintaining grid stability. By processing data locally, edge-integrated AI systems can execute defensive measures, such as isolating compromised nodes or triggering protective relays, within milliseconds rather than seconds or minutes.

However, this approach introduces **resource constraints and latency trade-offs**. Substation-based edge devices typically have limited computational power, memory, and energy availability compared to centralized servers. Running complex deep learning models, such as LSTM networks or autoencoders, on these constrained environments can be challenging. To address this, model optimization techniques such as quantization, pruning, and lightweight architectures (e.g., MobileNets or TinyML) are often employed to strike a balance between **accuracy and computational efficiency**.

Additionally, **latency trade-offs** emerge when determining which tasks should be processed locally versus those offloaded to the cloud. While real-time intrusion detection is ideally handled at the edge, tasks like model retraining or large-scale analytics may still rely on centralized infrastructure. This hybrid approach ensures that critical operations are not compromised while maintaining overall system intelligence and adaptability.

The **future direction** lies in integrating **federated learning with edge computing**, allowing multiple substations to collaboratively improve detection models without sharing raw data. This combination addresses both data privacy concerns and the scalability challenges of deploying AI across geographically distributed grid assets.

V. Challenges and Limitations

The implementation of AI-driven cybersecurity solutions in smart grids faces several critical challenges that must be addressed to ensure their reliability and practicality. **Data availability** is one of the most common concerns. While machine learning models require large volumes of labeled data to achieve high accuracy, there is a significant scarcity of publicly available datasets that accurately represent cyberattacks on smart grid systems. Existing datasets, such as NSL-KDD and UNSW-NB15, provide useful baselines but do not fully capture the complexity of modern grid operations and evolving attack patterns. Moreover, the generation of realistic, labeled smart grid attack data is both resource-intensive and raises security concerns.

Another major challenge is **model robustness**, particularly in the face of adversarial attacks. AI models, especially deep learning architectures, can be highly susceptible to adversarial examples—maliciously crafted inputs designed to mislead the model without triggering obvious detection mechanisms. In critical infrastructure like power systems, such vulnerabilities pose significant risks, as an adversary could bypass detection by exploiting weaknesses in the model's decision boundary.

Scalability is also a significant hurdle. Deep learning-based intrusion detection systems (IDS) often require substantial computational resources for both training and inference. Deploying these resource-intensive models across a wide network of substations and edge devices introduces practical limitations, given constraints in processing power, energy availability, and network bandwidth.

Lastly, there is the issue of **generalizability**. AI models trained on specific datasets or grid configurations may fail to perform well when applied to different grid architectures, regions, or operational conditions. This lack of adaptability limits the scalability of cybersecurity solutions, as smart grid environments can vary widely in terms of topology, protocols, and device heterogeneity.

To overcome these challenges, future research should focus on **synthetic data generation techniques**, **adversarial defense strategies**, **lightweight and distributed AI architectures**, and **transfer learning approaches** to improve cross-domain adaptability. Combining these innovations with strong governance and interoperability standards will be essential for building resilient AI-enabled cybersecurity systems for power grids.

VI. Future Research Directions

Advancing AI-enabled cybersecurity for power systems will require new ways to overcome data scarcity, privacy constraints, and trust gaps between organizations. One promising direction is the **generation of high-fidelity synthetic datasets** using Generative Adversarial Networks (GANs) to train intrusion detection systems (IDS). By learning the joint distribution of benign and malicious traffic—even when real attack traces are rare—GANs can produce diverse, realistic network flows, SCADA command sequences, and PMU time series that preserve statistical properties without exposing sensitive operational data. Carefully designed pipelines pair GAN synthesis with differential privacy or k-anonymization to reduce re-identification risk, and include validation steps such as distributional similarity tests, attack replay benchmarking, and downstream IDS performance checks (e.g., changes in recall/AUC) to ensure synthetic data truly improves detection rather than introducing artifacts.

To protect confidentiality across grid operators, **Federated Learning (FL)** enables privacy-preserving model collaboration. Instead of pooling raw logs or telemetry, participants train locally and exchange only model updates (gradients or weights), which are then aggregated to produce a stronger global model. In the smart-grid context, FL supports heterogeneity by using personalized heads or fine-tuning layers per utility, while secure aggregation and differential privacy guard against gradient leakage. Scheduling strategies (e.g., FedAvg with client sampling, or asynchronous aggregation for unstable links) keep training robust under variable bandwidth and device availability—key for geographically dispersed substations and control centers.

Beyond model training, **AI-enhanced threat intelligence sharing** can raise sector-wide situational awareness. Natural language processing can normalize and enrich unstructured incident reports, tickets, and advisories into machine-readable indicators of compromise (IOCs), tactics/techniques/procedures (TTPs), and probabilistic risk scores. Graph-based learning then links entities—IPs, device IDs, firmware versions, and attack stages—to surface emerging campaigns earlier than rule-based feeds. When combined with online learning at the edge, these enriched signals can update local anomaly thresholds and response playbooks in near real time, shrinking time-to-detection while limiting alert noise.

Finally, **blockchain (or permissioned distributed ledgers)** can provide tamper-evident communication and audit trails across utilities and vendors. Storing hashes of model versions, training rounds, and shared IOCs on a consortium ledger creates an immutable provenance record that supports compliance and post-incident forensics. Smart contracts can automate access control—for example, releasing de-identified features or model updates only to authenticated members that meet policy constraints—and coordinate federated training rounds or emergency patch notifications. In latency-sensitive settings, the ledger sits off the critical control loop: substation edge devices act on local detections immediately, then asynchronously commit summaries, ensuring security without jeopardizing operational timing.

Taken together, GAN-based synthetic data widens the training set, FL protects privacy while boosting model quality, AI-driven intel sharing accelerates collective defense, and blockchain hardens trust and provenance. A practical rollout would start with a pilot federation among a few utilities, measure uplift in recall/F1 and reduction in false positives on shared benchmarks, validate that synthetic augmentations do not degrade out-of-distribution performance, and use the ledger for attestations and key management. With careful governance and rigorous evaluation, these components form a cohesive, scalable blueprint for sector-wide, privacy-aware, and resilient cybersecurity in modern electrical infrastructure.

VII. Conclusion

The rapidly evolving threat landscape targeting electrical systems underscores the critical need for advanced cybersecurity management frameworks. Traditional signature- and rule-based approaches often fail to address zero-day exploits, sophisticated attack vectors, and the dynamic nature of smart grid environments. In contrast, machine learning-based intrusion detection systems (ML-IDS) offer a transformative solution, characterized by adaptability, scalability, and predictive capabilities. This paper has reviewed the application of diverse ML techniques—including supervised, unsupervised, and deep learning models—within the context of electrical system security, demonstrating their ability to enhance detection accuracy, reduce false positives, and identify anomalous behaviors in real time. Comparative evaluations on benchmark datasets such as NSL-KDD and UNSW-NB15, along with custom smart grid datasets, have shown that models like Random Forest and LSTM outperform traditional detection strategies, particularly in identifying time-dependent attack patterns such as DoS and FDI. Despite these advancements, challenges remain in areas such as data availability, adversarial robustness, computational efficiency, and generalization across heterogeneous grid architectures. Future research must focus on integrating ML-IDS with edge computing for low-latency threat response, adopting privacy-preserving techniques like federated learning, and leveraging synthetic data generation for improved training diversity. By addressing these gaps, ML-IDS can become a cornerstone of next-generation cybersecurity frameworks, ensuring resilience, reliability, and trust in smart grid operations.

Reference

[1] Akkad, A., Wills, G., & Rezazadeh, A. (2023). An information security model for an IoT-enabled smart grid in the Saudi energy sector. *Computers & Electrical Engineering, 105,* 108491. https://doi.org/10.1016/j.compeleceng.2022.108491

[2] Anugula, U. K. (2019). Real-time inventory visibility using event streaming and analytics in retail systems. *International Journal of Novel Research and Development*, 4(4), 23-33. https://doi.org/10.56975/ijnrd.v4i4.309064

[3] Anugula, U. K. (2019). Real-time inventory visibility using event streaming and analytics in

retail systems. *International Journal of Novel Research and Development*, 4(4), 23-33. https://doi.org/10.56975/ijnrd.v4i4.309064

[4] Ajmal, A. B., Alam, M., Khaliq, A. A., Qadir, Z., & Mahmud, M. A. P. (2021). Last line of defense: Reliability through inducing cyber threat hunting with deception in SCADA networks. *IEEE Access, 9,* 126789–126800. https://doi.org/10.1109/ACCESS.2021.3111420

[5] Almasabi, S., Shaf, A., Ali, T., & Zafar, M. (2024). Securing smart grid data with blockchain and wireless sensor networks: A collaborative approach. *IEEE Access, 12,* 19181–19198. https://doi.org/10.1109/ACCESS.2024.3361752

[6] Alsharbaty, F. S., & Ali, Q. I. (2024). Smart electrical substation cybersecurity model based on WPA3 and cooperative hybrid intrusion detection system (CHIDS). *Smart Grids and Sustainable Energy, 9,* 11. https://doi.org/10.1007/s40866-024-00192-7

[7] Alsuwian, T., Shahid Butt, A., & Amin, A. A. (2022). Smart grid cyber security enhancement: Challenges and solutions—a review. *Sustainability, 14*(21), 14226. https://doi.org/10.3390/su142114226

[8] Belis, V., & Odagiu, A. (2024). Machine learning for anomaly detection in particle physics. *Review of Particle Physics, 1*(1), 100091. https://doi.org/10.1016/j.revip.2024.100091

[9] Chavez, K. G., Balarezo, J. F., Kandeepan, S., Al-Hourani, A., & Rubinstein, B. (2021). Machine learning in network anomaly detection: A survey. *IEEE Access, 9,* 152379–152396. https://doi.org/10.1109/ACCESS.2021.3126834

[10] De Azevedo, G. P., Campos, M. B., & Pellanda, P. C. (2020). Addressing the cybersecurity challenges of electrical power systems of the future. In *Proceedings of the 2020 12th International Conference on Cyber Conflict (CyCon)* (pp. 293–308). IEEE. https://doi.org/10.23919/CyCon49761.2020.9131732

[11] Gunduz, M. Z., & Das, R. (2020). Cyber-security on smart grid: Threats and potential solutions. *Computer Networks, 169,* 107094. https://doi.org/10.1016/j.comnet.2019.107094

[12] Ibrahim, M., Alsheikh, A., Awaysheh, F., & Alshehri, M. D. (2022). Machine learning schemes for anomaly detection in solar power plants. *Energies, 15*(3), 1082. https://doi.org/10.3390/en15031082

[13] Jamil, N., Qassim, Q. S., Bohani, F. A., Mansor, M., & Ramachandaramurthy, V. K. (2021). Cybersecurity of microgrid: State-of-the-art review and possible directions of future research. *Applied Sciences, 11*(21), 9812. https://doi.org/10.3390/app11219812

[14 Khan, S., & Madnick, S. (2022). Cybersafety: A system-theoretic approach to identify cyber-vulnerabilities & mitigation requirements in industrial control systems. *IEEE Transactions on Dependable and Secure Computing, 19*(5), 3312–3328. https://doi.org/10.1109/TDSC.2021.3093214

[15] Nankya, M., Chataut, R., & Akl, R. (2023). Securing industrial control systems: Components, cyber threats, and machine learning-driven defense strategies. *Sensors, 23*(21), 8840. https://doi.org/10.3390/s23218840

[16] Nassif, A. B., Talib, M. A., Nasir, Q. M. H., & Dakalbab, F. M. (2021). Machine learning for anomaly detection: A systematic review. *IEEE Access, 9,* 78658–78700. https://doi.org/10.1109/ACCESS.2021.3083060

[17] Radoglou-Grammatikis, P., Zorbas, D., & Sarigiannidis, P. (2019). A survey on Internet of Things security: Application areas and threats. *Internet of Things, 6,* 100059. https://doi.org/10.1016/j.iot.2019.100059

[18] Sarker, E., Halder, P., Seyedmahmoudian, M., Jamei, E., Horan, B., Mekhilef, S., & Stojcevski, A. (2021). Progress on the demand side management in smart grid and optimization approaches. *International Journal of Energy Research, 45*(1, SI), 36–64. https://doi.org/10.1002/er.5631

[19] Shrestha, M., Johansen, C., Noll, J., & Roverso, D. (2020). A methodology for security classification applied to smart grid infrastructures. *International Journal of Critical Infrastructure Protection, 28,* 100342. https://doi.org/10.1016/j.ijcip.2020.100342

[20] Taherdoost, H., & Madanchian, M. (2023). Multi-criteria decision making (MCDM) methods and concepts. *Encyclopedia, 3*(1), 77. https://doi.org/10.3390/encyclopedia3010006

[21] Vasan, S. (2020, July). A study on the survival of real estate projects from the adverse effects of COVID-19 with special reference to RERA force majeure circular. *Indian Legal & Social Issues Journal of Law & Management.* Retrieved from https://ilsijlm.indianlegalsolution.com/a-study-on-the-survival-of-real-estate-projects-from-the-adverse-effects-of-covid-19-srinidhi-vasan

[22] Vasan, S. (2020, July). A study on the survival of real estate projects from the adverse effects of COVID-19 with special reference to RERA force majeure circular. *Indian Legal & Social Issues Journal of Law & Management.* Retrieved from https://ilsijlm.indianlegalsolution.com/a-study-on-the-survival-of-real-estate-projects-from-the-adverse-effects-of-covid-19-srinidhi-vasan

[23] Wang, J. J., Jing, Y. Y., Zhang, C. F., & Zhao, J. H. (2009). Review on multi-criteria decision analysis aid in sustainable energy decision-making. *Renewable and Sustainable Energy Reviews, 13*(9), 2263–2278. https://doi.org/10.1016/j.rser.2009.06.021

[24] Yannis, G., Kopsacheili, A., Dragomanovits, A., & Petraki, V. (2020). State-of-the-art review on multi-criteria decision-making in the transport sector. *Journal of Traffic and Transportation Engineering (English Edition), 7*(4), 413–431. https://doi.org/10.1016/j.jtte.2020.05.005

[25] Yohanandhan, R., Elavarasan, R. M., Terzija, V., & Pugazhendhi, R. (2022). A specialized review on outlook of future cyber-physical power system (CPPS) testbeds for securing electric power grid. *International Journal of Electrical Power & Energy Systems, 136,* 107720. https://doi.org/10.1016/j.ijepes.2021.107720