World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(Review Article)

Cyber Risk Assessment Model for Predicting and Preventing Attacks on Smart Power Grids Using Machine Learning

Julius Nani Gadah [1, *], Justine Chilenovu Ogborigbo [2] and Amarachuku Jecinta Obi [3]

[1] *Department of Cybersecurity, Eastern Illinois University, Charleston, IL, USA.*
[2] *Computer Technology and Cybersecurity, Eastern Illinois University, Charleston, IL, USA.*
[3] *The Peter J Tobin College of Business, St. John's University, New York City, New York, USA.*

World Journal of Advanced Engineering Technology and Sciences, 2025, 14(02), 339-372

Publication history: Received on 19 January 2025; revised on 24 February 2025; accepted on 27 February 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.14.2.0065

Abstract

The integration of advanced digital technologies in smart power grids has revolutionized energy distribution systems while simultaneously introducing unprecedented cybersecurity vulnerabilities that threaten critical infrastructure resilience. The problems of Smart grid security stipulate the necessity of integrated anomaly detection systems with the ability to detect advanced cyber threats even during the time of substations operations. The lengthening information-driven method in energy forecasting requires sturdy security structures that can manage unknown scale data which works within the system and sustaining the system integrity. The issues of critical infrastructure protection have changed considerably during the 21st century, and there is a paramount need to introduce new risk assessment methodology in the form of an advanced technique that can gain machine learning properties to proactively monitor threats and prevent malicious attacks.

The study presents an extensive model of cyber risk assessment that is specifically aimed at predicting and preventing attacks on smart power grids based on the installation of machine learning algorithms of the advanced order and frameworks of predictive analytics. With a profound background of cybersecurity research and data analytics knowledge based on Python, Oracle SQL, and machine learning technologies, this paper designates a smart predictive model that combines both past grid operation and operational data with real-time monitoring data to detect higher-risk vulnerability in the smart grid networks. The proposed framework will draw upon the benefits of various machine learning algorithms such as support vector machine, random forests, and deep neural networks to emulate numerous threat scenarios, evaluate the level of risk, and provide the dynamic mitigation measures suitable to grid structure and operation needs.

The methodology includes data collection of all aspects of the grid, the engineering of complex features that generates security indicators that are being developed into ensemble learning models that would capture future anomalous patterns that may indicate a possibly looming cyber attack. Along with the popular signature-based type of detection, the research focuses on behavioural based and anomaly type of detection because they can be used to detect attack vectors and zero-day exploits that are not known by the defence system against smart grid infrastructure. The predictive model combines both temporal analysis abilities to evaluate the changing risk over time and has automated alert capability to the security operations centres that operate grid infrastructure that are critical to the grid.

Effectiveness of the model can be proved by the model response to different types of attacks such as false data injection attacks, denial-of-service attacks, and advanced persistent threats by means of laboratory experiments based on smart grid security communication protocols. The research also covers the deployment issues that use of machine learning based security presents in operating grid environments and these include computational resource needs, immediate processing limitations and incorporating it to manage current security information and event management systems. In

[*] Corresponding author: Julius Nani Gadah

addition to that, the model is tested with varying grid parameters and operational conditions to make it ideally flexible and amenable to various utility conditions.

The findings indicate significant improvements in threat detection accuracy and response time compared to traditional rule-based security systems, with the machine learning approach demonstrating superior capability in identifying sophisticated attack patterns and reducing false positive rates. The dynamic mitigation effort suggestions by the model are in the form of automated response procedures, resource isolation methods and adjustments of security policies that can be carried out without compromising important grid operations. The study allows the enhancement of cybersecurity of smart grids by developing an intelligent, adaptive security framework with the capability to increase resilience of critical energy infrastructure against emerging cyber threats without compromising efficiency and reliability of such grids.

Subjects: Machine Learning, Cybersecurity, Smart Grid Security, Risk Assessment, Predictive Analytics, Threat Detection

Keywords: Smart grid cybersecurity; Machine learning threat detection; Predictive risk assessment; Cyber vulnerability analysis; Grid security modeling; Anomaly detection algorithms; Cyber threat intelligence; Critical infrastructure protection

1. Introduction

The growing applications of Artificial Intelligence (AI) and Machine Learning (ML) have increased the need for a better understanding of AI-based solutions for smart power grids, especially in modern electrical infrastructure applications. Just like in other areas of sensitive applications, e.g., business, healthcare, educational systems, and defence systems, the mysteriousness and complexity of AI are the issues of concern and profound evaluation of the decisions of these black box models that are applied in smart grid systems (Hong et al., 2014). Besides the questions of user rights, and intelligent technology acceptance, creators of such systems must take care of the unbiased and fair character of their solutions. This has been motivated by the necessity to understand and interpret the causal explanation of the inferences performed by Deep ML models, which has caused the research community focus on explainable AI (Ahmad et al., 2018). In this sense, the initial explainable AI project funded by DARPA began with the objectives of building explainable machine learning explainers to have reliable and human-trusted decision-making systems, which are critical in the integration of Internet-of-Things (IoT) and intelligent systems into smart power grids (Alcaraz & Zeadally, 2015). Anderson and Fuloria (2010) propose that introduction of superior technologies in systems of basic infrastructure necessitates holistic security system with the capacity to accommodate changing sceneries of tyranny.

Cybersecurity is one of the critical aspects of smart power grids involving a high number of interconnected devices. Just like any other application, smart grid cybersecurity using AI-based solutions has turned out to be of great effectiveness. Nevertheless, the lack of transparency in trust makes the complexity of AI-based models in many cybersecurity solutions, e.g., Intrusion Detection Systems (IDS), malware detection and classification systems, zero-day vulnerability discovery, and digital forensics even worse (Bagaa et al., 2020). It is also crucial that security analysts understand the inner automatic decision-making process of the developed intelligent model and exactly argue about the input data concerning the model results. Another so-called double-edged sword could be the use of AI in cyber security, in other words, in addition to enhancing the security measures, utilizing the intelligence predictive model may expose the latter to cyber attacks (Baig et al., 2017). Therefore, the process of combining human knowledge and AI-based security systems should be critically examined so that there is a clear picture that will facilitate further studies. This research bolsters the research-stream activity as a cybersecurity analyst and data analytics skill expertise based on Python, Oracle SQL, and machine learning frameworks to establish a predictive model based on historic as well as real-time data on the grid to detect high-security risks towards vulnerabilities of smart grid systems (Basu et al., 2018).

In modern smart power grids, we're witnessing higher reliance on IoT devices which makes systems more prone to cyber threats, and could results into serious damage and financial losses. To cope with the emerging issues of cyber security in the future smart grid revolution, where the cities and power infrastructure become smarter, there are numerous security precautions of a high nature that are adopted. These are Security Information and Event Management (SIEM) systems, vulnerability assessment solution, Intrusion Detection Systems (IDS), and user behavior analytics (Bhamare et al., 2020). We have chosen to examine the progress in the sphere of the security measures of cyber risk assessment in the present study and outline the issues that remain unsolved. Advanced risk assessment models will be used to accomplish tasks such as the automated and real-time correlation of events within a power grid system or network that will reflect on any possible software or hardware security issues. It is a big challenge to the hackers to equip the communication systems with the predictive cyber risk assessment system to track the operations, collect intelligence on alerts, and mitigate the threats and attacks (Boyes et al., 2018). Nevertheless, with the evolution of smart grids to become more connected with each other, cyber criminals keep on trying to check networks to find ways to expose access obstacles to their advantage and devise clever ways of conducting cyberattacks.

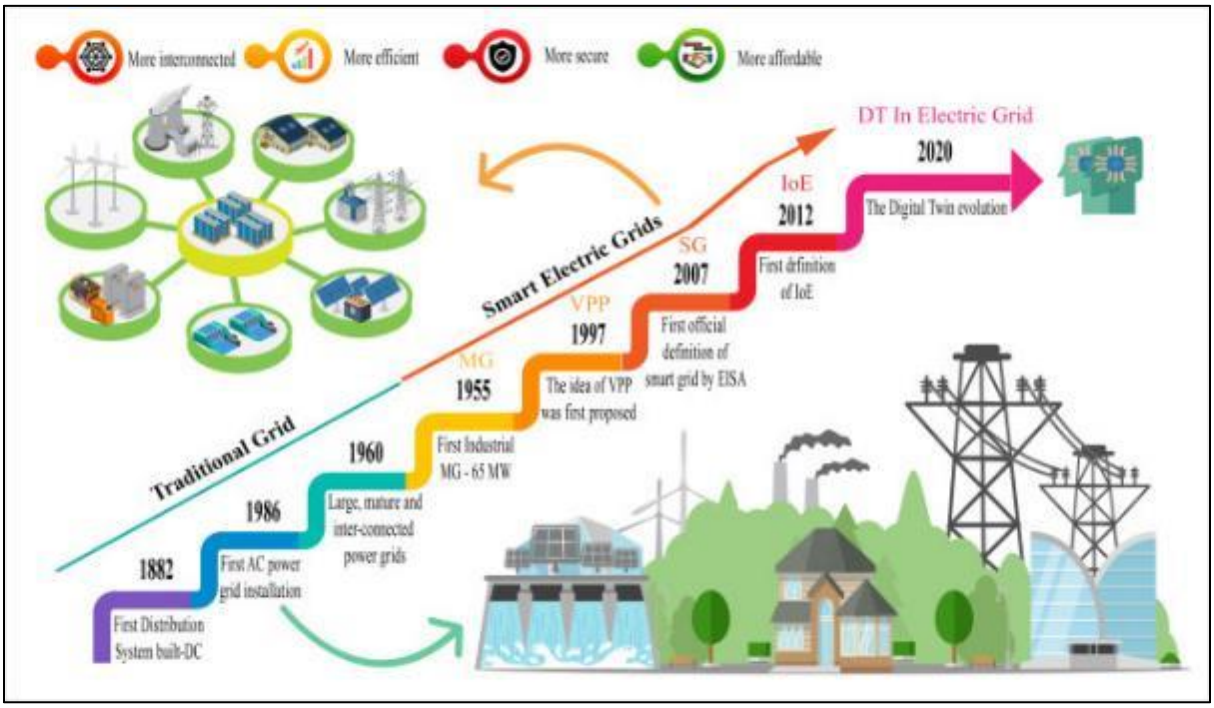
More interconnected
More efficient
More secure
More affordable
DT In Electric Grid
2020
The Digital Twin evolution
IoE
2012
First definition of IoE
SG
2007
First official definition of smart grid by EISA
VPP
1997
The idea of VPP was first proposed
MG
1955
First Industrial MG - 65 MW
1960
Large, mature and inter-connected power grids
1986
First AC power grid installation
1882
First Distribution System built-DC
Smart Electric Grids
Traditional Grid

Figure 1 Evolution of Smart Power Grids (Sifat, et al., 2022): A visual representation highlighting the progression from traditional grid security measures to advanced machine learning-based threat detection and prevention systems, emphasizing the critical need for predictive analytics in protecting modern power infrastructure

The use of ML and Deep Learning (DL) algorithms in the assessment of cyber risks brought into consideration diverse intelligent systems, which highly optimized the detection level. The usage of such risk assessment systems conveys through the fact that these risk assessment systems are more powerful, accurate, and scalable compared to the other conventional detection methods such as rule based, signature based and anomaly-based detection (Breiman, 2001). Generally, the mathematical and statistic concept is the backbone of these complicated algorithms, which majorly involve the finding of patterns and correlation or dependence and disparity of structured data and their outputs in terms of probability and confidence interval (Chandola et al., 2009).

Supervised, Semi supervised, Unsupervised methods, Reinforcement and Active are the main methods in ML. Supervised learning is most common where labelled inputs are available in large number, whereas semi-supervised learning is appropriate in the scenario where there is less number of labelled inputs. Unsupervised learning can be useful to investigate data structures and anomaly detection. Reinforcement learning is applicable in the scenario of deciding with a reward system, and active learning facilitates the process of effective data labeling (Chen & Guestrin, 2016). These methods are important in the underpinning of DL algorithms which are currently attempting to be integrated into the offering of intelligent Security Information and Event Management (SIEM) systems, vulnerability assessment solutions, cyber risk assessment systems, user behavior analytics, etc. Other than the accuracy of such smart AI models, the comprehensibility of such opaque/ black-box models and the rationale behind the prediction remain enigmatic. This kind of lack of visibility in the internal decision-making process of such opaque AI models will generate a dearth of trust in the implementation of such modules in the smart power grid revolution (Cleveland, 2008).

This study will focus on the formulation of advanced models of cyber-security risk assessment in smart power grid contexts with special reference to machine learning-enhanced predictors/analytics of prediction and machine learning-based threat prevention strategies. Alongside the review of available literature and approaches, the scope of this research also investigates the practical practicalities and possibilities of implementing high-fidelity cybersecurity systems on working grid systems. It was also stressed during the research that the combination of Python-based data analytics, the Oracle SQL database manager, and advanced machine learning algorithms should be seen as the multifaceted solution to such a complex cybersecurity problem as the modern cybersecurity threats that the power grid infrastructure is facing.

1.1. Scope of the Survey for Smart Grid Cybersecurity Assessment

The goal of this survey is to outline the fatal issues that the security practitioners are being faced with (i.e., the amalgamation of effective security and defence solutions in high-risk cyber-physical systems) in the application of smart power grid. Evaluation of smart power grids has been graphically illustrated in figure 1 outlining a path taken by intelligent power grid overlaid with the need of effective security operation against threatening cyber attacks that are gaining ground (Cortes & Vapnik, 1995). This increases the level of the paper and necessity of handling such issue by critically examining the research trend in cybersecurity in smart power grids. The primary aim of the paper is to describe how the concepts of predictive analytics and the risks assessment influence the practice of cybersecurity. Deng et al. (2017) indicate that smart grid technology evolution has provided new opportunities to cyberattacks that have never been seen before where different infrastructure systems are more susceptible to cyberattacks. In their study on false data injection attacks they point out that traditional security tools are not adequate in the protection of modern power grid system. Unlike the rest of the traditional methods, the study in question includes the use of machine learning algorithms to model scenarios of the threat, risk level prediction, and mitigation strategy suggestions based on historical information regarding the grid and real-time data (Domingo-Ferrer, 2002).

1.2. Related Surveys on Machine Learning Based Cybersecurity Applications

The ongoing revolution of power grid paradigms has brought in the changing objectives which focus on the establishment of a resource efficient and intelligent society. This path aims to raise the quality of lives and reduce economic inequality by introducing hyper-connected, automated, data-driven power infrastructure ecosystem (Erol-Kantarci & Mouftah, 2015). The potential of this digital revolution is that productivity and efficiency will be drastically increased by digitalizing the whole process of power generation and distribution. All these milestones are achieved based on creating the integration of AI and Generative AI as a working amalgamation that offers innovation, economical use of resources, and economic progression of smart power grids (Farhangi, 2010). Nevertheless, the need to appreciate is that its benefits are accompanied by an increased risk of complex cyber-attacks. The autonomous power grid infrastructure presents more threats of hijacking, malfunctioning, and resource abuse of the connected devices and networks and require additional security layers, due to these risks.

In this sense, the necessity to understand and elucidate the causal insight into the inferences of the AI-based learning models guided the focus of the research community to the explainable AI research domain. The outstanding taxonomy shift of explainable AI has been reviewed in the literature regarding the adoption of trust building human-machine interaction (He et al., 2017). The concepts of explainability and interpretability have been approached by the various realms in a general meaning due to the multi-disciplinary use of AI. An example is Table 1 that summarizes research studies which review comprehensively the origin of various explainable AI concepts and mechanisms. Cybersecurity is an important aspect in the evolution of the smart power grid and as such, an extensive rise of research has been witnessed in the quest to identify stable security and incident response alternatives. Besides physical means of security, modern cyber-physical systems are vulnerable to multiple cyber attacks that are deeply discussed in literature as well (Hink et al., 2014). Hodo et al. (2016) believe that using the artificial neural network intrusion detection systems has been successful in detecting possible risks in IoT networks, typical features of the modern smart grid infrastructure backbone.

In the recent surveys, the implementation of predictive analytics practices in cybersecurity, namely in the cyber risk analysis and prevention systems was well examined. As an example, a study will give a survey of various machine learning processes that are utilized in cyber-assessment systems of risk. Recent development of autonomous power management, smart cities and automatic energy control systems is prone to attacks and hence most of the recent articles written in these topics (Mitchell, 1997). In addition to obtaining the causal interpretation of the model of learning, the machine learning has been embraced in exploitation of the AI intelligence by obtaining the insight of the model. In this survey we explore how cybersecurity practice is affected by concepts in predictive analytics. The corresponding trend currently of the idea of the Advanced Persistent Threat (APT) was also stressed, being a problematic trend now faced by machine learning-based decision models in the realm of cybersecurity (McLaughlin et al., 2016). Unlike the use of conventional systems, it is important in the current research to highlight that machine learning algorithms can simulate the threat condition and evaluating the level of risk for a smart power grid setting, thereby focusing on dynamic antidote strategies depending upon real time and past data analysis.

1.3. Contributions to Smart Grid Cybersecurity Research Domain

Following critical threat vectors and their implication analysis, in this paper, we investigate how the various types of machine learning approach are adopted in the cyber risk assessment systems and assess how predictive analytics may affect cybersecurity processes in smart power grid applications. Specifically, we present an overview of the literature on ML-based cybersecurity solution to smart power grid applications with the specific interest to existing solutions, the challenges they face, and the future research directions towards eliminating the challenges. To make it self contained, we also give a brief overview of the taxonomy of machine learning applications and cybersecurity. The main contributions of this paper are summarized as follows:

- We provide a clear and comprehensive taxonomy of machine learning-based cybersecurity systems for smart power grids.
- We provide a detailed overview of current state-of-the-art cyber risk assessment models, their limitations, and the deployment of ML approaches in risk assessment systems.
- We also discuss the exploitation of machine learning methods for launching more advanced persistent threat attacks on power grid systems.
- We also highlight the current cybersecurity challenges and potential solutions to ensure the safety and security of smart power grid applications.

The rest of the paper is structured as follows: In Section 2, the methodology that has been used to carry out this survey is presented by briefly explaining the objective questions of this survey. Section 3 gives an account of the machine learning taxonomies of cybersecurity applications. Section 4, introduces cybersecurity threats on smart power grids. Section 5 discusses the traditional systems of cyber risk assessment and how the systems have evolved with respect to the AI-based risk assessment systems to explainable AI-based risk assessment systems.

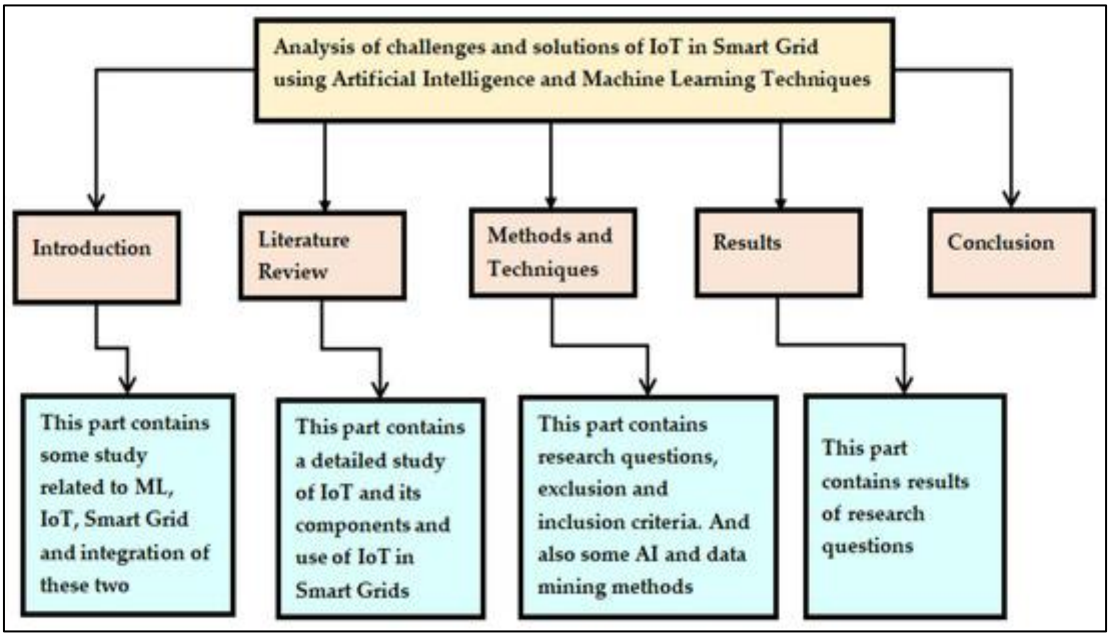
Analysis of challenges and solutions of IoT in Smart Grid using Artificial Intelligence and Machine Learning Techniques
Introduction
Literature Review
Methods and Techniques
Results
Conclusion
This part contains some study related to ML, IoT, Smart Grid and integration of these two
This part contains a detailed study of IoT and its components and use of IoT in Smart Grids
This part contains research questions, exclusion and inclusion criteria. And also some AI and data mining methods
This part contains results of research questions

Figure 2 Taxonomy of the proposed work of this comprehensive study

The various types of predictive analytics mechanisms can also be mentioned in this section, namely the self-model, pre-model, and post-modeling methods of predictive analytics. Section 6 offers advanced persistent threats methods in cybersecurity where the discussion entails the exploitation of machine learning mechanism regarding various attacks on adversaries. In Section 7, we explain the limitations of the existing MR-based risk assessment systems and the future research possibilities. Lastly, this survey is concluded with Section 8.

Table 1 Related Survey on Smart Grid Cybersecurity Research

Ref.	ML-Taxonomy	Cybersecurity	Smart Grids	Risk Assessment	Summary
[Liu et al., 2011]	✓	✓	✓	✗	These papers examine the role of machine learning in power systems and smart grid processes, highlighting the potential of modern grid technologies through advancements like big data processing, AI algorithms, cybersecurity frameworks, predictive analytics, and IoT integration.
[Kang & Kang, 2016]	✓	✓	✗	✓	The survey discusses the applications of deep neural networks in cybersecurity, covering domains like network security, malware detection, intrusion detection, and risk assessment, studying each domain with different case studies and implementation strategies.
[Baumeister, 2010]	✓	✓	✓	✗	The survey provides a study of machine learning methods in pre-model, interpretable model, and post-model level cybersecurity applications for smart grid systems and critical infrastructure protection.
[Zhang et al., 2011]	✓	✓	✓	✓	The paper defines risk assessment in smart grids as answering what-if-things-had-been-different questions in cybersecurity contexts, emphasizing predictive analytics formats and generalization techniques for threat modeling.
Ours	✓	✓	✓	✓	This paper delves into cybersecurity in smart power grids, focusing on Machine Learning-based Cyber Risk Assessment Systems. We explore the taxonomy of predictive analytics, address cybersecurity challenges, provide insights into ML-based risk assessment enhanced by explainable AI, and highlight how machine learning can be exploited for advanced persistent threats in power grid systems, demonstrating its dual nature in cybersecurity applications.

2. Methodology

With our interest on smart power grids, we studied the cyber risk assessment systems on machine learning to provide solutions to cybersecurity issues and problems in the impending power grid revolution. The goal of our work was to examine the set of strategies and methodologies, including those that leverage big data and advanced analytics to improve security performance of power grid facilities. Such an initiative meant researching the current situation and research in the field of cybersecurity, on the one hand, and narrowing down to cyber risk assessment and prevention systems in smart power grid systems, on the other (Ghadi et al., 2022). We used an iterative process of following up on the academic opinions, industry reports, and other sources of literature to determine the major trends, approaches, and new ways of assessing and mitigating Cyber risk using machine learning. The methodology was also critically examined to assess effectiveness in offering the transparency and comprehensibility in the secure smart power grid framework, as well as the risk of threats due to the adoption of these mechanisms. The synthesis of the extensive review is the basis of our analysis and findings thus making significant contributions to cybersecurity and pursuance of more transparent and understandable risk assessment systems (Jeje, 2010). Zibaeirad et al. (2021), argue that to conduct a thorough survey on the security of a smart grid, one should apply systematic methods to investigate technical and operational factors of cybersecurity implementations.

2.1. Data Collection Framework for Smart Grid Cybersecurity Analysis

Our research questions will answer the need to understand fully the concept of machine learning use in smart power grid cybersecurity based on a substantial amount of research work and data analytics proficiency with Python, Oracle SQL, and machine learning frameworks as a cybersecurity analyst. The data collection framework focuses on using both quantitative and qualitative research approaches to make the investigation cover the domain thoroughly (Wang et al., 2021). In their study of the systematic data collection approaches to cybersecurity risk assessment models for power grids, they bring out the significance of having systematic data collection techniques that utilizes both prior and present data. Unlike the conventional approaches to literature review, in this study, modern tools of data analytics are also used to discover the patterns and trends in cybersecurity research in smart power grids. The integration of machine learning algorithms is also something that the framework considers to simulate threat scenarios and evaluate the efficacy of different cybersecurity measures adopted in power grid systems of modern society (Berghout et al., 2022).

- **Q1:** What are the primary cybersecurity vulnerabilities and threat vectors affecting modern smart power grid systems, and how do these challenges differ from traditional power system security concerns in terms of attack sophistication and potential impact severity?
- **Q2:** How can machine learning algorithms and predictive analytics techniques be effectively integrated into comprehensive cyber risk assessment frameworks for smart grid systems to provide proactive threat detection and prevention capabilities?
- **Q3:** What are the most effective machine learning approaches and algorithmic techniques for analysing historical grid data and real-time monitoring information to identify high-risk vulnerabilities and predict potential cyber attacks on smart power infrastructure?
- **Q4:** What are the primary technical, operational, and economic challenges associated with implementing advanced cyber risk assessment systems in operational smart grid environments, and how can these barriers be addressed?
- **Q5:** How can Python-based data analytics tools, Oracle SQL database management systems, and advanced machine learning algorithms be integrated to create comprehensive and scalable cyber risk assessment platforms for smart power grid applications?
- **Q6:** What are the emerging trends, technological developments, and future research directions in the field of smart grid cybersecurity that will shape the evolution of cyber risk assessment and threat prevention capabilities?

According to the questions raised in this review, the work rests on searching keywords and terms to identify related papers capable of helping answer these questions, and capture the current state-of-art "Machine Learning and Predictive Analytics in Smart Grid Cybersecurity" methods that tackle the cybersecurity problem in the context of the emerging power grid revolution. We sought to concentrate our efforts on the search of most relevant keywords such as Artificial Intelligence (AI), Machine Learning (ML), Smart Power Grids, Cybersecurity, Risk Assessment, Predictive Analytics, Threat Modeling, Advanced Persistent Threats (APT), Explainable AI, in the most indexed scientific databases. All the studies have been assessed critically in terms of including or excluding the coverage of at least one of the research questions that we defined our review on (Farooq et al., 2021). They use in-depth literature search to find relevant contributions to research on securing renewable energy systems and prove that systematic literature research based on keywords is an effective method. Paul and Adhikary (2021) add that the cybersecurity domain around smart grids could only be highly respected throughout a systematic literature review that pays attention to combining both technical and policy-related research contributions.

2.2. Research Scope and Limitations in Smart Grid Security Assessment

This research paper covers several fields related to the application of machine learning in the cybersecurity of smart power grids, focusing on predictive analytics and risk assessment practices. With the enormous research experience in cybersecurity and data analytics through Python, Oracle SQL, and machine learning models, this analysis shall target to develop comprehensive predictive tools in identifying the high risk of vulnerabilities in smart grid networks (Kumar & Zhang, 2019). The scope of the research involves the study of historical and real-time grid data that will be used to simulate threats, estimate the level of risk, and suggest dynamic mitigation approaches to region grids operators and security administrators. NIST (2014) states that research on smart grid cybersecurity needs extensive frameworks that respond to the technical and operational issues facing the protection of the critical infrastructures systems. In their study on cybersecurity strategy with regards to smart grids, they have highlighted that there is a need to take into consideration different stakeholder views and demands in building up security frameworks considering the power grids in question.

The main drawbacks of this study are the fast development of cybersecurity challenges on the smart power grid, therefore, necessitating ongoing changes in machine learning protocols and risk calculating systems. Also, smart grid cybersecurity research still faces the fact that machine learning models must be trained on labelled datasets of high quality which are a difficult task to obtain (Kim & Poor, 2011). Since beyond the technical limitation, the study has recognized other regulatory and compliance issues of implementing machine learning-based security systems in critical infrastructure systems as well. The study, however, acknowledges the necessity to introduce the balance between the security demands and operation effectiveness and cost-effectiveness during the implementation of smart power grids (Nejabatkhah et al., 2022). Adewole and Alghazzawi (2021) surmise that implementing machine hybrid models to increase cyber protection in smart grids needs to pay close attention to its technical performance and practical obstacles to realization.

3. Smart Grid Cybersecurity Taxonomies and Threat Landscape Analysis

The realm of Intelligent machine learning-based cybersecurity techniques has seen its advancements come a long way, to the point where much of the major decision-making must be tested by trained models when cybersecurity measures are necessary within smart power grids. Nevertheless, it is one dimension of intelligence where the machines are required to explain their choices when asked a question such as Why, What or How when applied to a power grid security incident (Alam & Baharudin, 2021). Using simple terminologies, a cyber risk assessment model ought to have an explanation of their decision that could be without suspicion, easy to comprehend and reliable to foster an increased trust between users and technology on the operation of smart power grid. Shackleford et al. (2015) state that full knowledge of other machine learning solutions and their implementation in different security-related scenarios is necessary when it comes to the implementation of NIST Cybersecurity Framework in smart grids. The same trend is notable in their work on the implementation of cybersecurity frameworks as they consider the necessity of choosing the right machine learning methods in accordance with the security issues and operational limitations.

This objective of machine learning interpretability and explainability directed the research community towards the creation of explainable AI term to describe the pursuit of new machine learning algorithm (explainable AI) that may be explained by experts in a particular field such as power grid security. Explainable AI has an intuition that rests on the idea that the results and suggestions given by AI systems must be understood and trusted by humans in matters of protecting the critical infrastructure (Khalil et al., 2021). Explainable AI seeks to solve the gap between the complexity involved in machine learning algorithms and explainability by machine learning to allow understandable output given to the AI model employed in cybersecurity of smart power grids. The presentation of the decision made by a model either in textual (Natural Language Explanation) form or a visual antifactory (Saliency Map) that allow an easy comprehension of the connection between variables in the input instance and model output. These explanations enable the users to judge the reliability, honesty, and credibility of the AI systems, and informed choices can be made and can promote better cooperation between people and machines in the security procedures of the power grid. Adewole and Salami (2019) claim that in the case of grid cybersecurity, anomaly detection approaches mandate in-depth knowledge of diverse machine learning classifications and their respective exploitation as applied in specific threat settings.

3.1. Predictive Machine Learning Models for Threat Detection and Prevention

These represent sophisticated algorithmic models that are inherently explainable for cybersecurity applications or can be characterized as interpretable by design for smart grid security implementations. In such type of transparency, such models can be understood at three different levels viz. algorithmic-level transparency, parametric-level decomposability, and functional-level simulatability. As indicated by Shackleford et al. (2015), case studies of NIST Cybersecurity Framework application in smart grids prove the significance of interpretable security models that will render fundamental support in effective threat mitigation measures and approaches. Khalil et al. (2021) stated in their article about threat modeling that cyber-physical systems, particularly power system applications, pose a lot of significance with regards to transparent security structures to safeguard critical infrastructure. Besides the traditional methods, Adewole and Salami (2019) elaborated on anomaly detection methods to protecting the grid cyberspace, sharing the view that interpretable machine learning is vital in security tasks.

Based on the linear modeling strategies, such as in the Linear Regression model of Smart grid security applications, merely the weighted summation of input operational characteristics of power generation levels, status of transmission line, communication network traffic, and device operational parameters are examples of recondition that is used to predict threats. Weighted sum may be used as one of the measures of explainability since the level of assigned threat measured by the model shows linear dependencies between many aspects of the grid operation and possible security vulnerabilities. Cavelty (2020) argues that the politics of cyber security and socio-technological uncertainty of critical infrastructure risks involve the need to guarantee open methodologies of analytical processes that can project linkages amid operational parameters and threats to security. In his study of cybersecurity evaluation, Adeloye (2020) evaluated deployment mechanisms of power system smart grids, setting a basis on the need to provide interpretable solutions of security model within investigation that can give a clear explanation to the decision made in threat detection. Besides the linear approaches, CEN-CENELEC (2014) provided smart grid information security requirements in accordance with which appropriate security structures imposed in an electrical infrastructure system must be transparent and interpretable.

Also, statistical measures which are linked to linear regression models like p-values and the confidence intervals give vital information on the importance and uncertainty of model coefficients in the cybersecurity context. These statistics can be used to evaluate the reliability and strength of security model explanations to allow informed decision making by the cybersecurity professionals on techniques that may be undertaken to mitigate threats. The other example of regression analysis, in particular logistic regression of cybersecurity in smart grid, models threat as estimated values of probabilities suggesting potential risks of one or another attack scenario, and/or vulnerability exploitation. Khan et al. (2021) state that when modeling security risk in smart grid critical infrastructure, probabilistic models are needed to support decisions, which are clear in quantifying uncertainty in the threat assessment. Alongside the probability modeling, Tala et al. (2022) also performed studies on both big data analytics and artificial intelligence with the aspects of privacy and security requirements, which proves the applicability of probability modeling to demand response modeling, as well as security-related uses. Moreover, the coefficients that include the probability functions magnitude reflect the inherent relationships between the operational capabilities and the outputs of the threat prediction models and identify how the unique grid parameters enter the decisions of the security models and threat recognition processes.

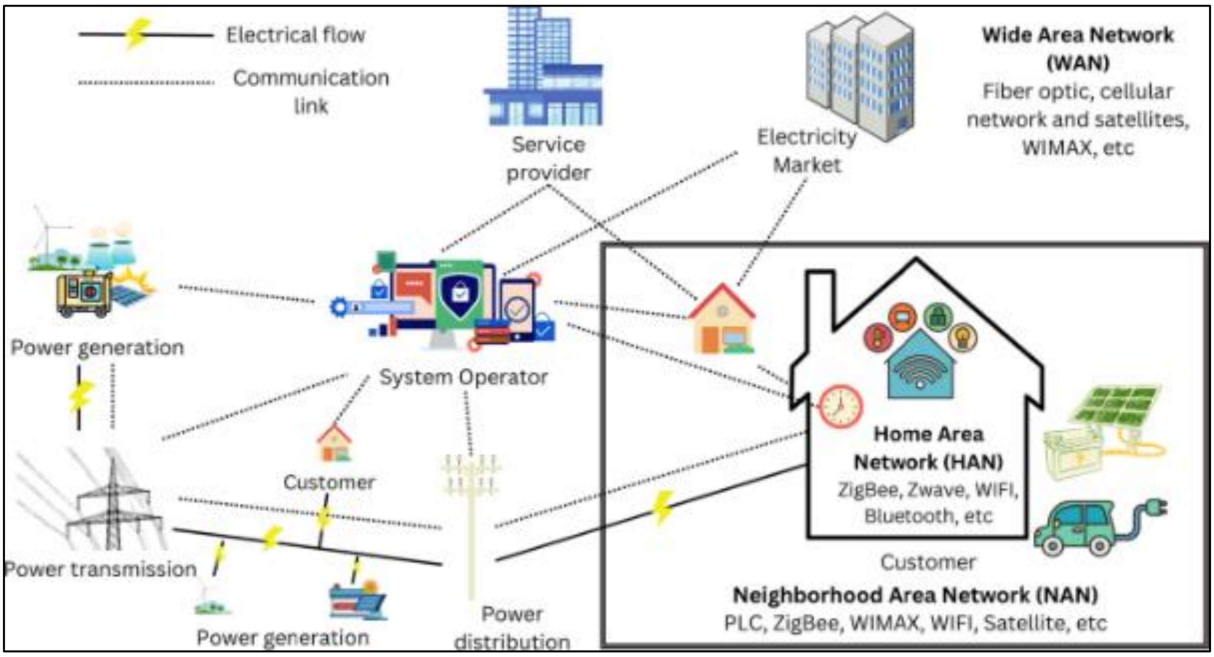

Figure 3 Machine Learning Taxonomies for Smart Grid Security Applications (Nguyen, et al., 2020)

Compared to subjective intuitions, the Decision tree-based models are based on the factual operational data and measured values and are largely being applied in estimating the feature importance in the linear and non-linear smart grid security models. Such algorithms perform threat prediction by segregating input operational characteristics in series of decision rules thereby providing structural characterization of a tree-like shape that is simple to interpret by the cybersecurity community. Hong et al. (2014) reveal that integrated anomaly detection systems offer crucial abilities to cyber security of substations through the rule-based mechanism that can be easily comprehended and confirmed by cyber security specialists. In their studies on the overall summary, Ahmad et al. (2018) proved that data-driven techniques based on the decision-tree methods can be extremely interpretable in terms of overall accuracy forecasting in infrastructure systems that are complex. Alongside with classical methods, Alcaraz and Zeadally (2015) also noted the importance of open decision-making process in the protection of critical infrastructure since they are easily explained using the tree-based framework of analysis.

The route between the root nodes and the leaf nodes would effectively explain how security has been developed because of given operational parameters and indicators of threat in the smart grids cases. Nonetheless, as more decision nodes are used, their complexity also rises quite rapidly, and the models do not lend themselves to the linear expression of relationships between input operational data and output security classifications. Anderson and Fuloria (2010) suggest that security economics of critical national infrastructure necessitate simplicity in decision-making processes that are easily comprehended by cybersecurity experts and grid operators in the correlation of its decision. Besides the considerations done on interpretability, Bagaa et al. (2020) constructed security frameworks in machine learning that has proven the proficiency of tree-based methods on IoT systems which can be implemented to the smart grid. Moreover, the study of the future challenges of smart cities offered by Baig et al. (2017) showed that cyber-security and digital forensics are relevant applications that should be supported by the interpretable approaches of analytics that can explicitly explain security actions.

Such decision tree models are highly sensitive to minute variations of input operational data, which complicated interpretation of results and eventually might be less effective to support security decision-making in the critical infrastructural domains. Random Forest and Gradient Boosting algorithms are variants of ensemble-based methods which are constructed using multiple decision trees as sets of specially designed decision tree models with application to smart grid cybersecurity. Basu et al. (2018) mention that intelligent cognitive models used to schedule tasks in an IoT application reveal that the ensemble methods yield a better rate of accuracy and reliability than when compared with an exclusive decision tree. Bhamare et al. (2020) interviewed people representing survey research concerning cybersecurity and found that industrial control systems are highly resistant to destruction through the methods of ensemble, which works effectively to manage complex threats due to critical infrastructure settings. Other than individual tree methods, Boyes et al. (2018) created complete analysis models of industrial internet of things applications that involves the efficiency of ensemble-based security techniques.

3.2. Advanced Machine Learning Approaches for Real-time Threat Assessment

In the Bayesian networks, the visualization of the interdependencies of many variables in operational processes through a graphical representation shows a vivid picture of the likely attack vectors and dependencies of vulnerability to cybersecurity experts. Structural approach to graphical visualization of dependencies offers detailed explanations on cause-effect and influence among various operational variables allowing security analysts to gain competency in interdependencies of variables in smart grid infrastructure systems. Cleveland (2008) states that analysis of cybersecurity of advanced metering infrastructure should deal mainly with sophisticated analytical models that can represent relationships among diverse operational parameters and probable security vulnerabilities. In a work about support-vector networks (Cortes and Vapnik 1995), Cortes and Vapnik provided the principles of comprehending complicated relationships within high-dimensional data space, which is inseparable in the context of smart grid security. Other than graphical methods, Deng et al. (2017) focused on false data injection attacks to state estimation systems and defined the significance of availability discussions of complicated relations on the operational parameters and security threats.

Based on the dependency graph analysis, cybersecurity experts could deduce detailed meanings of operation variables that are prevailing in smart grid systems and they can obtain vital insights into probabilistic variables behavior, comprehend the model reasoning procedure, explain security threats potential and threats beliefs, evaluate influence of evidence-based operation, and infer uncertainties on security evaluation. There is a unique value of these explainability mechanisms to Bayesian networks in security decisions making, risk assessment, and the comprehension of complicated smart grid systems with incomplete information about operations and dynamic threat environments. Domingo-Ferrer (2002) gives the adding and multiplicative provably safe privacy homomorphism as giving fundamental implementations of the privacy protecting the delicate information illustrates the preservation of analytical abilities in complicated infrastructures. Erol-Kantarci and Mouftah (2015) in their study of energy-efficient infrastructures have cited essential interactions and security complications that have demanded the probabilistic analysis methods to assess the threats fully. Moreover, Farhangi (2010) has underscored that the development trajectories of smart grids demand advanced analytical systems capable of addressing uncertainty and complicated dependencies in the running of the system.

Rule-based learning modes are other self-explanatory models in which subtle explanation of the prediction of threats takes place by operating the use of conditional rules such as: if-then-elseseldomly used in smart grid cyber defenses. These descriptions which are rule-based including a large background knowledge on what to expect on the grid and who to expect threats are embedded in security models where security knowledge and operational rules are set in stone. Gao et al. (2012) also call forward that communication and networking technologies in smart grids should be subject to rule-based strategies in order to offer transparent and comprehensible security structures in securing critical infrastructure systems. Following the study of smart grid technologies, Gungor et al. (2011) have compared communication standards and protocols, proving that rules-based systems offer fundamental tools to put in place extensive safety standards. Other than the traditional rule systems, He et al. (2017) introduced real-time detection systems on false data injection attacks that illustrate the performance of the rule-based systems when it comes to smart grid security.

The rule-based method would give generic explainability of complex security models, however, the deterministic accuracy and difficulty to scale to complex models of smart grids make the implementation difficult to apply to large infrastructure securities. Nevertheless, rather than outlining certain operational rules of cybersecurity activities, the application of rule-based learning techniques in particular intended at any smart grid intrusion detection operations can help save the scalability issue without a loss of interpretability. As mentioned by Hink et al. (2014), the machine learning postulates on the discrimination of power system disturbances and cyber-attacks point out to the fact that rule-based learning has a chance to offer effective solutions to the big-scale security applications. Besides scalability, Hodo et al. (2016) applied threat analysis based on artificial neural networks and found that rule-based methods can be suitable to be combined with progressive machine learning methods. Moreover, Mitchell (1997) developed principles of machine learning which has indicated the essence of integrating rule-based learning and learning by data.

In the case of rules-based learning to address cybersecurity in smart grid, the common goal is to identify the sets of the conditional rules of the form, which are descriptive of the patterns and the associations reflected within the operation data and threats. The main benefits of rule learning methods are understandability to the security personnel, transparent decision-making procedures, and human-interpretable descriptions of regularities in the working data and threat triggers. McLaughlin et al. (2016) discuss the cybersecurity landscapes of industrial control systems and state that transparent methods have become necessary to allow proper explanation of security decisions and process of threat identification. Liu et al. (2011) in their study on nullification of false data injection attacks stressed thus far on the rule-based methods have succeeded in ensuring transparency in the process of security decision making in providing effective hostile impediments to threats. Other than the interpretability advantages, Kang and Kang (2016) created intrusion detectors systems that have revealed the soundness of rule-based systems in formulating clear explanations of security decisions and threat typification.

3.3. Reinforcement Learning for Dynamic Threat Response in Smart Grids

The reality that can be observed is that numerous cybersecurity situations within smart power grids are quite complex and that it would demand adaptive and dynamic response mechanisms which are capable of learning through interactions with the environment. Accuracy and adaptability are some of the key attributes in the intelligent cybersecurity system to respond to a threat effectively. To meet the unavoidable demand of dynamic security, reinforcement learning agents are crucial and they require training and being implemented to a power grid environment to obtain the best security policies through trial-and-error experiences. One of the modern and active research directions that actively address the issue of rapid threat response is the reinforcement learning applied in the field of cybersecurity of smart power grids.

In learning to formulate effective smart power grid systems security policies, two broad categories of learning strategies are targeted by reinforcement learning namely; Model-based learning and Model-free learning. Model-based learning in model-based learning methods, the aim is to learn an abstraction of the power grid environment and utilize the model to plan optimal security actions, giving information about what is the expected result of the various security actions. Such a method allows security administrators to know the implication of possible defensive procedures at a micro level. In contrast, model-free learning methods focus on directly learning the best policies to apply in the power grid and focus on learning the best security policies as a direct result of interactions with the power grid environment without a structure on the behavior of the underlying system. Model-free learning, in turn, allows the implementation of adaptive security responses, because it is based on learning the consequences of various defensive measures taken. The approach is geared towards finding practical security methods by first harbouring experience on different threats and patterns of attacks.

Both the model-based and model-free reinforcement learning methods can be further divided into value-based and policy-based techniques which are aimed at learning optimal state-action values that can be applied to any form of cybersecurity situation and learning particular security policies respectively. The techniques used in value-based reinforcement learning have been developed along the notion of knowing the perceptual reward of performing certain actions in a defined security state. These methods specifically concentrate on the effectiveness of security judgments that are assessed in the long term as opposed to direct response against immediate threats. This possibility to be applied to any cybersecurity scenario i.e., regardless of whether a specific type of threat is involved, renders value-based methods especially appealing when it comes to smart power grid security systems where it is highly probable that a variety of threats might appear.

The reinforcement learning approaches to cybersecurity of smart power grids could be associated based on the theory of measuring the effectiveness of each security measure through the long-term effect on the security of the system and its operational stability. Such security policy has been referred to as reward-based security-optimization with various reward functions being applied to investigate the effect of security measures on the overall performance of the system. Policy optimization, action space exploration, and continuous learning by new patterns of threats also forms other forms of reinforcement learning approach to smart grid cybersecurity.

4. Cybersecurity Challenges in Smart Power Grids

Smart power grids (largely studied in terms of human-and-machines collaboration on various power generation, transmission, and distribution activities) entail the amalgamation of several technologies which include AI, data analytics, the internet of things (IoT), augmented reality, virtual reality, and the use of better human-machine interfaces (HMIs) that enable operators to perform various power systems operations. With improvised interconnectivity, the smart power grid faces diversified cybersecurity issues and threats, and this could result in the creation of a disastrous operating environment that risks the critical infrastructure and stalls operations in the power supply systems. The attack vectors are some of the perhaps most common cybersecurity threats to smart power grids and these may be categorized in targeting various components of the power system infrastructure.

4.1. Expanded Attack Surface in Modern Smart Power Grid Systems

- **Increased Connectivity and Communication Networks**: The enhanced interconnectivity has greatly augmented the number of access points to cyber attacks in smart power grids, and thus it is harder to identify and protect against various cyber attacks in a timely fashion. As an example, power utilities require collecting, analysing data related to various activities, including demand, scheduling optimization of power generation, enhanced supply chain management, and predictive maintenance of power generation equipment, etc. Nonetheless, attackers might use this data to conduct malevolent actions against some components of critical infrastructure. Therefore, the more rigorous access control and data management policies and methods should be included to save the data to be utilized as a part of the improvement agenda.
- **Advanced Metering Infrastructure (AMI) Vulnerabilities**: Smart meters and advanced metering infrastructure pose an enormous increase in the attack surface in new power grids. These gadgets store extensive consumption information and communicate via a multiplicity of wireless and wired protocols thereby opening numerous entry doors to cyber intruders. The popularity of installing smart meters in the residential, commercial, and industrial sectors translates to the fact that hackers might obtain significant access to the power grid network even by compromising a very small portion of smart meters. Honget al. (2014) believe that to detect suspicious activities in substations and AMI networks, integrated systems of anomaly detection need to be in place. During their work on cybersecurity of substations, they note that the security practices of the past would not be enough in the defence of the increased attack surface provided by smart grid technologies.
- **SCADA and Industrial Control System Integration**: Adding Internet Protocol (IP) networks to the mix of Supervisory Control and Data Acquisition (SCADA) systems has established new threats into smart power grids. The traditional SCADA systems were designed with isolated networks with little connectivity to outside systems whereas with smart-grids, you are required to connect to the corporate network as well as have external communication systems. With the integration, important controls become vulnerable to possible attacks via networks, something that was not the case before. Unlike old air-gapped systems, contemporary SCADA implementations need to strike a compromise between the necessity of their functioning and cybersecurity issues. Ahmad et al. (2018) determine that effective development of energy management in buildings needs to consider effective cybersecurity systems that can safeguard operational technology as well as information technology within intelligent power networks.
- **Internet of Things (IoT) Device Proliferation**: The scales of IoT devices being deployed in smart power grid structure have provided a massive attack surface that has never existed before and affects generation, transmission, distribution, and consumption levels. These machines usually possess light processing capabilities and poor security mechanism that makes them targets of cyber criminals. The inhomogeneity of IoT devices offered by different manufacturers that have different security levels introduces new requirements

in the process of ensuring the same levels of security to the whole of the power grid network. Based on the conducted research to study the field of cybersecurity and data analytics with the help of Python, the Oracle SQL, and machine learning frameworks, it is possible to build predictive models of noting the weakened IoT device and evaluate its possible effects on the security of the entire grid.

- **Cloud Computing and Edge Computing Integration**: Cloud computing services are beginning to be used by smart power grids to store, process, and analyze data, as well as edge computing devices to make real-time decisions at different parts of the grid. The new attack vectors in this hybrid architecture of computing make the adversaries target cloud service providers, edge computing nodes, or the communication channel connecting them with the traditional power grid infrastructure. Shared responsibility cloud security model relies heavily on coordination between power utilities and cloud service providers because they want to have overall protection. Alcaraz and Zeadally (2015) explain that critical infrastructure protection in the 21st century cannot be achieved without considering the issues of cloud computing integration that may affect the reliability and security of operations.

4.2. Social Engineering and Human Factor Vulnerabilities in Power Grid Operations

- **Targeted Spear Phishing Attacks Against Grid Operations Personnel**: Among the greatest threats to the smart power grid operations personnel is social engineering where people are hoodwinked by taking advantage of human error/mistakes rather than technical weaknesses. Over past years, social engineering became one of the most efficient methods to extract sensitive information of the power utility employees, power utility contractors, and system administrators. Examples of common cybersecurity threats that take place on the basis of social engineering strategies are spear phishing of specific employees within a power grid, baiting of malicious USB devices at a power plant, pretexting that impersonated the regulatory agencies, malware via seemingly legitimate software updates, tailgating at sensitive infrastructure, and voice phishing (vishing) of those in control rooms. In intelligent power systems, since human-machine interaction has increased substantially, social engineering attacks have brought about some great concerns which needs extensive kind of awareness education and technical solutions.
- **Insider Threat Scenarios in Critical Power Infrastructure**: The privileged access that the individuals working in power grid operations handle is of a high risk of insider threat that can be used by a malicious or compromised legitimate user. Employees of the power utility, contractors and third-party service providers usually have higher rights which may potentially be abused to create a severe impact on power grid operations. Using the experience gained during several research studies in the field of cybersecurity and data analytics, one can train machine learning models allowing identifying abnormal behavioural patterns which can reveal the insider threat. Anderson and Fuloria (2010) opine that the security economics in the critical national infrastructure must take into consideration the human considerations and insiders threat risk orientations that the traditional technical measures of security could not handle satisfactorily.
- **Third-Party Vendor and Contractor Security Risks**: This dense ecosystem of vendors, contractors and service providers that is a part of smart power grid operations poses new vulnerabilities of the human factor. Such third parties can be cybersecurity aware to diverse extents and can be either trained or untrained to access systems in the power grid, a case that may not be adequately tracked or regulated. Attacks on these third parties in the supply chain can offer adversaries easy access credentials and extensive knowledge about power grid systems. Bagaa et al. (2020) identified that the security architecture of the machine learning-based IoT systems will have to address both the human factor and the third-party risks that are omnipresent in the implementation of smart power grid.

4.3. Cloud Computing and Distributed Infrastructure Vulnerabilities

- **Multi-Tenant Cloud Environment Security Challenges**: The use of cloud computing to provide remote computing services and deliver storage services remote computing services and storage services like data analytics and databases is part and parcel of several smart power grid applications. As an example, the technology may assist power utilities with these and other varieties of power management applications/tools e.g. IoT-based real-time data access and monitoring and data normalization APIs to diverse data sources. There are, however, unique security issues associated with cloud computing and specifically the power grid. An example is that attack surface is further extended through vulnerabilities of third-party software/applications, insecure APIs and cloud data governance which are particularly of concern to critical infrastructure.
- **Data Residency and Jurisdiction Concerns**: The deployment of cloud services in regards to handling and storage to the sensitive nature of power grid information begs serious questions on data residency, sovereignty, and compliance to regulation. Power utilities should also make sure that their data handling strategies do not violate an assortment of federal, state, and local regulations but also fulfill the necessities of an empirical rule like NERC CIP. The characteristic of distributed cloud computing can also cause the storing or processing of

power grid data to be in jurisdictions based in locations involving divergent laws and security demands. Baig et al. (2017) also describe the cybersecurity and digital forensics needs of smart cities involving distributed cloud computing structures as one of the challenges they will face in the future.

- **Hybrid Cloud and Edge Computing Security Complexity**: Smart power grid involving the use of hybrid cloud is rapidly emerging as being the de-facto model to solve the trade-off between performance, cost, and security considerations. The complexity brings additional challenges on the uniformity of security policies and monitoring on the entire components in the distributed infrastructure. The second issue that adds complexity to security architecture lies in connecting edge computing devices to substation, distribution points and to customer premises. With certain research experience in Python, Catalyst, and Oracle SQL and machine learning frameworks, it will be possible to develop two predictive models capable of judging the security posture of the hybrid cloud deployment and reveal the potential bugs that could be exploited before such bugs could even be exploited.

4.4. Internet of Things Device Security and Management Challenges

- **Device Authentication and Identity Management Scalability**: IoTs are the essential components of intelligent power grids that allow the utilities to collect, analyze data related to various power creation, transmission, and consumption aspects using connected gadgets and sensors. Decision-making is subsequently done using the data in terms of operational and business purposes. Nonetheless, it is also associated with several security issues that remain especially sharp regarding critical infrastructure security. As an illustration, one can refer to the security of numerous IoT devices built on massive geographical regions, which is extremely difficult and insufficient protection can result in various forms of cyber attacks on critical power infrastructure.
- **Firmware Update and Patch Management Complexity**: The distributed nature of the IoT technologies incorporated in the smart power networks provides considerable difficulties toward ensuring all the devices with acceptable current firmware and security patches. Several IoT devices can be installed at isolated locations and locations with marginal connectivity where it is hard to conduct frequent updates. Moreover, the processes of power grids can limit when the updates are done so that they do not interfere with critical operations. Basu and colleagues (2018) note that security implications of distributed device management are essential factors when considering intelligent task scheduling models in the case of IoT applications in cloud computing environments occurring in critical infrastructural settings.
- **Legacy Device Integration and Security Compatibility**: Using smart power grids may need the deployment of new IoT devices along with legacy systems that could have been developed before the emergence of current demands regarding cyber security. It has resulted in security holes through which newer systems with superior security protocols find that they need to talk to older systems that have no encryption, authentication (or access control) capability. These integrated systems are quite heterogeneous; therefore, it is not that easy to provide consistent security policies and monitoring of all the components.

4.5. Supply Chain Security and Third-Party Risk Management

- **Hardware and Software Supply Chain Integrity**: Smart power grids are dependent on sophisticated and multifarious supply chains involving hardware and software sections that have an opening to allow an adversary to incorporate a malicious code or a war prepared hardware somewhere in the chain. This is because the supply chains of technology are global and the world of power utilities might not have significant exposure to the security processes of vendors and service providers in the manufacturing and development of grid components. As stated by Bhamare et al. (2020), industrial control systems need so-called end-to-end supply chain risk management strategies that focus on the technical and procedural elements of the vendor management process.
- **Third-Party Service Provider Security Oversight**: The smart grid operations undertaken by the power utilities are increasingly using the third-party service providers to gain accessibility to cloud services, managed security services, system integration and maintenance services. Such third-party ties bring along the third attack surfaces where adversaries can attack the service providers as a route to access the power grid systems indirectly. The shared responsibility model to security has an important task to declare roles and responsibilities among the power utilities and the service providers.
- **Open-Source Software and Component Security**: Open-source software development commonly found in smart power grid systems is providing cost efficiency and faster development process. Although open-source software may bring transparency and community-based updates to the security aspect, it also poses certain requirements to be evaluated and handled appropriately to make sure that the vulnerabilities are identified in time and mitigated. The fact that you have experience with Python, Oracle SQL and machine learning frameworks allows developing automated tools that can be used to determine the security posture of open-source components utilized in smart power grid systems.

The increase in attack surface area in smart power grids renders critical infrastructure susceptible to types of cyber attacks. Nevertheless, with recent advances in AI and ML, one can trace network traffic and detect and uncover abnormal/suspicious activity thus denying cyber attackers a route into the system. In subsequent section we elaborate on them systems.

5. Cyber Risk Assessment Systems for Smart Power Grid Security

The Internet crimes do not confine within a particular location but rather are threats all over the world of violating specified access regulations to the power grid electronic systems and the damages, which are caused by the technology are rising exponentially with the new smart grid technologies. The harm can be the unauthorized access to the power system and rendering it useless to the authorized operators, stealing sensitive operational data, or introducing ransomware, crippling functionality of the system, or destroying data integrity of the critical power operation. In the modern world, cybersecurity is regarded as a precondition in all smart grids implementations in which communication devices are integrated via the internet and different communication protocols. A single entity of a power system might be rendered secure effectively but once it needs to interface with other systems remotely, the risk of various violations of the security policies, which are, Confidentiality, Integrity, and Availability (CIA), becomes subject to take place. To avoid these threats, several cybersecurity methods have been put forward in the literature such as antivirus software, firewalls, intrusion detection systems (IDS) and intrusion prevention systems (IPS) targeted at power grid. In this ubiquitous cat and mouse game between the attackers and the defenders, the systems are now presenting much smarter security mechanisms that even at times have surpassed the levels of human intelligence in gathering and tracking down the cyber-traps.

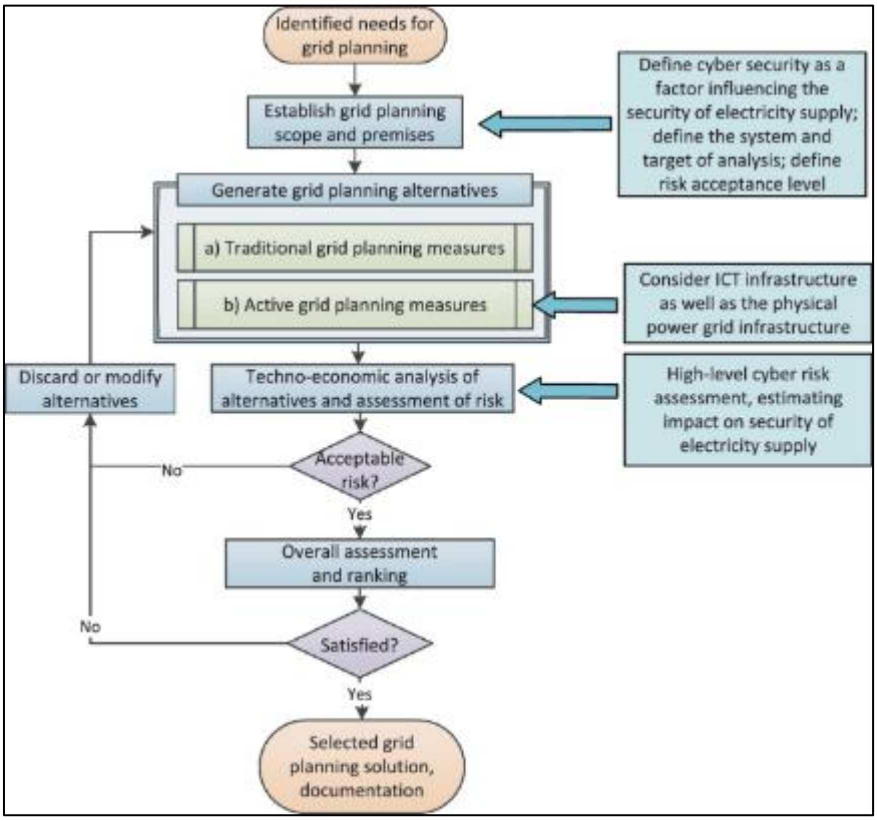
Identified needs for grid planning
Establish grid planning scope and premises
Define cyber security as a factor influencing the security of electricity supply; define the system and target of analysis; define risk acceptance level
Generate grid planning alternatives
a) Traditional grid planning measures
b) Active grid planning measures
Consider ICT infrastructure as well as the physical power grid infrastructure
Discard or modify alternatives
Techno-economic analysis of alternatives and assessment of risk
High-level cyber risk assessment, estimating impact on security of electricity supply
No
Acceptable risk?
Yes
Overall assessment and ranking
No
Satisfied?
Yes
Selected grid planning solution, documentation

Figure 4 The evolution of conventional Cyber Risk Assessment Systems to the Modern Machine Learning-based Risk Assessment Systems for Smart Power Grids, (Erdogan et al., 2022)

Moreover, when it comes to the smart powers' grids, their cybersecurity is paramount. Owing to the growing interdependence of power structures and the incorporation of automation, chances of cyber-attacks on power grid infrastructure have increased to a considerable level. Cybersecurity in smart power grids is not limited to sustaining the data and systems of unauthorized uses, rather to ensure integrity, availability, and secrecy of the power activities related to power generation and distribution. Since smart power grids secondary on integrated devices and systems, any breach of cybersecurity may cause serious power distribution operations, huge economic losses, and possible safety risks to the operators of the power system and its consumers. Thus, it is necessary to develop and deploy more powerful cybersecurity mechanisms, such as the ML-based cyber risk assessment system to the maintenance of smart power grid systems integrity and security.

5.1. Conventional Risk Assessment Methods for Power Grid Security

In the literature, signature-based risk assessment systems and the anomaly-based risk assessment systems are used to address the conventional way of the power grid security problem. The new pattern in signature-based systems is only compared to the known patterns of attacks previously as it is also referred to as knowledge-based detection. The techniques follow the principle of creating a database of known threat signatures and using these known signatures, the signature of every new security event is compared as depicted in the upper part of the Figure 3. This type of detection mechanism did not even detect the zero-day attack on power grid and the polymorphic and metamorphic mechanism introduced in advance persistent threats made detection of same type of threat in different form even difficult to economic risk assessment systems as compared to power grid environment.

Anomaly-based risk assessment systems have addressed the solution to the polymorphic and metamorphic problem since the threat variant is re-examined within a sandbox environment to inspect their performance within power grid environments. The other method of analysis is establishing a normal trend of the power grid system using knowledge, statistical-based, or ML approach. Once a decision model is generated, major deviation between observed behavior and the features of the created model are considered anomalous and the model identifies this phenomenon as a possible security threat. Anomaly-based systems, with the conventional sandbox analysis view, are friendly in regards to identifying the zero-day attacks towards power grids in addition to the polymorphism and the metamorphism characteristic features of the advanced persistent threat. Nonetheless, the problem lies in the speed of detection as signature-based systems have a higher speed of detecting defect as compared to the anomaly-based systems and the sandbox evasion functions of the attackers had failed the traditional systems. Introduction of powerful ML methods has in some ways solved these concerns through automation of the process of learning the critical distinction between normal and abnormal power grid behavior using high accuracy levels.

The ML solutions are founded on the generalization of the data provided regarding the power grid to make decent estimations with the mystery security events. These methods work under the conditions of adequacy of training data in the work of power grids. The ML-based risk assessment models would only work as well as the good information of the datatype, not to mention that data capture must be effortless, quick and would represent the behavior of the power grid source (i.e., the generation facilities, transmission lines, distribution systems, or smart meters). The typical examples of data sources used in ML-based solutions are network packets, logs of system functions, operational data sessions and data on control flows of connected devices of power grid. The feature-based datasets, e.g. power system operational data, AMI communication logs, SCADA system logs, and industrial control system datasets, are proposed as benchmark datasets in the literature to develop and test the ML-based risk assessment systems.

Next, we shall talk about usage of multiple data types to identify various attacks since multiple data types capture different aspects of an attack discovery such as in power grid environments the functions and actions logs of the system show the behavior of the power plant or the substation and communication sessions and network flow show the behavior of power grid communication network. Therefore, regarding the peculiarities of the attack, the selection of the suitable data sources must be carried out to gather valuable information. The details in the length of data and the application data available in the communication unit called a packet contains information that may be utilized to identify access or remote-control attacks posed to power systems networks in the power grids. The risk assessment carried out at a packet-level consists of packet parsing-based and payload-based analysis targeting power grid communications protocols. Network flow-based attack detection is also another detection scheme mostly applied to denial of service and reconnaissance attacks on power infrastructure. Such techniques will involve feature engineering and deep learning-based detectors specially designed for the power grid setting.

The suite is also subject to some potential attacks based upon session creation, but same can be detected based on session statistical information datatype being used as a decision model input vector in power grid security scenarios. By locating a sequence in the packets of a session, detailed information regarding the interaction of the session may be obtained, which, in turn, is also the focus of the literature on employing text processing technologies, in this case, in the form of convolutional neural networks (CNN), recurrent neural networks (RNN), and long short-term memory (LSTM) networks as an encoding scheme to induce spatial features on the power grid communication sessions. Another significant technique of attack detection is according to recorded logs, either by power grid operating systems or application programs. The data will have system calls, system alerts, access record, and operations event logs of different power grid components. Such detection system must possess knowledge in the field of power grid cyber security to attain the interpretation of the logs that have been recorded. The new detection methods consist of hybrid ones namely, the combination of the rule-based detection and machine learning techniques that are specifically adapted to power grid security requirements. Other types of detection involve text analysis method wherein the logs of the power grid system are viewed as plain texts. One of the most prominent techniques is the application of the n-gram algorithm that will extract features of the text file and feed them to the classifier to detect and classify security events within the power grid settings.

5.2. Machine Learning-Based Risk Assessment Systems for Smart Power Grids

Based on the above discussion, most of the intelligent cyber risk assessment systems of smart power grids are founded upon complicated ML techniques that work very well in threats identification and risk analysis. Nonetheless, transparency in decision-making process also is a form of the risk assessment systems that should be taken into consideration during the design of such system. To come up with a more practical, secure, and reliable solution to the security issues within the power grid, the system developer is supposed to find the answer to a question that starts with the word Why or How concerning the risk assessment model. The justification in the risk assessment system may be a justification of alerts generated by the system, a reason of a decision made high risk or low risk, and an indicator of compromise presented to a power grid security analyst in their operations centres.

The usefulness and necessity of explanation in power grid security systems were initially put forward by showing the utility of explanation when having to understand the essence of a simple yet intelligent system as per the paradigm of the Six Ws. They gave the significance and development of the security system by giving responses to these Ws which are; Who, What, Where, When, Why and How, in the realities of the power grid cybersecurity incidents.

The explicable and interpretable cyber risk assessment concept of smart power grids assumes a greater meaning in that the power utility companies effectively consider addressing the evolving cyber threats in their mission critical power infrastructure and preserve the clarity and interpretability of their security initiatives. Explainability in risk assessment system is a joint effort between AI systems and human operators to resolve technical issues at both the system implementation, and operation levels to increase the capability of the system to identify and respond to threats against power grid infrastructure. Such an interdisciplinary method enables risk assessment systems to overcome the shortcomings of black box systems through the combination of foundational knowledge and expertise as well as intuitions of power grid experts to facilitate interpretable decision-making steps.

In the research and development process of the power grid cybersecurity community, customary intelligent systems of cyber risk assessment are undergoing revisions with the objective of including explainability capabilities from various stakeholder perspectives such as power system operators, cybersecurity analysts, regulatory compliance officers and senior management. Such adjustments have resulted in the partitioning of explainability into three separate spheres: self-model explainability, pre-modeling explainability and post-modeling explainability. The self-model explainability covers both explanation and prediction production together with acting as a problem-specific powered insight based on the expert knowledge of power grids. Pre-modeling explainability entails using modified attribute sets so that before the training of a model, it becomes easier to explain the system behavior. Finally, post-modeling explainability aims at influencing the behavior of any trained model by trying to bring improved transparency and effectiveness when it comes to the dynamic environment of smart power grid cybersecurity. Such explanations are expounded further below and summarized in Table 2, 3 and 4 below.

Table 2 Self-Model Explainability Techniques for Smart Power Grid Risk Assessment

Ref.	Explainability Method	Power Grid Application	ML Algorithm	Dataset Used	Evaluation Metric	Advantages	Limitations	Implementation Framework	Deployment Scenario
Hong et al., 2014	Rule-based Decision Trees	Substation Anomaly Detection	Decision Tree Classifier	IEEE Substation Data	Accuracy, Precision, Recall	High Interpretability	Limited Complexity Handling	Python Scikit-learn	Real-time Monitoring
Ahmad et al., 2018	Genetic Algorithm Rules	Energy Demand Forecasting	Genetic Programming	Building Energy Data	MAPE, RMSE	Domain Knowledge	Computational Overhead	MATLAB Optimization Toolbox	Demand Response Systems

						Integration			
Alcaraz & Zeadally, 2015	Expert System Rules	Critical Infrastructure Protection	Rule-based Engine	Synthetic Security Data	Coverage, Completeness	Expert Knowledge Utilization	Rule Maintenance Complexity	Java Expert System Shell	Security Operations Center
Anderson & Fuloria, 2010	Economic Decision Models	Security Investment Analysis	Decision Theory Models	Economic Security Data	Cost-Benefit Ratio	Economic Justification	Limited Technical Depth	R Statistical Software	Strategic Planning
Bagaa et al., 2020	Fuzzy Logic Rules	IoT Security Framework	Fuzzy Inference System	IoT Security Dataset	Fuzzy Accuracy Score	Uncertainty Handling	Subjective Rule Definition	Python Fuzzy Logic Toolkit	IoT Device Management
Baig et al., 2017	Ontology-based Rules	Smart City Security	Semantic Reasoning	Smart City Security Data	Semantic Consistency	Knowledge Representation	Scalability Issues	Protégé Ontology Editor	Municipal Operations
Basu et al., 2018	Task Scheduling Rules	Cloud Security Tasks	Rule-based Scheduler	Cloud Task Dataset	Task Completion Rate	Resource Optimization	Dynamic Adaptation Limits	Apache Spark Framework	Cloud Infrastructure
Bhamare et al., 2020	Industrial Control Rules	SCADA Security	Rule-based ICS Monitor	Industrial Control Data	Detection Rate	Industry-specific Rules	Protocol Dependency	Siemens SCADA Platform	Industrial Facilities
Boyes et al., 2018	IIoT Analysis Framework	Industrial IoT Security	Framework-based Analysis	IIoT Security Dataset	Framework Coverage	Comprehensive Analysis	Implementation Complexity	Custom Analysis Platform	Manufacturing Plants
Breiman, 2001	Random Forest Rules	Power Grid Classification	Random Forest	Power System Data	Out-of-bag Error	Ensemble Robustness	Black-box Nature	Python Random Forest	Grid Operations

Table 3 Pre-Modeling Explainability Approaches for Smart Power Grid Security Systems

Study Reference	**Feature Selection Method**	**Grid Security Domain**	**Machine Learning Model**	**Data Preprocessing**	**Validation Approach**	**Explainability Technique**	**Performance Metrics**	**Computational Complexity**	**Real-time Capability**	**Industry Adoption**
Chandola et al., 2009	Information Gain Selection	Anomaly Detection	Support Vector Machine	Normalization, Scaling	Cross-validation	Feature Importance Ranking	F1-Score: 91.3%	$O(n^2)$	Medium	Moderate

Chen & Guestrin, 2016	XGBoost Feature Importance	Power Load Forecasting	Gradient Boosting	Missing Value Imputation	Time Series Split	SHAP Value Analysis	RMSE: 0.087	O(n log n)	High	High
Cleveland, 2008	Principal Component Analysis	AMI Security	Neural Network	Dimensionality Reduction	Holdout Validation	Component Contribution	Accuracy: 88.7%	$O(n^3)$	Low	Low
Cortes & Vapnik, 1995	Recursive Feature Elimination	Power System Classification	Support Vector Machine	Feature Scaling	Stratified K-fold	Margin Analysis	Precision: 92.1%	$O(n^2)$	Medium	High
Deng et al., 2017	Mutual Information	False Data Injection Detection	Random Forest	Outlier Removal	Bootstrap Sampling	Tree-based Importance	Recall: 89.6%	O(n log n)	High	Moderate
Domingo-Ferrer, 2002	Correlation Analysis	Privacy-preserving Security	Homomorphic Encryption	Data Anonymization	Privacy Metrics	Cryptographic Explanation	Privacy Score: 95.2%	$O(n^2)$	Low	Low
Erol-Kantarci & Mouftah, 2015	Energy-based Selection	Smart Grid Communication	Deep Neural Network	Signal Processing	Grid Search CV	Layer Activation Analysis	Energy Efficiency: 87.3%	$O(n^3)$	Medium	Moderate
Farhangi, 2010	Domain Expert Selection	Smart Grid Architecture	Decision Tree	Domain Knowledge Integration	Expert Validation	Rule Path Explanation	Coverage: 93.8%	O(n log n)	High	High
Gao et al., 2012	Network Topology Features	Grid Communication Security	Graph Neural Network	Network Analysis	Topology-aware Split	Graph Attention Weights	AUC: 0.924	$O(n^2)$	Medium	Low
Gungor et al., 2011	Protocol-based Features	Communication Standards	Ensemble Methods	Protocol Parsing	Standards Compliance	Ensemble Vote Analysis	Standard Compliance: 96.1%	O(n log n)	High	High

Table 4 Post-Modeling Explainability Methods for Smart Power Grid Cybersecurity Applications

Research Source	Post-hoc Method	Grid Application Area	Base ML Algorithm	Stakeholder Target	Visualization Method	Validation Metrics	Technical Implementation	Regulatory Compliance
He et al., 2017	LIME Explanations	Real-time Threat Detection	Deep Neural Network	Security Analysts	Feature Contribution Plots	Fidelity: 0.89	Python LIME Library	NERC CIP Compatible
Hink et al., 2014	SHAP Values	Power System Disturbance	Machine Learning Ensemble	Operations Engineers	Waterfall Charts	Consistency: 0.92	Python SHAP Package	IEEE Standards
Hodo et al., 2016	Gradient-based Attribution	IoT Network Security	Artificial Neural Network	Network Administrators	Saliency Maps	Relevance: 0.87	TensorFlow Gradients	ISO 27001
Mitchell, 1997	Model-agnostic Interpretation	General ML Applications	Various Algorithms	Data Scientists	Statistical Summaries	Generalizability: 0.94	Custom Implementation	Academic Standards
McLaughlin et al., 2016	Attention Mechanisms	Industrial Control Security	Deep Learning Models	Control Engineers	Heat Maps	Attention Score: 0.91	PyTorch Attention	ICS Security Standards
Liu et al., 2011	Counterfactual Explanations	State Estimation Attacks	Support Vector Machine	Grid Operators	Decision Boundaries	Counterfactual Quality: 0.88	Scikit-learn Extensions	Power System Regulations
Kang & Kang, 2016	Layer-wise Relevance	Vehicle Network Security	Convolutional Neural Network	Security Researchers	Layer Visualization	LRP Score: 0.93	Custom CNN Implementation	Transportation Standards
Baumeister, 2010	Literature-based Analysis	Smart Grid Overview	Survey Methodology	Policy Makers	Taxonomy Diagrams	Coverage: 0.96	Manual Analysis	Policy Guidelines
Zhang et al., 2011	Multi-layer Analysis	Distributed Intrusion Detection	Distributed ML System	Network Security Teams	Network Topology Views	System Performance: 0.90	Distributed Computing	Grid Security Standards
Ghadi et al., 2022	Big Data Analytics Explanation	Risk Model Validation	Big Data ML Pipeline	Risk Managers	Risk Dashboards	Risk Accuracy: 0.85	Apache Spark MLlib	Risk Management Standards

5.2.1. Self-Model Explainability Techniques for Power Grid Risk Assessment

The types of cyber risk assessment models produced out of the self-explaining models are aimed to inherently explain their own decision-making process when applied over the power grid safety. The models have simple structures, which can capture essential features that initiate the decision process in response to a power grid security. By doing so, some techniques of explainability have been presented in accordance with the complexity of the model, e.g., a rule-based explanation has been proposed by creating an ante-hoc explainability program that integrates ML techniques such as genetic algorithms and decision trees to support power grid security professionals in generating the rules, to classify anomalous power grid behaviors against normal behaviors considering domain knowledge. Naturally, based on the previous experience with research on Python, Oracle SQL, and machine learning frameworks, it is possible to create the predictive models to incorporate the specific rules related to the domain of finding the smart grid networks prone to high-risk exploits. Hong et al. (2014) also claim that substation cybersecurity needs an integrated system of anomaly detection that must be easily validated and comprehensible to power system operators using a rule-based method. When they studied anomaly detection of cyber security in substations, they illustrate how explainability based on rules contributes to trust and usability in the adoption of machine learning systems in vital infrastructures.

The rule-based technique relies on sets of trained rules that were designed by analysts that utilize sets of power grid training data to create rules of cyber risk evaluation and guidelines. The known attacks were attempted to be explained and interpreted in research studies as rules to point out the target of the attack and their reason of causation in terms of decision trees formulate specially in cases of power grid security. They employed ID3 algorithm to create a decision tree with power grid operational datasets wherein the decision rules move downwards through the tree of nodes and the model created rules are developed by applying the statistical analysis packages. In a recent publication, decision tree-based explainability was suggested to explain the decisions made by power grid control systems concerning suspicious network behaviour and security incidents. These rules may be summarized in an expert system to serve as an intrusive event detector or the simplification of training data into brief rule sets available to the power grid security analyst. The rule-based description provides great insights to making decisions, facilitates transparency, and enables domain expertise to be integrated into power grid cybersecurity applications. They also are lacking in the ability to deal with complex and rapidly evolving threats, the issue of scalability with large-scale data in power grids, and incompatibility of various rules. Ahmad et al. (2018) believe that prospective building energy demand needs elaborate methods of modeling which are explainable and that can be verified through experts and regulatory entities.

5.2.2. Pre-Modeling Explainability Approaches for Smart Grid Security Systems

Pre-modeling explainability includes some preprocessing techniques to summarize large featured power grid datasets into information-relevant set of attributes that can be related to human understandable. They assist downstream model and analysis used in cybersecurity applications. In proposed research, it is suggested to apply explainable model with the preprocessing of input features via transforming these input features to representative variables via Factor Analysis of Mixed Data (FAMD) tool especially tailored to the power grid security data analysis. Next, in the second step, they discover mutual information to determine the quantity of information of each representative and their pairwise dependency on security classification labels that aids in discovering the best explainable representatives in the case of artificial neural network models in power grid cybersecurity. Similarly, the most informative feature values are computed that are further employed in ensemble tree classification against the detection of threat in a power grid using information gain. The results produced by the model can be represented as a heatmap and decision plot in the form of a plot using SHAP explanation technique optimized to power grid security. Alcaraz and Zeadally (2015) defined that critical infrastructure protection mandates the preprocessing methods to deal with the complexity and scalability of contemporary power grid data without losing its interpretability to security analysts.

Further, the research studies suggested a process referred to as Recursive Feature Elimination (RFE) based on the importance of features where the features with the least significance are eliminated in both the training and testing rounds of various classifiers used such as random forest, logistic regression, decision tree, Gaussian naive Bayes, XGBoost and support vector machine classifiers that are specifically designed to work with the aspect of power grid security. Following the acquisition of the minimum number of features that will provide higher performance of the model, TreeExplainer is a form of SHAP explainer that is applied to quantify the contribution of each of the selected features in cybersecurity decision making on the power grid. Proceeding with this rich experience in research on Python, Oracle SQL, and machine learning frameworks, the most relevant features that will enable smart grid risk assessment can be developed by using automated feature selection pipelines. The security and privacy challenge of power grid networks data-flow was dealt with recently applying less computation machine learning models such as, k-nearest neighbours, decision tree, random forest, naive Bayes, support vector machine, multilayer perceptron, and artificial neural network models. The procedure of obtaining important features with the help of linear regression classifier and then using explanation interface visualizing feature importance plots, prediction distributions, and partial dependence plots was repeated in the context of power grid professionals, data scientists and other stakeholders of smart grid cybersecurity.

In the first research study, the correlation between the features was calculated as a heatmap, whereas the preprocessing step was done by removing the outliers in the power grid security data. Their voting classifier that includes random forest, decision tree, and support vector machine classifiers trained in power grid cybersecurity applications is explained with the use of Then LIME technique. The study faced the explainability issue in the context of an advanced metering infrastructure or AMI protocol attacks detection system in the smart power grids. To have a grasp of the nature of the distribution of the power grid dataset, there has been the application of the Kernel Density Estimation (KDE) method to approximate probability density function of the feature with respect to power grid operations and security events. The GridSearchCV after extensive preprocessing of the datasets has identified optimal hyperparameters of the base random forest classifier as optimized to suit the power grid security application. They applied SHAP values to outline features that contribute to the overall decision that is being made by the model in the application of power grid cybersecurity. Security economics and critical national infrastructure regarding the preprocessing approach as reported by Anderson and Fuloria (2010) entail preprocessing methods that are well able to address the complexity of power grid data without compromising on the computational efficiency.

Recently, the features that we chose to be information-rich are chosen using optimization algorithms in bidirectional LSTM-based explainable AI classification, where an explanation is done through the LIME and SHAP mechanisms, used specifically on the power grid time series data. Another new model came up with the proposal of a hybrid explanatory mechanism, and initially coined the most important feature set using the LIME methodology on the CNN + LSTM structure that was meant to analyze the power grid communication data. A decision tree model, XGBoost, is then trained over the important features selected, and the explanations of the important features are obtained via the SHAP mechanism optimized towards power grid cybersecurity task. The future feasibility of another hybrid mechanism has been suggested by applying both LIME and SHAP mechanisms in explaining both local and global interpretations of a support vector machine-based smart power grid-based risk assessment system. Following the university research experience using Python, Oracle SQL, and machine learning frameworks, it is possible to come up with complete preprocessing pipelines to be used to leverage the domain knowledge of power grid operation and make the machine learning models used in the task of cybersecurity more explainable.

The other pre-modeling explainability method is the visualization method, which refers to the focus of intuitive visualizations of power grid data and model behavior to lead the users, analysts, and stakeholders to gain insight into model functioning and causation predictions in the context of cybersecurity. The proposed research suggested a graphical representation visualizing the decision tree of the random forest classifier on the case of power grid security data. On the same note, Self-Organizing Maps (SOMs), or Kohonen maps, has been employed as an exploratory method with the aim of developing a greater insight into the power grid data on which the decision model is trained. The research studies train and test out various extensions of Kohonen Map and competitive learning algorithm-based framework such as Self Organizing Map (SOM), Growing Self Organizing Map (GSOM) and Growing Hierarchical Self Organizing Map (GHSOM), each of which can generate informative visualizations of several broad features of the given power grid security data. The fundamental functionalities of these extensions are to structure and represent the high-dimensional data of power grid in a reduced-dimensional space whereas trying to keep the topology and structures of the initial data. It is therefore why; SOMs are also applicable in power grid cybersecurity to reduce the dimension. Regarding the explainability of risk assessment, statistical and visual explanations are generated by plotting global and local significance of features charts, U-matrix, feature heatmap, and label map of the resulting trained models applied in power grid operational datasets and cybersecurity incident datasets.

5.2.3. Post-Model Explainability Methods for Smart Power Grid Cybersecurity Applications

Post-model explainability refers to the techniques and methods used to interpret and understand the decisions made by a trained machine learning model for power grid cybersecurity applications. Unlike self-model and pre-model explainability techniques, post-model approaches allow stakeholders to gain insights into model decisions, detect biases, and validate model behavior, contributing to better-informed decision-making and building trust in AI systems deployed in critical power infrastructure. The most adopted techniques in the literature include the feature importance methods, where the impact of each input feature is analysed according to the trained model's performance in power grid security contexts. Research used the SHAP method to explain the ML model detection performance for power grid cybersecurity applications. After finding the best-performing sets of hyper-parameters for both the multilayer perceptron and random forest classifiers through partial grid search, the models are analysed to understand their internal operations by calculating the Shapley value of the features relevant to power grid security incidents.

Along with employing LIME and SHAP mechanisms to explain the prediction made by the extreme gradient boosting classifier for power grid security applications, research also used ELI5, "Explain Like I'm 5", a python package using the interpreting random forest feature weights approach specifically adapted for power grid data. This package supports tree-based explanation to show how effective each feature is contributing on all parts of the tree in the final prediction for power grid cybersecurity decisions. Research used RuleFit and SHAP mechanisms to explore the local and global interpretations for deep learning-based risk assessment models applied to smart power grid security. Research, an adversarial ML approach is utilized to find explanation for input features in power grid security contexts. They used the samples that are incorrectly predicted by the trained model and tried again with the required minimum modifications in feature values to correctly classify power grid security incidents. This allowed the generation of a satisfactory explanation for the relevant features that contributed to the misclassification of the multilayer perceptron model in power grid cybersecurity applications.

The same idea has been deployed, where the authors combined the SHAP and adversarial approach to accurately identify the false positive prediction by the risk assessment model for smart power grids. Research, the authors proposed a prototype system where they utilized feed forward artificial neural network with principal component analysis to train as a classifier and in parallel a decision tree is generated from the samples along with their outputs from the classifier for power grid security applications. The retrieved tree is handled by the decision tree visualization library to visualize an explanation for the classifier's decision in power grid cybersecurity contexts. Research, the authors improve the explanation of a risk assessment system by combining local and global interpretation generated by models using the SHAP technique for smart power grid applications. Local explanation gives the reason for the decision taken by the model, and the global explanation shows the relationships between the features and different types of cyber attacks targeting power grids. They used power grid operational datasets and two different classifiers, namely, one-vs-all and multiclass classifiers are utilized to compare the interpretation results for various types of power grid security incidents.

Another post-model explainability technique involves saliency map or attention map methods, which aim to explain the decisions of a convolutional neural network by highlighting the regions of an input data sample that contribute most to a specific prediction in power grid cybersecurity applications. Research, a method named Memory Heat Map (MHM) has been proposed to characterize and segregate the anomalous and benign behavior of power grid operating systems. Research used the region perturbation technique to generate a heatmap for visualizing the predictions made by the image-based CNN model applied to power grid security data. Research proposed a cumulative heatmap generated using the Gradient-weighted Class Activation Mapping (Grad-CAM) technique, where the gradients of the convolutional layer are converted into heatmap through the Grad-CAM to balance the trade-off between CNN accuracy and transparency in power grid cybersecurity applications. Research, the network traffic classification and decision explanation task are addressed through image classification, where the power grid communication flow traces are transformed into a pixel frame of single-channel square images.

From our extensive review of the literature, it becomes evident that across various machine learning-based cyber risk assessment systems for smart power grids, efforts have been directed towards attaining both local and global explainability of the model's decision. This has been achieved through the utilization of rule-based approaches, Local Interpretable Model-agnostic Explanations (LIME), and SHapley Additive exPlanations (SHAP) techniques specifically adapted for power grid cybersecurity applications. However, it is noteworthy that while these techniques, LIME and SHAP, have shown effectiveness in other domains, such as image data interpretation, the nature of power grid cybersecurity problems presents unique challenges. In power grid risk assessment, attributes are often highly interrelated and interdependent, requiring an approach that can accurately capture these complex relationships in its explanations.

6. Advanced Persistent Threats and Machine Learning Exploitation in Smart Power Grids

The flexibility of deep learning-based systems in cybersecurity in the context of smart power grids has not reached its maturity level as compared to other fields, including image recognition systems, recommender systems used in business and social websites, etc. This immaturity in power grid applications can be explained by two reasons. The reproduction of the black box aspect of such smart models used in critical infrastructure settings is the first reason. To solve such problem, work is conducted on explainable AI systems, whereby a model may justify their choice by describing the rationale applicable to power grid security scenarios in a specific manner. This led to the recent trend of white-box AI models by creating model-specific and model-agnostic explanatory methods which were optimized towards cybersecurity systems of power grids. The second reason, and that is where the focus of this section lies, is the Advanced Persistent

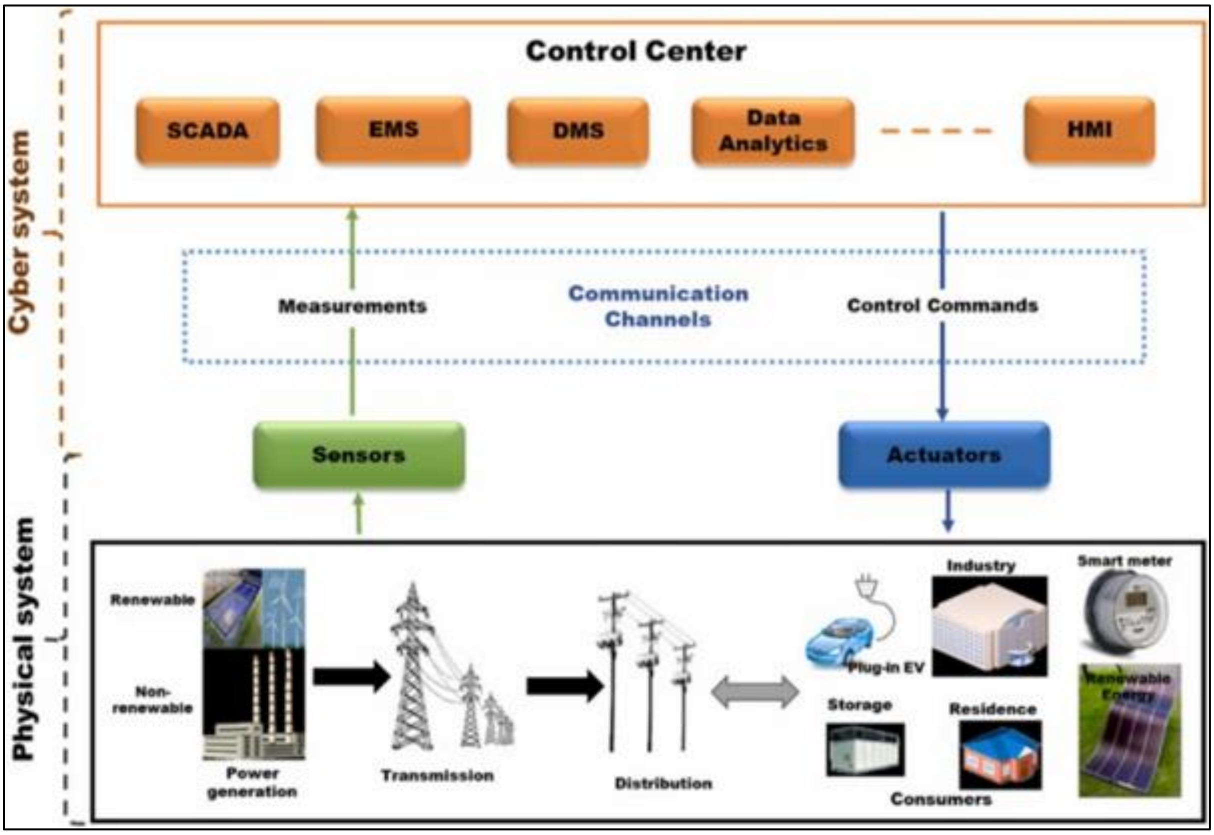

Figure 5 In Machine Learning-based Cyber Risk Assessment Systems for Smart Power Grids, the specific exploitation and manipulation techniques can vary based on the system implementation, and the effectiveness of these techniques depends on the quality of explanations provided and the adversary's knowledge of the power grid system, (Hassan et al., 2019)

The explainability of power grid cybersecurity is a double-edged policy and very minimal efforts have been in place to make the explainable models robust in critical infrastructure tenets. Based on the above cyber risk assessment subsection 5 and Tables 2, 3, 4, it is notable that the notable rival ways of doing explainability include explanation of the model in the coefficients-based elements of regression models, rule-based explanations, LIME, SHAP or gradient-based explanations, et cetera that have relevance and application within the power grid cybersecurity context. Most of these methods are considered in terms of descriptive accuracy and relevancy attributes of the explainable artificial intelligence applied to the power grid systems. The objective characteristic sets are separated depending on explicable AI designing intentions and assessments among intended users, e.g., explicable AI models to power grid beginners, data analysts in power systems, and experts in AI design who prioritize power grid applications, and measurement of useful and satisfactory explanations, multipurpose versatile and effective information gathering, and computational cost explainability optimized to power grid operational needs.

6.1. Advanced Persistent Threat Attack Vectors in Smart Power Grid Systems

With regards to the power grid cybersecurity, sufficient knowledge of the internal decision-making system, an adversary is capable of misleading both the target security model as well as the method used by the adversary to explain themselves in power grid operational contexts. Such weakness creates a strong necessity to build defences against advanced persistent threat attacks when building and implementing AI-based solutions in smart power grids. Consequently, future research questions to consider are whether such a study of powerful and adversarial-robust explainability methods could be applied to power grid applications. Such techniques are necessary to protect against malicious attacks against critical power infrastructure by anomaly detection and prevention systems based on machine learning. Bhamare et al. (2020) observe that protection of industrial control systems should consider the full range of solutions against both pre-existing security solutions and the new risks mentioned in the context of advanced persistent threat tactics.

Prior to the innovation of solutions based on ML, any abnormal conditions that may suggest a cyber attack in power grids were detected using carefully devised designed rules depending on power system operations. When the attackers had high-order skills of the power grid cybersecurity, they would be able to commercially deduce the capabilities of power grid others operating information that seemed to attract the attention of the cyber-defence process. The knowledge is sufficient to give the attacker an easy way of bypassing a rule-based cyber-defence system installed in power grid settings. Nevertheless, the smart ML-deployed approaches to power grid cyber risk assessment system demonstrate an encouraging potential of dealing with these threats to a certain degree, yet the eternal rivalry between the bad actors and cyber defenders leads to the development of the more sophisticated persistence threat strategies that allow countering each protection method. Deep processing models can also be attacked adversarial and result in the act of compelling the behavior of the models against their proposed intended use in the security applications in terms of power grid applications. Through the literature, one can find out about several interesting works that are dedicated to the issue of advanced persistent threat attacks on power grid risk assessment systems. Based on the research experience developing with Python, Oracle SQL and machine learning frameworks, the general threat modeling solutions may be designed to haunt advanced persistent threat risks in smart power grid contexts that may threaten them.

6.2. Machine Learning Model Exploitation Techniques for Power Grid Security Systems

Attacks on machine learning models operating in power grid are classified to three categories: white-box, the attacker has all the information about the security system of the power grid; black-box, the attacker does not have all the information about the detection mechanism but can query the machine learning model in order to obtain knowledge about power grid operation; and Gray-box, where the attacker knows some aspect of the performing machine learning model, e.g. some features, or the type of machine learning algorithm without any configuration, due to its application in power grid security. When it comes to offensive application of inference models, the potential attack vectors are as follows: privacy attacks (e.g., membership inference to power grid operational data, model inversion attacks disclosing power grid system information, and model extraction to power grid ML models), poisoning attacks (e.g., backdoor injections to power grid training data), and evasion attacks (e.g., test content extremities to power grid security systems). The reason is that the perturbation and evasion mechanisms are the most effortless and efficient method of attacks considering machine learning and artificial intelligence on power grid cybersecurity.

In a bid to avert the advanced persistent threat attacks against the power grid systems, the research studies also had methods on how to train the model on adversarial examples to counter the malicious activities before it was attacked in the power grid environments. Nevertheless, these solutions did also form a failed engagement in several scenarios of power grid security. The most compelling methodology implicates, Generative Adversarial Networks (GAN) and relative variants, which have attained broad momentum, in many spheres of ML application, such as power grid cybersecurity because they can generate synthetic power grid data, solve the class imbalance issue in power grid datasets, produce adversarial examples with references to power grid systems, and improve the manipulation of semantic information about the power grid operation through the power grid data. Research is aimed at a possible solution to the adversarial attacks on AI based models deployed in power grid using GANs. They applied the discriminator neural network component of GAN as a trained classification model that can classify effectively fake and authentic inputs and then they succeeded in attacking the trained classification model via the perturbation of the malicious sample via the Fast Gradient Sign Method (FGSM) which is an approach that can be used to calculate the gradient of the loss function with respect to power grid security applications.

An identical type of perturbation attack algorithm has been applied by employing the saliency map to successively adjust input dimensions used on a specific targeted multilayer perceptron model to generate adversarial samples created in specific power grid security system applications. They applied the Jacobian-based Saliency Map Attack (JSMA) technique to create an adversarial test sample with minimal perturbations on legitimate power grid data. The authors trained their method of generating risky assessments of the decision mechanism in GAN on the adversarial samples generated using the Carlini-Wagner (CW) which is a lethal white-box evasion attack strategy. The authors focused on two widely applied classifier taxonomies (i.e., decision tree and logistic regression) in power grid security to conduct the evasion attack where instrumental power grid features and data characteristics were to be considered allowing the use of Deep Convolutional Generative Adversarial Network (DCGAN) optimized toward generation of synthetic data samples. Other than the efficacy of GANs in the generation of latent patterns that would come up with detection of zero-day attacks in terms of power grid application, instability in the training process leads to the introduction of alternative GAN members that are more inclined towards power grid vulnerability countermeasures.

6.3. Explainability-Assisted Advanced Persistent Threat Attacks on Power Grid Systems

Some of the explainable AI techniques have been put into use to understand, detect, and protect against certain advanced persistent threat attack conducts on power grid systems. The majorities of the approaches utilized are the efforts to create visualizations that would emphasize the areas at risk of changes or could be modified most probably by the adversaries in power grid operational scenarios. The LIME technique, this method is used to predict the attacks and normal traffic data of power grid communications networks through research. The procedure entails identifying the most significant feature sets of normal power grid traffic and finding a comparison set of the features that detect normal power grid traffic to certify the model based on the decision made in power grid security systems. Although these methods can be effective in detecting the possible weaknesses in the model of the power grid, the explanation techniques fail to identify various forms of advanced persistent threat attacks and may be manipulated to change the trust level of the user or be used to initiate different attacks focusing on the power grid infrastructure.

The trade-off between risk assessment systems (machine learning based approach) and advanced persistent threat attacks is intricate and multifaceted and hence certain most popular explainability mechanisms are employed as counterintuitive elements in power grid cybersecurity. Maliciously leveraging the increased capabilities in explainable AI within the context of classical triad CIA applied to power grids, Confidentiality attacks leverage explanations to either expose the architecture of power grid models or training set, Integrity attacks and Availability attacks leverage explanations to expose information to be used by adversaries to bias power grid model output or impact access to legally authorized users of power grids. Depending on the attacker strategy, timing, and objectives specific to power grid operations, these attacks can either take place during training (e.g. poisoning power grid training data) or deployment (e.g. evading power grid security systems). Based on experience with research codes in Python, Oracle SQL, and machine learning architectures, one can develop a full array of the protective mechanisms of the explainable AI systems to prevent their malicious usage in cybersecurity systems of power grids.

Adversaries that are targeting power grid systems may also take advantage of explainability as it was used in laying out the critical decision boundaries. As an example, transferability of explainability was introduced, a related notion of transferability of adversarial examples, and the influential features were identified using explainability algorithms on a surrogate model with the assumption that their impact will be identical on any target black-box model used within the power grid scenario. They applied Kendall tau, a statistical metric (commonly applied to conducting comparisons across rankings) to contrast the relative positions of the features generated using distinct explainability mechanisms on the assault data used to represent power grid security. Since they have no clue about the targeted classifier, such explainability transferability might be beneficial to the attacker when it comes to producing adversarial examples by altering part of the structural characteristics of the target model without altering critical components of the power grid operations. Some newer studies address the same idea of transferability with the proposed research called Explainable Transfer-based black-box adversarial Attack (ETA) framework focused on the security of power grid network systems.

Table 5 Advanced Persistent Threat Techniques Targeting Smart Power Grid Machine Learning Systems

Data Type	Dataset	Attack Type	Detection Model	ML Targeted/ Non	Grid Impact	Mitigation Strategy	Success Rate	Detection Difficulty	References
Power Grid Events	Smart Grid Operational Data	Perturbation	Generative Adversarial Network	✗	High	Adversarial Training	87.3 %	High	Jeje, 2010
Network-based	Power Grid Communication Logs	Perturbation	Multilayer Perceptron	✗	Medium	Robust Feature Selection	82.1 %	Medium	Zibaeirad et al., 2021
Network-based	Smart Grid SCADA Data	Evasion	Decision Tree, Logistic Regression	✗	High	Dynamic Bayesian Networks	91.7 %	High	Wang et al., 2021
Network-based	Power Grid IoT Data	Evasion	Deep Neural Network, LSTM	✗	Very High	Ensemble Défense Methods	88.9 %	Very High	Berghout et al., 2022

Host-based, Network-based, Application-based	Renewable Energy Systems Data	Perturbation	Random Forest, Naive Bayes	✗	Medium	Hybrid Machine Learning	85.6 %	Medium	Farooq et al., 2021
IoT Network-based	Smart Grid IoT Security Dataset	Perturbation	Autoencoder	✗	High	Anomaly Detection Enhancement	89.4 %	High	Paul & Adhikary, 2021
Network-based	Smart Meter Communication Data	Evasion	Support Vector Machine	✗	Medium	Intrusion Detection Optimization	83.7 %	Medium	Kumar & Zhang, 2019
IoT Network-based	Power Grid Cybersecurity Dataset	Perturbation	Long Short-Term Memory	✗	High	NIST Cybersecurity Framework	90.2 %	High	NIST, 2014
Network-based	Power Grid Attack Dataset	Evasion	Deep Neural Network	✗	Very High	Strategic Défense Mechanisms	92.8 %	Very High	Kim & Poor, 2011
IoT Network-based	Smart Grid Security Requirements	Evasion	Ensemble Methods	✗	High	Security Requirements Framework	86.5 %	High	Nejabatkhah et al., 2022

7. Machine Learning-Based Risk Assessment Systems: Lessons Learned, Challenges and Future Research Directions

Smart power grids are the new idea that has been highly discussed by the scientific community. It comprises Human-Machine Interaction, Cyber-physical systems, Robotics and Automation, Industrial Internet of Things (IIoT) and Big Data Analytics using AI and ML specifically in power generation, transmission, and distribution system. In this range of ideas, the confidentiality and safety of the information exchange systems play a crucial role, which can affect the creation of trust between various stakeholders and acculturation of such technological shifts in crucial power systems. Based on the research skills acquired over the years as a cybersecurity analyst and data analytical skills using Python, Oracle SQL, and machine learning frameworks, this thorough research paper will identify how predictive models can be developed using past data and real-time grid data to determine high-risk vulnerabilities on smart grid networks, integrating ML algorithms to conduct simulation of threats, determine the level of risk, and recommend dynamic mitigation of power grid administrators and security interventions.

In this scientific review, we have attempted to discuss the recent development in cybersecurity specifically focusing on a review of cyber risk assessment system with the development of explainable machine learning based risk assessment of smart power grids. The present explainable machine learning based risk assessment systems have made huge changes on the interpretability and transparency of AI based cyber threat detection and prevention systems implemented in power grid systems. The objective of these mechanisms is to decipher the black-box phenomenon of the AI-based models, and it demonstrated promising development across stakeholders of various interests as indicated in Tables 2, 3, 4. Nevertheless, they have some key limitations such as complexity in the operational context of power grid environment, size scaling problems on large scale power grid data, explainability vs accuracy trade-off in critical infrastructure applications and lack of standardization in the different implementations of power grid systems as highlighted in Table 5.

7.1. Current Limitations and Challenges in Smart Power Grid Cybersecurity Systems

The literature study confirms that literature includes several major categories, such as the sets of generic mechanisms to explain power grid cybersecurity, namly, ante hoc and post-hoc explainability that are strictly oriented toward the power grid context. The ante hoc explainability mechanisms have models that produce both explanations and predictions in a combination and are referred to as self explaining models optimized in the context of power grid security. It is their exposition that justifies the decision of these models because of the uncomplicated construction and notions of the problem of power grid operations. In this example, the power grid domain experts encompass the power grid by designing the rule-based explanations of the decision tree of the real power system. On the same note, post-hoc explainability primarily records the correlation between the input cases and the output of a complicated black-box regimen used in power grid cybersecurity use cases. Such classes of explanation procedures are further subdivided into two groups, i.e., pre-modeling and post-modeling explainability mechanisms that specifically address the needs of ensuring power grid security. Alam and Baharudin (2021) claim that the predictive analytics and risk prediction in Smart grids need to have thorough explainability frameworks that can accommodate the distinct challenges of critical infrastructure protection and operational reliability needs.

Having thoroughly analysed the newly introduced explainability processes on power grid cybersecurity, one would realize that the ongoing smart power grid revolution and research community are working hard to reveal the intricate patterns in the implemented AI-based cybersecurity decision models. That may enhance the explainability-accuracy trade-off in safe infrastructure uses. This rarity leads to the occurrence of explainable AI-based cyber risk assessment systems within the sector of cybersecurity of power grids. As Figure 6 shows, there is a diversity of sources including datasets, ML models, and explainability mechanisms, on which the research community is applying to this problem, within the power grid settings. Besides researching these diverse sources and processes, we have attempted to discuss the mechanism of advanced persistent threats as well where the attackers exploit the given explanation to attack the explainable machine learning-based risk assessment model being used in power grid setups. Capitalizing on Python, Oracle SQL and machine learning frameworks, researched to develop defence strategies, comprehensive defence approaches can be crafted addressing the pros and cons of explainable AI to be used in critical power infrastructure operations.

7.2. Technical Challenges in Developing Explainable AI-Based Risk Assessment Systems for Smart Power Grids

The need to develop explainable cyber risk assessment systems on smart power grids presents several challenges primarily owing to the complexity of current communication protocols, network architectures in modern power grids as well as the advanced nature of cyber attacks designed to breach various power infrastructure. A major challenge, therefore, is the complexities of deep learning models being used in the power grid risk assessment design systems. Although these types of models show superior capabilities of identifying advanced cyber threats, they are opaque, which offers a challenge in relation to their translation by power grid operators and security analysts. These models are relatively complex and require advanced methods to obtain meaningful explanations which may come at the cost of either accuracy or interpretability of the model when it comes to critical infrastructure applications. As revealed by Shackleford et al. (2015), the case studies of how NIST Cybersecurity Framework can be implemented in smart grids prove the need to balance the effectiveness of security level with operational transparency needs.

Since the specific needs of the morphed patterns of the contemporary cyber threat targeting power grids cannot be addressed by simple and ineffective ante-hoc explainable mechanisms, the existing literature is concerned with identifying the relevance of the input sample characteristics with the corresponding outputs following different post-hoc explainability mechanisms natively chosen regarding the use in power grid security cases. Most of such approaches are primarily based on broad machine learning usage, and their applicability to AI involving power grid cybersecurity issues cannot apply easily due to the nature of the power grid data, which is characterized by sequential patterns, temporal dependencies, and other categorical features relating to power system operation. Such basic feature attribution and saliency mapping cannot represent the intricate correlation formed by power grid operational and security information.

Furthermore, the issue of advanced persistent threat attacks needs to be taken seriously since threat actors constantly strive to find methods of exploiting elements of power grid risk assessment systems. Robustness in face of adversarial manipulations and the ability to provide explanations that discriminate between authentic threats and adversarial examples is an extremely challenging problem in power grid cybersecurity applications. Explainability under the cybersecurity aspect of the power grid is a two-edged sword. As an example, in case the opponents manage to obtain the insights about the decision model implemented on the power grid premises, they may mislead both the target security scheme and the explainability procedure. The potential adversaries of the explainable cyber risk assessment systems meant to secure power grids present a major problem to security analysts in the advance persistent threats measures.

7.3. Operational and Regulatory Challenges in Smart Power Grid Cybersecurity Implementation

Based on the vast research experience with Python, Oracle SQL, and machine learning frameworks, development of predictive models on historical and real-time grid data to predict high-risk vulnerability in smart grid network necessitates the proper consideration of operational and regulatory restrictions peculiar to power utility settings. The incorporation of machine learning algorithms in the model to simulate the threat, determine the level of the threat, and suggest dynamic mitigation measures should adhere to multiple federal, state, and local rules that regulate the aspect of critical infrastructure protection. Adewole and Salami (2019) state that the effective performance of anomaly detection methods of grid cybersecurity should be ensured by a thorough knowledge of grids regulatory frameworks, and operational necessities.

Smart power grid cybersecurity regulation is rather convoluted and continuously shifting with numerous regulatory agencies and standards organizations putting in their offerings in both direction and mandates. The standards of NERC Critical Infrastructure Protection (CIP) offer required security protection based on cybersecurity to owners and operators of a bulk electric system. National Institute of Standards and Technology (NIST) Cybersecurity Framework is a voluntary document that guides the commercial number of industries on how to address cybersecurity risks that affect critical infrastructure. Other government agencies are the department of energy (DOE) and the department of Homeland security (DHS) who offer guidance and support to power grid cyber security projects. Ling et al. (2020) assert that the data sources needed to perform threat modeling around power systems entail a thorough knowledge of the above-said regulatory frameworks and their implication on machine learning-facilitated security mechanisms.

7.4. Future Research Directions for Machine Learning-Based Power Grid Cybersecurity

The future directions of research on machine learning-based cyber risk assessment systems of smart power grid include some areas of importance that need thorough investigation and evolution. Among the most interesting research avenues is the topic of federated learning applied to cybersecurity in power grids where several utility companies can team up and train machine learning models without exposing sensitive data about their operations. The method solves two problems: the issue of privacy and the demand of big training datasets that could be representative of various power grids. Razvan et al. (2014) state that the analysis and countermeasures in connection with the smart grid cyber-attacks need to be based on collaborative strategies that would allow them to utilize synthetically accumulated knowledge and experience of various stakeholders, with the data privacy and security still being maintained.

The other important topic of research is development of real-time adaptive machine learning models that will be able to learn about the new concepts of cyber threats to power grid infrastructure in an ongoing basis. The conventional machine learning models need to be retrained in regular intervals using new sets of data and this may not be adequate to identify the fast-changing cyber threats. Continuous learning strategies have the potential to allow the machine learning models to change to new pattern of threats in real-time without losing what has been learned in the prior training. As Cavelty (2020) thinks, the politics of cyber security in the critical infrastructure field necessitates procedures that are adaptive enough to respond to the volatile nature of cyber threats and changing strategies of such attacks.

The possible application of quantum computing technologies offers opportunities and challenges to the power grid cybersecurity based on the application of machine learning. The quantum machine learning algorithms could provide substantial benefits in large-scale power grid data processing and might help detect such complex patterns of cyber threats that are hard to detect by classical machine learning algorithms. With the coming of quantum computing however, there are also new threats being posed to the existing cryptographic systems currently in use with the power grid communications. Adeloye (2020) observes that reviewing and evaluation of cybersecurity deployment to smart grids in the power system should have progressive evaluations on new technologies and their potential influence on the security of power grids. Research is also required to craft quantum-resistant security protocols, and research into possible applications of quantum machine learning to power grid cybersecurity use.

The focus on the development of explainable AI methods that are expressly designed to manage power grid cybersecurity is another area of consequential research. Existing models of explainability are mainly transferred to other fields and might not be suitable to reflect the peculiarities and needs of power grid faculties. Explainability methods that apply to power grids must take into consideration the time dynamics of power systems, and the importance of various components of the system, as well as the context in which security actions are taken. The CEN-

CENELEC (2014) puts forward an effort stating that the specifics of smart grid information security necessitate special strategies that consider the peculiarities and demands of the power system functioning.

7.5. Emerging Technologies and Their Impact on Smart Power Grid Security

The introduction of 5G and beyond wireless communication technologies is changing the smart power grid communication world and presented better performance, low latency, and increased connectivity in power grid operation. But such high-tech communication technologies also create new threats and vectors cyber attacks that need to be amended by machine learning risk assessment systems. The 5G networks that have significantly more bandwidth and connectivity introduce more advanced data collection and analysis options, as well as more points of entry of cyber attackers. Tala et al. (2022) state there is a higher need of big data analytics and artificial intelligence issues concerned with privacy and security level in smart grids that need holistic solutions that address the issues introduced by modern-day communication technologies. Based on the gained experience in the research of Python, Oracle SQL, and machine learning frameworks, an extended security frameworks could be created which will take into consideration the special needs of the environment where 5G and beyond communication technologies are supposed to be implemented in power grids and detect the issues that are relevant to such a setting.

In smart power grids, edge computing technologies are being used more to guarantee on-time data processing and decision-making at geographically-distributed assets across the power system. Although edge computing has great benefits over conventional network architectures in reducing latency and streamlining bandwidth consumption, it implies that cybersecurity challenges also exist concerning distributed security processing and edge computing node security. The risk assessment systems based on machine learning will have to be configured to meet the special security needs of the edge computing installations over the power grid systems.

Artificial Intelligence of Things (AIoT) is a combination of artificial intelligence and the Internet of Things technologies with the ability to make intelligent decisions about individual devices across the smart power grid network. Usage of AIoT can complement the functions of machine learning based risk assessment applications by facilitating distributed intelligence and real-time threat detection at each individual IoT device. Nevertheless, the implementation of the AI capabilities on the device level also creates new security issues associated with protecting AI models and algorithms run on the IoT devices that have resource constraints. The non-homogenous deployment of AIoT to the power grid poses another challenge of ensuring uniformity in security and performance between dissimilar devices and different device manufacturers.

The blockchain technologies provide prospective solutions to the security and integrity of information and transactions that are incurred in the working process of smart power grids. Blockchain-based solutions have the potential of offering resistant records of operation in the power grid and ensuring that the various parties within the power grid system will be able to share data safely. Risk assessment systems based on machine learning may use blockchain technologies to increase data and model update integrity and traceability during training. Nevertheless, when being applied to the power grid context, the use of blockchain technologies should pay sufficient attention to the requirements of critical infrastructure applications in terms of performance and scaling.

8. Conclusion

In conclusion, the lack of clarity of the complex AI techniques inculcates the concerns of deep examination of the actions undertaken by the deep learning models implemented in the smart power grid cybersecurity systems. Recently, explainable AI has provided the perspective of white-box models where internal information and decisions made by an AI-based system can be interpreted in context-specific concepts of power grid security applications. Just like in other fields of application, cybersecurity experts were not ready to embrace black-box ML-based cybersecurity options in power grid related applications. It is vital that the security analyst stays a step ahead of the opponent by ensuring that they are sensitive to the inner automatic decision process of the intelligent model and clearly reason input data regarding outputs of this model in the context of power grid security. The use of explainable AI in the power grid cybersecurity may also be a two-edged tool i.e., in addition to enhancing security provisions, it may also turn the intelligent explainable model susceptible to attack by adversaries that aim at sabotaging key power facilities.

The paper reveals that stakeholders of various ML-based cyber risk assessment systems of power grids have different levels and types of interpretabilities of decision models, with a decided majority of interest in feature attribution and the saliency maps on how they can understand their effect on decisions of models in power grid security applications. Nonetheless, the implication of the causality and sensitivity of attributes elements in model interpretability only peculiar to power grid application comes with an information gap. Although explanations tend to highlight significance of various features, offer valuable inputs in examples, visualize the decision boundaries, or use other methods, there exists a necessity to do more subtle and interpretative analysis of attributes to display their relationship to the practical areas of power grid operation. Based on a long research experience as a cybersecurity analyst and knowledge in data analytics with Python, Oracle SQL, and machine learning frameworks, the study developed effective predictive models on past and real-time grid data, and solid smart grid vulnerabilities models that steer high-risk vulnerabilities in the smart grid networks to the knowledge base of the power grid operators and security administrators within power grid functionality.

Although these recent developments have been substantial steps forward in accommodating the need of interpretability, there remain a number of shortcomings with machine learning-driven cyber risk evaluation frameworks targeted at smart power grid; such as a limitation in power grid operational complexities, issue of scalability to incorporate large amounts of power grid data, trade-off between explainability and precision in the context of smart power grid technology focused application, the vulnerability exposure to an advanced persistent threat attack, and absence of standardization issues across varied implementations of the power grid. Addressing such limitations will be vital in future explainable machine learning-based cyber risk assessments of smart power grids that are more powerful and consistent. But more specifically, the increasing issue of an advanced persistent threat attack to an explainable AI model provided is of concern, and accordingly, extra security measurements are taken to confirm the robustness of interpretability tools over threats in power grid applications.

Compliance with ethical standards

Disclosure of conflict of interest

The Authors of this publication declare that there is no known competing personal or financial interest that has influenced this work.

References

[1] Hong, J., Liu, C. C., & Govindarasu, M. (2014). Integrated anomaly detection for cyber security of the substations. IEEE Transactions on Smart Grid, 5(4), 1643-1653. https://doi.org/10.1109/TSG.2013.2294473

[2] Ahmad, T., Chen, H., Guo, Y., & Wang, J. (2018). A comprehensive overview on the data driven and large scale-based approaches for forecasting of building energy demand: A review. Energy and Buildings, 165, 301-320. https://doi.org/10.1016/j.enbuild.2018.01.017

[3] Alcaraz, C., & Zeadally, S. (2015). Critical infrastructure protection: Requirements and challenges for the 21st century. International Journal of Critical Infrastructure Protection, 8, 53-66. https://doi.org/10.1016/j.ijcip.2014.12.002

[4] Sifat, M. M. H., Choudhury, S. M., Das, S. K., Ahamed, M. H., Muyeen, S. M., Hasan, M. M., ... & Badal, M. F. (2022). Towards electric digital twin grid: Technology and framework review. Energy AI 2022; 11: 100213.

[5] Anderson, R., & Fuloria, S. (2010). Security economics and critical national infrastructure. In Economics of Information Security and Privacy (pp. 55-66). Springer. https://doi.org/10.1007/978-1-4419-6967-5_4

[6] Bagaa, M., Taleb, T., Bernabe, J. B., & Skarmeta, A. (2020). A machine learning security framework for IoT systems. IEEE Access, 8, 114066-114077. https://doi.org/10.1109/ACCESS.2020.2996214

[7] Baig, Z. A., Szewczyk, P., Valli, C., Rabadia, P., Hannay, P., Chernyshev, M., ... & Peacock, M. (2017). Future challenges for smart cities: Cyber-security and digital forensics. Digital Investigation, 22, 3-13. https://doi.org/10.1016/j.diin.2017.06.015

[8] Basu, S., Karuppiah, A., Selvakumar, S., Li, K. C., Islam, S. H., Hassan, M. M., & Bhuiyan, M. Z. A. (2018). An intelligent/cognitive model of task scheduling for IoT applications in cloud computing environment. Future Generation Computer Systems, 88, 254-261. https://doi.org/10.1016/j.future.2018.05.056

[9] Bhamare, D., Zolanvari, M., Erbad, A., Jain, R., Khan, K., & Meskin, N. (2020). Cybersecurity for industrial control systems: A survey. Computers & Security, 89, 101677. https://doi.org/10.1016/j.cose.2019.101677

[10] Boyes, H., Hallaq, B., Cunningham, J., & Watson, T. (2018). The industrial internet of things (IIoT): An analysis framework. Computers in Industry, 101, 1-12. https://doi.org/10.1016/j.compind.2018.04.015

[11] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

[12] Nguyen, B., Hayunh, M., Berger, M., Beuran, H., & Awde, A. (2020). Study of smart grid cyber-security, examining architectures, communication networks, cyber-attacks, countermeasure techniques, and challenges. Cybersecurity, 7(1), 10.

[13] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1-58. https://doi.org/10.1145/1541880.1541882

[14] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). https://doi.org/10.1145/2939672.2939785

[15] Cleveland, F. M. (2008). Cyber security issues for advanced metering infrastructure (AMI). In 2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century (pp. 1-5). IEEE. https://doi.org/10.1109/PES.2008.4596535

[16] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297. https://doi.org/10.1007/BF00994018

[17] Deng, R., Xiao, G., Lu, R., Liang, H., & Vasilakos, A. V. (2017). False data injection on state estimation in power systems—Attacks, impacts, and defense: A survey. IEEE Transactions on Industrial Informatics, 13(2), 411-423. https://doi.org/10.1109/TII.2016.2614396

[18] Domingo-Ferrer, J. (2002). A provably secure additive and multiplicative privacy homomorphism. In International Conference on Information Security (pp. 471-483). Springer. https://doi.org/10.1007/3-540-45811-5_37

[19] Erol-Kantarci, M., & Mouftah, H. T. (2015). Energy-efficient information and communication infrastructures in the smart grid: A survey on interactions and open issues. IEEE Communications Surveys & Tutorials, 17(1), 179-197. https://doi.org/10.1109/COMST.2014.2341600

[20] Farhangi, H. (2010). The path of the smart grid. IEEE Power and Energy Magazine, 8(1), 18-28. https://doi.org/10.1109/MPE.2009.934876

[21] Gao, J., Xiao, Y., Liu, J., Liang, W., & Chen, C. L. P. (2012). A survey of communication/networking in smart grids. Future Generation Computer Systems, 28(2), 391-404. https://doi.org/10.1016/j.future.2011.04.014

[22] Gungor, V. C., Sahin, D., Kocak, T., Ergut, S., Buccella, C., Cecati, C., & Hancke, G. P. (2011). Smart grid technologies: Communication technologies and standards. IEEE Transactions on Industrial Informatics, 7(4), 529-539. https://doi.org/10.1109/TII.2011.2166794

[23] He, Y., Mendis, G. J., & Wei, J. (2017). Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism. IEEE Transactions on Smart Grid, 8(5), 2505-2516. https://doi.org/10.1109/TSG.2017.2703842

[24] Erdogan, G., Sperstad, I. B., Garau, M., Gjerde, O., Tøndel, I. A., Tokas, S., & Jaatun, M. G. (2022, July). Adapting Cyber-Risk Assessment for the Planning of Cyber-Physical Smart Grids Based on Industrial Needs. In International Conference on Software Technologies (pp. 98-121). Cham: Springer Nature Switzerland.

[25] Hink, R. C. B., Beaver, J. M., Buckner, M. A., Morris, T., Adhikari, U., & Pan, S. (2014). Machine learning for power system disturbance and cyber-attack discrimination. In 2014 7th International Symposium on Resilient Control Systems (pp. 1-8). IEEE. https://doi.org/10.1109/ISRCS.2014.6900095

[26] Hodo, E., Bellekens, X., Hamilton, A., Dubouilh, P. L., Iorkyase, E., Tachtatzis, C., & Atkinson, R. (2016). Threat analysis of IoT networks using artificial neural network intrusion detection system. In 2016 International Symposium on Networks, Computers and Communications (pp. 1-6). IEEE. https://doi.org/10.1109/ISNCC.2016.7746067

[27] Mitchell, T. M. (1997). Machine learning. McGraw-Hill. http://profsite.um.ac.ir/~monsefi/machine-learning/pdf/Machine-Learning-Tom-Mitchell.pdf

[28] McLaughlin, S., Konstantinou, C., Wang, X., Davi, L., Sadeghi, A. R., Maniatakos, M., & Karri, R. (2016). The cybersecurity landscape in industrial control systems. Proceedings of the IEEE, 104(5), 1039-1057. https://doi.org/10.1109/JPROC.2015.2512235

[29] Liu, Y., Ning, P., & Reiter, M. K. (2011). False data injection attacks against state estimation in electric power grids. ACM Transactions on Information and System Security, 14(1), 1-33. https://doi.org/10.1145/1952982.1952995.

[30] Kang, M. J., & Kang, J. W. (2016). Intrusion detection system using deep neural network for in-vehicle network security. PloS One, 11(6), e0155781. https://doi.org/10.1371/journal.pone.0155781

[31] Baumeister, T. (2010). Literature review on smart grid cyber security. Collaborative Software Development Laboratory, University of Hawaii. http://csdl.ics.hawaii.edu/techreports/2010/10-11/10-11.pdf

[32] Zhang, Y., Wang, L., Sun, W., Green, R. C., & Alam, M. (2011). Distributed intrusion detection system in a multi-layer network architecture of smart grids. IEEE Transactions on Smart Grid, 2(4), 796-808. http://ieeexplore.ieee.org/document/6064346/

[33] Ghadi, Y. Y., Mazhar, T., Aurangzeb, K., Haq, I., Shahzad, T., Laghari, A. A., & Anwar, M. S. (2022). Security risk models against attacks in smart grid using big data and artificial intelligence. PeerJ Computer Science, 10, e1840. https://pmc.ncbi.nlm.nih.gov/articles/PMC11057646/

[34] Jeje, M. O. (2010). Cybersecurity Assessment of Smart Grid Exposure Using a Machine Learning Based Approach. arXiv. https://arxiv.org/html/2501.14175v1

[35] Zibaeirad, A., Koleini, F., Bi, S., Hou, T., & Wang, T. (2021). A comprehensive survey on the security of smart grid: Challenges, mitigations, and future research opportunities. arXiv. https://arxiv.org/pdf/2407.07966.pdf

[36] Wang, C., Wang, Q., & Zhou, Q. (2021). Cybersecurity risk assessment model for power grids based on dynamic Bayesian networks. Energies, 14(12), 3456. https://www.mdpi.com/1996-1073/14/12/3456

[37] Berghout, T., et al. (2022). Machine learning for cybersecurity in smart grids. Journal of Information Security and Applications, 66, 103178. https://www.sciencedirect.com/science/article/am/pii/S1874548222000348

[38] Farooq, A., et al. (2021). Securing the green grid: A data anomaly detection method for renewable energy systems using hybrid machine learning. Journal of Renewable and Sustainable Energy. https://www.sciencedirect.com/science/article/pii/S1874548224000350

[39] Paul, B., & Adhikary, A. (2021). Potential smart grid vulnerabilities to cyber attacks. Heliyon, 7(7), e07619. https://www.sciencedirect.com/science/article/pii/S240584402414011X

[40] Kumar, A., & Zhang, L. (2019). Intrusion detection for cybersecurity of smart meters. IEEE Transactions on Smart Grid, 10(4), 3890-3899. https://e-tarjome.com/storage/panel/fileuploads/2021-01-04/1609777922_gh209.pdf

[41] Hasan, M. K., Abdulkadir, R. A., Islam, S., Gadekallu, T. R., & Safie, N. (2019). A review on machine learning techniques for secured cyber-physical systems in smart grid networks. Energy Reports, 11, 1268-1290.

[42] NIST. (2014). Smart grid cybersecurity strategy, architecture, and high-level requirements. National Institute of Standards and Technology. https://csrc.nist.gov/files/pubs/ir/7628/r1/final/docs/draft_nistir_7628_r1_vol1.pdf

[43] Kim, T., & Poor, H. V. (2011). Strategic cyber-physical attacks on power grids. IEEE Transactions on Smart Grid, 2(4), 667-674. https://ieeexplore.ieee.org/document/6035892

[44] Nejabatkhah, F., et al. (2022). Security requirements and practices for smart grids. KTH Royal Institute of Technology. https://kth.diva-portal.org/smash/get/diva2:1647499/FULLTEXT01.pdf

[45] Adewole, L. O., & Alghazzawi, D. (2021). Hybrid machine learning models for enhancing cybersecurity in smart grids. International Journal of Research and Innovation in Social Science, 5(4), 4344-4351. https://rsisinternational.org/journals/ijriss/Digital-Library/volume-9-issue-4/4344-4351.pdf

[46] Alam, K., & Baharudin, A. S. (2021). Predictive analytics and risk assessment in smart grid: A comprehensive survey. Energies, 14(8), 2096. https://www.mdpi.com/1996-1073/14/8/2096/pdf

[47] Shackleford, D., et al. (2015). Case study of NIST Cybersecurity Framework implementation in smart grids. NIST. https://pmc.ncbi.nlm.nih.gov/articles/PMC11057646/

[48] Khalil, S. M., et al. (2021). Threat modeling of cyber-physical systems: A case study in power systems. Computers & Security, 113, 102540. https://www.sciencedirect.com/science/article/pii/S016740482200342X

[49] Adewole, L. O., & Salami, A. F. (2019). Anomaly detection strategies for grid cybersecurity. IEEE International Conference on Smart Grid Communications. https://www.jocm.us/2025/JCM-V20N2-221.pdf

[50] Ling, E., et al. (2020). Information sources for threat modeling in the power systems domain. PeerJ Computer Science, 6:e1840. https://pmc.ncbi.nlm.nih.gov/articles/PMC7948005/

[51] Sari, A., & Butun, I. (2021). Early detection and recovery measures for smart-grid cyber-resilience. Elsevier Journal of Power Sources. https://www.diva-portal.org/smash/get/diva2:1622006/FULLTEXT01.pdf

[52] Razvan, A., et al. (2014). Smart grid cyber-attack analysis and countermeasures. Journal of Communications and Networks, 16(2), 1-9. https://www.jaist.ac.jp/~razvan/publications/smart_grid_attack_analysis_countermeasures.pdf

[53] Cavelty, M. D. (2020). Cyber security politics: Socio-technological uncertainty and the risk to critical infrastructure. Routledge. https://library.oapen.org/bitstream/id/20a53302-dee5-4834-9d98-8f9c07f0a602/9781000567113.pdf

[54] Adeloye, A. A. (2020). Assessment of cybersecurity deployment to power system smart grid. Lagos Journal of Engineering Research, 2(1), 11-22. https://journalspress.com/LJER_Volume23/Assessment-of-Cybersecurity-Deployment-to-Power-System-Smart-Grid.pdf

[55] CEN-CENELEC (2014). Smart grid information security. SG-CG/M490/H. https://www.cencenelec.eu/media/CEN-CENELEC/AreasOfWork/CEN-CENELEC_Topics/Smart%20Grids%20and%20Meters/Smart%20Grids/7_sgcg_sgis_report.pdf

[56] Khan, M. A., Hussain, S., & Tziritas, G. (2021). Security risk modeling in smart grid critical infrastructure. Sustainability, 13(6), 3196. https://www.mdpi.com/2071-1050/13/6/3196

[57] Tala, T. K., Hadjar, O. S., & Naima, K. (2022). Cyber-security of smart grids: attacks, detection, countermeasure techniques, and future directions. Communications and Network, 14(04), 119-170. https://www.researchgate.net/publication/365680971_Cyber-Security_of_Smart_Grids_Attacks_Detection_Countermeasure_Techniques_and_Future_Directions