

Ensemble learning based fault detection using PMU data in imbalanced data condition

Kiruthika Krishnan¹, Srivani Iyengar²

¹Department of Electrical Engineering, Rajarajeshwari College of Engineering, Bangalore, India

²Department of Electrical Engineering, R. V. College of Engineering, Bangalore, India

Article Info

Article history:

Received Jan 11, 2024

Revised Mar 7, 2025

Accepted Mar 29, 2025

Keywords:

AUC-ROC curve

Ensemble learning

Logistic regression

PMU fault detection

SMOTE

Stack ensemble learning

ABSTRACT

Significant advancements in the electrical grid include enhanced regulation, communication, metering, and customer interaction, driven by information communication technologies (ICTs) and cyber-physical systems (CPS). The adaptation of synchro phasor devices like phasor measurement units (PMUs) enables real-time monitoring and control, aiding in power system security assessment. PMUs record voltage and current phasors with GPS time stamps, transmitting data to phasor data concentrators (PDCs) for decision-making. However, ensuring the stability and security of this method against cybersecurity threats is crucial due to its reliance on Internet Protocol (IP) networks. Dynamic security assessment utilizes PMU data, reported up to 30–60 times per second, to evaluate power system safety. To address security issues, a Python-based fault detection system employing a stack ensemble learning algorithm is developed. This approach consistently outperforms traditional methods, producing satisfactory results with superior AUC-ROC curves, validated through correctness checks and graphical analysis. The dataset includes both natural and man-made security threats, facilitating comprehensive assessment and mitigation strategies. The ensemble learning algorithm performed better than the individual algorithms by obtaining 95% in the AUC-ROC curve.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Kiruthika Krishnan

Department of Electrical Engineering, Rajarajeshwari College of Engineering

Ramohalli Cross, Mysore Rd, Kumbalgodu, Bangalore, Karnataka 560074, India

Email: kiruthi.km21@gmail.com

1. INTRODUCTION

Cyberattacks originating from global hackers have had a substantial impact on the financial well-being of individuals. In recent times, cyberattacks have emerged as a significant tool for disrupting various government agencies. This threat extends to power systems that utilize phasor measurement units (PMUs), as these devices are connected to the internet. Notably, machine learning algorithms have displayed a crucial role in advancing the detection of cyberattacks within the power system environment. A report on cyber-attack suggests that 26% of the incidents of cyber-attacks are based on spear phishing, a way of making the victim believe that the mail is coming from the original source, but actually not from the original source [1]. A record of 159700 cyber incidents occurred in the year 2017 alone. A cyber-physical system is created by combining cyber systems with physical power grids in modern smart grids. Time-synchronized measurement data is transmitted to the cyber system from the physical grid by using phasor measurement units (PMUs). By returning the required commands to the PMUs, the system operators (SO) at the cyber band evaluate both the on and offline formats of the generated data and guarantee the grid's reliability and security. Nevertheless, a variety of physical occurrences, including cyberattacks, frequency events, transformer events, and line-to-

ground failures, can result in deviations in the measurements that the SO receives, which is known as "bad data." The SO may then choose the incorrect restorative or mitigation approach as a result of these inaccurate data. For the grid to operate safely and optimally, precise bad data detection and identification of the proper bad data kind are therefore essential. Microgrid integration with smart infrastructure, such as sensors, communication, and monitoring devices, has led to the evolution of the concept of the smart grid (SG), which offers additional levels of robustness, dependability, and efficient operation. The physical layer of an MG is made up of interconnected components, while the cyber layer is made up of interconnected smart communication systems, metering devices, and monitoring equipment. Since the physical layer is constructed upon the cyber layer, the entire SG is a cyber-physical system (CPS). Phasor measurement units (PMU), which offer quicker reporting rates and more dependable and more secure system monitoring than traditional supervisory control and data acquisition (SCADA) systems, are one of the crucial parts of the SG. However, the introduction of smart devices like PMUs into MGs necessitates safe data storage and processing techniques and increasing reliance on communication linkages across the many CPS levels. Cyberattack risk is increased by this reliance on communication channels and data storage systems, especially for vital infrastructures like data centers, hospitals, and military installations. The fundamental idea behind the SG is modernization of electrical network through integration of artificial intelligence, signal processing, improved automatic control, communications, and information technology. Various levels of the grid are monitored by the smart meter, FDR, PMU, SCADA, WAMS, and other monitoring and measuring systems. As a result, a smart grid must store and distribute enormous amounts of real-time data among users, control centers, and utilities. As a result, data analytics will be very helpful for processing and evaluating this enormous volume of power system data [2].

Major blackouts that have occurred in several power systems throughout the globe have made it clear how valuable PMU data is, and installing PMUs on the networks of power transmission that relate to most major power systems has become a crucial present endeavor. This article addresses the applications of wide-area measurement system (WAMS) and PMU technology for better power network monitoring, protection, and control. It also offers a brief introduction to these technologies [3]. The potentials of wide area technologies i.e. wide area monitoring, protection, and control, or WAMPAC—are discussed in this study. PMUs must be positioned appropriately based on the real-time application as WAMPAC deployment necessitates distributed phasor measurements across the system. The purpose of phasor measurement unit (PMU) technology and its use in the power system are discussed in this study [4]. This literature presents the results of an experimental study that shows how malicious assaults affect the PMUs in smart grids. A simulated attack environment architecture is suggested, and a physical test-bed equipped with a network attack environment, complete with mainstream PMUs, is established. Tests for cyberattacks are conducted, including deceitful communication and interference with GPS signals. The experimental study findings have shown PMU's weaknesses and vulnerabilities to malicious assaults.

Additionally, the foundational research for improving cybersecurity protection for wide area measurement systems (WAMS) in smart grids has been established [5]. This work presents a unique density-based spatial clustering method for data manipulation assaults on PMU measurements, including online detection, classification, and data recovery. The suggested approach is entirely data-driven and can handle many measurement assaults at once without adding more hardware to the current setup. Additionally, the suggested method does not rely on traditional state estimate [6]. Using the subspace identification approach, a data-guided design scheme of untraceable fake data-inculcation attacks to cyber-physical systems is initially presented in this work. Next, by solving a restricted optimization problem and considering the limitations of energy limitation and undetectability, the effects of unnoticeable bad data-inducing assaults assessed. Furthermore, coding theory is used to study the detection of planned data-driven fake data-inducing attacks. Ultimately, simulations conducted on a flying vehicle model are shown to confirm the efficacy of the suggested techniques [7]. Attackers might use the communication flaw in wide-area monitoring systems (WAMS) to target WAMS records with malicious data integrity assaults, which could have disastrous results. In response to the cybersecurity issues brought to light by WAMS, specific machine learning-based methods have recently been created to verify the source information of WAMS data. Most of the methods of source authentication now in use aim to verify WAMS data from a limited quantity of sites distributed across an expansive geographic region, which could not fully reflect WAMS's operating condition in real-world networks. This study's objective is to ascertain if machine learning-based methods can be used to real-world power grids in order to develop reliable source authentication of WAMS data. Four machine learning-based "state-of-the-art" techniques that combine shallow and deep learning [8]. To make the smart grid completely visible, the not many PMUs are arranged in phases in line with intended reinforcement-based learning method. Most susceptible buses that have the ability of getting compromised by adjusting the fewest amount of measures are identified using a multistage optimum PMU placement method that combines a least-effort attack model with a reinforcement learning technique [9]. The literature offers an innovative strategy for creating and identifying threats to data integrity in smart grids. Additionally, it offers a way to optimize the creation of FDIA against

the control center's state estimation techniques. The technique for generating AC state estimate attacks with both entire and partial information is presented together with DC state estimation assaults. It also recommends incorporating methods for the voting-based ensemble learning approach (MVCC) to detect FDIA in smart grids. Next, a 39 bus New England system and an IEEE 24 bus system are employed as test systems for the model, and fictitious data injection attacks are created and detected. The detection approach is compared against ensemble methods, classical weighted least squares, and most modern machine learning algorithms currently in use [10].

A real-time sequential approach for detecting and classifying faulty data was presented in the literature. Initially, the Hankel-matrix's low rank characteristic to quickly identify erroneous data is implemented. Second, step is to categorize malicious data into two groups: real-world occurrences and online attacks. The method utilizes the multi-channel Hankel-matrix's low rank approximation error before to and later random column permutations on going physical events. In the improbable event that compromised data is discovered to be the result of a cyberattack, our suggested method then moves on to identify the source of attacks. Two potential cyberattack routes examined are GPS spoofing and fake data injection attacks (FDIA and GSA). To differentiate between them, the approach leverages the rank-1 closeness error of the single-channel Hankel matrix with unwrapped phase angle data [11]. Applying machine learning-based approaches to PMU data is one of the most crucial attack detection measures. Analyzing the residue of the observers and estimators is another method. Using PMU data, this study attempts to detect assaults on power systems using both techniques. The style of attack, such as man-in-the-middle (MitM) or a potential denial-of-service (DoS), is identified using an algorithm. Lastly, the suggested procedure is replicated using an example IEEE power system, and encouraging outcomes that confirm the approach's effectiveness is explained in detail [12].

A new voting-based technique for detection for systemic cyber intrusions is developed within this work. The attacker in the cyberattack under examination introduces a span of false data in PMU in an attempt, replicate fictitious short circuit occurrences in the system. The suggested spotting procedure makes use of a variety of machine learning (ML) techniques, such as ensemble learning, recurrent neural network (RNN), feedforward neural network (FNN), decision trees, discriminant analysis, k-nearest neighbors (KNN) classification, support vector machine (SVM), and naive bayes. By determining the average output depending on detector performance, the voting-based technique may be able to differentiate between FDI attacks and actual short circuit failures. To minimize redundancy and enhance relevance, the mechanical and electrical components of the system are optimally selected for training objectives [13]. Literature presents an ensemble bagged tree for relatively accurate real-time attack and defect detection. This suggested structure is predicated on data from the phasor measurement unit (PMU) and relays during normal, cyberattack, and failure conditions. This study compares the effectiveness of the recommended method against several machine learning methods and validates it in a MATLAB/Simulink testing environment [14]. Literature has conducted a comprehensive investigation of big data analytics applications, current issues, and solutions [15].

Intrusion detection systems (IDS) are critical to oversee the security of cyber-physical energy and power systems present in SG with increasing machine-to-machine connections. Still, IDS is finding it very challenging to reliably differentiate between benign and malevolent events due to the many-sourced, large, linked, and often noise-containing unwanted data that saves a range of concurrent cyber and physical activity. To deal with these and similar issues, here, a robust start-to-finish framework in line with the ensemble machine learning and stacked denoising autoencoder (SDAE) to extract new characteristic sets informed by attacks and noise from cyber-physical system data and integrate multiple information sources for authentic event categorization. The put forwarded methodology influences stochastic difference of anomaly extraction (SDAE) to first provide smaller-dimensional attributes that permit the restoration of a clutter-free input from clutter-damaged perturbations. Novel characteristics that will maintain and update information as normal, fault, and attack events against a range of synthetic attack data, with the goal of improving categorization by integrating attack and noisy inputs. In addition, ensemble learning-based multi-classifier classification, considering the heterogeneous nature of the inputs such as PMU measurements, system logs, and IDS alerts, and classifying the specimens based on the SDAE-extracted characteristics using the extreme gradient boosting (XGBoost) technique, is developed. Moreover, normalization and oversampling were used to enhance the data's balance and homogeneity. The present SDAE+XGBoost approach attains more than 90% classification correctness on a practical dataset comprising 37 sub-types of normal, fault, and attack obtained via co-simulations on a hardware-in-the-loop (HIL) testbed security testbed [16]. Specifically, in the context of cyber threats, a unique "greedy" method for PMU placement is developed. According to this research, cyber risk may greatly raise a power system's unobservability risk, necessitating the inclusion of PMU allocations [17].

To develop a complete architecture that is resistant to cyberattacks and uses strategically positioned phasor measurement units (PMUs) to counteract structural weaknesses in smart grids, a brand-new hybrid betweenness centrality (HBC) metric is put forth that successfully pinpoints a system's most crucial lines. A distinct objective function is created with the purpose of deliberately inserting PMUs into the system in order to strengthen its defenses against any attacks by bogus data injection on these susceptible lines. Finding the

best PMU location results in the fewest sets of measurements required to defend the state variables against all kinds of attacks on data integrity. This design's effectiveness is illustrated with the IEEE 14-bus system [18]. Data-driven hacking techniques like the fake data injection attack (FDIA) seriously jeopardize the states of the grid. Literature [19] presents an effective formulation method for blind FDIA that requires exact measurement subspace information. In order to allay this worry, efficient implementation of new, robust, nonlinear deep learning models that can, in addition to effectively detecting the existence of blind attack intrusions in real time, pinpoint their precise locations is to be implemented. These versions can work in conjunction with conventional bad data detectors to offer a practical and affordable solution. By identifying the discrepancy with the co-occurrence dependency of the attack vectors added to the raw data, these neural network models also demonstrate a multilabel classification technique. Moreover, it is demonstrated that these deep learning structures are model-free, suggesting that assaults might be identified without requiring statistical knowledge of the grid. On the standard IEEE test bench, the suggested framework's performance is assessed under a range of assault and noise scenarios [19].

One of the biggest risks to the safety, dependability, and cost-effective operation of power systems nowadays is cyberattacks. It is challenging to identify and classify various cyberattacks while maintaining the stability and security of the power infrastructure. An automated technique based on the convolutional neural network for the recognition and categorization of various cyberattacks to address this problem. The convolutional neural network collects temporal information and spatial interactions between various nodes from the prior operational state of the sent data packets. The suggested structure's capsules have significant effects on preserving the measurement matrix's topological consistency. Additionally, the suggested approach eliminates the influence of uncertainty in system characteristics on detection performance and is model-free. In this study, many types of common cyberattacks are examined and modeled, such as replay, denial of service, bogus data injection, time-delay, and deception assaults. The suggested solution may achieve 99.97% detection accuracy on a single cyberattack and 96.25% detection accuracy on multiple cyberattacks, according to numerical findings on the IEEE 39-bus test system. The results of comparison show that the suggested approach performs better than conventional neural networks. The issue of multiple attacks detection and categorization is resolved by this technique [20]. Literature aims to carefully explain several techniques and processes for cyber-security in energy systems and examine relevant solution approaches. Additionally, a technical examination and debate of the traits and relevance of several cyber-attack models is carried out. The most recent research topics are discussed, along with cutting-edge cyber security methods for power systems and super grids, such blockchain and quantum computing. The talk covers essential protection mechanisms and problem-solving strategies. Finally, some thoughts on SGs' cyber-security in the future are expressed [21]. The new power system will face significant risk and security concerns because of the extensive integration of cyber and physical systems. To tackle this issue, a game-theoretic optimum defense resource allocation strategy is put forth to proactively guard against possible cyberattacks on smart grids. Using this technique, an ideal resource allocation for a 2-layer game model is produced. While the other tier involves several defense nodes in a noncooperative game, the first layer involves attackers and defensive nodes in an evolutionary game. After analyzing the offensive and defensive evolution outcomes of every scenario, a solution to the multi-node resource allocation problem is generated. In contrast to earlier research, the attacker's constrained rationality is considered based on the integrity, usability, and confidentiality indices. To measure player gains, quantum response equalization is added in the interim. Lastly, algorithms are used to show that the approach suggested in this research is both practical and efficient [22].

Distributed denial-of-service (DDoS) assaults are one kind of cyberattack that frequently targets smart grid networks. Furthermore, synchro phasor technology shields the wide-area measurement system (WAMS) from complicated difficult situations by managing several concerns in a grid. Because of communication protocols, vendor restrictions, and the complexity of the assault, detecting DDoS attacks is difficult. For measurement PMU data, attackers target the phasor data concentrator (PDC) database in WAMS. Nevertheless, design makes sure that the end application makes use of the regular PDC data stream even during the intrusion. PMU-generated data in WAMS is quickly verified using the suggested attack detection technique. Various machine learning algorithms are employed to identify DDoS assaults; yet the most effective detection model remains unclaimed. This study aims to determine (a) the best machine learning method for detecting DDoS attacks and (b) the accuracy of the algorithms that are taught. This study offers a hybrid approach based on machine learning that yields 83.23% accuracy. The suggested model is created using the Python compiler, and the outcome demonstrates how effectively the suggested detection technique raises the accuracy of DDoS assault detection [23]. The research suggests a novel approach combining data monitoring and fuzzy machine learning model classification to identify smart grid problems. Here, data from the smart grid has been tracked using improved smart sensor metering, which runs in the cloud at the edge of the network. Afterwards, the monitored data was categorized using an adversarial neural network with fuzzy reinforcement encoder. Throughput, mean average precision, accuracy, scalability, and dependability are all taken into consideration

while doing experimental study. Improved monitoring and prediction approaches can boost the existing grid's potential utilization while reducing fault frequency. The suggested method achieved 93% accuracy, 94% throughput, 81% dependability, 89% mean average precision, and 94% scalability [24]. The literature [25] offers an aggregated integer linear programming method utilizing micro-synchro phasor unit placement for machine learning to accomplish complete observability of the automated smart grids, keeping in mind that the distribution systems are reconfigurable. The suggested stochastic approach provides a multi-stage mechanism for micro-synchro phasor unit placing depending on the demand and load size of the system, in addition to pre-planned sectionalizing and tie switches. This method may also be used to apply the no-injection limits of the representation to narrow finding space for the issue. In addition, a new approach derived from the whale optimization technique (WOM) is presented for finding best design for every phase while concurrently increasing the reliability indices and lowering the expenses associated with customer disruptions and power outages. The WOM heuristic serves as the foundation for the restructuring process, and an integer linear programming framework is used to arrange the micro-synchro phasors. To manage the uncertainty effects, a stochastic framework derived from on point estimation is constructed, accounting for prediction error or metering device uncertainty.

The suggested approach ensures distinctness of the distribution network both before and after rearrangement within allotted time limit, as confirmed by simulation and numerical results on an actual system. Additionally, the results demonstrate that even when the system is subjected to various reconfigurations and topologies, system observability may be assured at varying load levels [25]. The anomaly detection and identification module (ADIM), a unique component proposed in the study, is designed to identify anomalies or erroneous data points before the state estimation process begins. We provide a deep learning technique that demonstrates exceptional precision in detecting irregularities within an ongoing data stream. Comprehensive testing on a range of test scenarios that effectively cover a variety of network topologies, transformer types, and load conditions demonstrates the capability of ADIM. It is demonstrated that anomalies may be efficiently found and recognized with ADIM, hence lowering the requirement for the component that detects faulty data and enhancing state estimation's overall responsiveness. Our study establishes the groundwork for creating an anomaly system for power system measurements that is based on detection and identification [26]. The data sources and SG architecture are briefly summarized in the paper's first part. Furthermore, examples of fraudulent data attacks and data security requirements are shown. The most recent ML-based detection methods are then summarized using the three primary detection scenarios: non-technical losses, state estimation, and load forecasting. Finally, considering the limitations of the current machine learning-based techniques, we investigate additional research opportunities at the conclusion of the project. We specifically cover intrusion detection against hostile assaults, a cooperative and decentralized detection framework, privacy-preserving detection, and a few possible cutting-edge machine learning algorithms [27]. A hybrid deep learning system that targets denial-of-service assaults on the Smart Grid's communication network. The gated recurrent unit and convolutional neural network approaches hybridize the suggested approach. The benchmark cyber security dataset from the Canadian Institute of Cybersecurity Intrusion Detection System is used in simulations. With a comprehensive accuracy rate of 99.7%, the simulation results show that the suggested approach works better than the existing intrusion detection systems [28].

There are several obstacles to the cyber security of the smart grid, most of them stem from malevolent assaults on devices connected to the system. These attacks could try to jeopardize the privacy of PMU and IP camera sensor data, or they might try to undermine the power supply to specific customers. Therefore, using security measures like network intrusion prevention systems (NIPS) and firewalls equipped with each distributed system linked into the grid is the easiest method to prevent such cyber-attack issues. The grid is nearly impenetrably protected from hackers thanks to several attack and defense mechanisms. Thus, employing FDI and MitM attack scenarios, the aim of this study is to give an analysis of PMU and IP camera sensor assaults at the component level when coupled to an IEEE 13-node system. Grid Attack Analyzer, a smart grid attack analysis tool, is used to collect data and simulate the research [29]. Such attacks are concealed by covertly modifying the SCADA and PMU metrics. This paper investigates cyber-physical assaults that are covert and target power systems. By flicking the corresponding switches or circuit breakers, one or many lines and buses can be interrupted during one of these assaults. The research develops a framework based on the non-linear power flow model to characterize such attacks and suggests a method for spotting such cyber-physical attacks using switching transients. The detection technique takes use of the fact that a physical separation will result in a high transient rate for the system. These PMU-observed transient components are utilized to detect covert line disconnection and bus blackout assaults, as well as to validate the correctness of the steady-state values of SCADA and PMU measurements. The suggested method may be able to identify cyber-physical assaults that mask frequent line disconnections and bus failures under various load scenarios, according to experiments conducted on the IEEE 30 bus system [30].

The literature discusses the challenge of guaranteeing precise online transient stability prediction in contemporary power systems, which rely increasingly heavily on smart grid technologies and are thus more

vulnerable to cyberattacks. Despite the rapid growth of technology, machine learning algorithms for stability prediction presently lack the resilience needed to effectively resist the complex and ever-evolving nature of cyberattacks. The study also evaluates the impact of topological modifications and the incorporation of renewable energy on these machine learning-based techniques, as well as cyberattacks. Transient stability prediction techniques used online are essential for tracking and predicting power system behavior in real time during disruptions. To assess the robustness of the proposed algorithms in the context of the potential for attackers to disrupt communication and hence affect the power system, the study reproduces many scenarios. The results show that machine learning algorithms perform worse during cyberattacks, resulting in a large drop in the correctness of transient stability forecasts when compared to normal working settings. This demonstrates how vital it is to have cutting-edge cybersecurity defenses to maintain power systems' capacity for prediction [31]. Today's cyber-physical grids are heavily integrating cost-effective communication networks and the internet of things (IoT), which has led to serious security problems. More specifically, network security becomes more vulnerable due to wireless communication technology. In addition to the well-researched cybersecurity challenges, we also need to consider physical layer security. As a result, a lot of work has gone into creating a solution to cope with cybersecurity problems. But there hasn't been much work done on creating encroachment finding systems for physical security. This work provides a sharp model that detects and classifies assaults using a combination of machine learning techniques, including identifying the kind of attack at the physical band. Additionally, the suggested method localizes the attack or vulnerability to certain system parameters or attributes, which can assist cybersecurity specialists in reducing the effect of attacks on communication grids.

The proposed model is compared with conventional machine learning classifiers using an SG dataset that is simulated at Oak Ridge National Laboratories. By dividing the data and calculating the comparison between the confined metrics generated by the suggested model, the confinement of errors and attacks is verified. When contrast to peer methodologies, the results show how good this method is at classifying threats and confining them [32]. The latest wave of resilient fake data injection attack methods necessitates a deep comprehension of the associated power grid network's structure. This study suggests three methods that are independent of network architecture for introducing fictitious data into the smart grid. These methods include delta thresholds, linear regression, and linear regression with timestamp. It is intended to close the gaps in real-time data measurements, hence increasing the probability that tampered data won't be detected. Modern defense strategies like AC state estimation, temporal behaviors based on false data detection, bad data detection, and SVM demonstrate the resilience of the suggested attack approaches [33].

Based on the attacker's perspective and using the PMU as the attack-defense target in the power system, a multi-level game model for FDIA is suggested. In a multi-stage game, special attention is paid to data manipulation, strategic modifications, and multi-path attacks. The PMU setup from the intruder's view is used to generate fake data, the strike range is optimized, and the attack repercussions are assessed. Second, the Nash equilibrium point may be established by using the ideas of two-player zero-sum game theory and accounting for both total income and multi-path attacks in multi-stage games to determine the best attack-defense combination. Lastly, a discussion of the experiment findings for both single- and multi-stage games follows.

The simulation's findings show that attackers can more effectively and efficiently employ the suggested multi-stage game approach [34]. Literature presents a realistic bi-level mixed-integer linear programming (BMILP) model to replicate fraudulent data injections (FDIs), which attempt to overload several lines of transmission and create a collapse of power supply in very big grids. In contrast to previous studies, this model accounts for the possibility that attackers may only have restricted access to measurement buses and simulates problems on certain lines that are overlooked by current DC state estimation. Furthermore, it is demonstrated that stealthy FDIs may be identified using an observation framework based on recursive weighted least-squares (WLS) state estimation, but classical WLS estimation is unable to detect FDIs. This helps protect the system against many kinds of threats. Two benchmarks are utilized to verify the efficacy of the suggested attack model and detection system on the real grid: the IEEE 118-bus benchmark and a 2000-bus artificial grid that mimics the Texas, USA, electrical network [35]. Literature [36] proposes a machine learning method that uses point of common coupling (PCC) sensors alone to identify cyberattacks in photovoltaic (PV) farms. First, a thorough cyber-attack model for a photovoltaic farm is created, considering the variety of operational circumstances. Two cyberattack kinds that are often harder to identify are particularly included in the attack model.

We present and contrast a convolutional neural network (CNN) using micro-phase measurement units (μ PMU) and raw electric waveform with figures of merit in relation to existing machine learning techniques. In the end, a distributed grid consisting of IEEE 37 buses and solar farms is established as a testbed for cyber-physical security. A framework for real-time simulation, detection, and visualization is created to show how the suggested approach works in an actual setting. The findings demonstrate that the suggested machine learning techniques are capable of achieving sufficient resilience and detection accuracy in a range of assault situations [36]. Literature [37] builds an IEEE-118-node power network and a 200-node scale-free information network

using the concept of an intricate grid to create a self-sustaining model. Following the process of information cataloging for detecting and ranking the network's junctions, a vulnerability analysis is conducted, and a stream model of the power cyber-physical system is built based on the node carrying capacity. Simulation demonstrates that intentional assaults cause the system structure to break down more quickly than random attacks do, and that failures may be successfully stopped from spreading by raising the node threshold. [37]. Literature [38] put forward a "quantum dwarf mongoose optimization with ensemble deep learning based intrusion detection (QDMO-EDLID) technique "in the context of CPS. The QDMO-EDLID method that is being discussed uses ensemble learning and feature selection (FS) to identify intrusions. The QDMO algorithm is used by the QDMO-EDLID method to choose feature subsets. Additionally, a mix of deep belief networks (DBN), convolution residual networks (CRN), and deep autoencoder (DAE) models is utilized to categorize intrusions. Using benchmark intrusion datasets, the QDMO-EDLID technique's experimental results were evaluated.

The simulation's outcomes demonstrated the QDMO-EDLID approach's increased effectiveness in relation to several performance metrics [38]. It is evident from the preceding paper that a large body of research has demonstrated how the SMOTE and Ensemble learning algorithms enhance categorization. The subject of cyberattack detection and classification is addressed in this paper by the application of the ensemble learning method and the SMOTE sampling strategy. The ensemble learning algorithm is used to compare a variety of methods, including logistic regression, multilayer perceptron modeling, support vector classifiers, and decision tree classifiers. The technique of ensemble learning uses SVC, MLP, logistic regression, and decision tree algorithms as core classifiers, with logistic regression serving as the metaclassifier. Several machine learning algorithms are implemented and contrasted with the stack ensemble learning strategy. An ensemble learning-based technique for identifying PMU cyberattacks is put forward in this paper. The specified ensemble learning approach is contrasted with the outcomes of many single learning techniques, such as logistic regression, decision trees, MLP classifiers, and support vector machines. PMU information on the issue and normal operations is collected. Preprocessing is done on the dataset to remove missing values and outliers. The ensemble learning algorithm is compared in an unbalanced environment with other single learning techniques, including logistic regression, decision trees, MLP classifiers, and support vector machines, in order to detect PMU cyberattacks.

2. ENSEMBLE LEARNING-BASED PMU CYBER ATTACK DETECTION USING SMOTE SAMPLING

A collection of data called "PMU cyber-attack Detection" was created specifically to help machine learning models identify allegations of cyberattacks. A variety of PMU data observed at various places in the power network are included in the dataset. In addition, only few re-researches have tested the isolation and location of assaults, which is crucial for defenders to implement the appropriate countermeasures to ensure the system continues to function normally even in the face of cyberattacks should be both fault- and attack-tolerant, so that even in the worst-case scenarios, it can continue to function as intended. Therefore, to increase the resilience of smart grids, the defenders must leverage fault-tolerant control. Thus, a device that can reliably detect or anticipate an attack is desperately needed. Figure 1 (see Appendix) [39] contains generators G1 and G2 are used. intelligent electronic devices (IEDs) R1 to R4 can turn on and off the breakers. The labels on these breakers are BR1 to BR4. There are double lines as well. Lines 1 and 2 extend from BR1 to BR2 and BR3 to BR4, respectively. One breaker is automatically controlled by each IED. Consequently, R1 governs BR1, R2 governs BR2, and so forth. The IEDs relay on a faraway protection technique that trips the breaker on detected faults since they lack inner validation to discern between real and counterfeit faults. Operators can manually instruct the IEDs R1 through R4 in addition to manually tripping the breakers BR1 through BR4. When doing repairs on the lines or other system parts, the manual override is employed.

3. METHODOLOGY

The dataset used for classification and the different classes are as given in the following: Table 1 provides an explanation of the 128 characteristics. Every phasor measurement unit (PMU) has 29 different kinds of measurements. A PMU or synchro phasor is a device that computes the electrical waves on an energy grid by synchronizing with a usual time source. Our system has four PMUs measuring 29 attributes, or 116 PMU measurement columns in total.

The events that need to be predicted using machine learning algorithms are natural events that occur in the power system, which are shown in Table 2, like the single line to ground (SLG) fault and line maintenance. The normal operation, as shown in Table 3, is due to the load changes in the power system. Table 4 details the different attack event scenarios in the power system due to a cyber-attack, which include data injection to trip the relay and remote tripping of the relay. These data are used to train the machine learning models for fault and non-fault conditions.

Table 1. PMU data specifications

Feature	Description
PA1:VH- PA3:VH	Voltage phase angle for A, B, and C phases
PM1:V- PM3:V	Voltage magnitude for A, B, and C phases
PA4:IH- PA6:IH	Current phase Angle for A, B, and C phases
PM4:I- PM6:I	Current magnitude for A, B, and C phases
PA7:VH- PA9:VH	Voltage phase angle for positive, negative, and zero sequence components
PM7:V- PM9:V	Voltage phase magnitude for positive, negative, and zero sequence components
PA10:VH- PA12:VH	Current phase angle for positive, negative, and zero sequence components
PM10:V- PM12:V	Current phase magnitude for positive, negative, and zero sequence components
F	Relay frequency
DF	Relay frequency delta (dF/dt)
PA:Z	Relay appearance impedance
PA:ZH	Relay angle of appearance impedance
S	Relay status flag

Table 2. Natural event scenarios

Scenario	Natural events (SLG faults)
1	Line1 SLG fault from 10-19%
2	Line1 SLG fault from 20-79%
3	Line1 SLG fault from 80-90%
4	Line2 SLG fault from 10-19%
5	Line2 SLG fault from 20-79%
6	Line2 SLG fault from 80-90%
	Natural events (line maintenance)
13	Line1 maintenance
14	Line2 maintenance

Table 3. No event scenarios

Scenario	No events (normal operation)
41	Normal operation due to load changes

Data was entered into the ARFF format for the initial multiclass dataset, which included fifteen datasets with around 5,000 data items per. The 128 aspects or variables that make up these data are mostly derived from synchro phasors or phasor measuring units (PMUs). The data was evaluated at 120 samples per second, and each scheme was simulated for 17 seconds [39]. Due to various capacity problems that are specific to each approach, generalization suffers in independent machine learning techniques. The generalization problem is mostly resolved by an algorithm that can combine the benefits of many machine learning techniques. An effective anomaly detection technique is the isolation forest algorithm. An improved implementation of the extended isolation forest technique for anomaly identification is included. The general execution particulars of suggested approach are shown in Figure 2. There are four main components to the implementation of fault prediction. data preprocessing automation, outlier detection and feature engineering, training and testing, model evaluation.

Figure 2 illustrates how the whole data preparation process is automated, leading to a machine learning paradigm with no need for human participation. The SVM learning method increases the generality of the learning process. Anomaly detection, data cleaning, and data classification into balanced and unbalanced data are all included in the automation of data preparation. This establishes the sample plan for the proposed implementation. Using the mean value as a stand-in, automatic impurity cleaning and missing value imputation are performed on the categorized data. As the most significant electrical applications are fault prediction algorithms, the primary factor influencing the research done in this way is the prediction algorithm's reliability. For a method to handle millions of data points, it must have higher generality and highly orthogonal input. In order to increase prediction performance, more sophisticated feature engineering techniques could be required if there is a larger connection between the sample data. It is challenging to acquire the data-aware preprocessing program as the recommended aim. The block diagram requires that the chosen fault diagnostic problem be subjected to an extended isolation procedure. When the extended isolation forest is used for fault diagnosis, the outlier identification shows improved orthogonality. The entire dataset is divided into training and testing datasets, each including the aforementioned CSV files. PMU cyber-attack training data has the information on whether there is attack or not which can be used as the target variable. The target attribute is fixed to whether the measured PMU data is cyber-attack or not since the machine learning algorithm has to predict the same.

Table 4. Attack event scenarios

Scenario	Attack type
	Data injection
	Attack Sub-type (SLG fault replay)
7	Line1 with tripping command from 10-19%
8	Line1 with tripping command from 20-79%
9	Line1 with tripping command from 80-90%
10	Line2 with tripping command from 10-19%
11	Line2 with tripping command from 20-79%
12	Line2 with tripping command from 80-90%
	Remote tripping command Injection
	Attack sub-type (command injection against single relay)
15	R1 relay command injection
16	R2 relay command injection
17	R3 relay command injection
18	R4 relay command injection
	Attack sub-type (command injection against single relay)
19	Relay1 and Relay2 command injection
20	Relay3 and Relay4 command injection
	Relay setting change
	Attack sub-type (disabling relay function-single relay disabled and fault)
21	Line1 SLG fault from 10-19% R1 disabled and fault
	Line1 SLG fault from 20-79% R1 disabled and fault
	Line1 SLG fault from 10-49% R2 disabled and fault
	Line1 SLG fault from 50-79% R2 disabled and fault
	Line1 SLG fault from 80-90% R2 disabled and fault
	Line2 SLG fault from 10-19% R3 disabled and fault
	Line2 SLG fault from 20-79% R3 disabled and fault
	Line2 SLG fault from 10-49% R3 disabled and fault
	Line2 SLG fault from 50-79% R4 disabled and fault
	Line2 SLG fault from 80-90% R4 disabled and fault

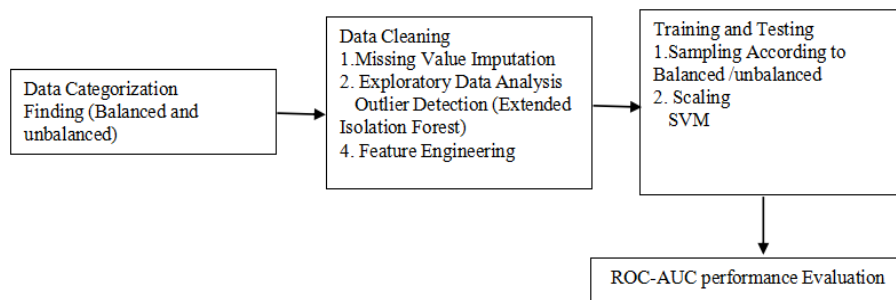


Figure 2. Overall block diagram of data preprocessing automated SVM learning

4. RESULTS AND DISCUSSION

Python is used in conjunction with the Scikit-learn (sklearn), synthetic minority oversampling technique (SMOTE), and Pandas toolboxes to create cyber-attack prediction. Using MLP, SVM, and decision tree algorithms as the base classifiers and logistic regression as the meta classifier, an ensemble learning code is created to find cyber attacks. Figure 3 displays the graph that was created to illustrate the various input variables. Since there is a noticeable disparity between the amount of cyberattack data and regular data, the data appears to be unbalanced.

The individual and ensemble learning paradigms are used to design the categorization issue. Data is resampled, and data imbalance is verified. Numerous attributes that are not necessary for categorization are removed from the dataset prior to resampling. The significant characteristics from the dataset remain after the insignificant features have been eliminated. Table 5 lists the parameters for the various machine learning algorithms utilized in the ensemble learning technique. An ML approach called ensemble learning combines many base classifiers to produce a more powerful prediction model. The decision tree algorithm, multi-layer perceptron (MLP), and support vector classifiers (SVC) are the base classifiers in this instance. Training each base classifier separately on the training set of data is the first stage in the ensemble learning process. For the test data, every base classifier will provide a unique set of predictions. A meta-classifier will be used to aggregate these predictions and provide a final prediction. In this instance, logistic regression serves as the meta classifier. It generates a final prediction by using the input predictions from each of the base classifiers. In order to generate the most accurate prediction possible, the meta-classifier integrates the knowledge gained

from the basis classifiers' predictions. Combining the meta-classifier's predictions with the basic classifiers' predictions yields the final ensemble model.

The way this combination is used optimizes the forecast accuracy. An ensemble learning approach can boost the model's accuracy by combining the benefits of many basic classifiers. The drawbacks of separate classifiers can be mitigated by combining the benefits and drawbacks of each fundamental classifier. All things considered, the ensemble learning algorithm is an important machine learning technique that may yield extraordinarily accurate predictions. It employs logistic regression as a meta-classifier and SVC, MLP, and decision tree algorithms as base classifiers. Figure 3 displays the graphs of various input variables represents Figure 3(a) voltage phase angle for C phase, Figure 3(b) voltage magnitude for B phase, Figure 3(c) voltage phase angle for B phase, and Figure 3(d) voltage magnitude for A phase. Figure 4 represents an AUC-ROC curve that was produced by the individual and group learning (stack distribution) technique. On comparison with other distinct machine learning techniques, the stack distribution approach yields the highest AUC-ROC curve performance. Out of all the algorithms, logistic regression demonstrated good performance. All the separate methods are outperformed by the ensemble learning algorithm. Given that the SMOTE algorithm was utilized as the sampling technique, it is concluded that the implementation's accuracy is good. Since there are several SMOTE methods available, the diversity of SMOTE implementation on up-sampling may be used to enhance the accuracy and AUC-ROC curve for the up-sampled input characteristics.

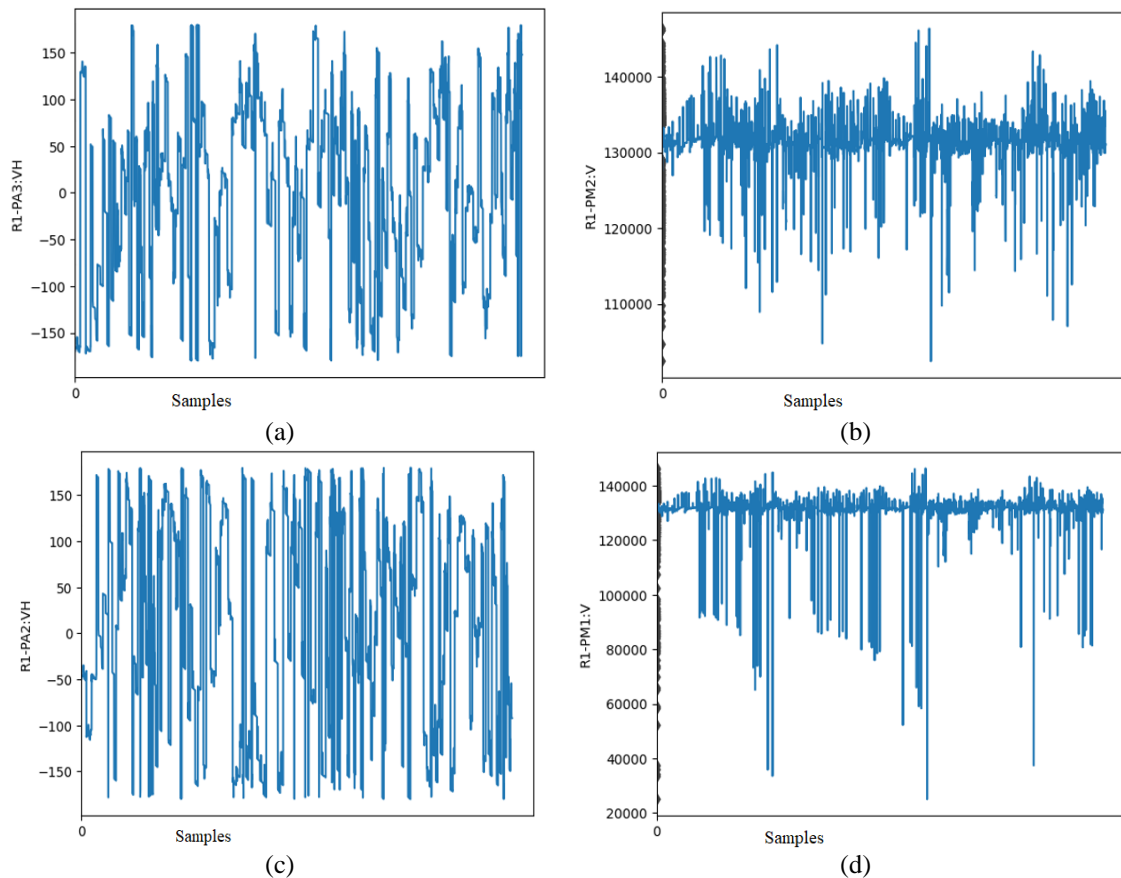


Figure 3. Input variables from the PMU (4 among the 128 variables): (a) voltage phase angle for C phase, (b) voltage magnitude for B phase, (c) voltage phase angle for B phase, and (d) voltage magnitude for A phase

Table 5. Ensemble learning parameters

Machine learning model	Parameters
MLP classifier	Activation = "relu", alpha = 0.1, hidden_layer_sizes = 30, learning_rate = "invscaling", max_iter = 50000, random_state = 1000
Decision tree classifier	max_depth = 5, max_features = "auto", min_samples_leaf = 0.005, min_samples_split = 0.005, random_state = 2000
SVC	C=85, degree = 15, gamma = .8, kernel = "rbf", probability = True
Logistic regression	random_state = 42

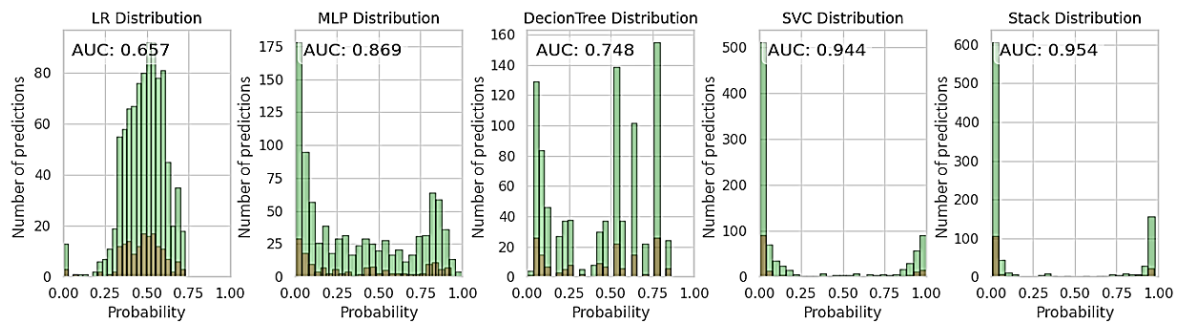


Figure 4. AUC-ROC curve for individual and stack ensemble learning algorithms

5. CONCLUSION

This research develops an ensemble learning-based method for cyberattack detection. We investigated how different ensemble techniques performed and contrasted with different single-learning algorithms. Our findings show for attack detection, the ensemble learning strategy performs much better than single learning algorithms. With a 95.4% accuracy rate, our method shows promise for effective cyberattack detection in the domain. Future research will examine different ensemble learning methods and expand the methodology to new fields. The outcomes of our experiments demonstrate that when it comes to cyberattack detection, the ensemble learning strategy performs noticeably better than single learning algorithms.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Kiruthika Krishnan	✓	✓	✓	✓	✓			✓	✓	✓			✓	
Srivani Iyengar	✓	✓							✓	✓	✓	✓		

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

No datasets were generated or analyzed during the current study.

REFERENCES

- [1] Online Trust Alliance, "Cyber incident and breach trends report," 2018. [Online]. Available: <https://otalliance.org/incident%0Ahttps://otalliance.org/resources/cyber-incident-breach-response>.
- [2] S. M. A. Bhuiyan, J. F. Khan, and G. V. Murphy, "Big data analysis of the electric power PMU data from smart grid," in *SoutheastCon 2017*, IEEE, Mar. 2017, pp. 1–5, doi: 10.1109/SECON.2017.7925277.
- [3] J. De La Ree, V. Centeno, J. S. Thorp, and A. G. Phadke, "Synchronized phasor measurement applications in power systems," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 20–27, 2010, doi: 10.1109/TSG.2010.2044815.
- [4] M. K. Penshanwar, M. Gavande, and M. F. A. R. Satarkar, "Phasor measurement unit technology and its applications-a review," *International Conference on Energy Systems and Applications, ICESA 2015*, pp. 318–323, 2016, doi: 10.1109/ICESA.2015.7503363.
- [5] Y. Yang, J. Q. Ju, Q. H. Li, and Q. Wang, "An experimental research on impacts of malicious attacks on PMU in smart grids," in *2018 International Conference on Power System Technology, POWERCON 2018 - Proceedings*, 2018, pp. 3035–3041, doi: 10.1109/POWERCON.2018.8602008.
- [6] X. Wang, D. Shi, J. Wang, Z. Yu, and Z. Wang, "Online identification and data recovery for PMU data manipulation attack," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 5889–5898, 2019, doi: 10.1109/TSG.2019.2892423.

Ensemble learning based fault detection using PMU data in imbalanced ... (Kiruthika Krishnan)

- [7] Z. Zhao, Y. Huang, Z. Zhen, and Y. Li, "Data-driven false data-injection attack design and detection in cyber-physical systems," *IEEE Transactions on Cybernetics*, vol. 51, no. 12, pp. 6179–6187, Dec. 2021, doi: 10.1109/TCYB.2020.2969320.
- [8] Q. He, F. Bai, Y. Cui, and M. Zillmann, "Machine learning-based cybersecurity defence of wide-area monitoring systems," in *2022 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)*, IEEE, Jul. 2022, pp. 991–996, doi: 10.1109/ICPSAsia55496.2022.9949686.
- [9] Z. Wu *et al.*, "Reinforcement learning based multistage optimal PMU placement against data integrity attacks in smart grid," in *Proceedings - 2021 4th IEEE International Conference on Industrial Cyber-Physical Systems, ICPS 2021*, 2021, pp. 572–577, doi: 10.1109/ICPS49255.2021.9468170.
- [10] H. Goyel and K. S. Swarup, "Data integrity attack detection using ensemble-based learning for cyber-physical power systems," *IEEE Transactions on Smart Grid*, vol. 14, no. 2, pp. 1198–1209, 2023, doi: 10.1109/TSG.2022.3199305.
- [11] I. Khan and V. Centeno, "Realtime detection of PMU bad data and sequential bad data classifications in cyber-physical testbed," *IEEE Access*, vol. 11, pp. 71235–71249, 2023, doi: 10.1109/ACCESS.2023.3292059.
- [12] H. Varmaziari and M. Dehghani, "Cyber attack detection in PMU networks exploiting the combination of machine learning and state estimation-based methods," in *2021 11th Smart Grid Conference (SGC)*, IEEE, Dec. 2021, pp. 1–6, doi: 10.1109/SGC54087.2021.9664189.
- [13] A. Jafari, H. Ergun, and D. Van Hertem, "A voting-based machine learning strategy to detect false data injection attack in cyber-physical power systems," in *2022 57th International Universities Power Engineering Conference (UPEC)*, IEEE, Aug. 2022, pp. 1–6, doi: 10.1109/UPEC55022.2022.9917789.
- [14] T. S. Damarla and A. Yadav, "Detection of cyber – attacks in digital power system using ensemble bagged trees," in *2023 4th International Conference for Emerging Technology (INCET)*, IEEE, May 2023, pp. 1–6, doi: 10.1109/INCET57972.2023.10170685.
- [15] J. Moradi, H. Shahinzadeh, H. Nafisi, M. Marzband, and G. B. Gharehpetian, "Attributes of big data analytics for data-driven decision making in cyber-physical power systems," in *2020 14th International Conference on Protection and Automation of Power Systems (IPAPS)*, IEEE, Dec. 2019, pp. 83–92, doi: 10.1109/IPAPS49326.2019.9069391.
- [16] C. Hu, J. Yan, and C. Wang, "Robust feature extraction and ensemble classification against cyber-physical attacks in the smart grid," in *2019 IEEE Electrical Power and Energy Conference (EPEC)*, IEEE, Oct. 2019, pp. 1–6, doi: 10.1109/EPEC47565.2019.9074827.
- [17] W. Ding, M. Xu, Y. Huang, P. Zhao, and F. Song, "Cyber attacks on PMU placement in a smart grid: Characterization and optimization," *Reliability Engineering and System Safety*, vol. 212, p. 107586, Aug. 2021, doi: 10.1016/j.res.2021.107586.
- [18] S. De and R. Sodhi, "A PMU assisted cyber attack resilient framework against power systems structural vulnerabilities," *Electric Power Systems Research*, vol. 206, p. 107805, May 2022, doi: 10.1016/j.epr.2022.107805.
- [19] D. Mukherjee, "Detection of data-driven blind cyber-attacks on smart grid: A deep learning approach," *Sustainable Cities and Society*, vol. 92, p. 104475, May 2023, doi: 10.1016/j.scs.2023.104475.
- [20] G. Zhang, J. Li, O. Bamisile, Y. Xing, D. Cao, and Q. Huang, "Identification and classification for multiple cyber attacks in power grids based on the deep capsule CNN," *Engineering Applications of Artificial Intelligence*, vol. 126, p. 106771, Nov. 2023, doi: 10.1016/j.engappai.2023.106771.
- [21] M. Ghiasi, T. Niknam, Z. Wang, M. Mehrandezh, M. Dehghani, and N. Ghadimi, "A comprehensive review of cyber-attacks and defense mechanisms for improving security in smart grid energy systems: Past, present and future," *Electric Power Systems Research*, vol. 215, p. 108975, Feb. 2023, doi: 10.1016/j.epr.2022.108975.
- [22] H. Ge *et al.*, "A game theory based optimal allocation strategy for defense resources of smart grid under cyber-attack," *Information Sciences*, vol. 652, p. 119759, Jan. 2024, doi: 10.1016/j.ins.2023.119759.
- [23] A. K. M. A. Habib, M. K. Hasan, R. Hassan, S. Islam, R. Thakkar, and N. Vo, "Distributed denial-of-service attack detection for smart grid wide area measurement system: A hybrid machine learning technique," *Energy Reports*, vol. 9, pp. 638–646, Oct. 2023, doi: 10.1016/j.egy.2023.05.087.
- [24] T. Chen and C. Liu, "Soft computing based smart grid fault detection using computerised data analysis with fuzzy machine learning model," *Sustainable Computing: Informatics and Systems*, vol. 41, p. 100945, Jan. 2024, doi: 10.1016/j.suscom.2023.100945.
- [25] L. Min, K. A. Alnowibet, A. F. Alrasheedi, F. Moazzen, E. M. Awwad, and M. A. Mohamed, "A stochastic machine learning based approach for observability enhancement of automated smart grids," *Sustainable Cities and Society*, vol. 72, p. 103071, Sep. 2021, doi: 10.1016/j.scs.2021.103071.
- [26] A. Akagic and I. Džafić, "Enhancing smart grid resilience with deep learning anomaly detection prior to state estimation," *Engineering Applications of Artificial Intelligence*, vol. 127, p. 107368, Jan. 2024, doi: 10.1016/j.engappai.2023.107368.
- [27] L. Cui, Y. Qu, L. Gao, G. Xie, and S. Yu, "Detecting false data attacks using machine learning techniques in smart grid: A survey," *Journal of Network and Computer Applications*, vol. 170, p. 102808, Nov. 2020, doi: 10.1016/j.jnca.2020.102808.
- [28] S. Y. Diaba and M. Elmusrati, "Proposed algorithm for smart grid DDos detection based on deep learning," *Neural Networks*, vol. 159, pp. 175–184, Feb. 2023, doi: 10.1016/j.neunet.2022.12.011.
- [29] A. M. Vidya, D. D. Sai, G. Sarveshwaran, S. S. Mukesh, and K. R. M. V. Chandrakala, "Identification of false data injection and man in the middle cyber-attacks impact on smart grid," in *2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC)*, IEEE, May 2023, pp. 678–683, doi: 10.1109/ICSCCC58608.2023.10176782.
- [30] R. J. R. Kumar and B. Sikdar, "Detection of stealthy cyber-physical line disconnection attacks in smart grid," *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 4484–4493, Sep. 2021, doi: 10.1109/TSG.2021.3082543.
- [31] K. Aygul, M. Mohammadpourfard, M. Kesici, F. Kucuktezcan, and I. Genc, "Benchmark of machine learning algorithms on transient stability prediction in renewable rich power grids under cyber-attacks," *Internet of Things*, vol. 25, p. 101012, Apr. 2024, doi: 10.1016/j.iot.2023.101012.
- [32] J. Sakhnini, H. Karimipour, A. Dehghantanha, and R. M. Parizi, "Physical layer attack identification and localization in cyber-physical grid: An ensemble deep learning based approach," *Physical Communication*, vol. 47, 2021, doi: 10.1016/j.phycom.2021.101394.
- [33] R. Nawaz, R. Akhtar, M. A. Shahid, I. M. Qureshi, and M. H. Mahmood, "Machine learning based false data injection in smart grid," *International Journal of Electrical Power and Energy Systems*, vol. 130, p. 106819, Sep. 2021, doi: 10.1016/j.ijepes.2021.106819.
- [34] N. Yi, Q. Wang, L. Yan, Y. Tang, and J. Xu, "A multi-stage game model for the false data injection attack from attacker's perspective," *Sustainable Energy, Grids and Networks*, vol. 28, p. 100541, Dec. 2021, doi: 10.1016/j.segan.2021.100541.
- [35] J. Khazaei and M. H. Amini, "Protection of large-scale smart grids against false data injection cyberattacks leading to blackouts," *International Journal of Critical Infrastructure Protection*, vol. 35, p. 100457, Dec. 2021, doi: 10.1016/j.ijcip.2021.100457.
- [36] J. Zhang *et al.*, "Machine learning-based cyber-attack detection in photovoltaic farms," *IEEE Open Journal of Power Electronics*, vol. 4, pp. 658–673, 2023, doi: 10.1109/OJPEL.2023.3309897.
- [37] R. Zhou, M. Peng, and X. Gao, "Vulnerability assessment of power cyber-physical system considering nodes load capacity," in *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*, IEEE, Apr. 2021, pp. 1438–1441, doi: 10.1109/ICSP51882.2021.9408825.

- [38] L. Almutairi, R. Daniel, S. Khasimbee, E. L. Lydia, S. Acharya, and H.-I. Kim, "Quantum dwarf mongoose optimization with ensemble deep learning based intrusion detection in cyber-physical systems," *IEEE Access*, vol. 11, pp. 66828–66837, 2023, doi: 10.1109/ACCESS.2023.3287896.
- [39] A. Afroz, "Smart grid false data injection attack prediction," 2023. [Online]. Available: <https://www.kaggle.com/code/pythonafroz/smart-grid-false-data-injection-attack-prediction>.

APPENDIX

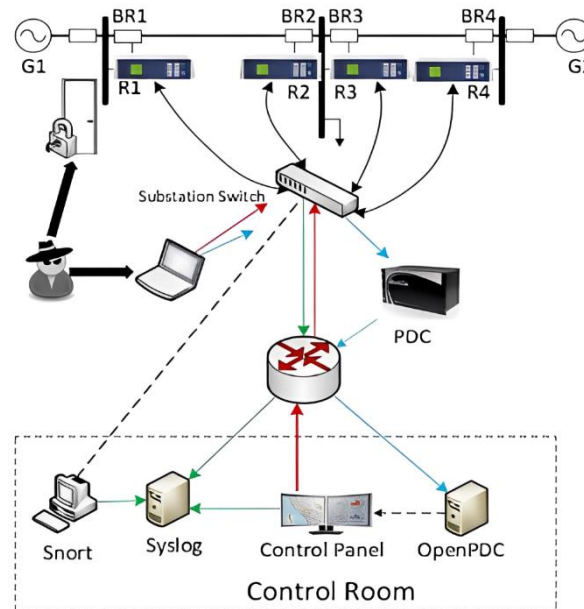








Figure 1. Proposed system with PMU [39]

BIOGRAPHIES OF AUTHORS



Kiruthika Krishnan    is Assistant Professor in the Electrical and Electronic Engineering Department, Rajarajeshwari College of Engineering, Bangalore since 2019. She received the B.E. (I&C) degree from Madras University, 2004, M.Tech. (power electronics) degree from Visveswaraya Technological University in 2013. She has published three papers in international conference proceedings. Her areas of interest are power system protection and control, fuzzy logic, electrical & electronics measurement, sensors & instruments analysis, and renewable energy systems. She can be contacted at email: kiruthi.km21@gmail.com.



Dr. Srivani Iyengar    received a B.E. (E&E) degree from Bangalore University, Bangalore in 1986, an M.E. (power system) degree from Bangalore University, Bangalore in 1990, and a Ph.D. in 2011 from National Institute of Technology, Karnataka (NITK) Surathkal, Mangalore, India. She joined RV College of Engineering, Bangalore, India, in 1990 as a lecturer and is presently working as Professor and HOD in the department of Electrical and Electronics Engineering. She is a life member of the Indian Society for Technical Education (ISTE) and a member of IEEE. She has published more than 120 technical research papers in various National and International Journals and conferences. Her areas of interest include power system protection, signal processing, power quality, renewable energy sources, grid integration, smart grid, power electronics applications, industrial drives and energy harvesting, fuzzy logic, and ANN applications to power systems. She can be contacted at email: srivanisg@rvce.edu.in or srivani.sg@gmail.com.