Hindawi
Security and Communication Networks
Volume 2022, Article ID 6200121, 9 pages
https://doi.org/10.1155/2022/6200121

Research Article

Cyber Security of Smart Metering Infrastructure Using Median Absolute Deviation Methodology

Priti Prabhakar,[1] **Sujata Arora**,[2] **Anita Khosla**,[3] **Rajender Kumar Beniwal**,[1] **Moses Ndole Arthur**,[4] **José Luis Arias-Gonzáles**,[5] **and Franklin Ore Areche**[6]

[1]*Electrical Engineering Department, Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India*
[2]*Noida International University, Gautam Buddha Nagar, Noida, India*
[3]*EEE Department, Faculty of Engineering and Technology, Manav Rachna International Institute of Research and Studies, Faridabad, Haryana, India*
[4]*Department of Biomedical Engineering, School of Engineering Sciences, College of Basic and Applied Sciences, University of Ghana, Accra, Ghana*
[5]*Department of Business, Pontifical Catholic University of Peru, Lima, Peru*
[6]*Academic Department of Agroindustrial Engineering, National University of Huancavelica, Huancavelica, Peru*

Correspondence should be addressed to Anita Khosla; anitakhosla.fet@mriu.edu.in and Moses Ndole Arthur; mnarthur@st.ug.edu.gh

Received 9 August 2022; Revised 5 September 2022; Accepted 14 September 2022; Published 7 October 2022

Academic Editor: Keping Yu

To combat cyber threats in the smart grid, an intrusion detection system can be integrated into the advanced metering infrastructure. Anomaly-based intrusion detection can detect even the tiniest changes in the parameter under investigation, whereas signature-based intrusion detection only recognises known attacks. The growing usage of smart grids necessitates the classification, identification, and implementation of countermeasures to threats. At the absolute least, smart grids must be protected against cyberattacks; thus, the highest level of information security must be offered. As a result of digitisation and the usage of more smart applications, the research looked at a variety of attack types, smart grid assaults, and major cyber threats on the voltage regulation. Machine learning techniques that analyse data in real time and formulate patterns to recognise an attack and scan through huge data for anomalies can be implemented into the advanced metering infrastructure (AMI) for intrusion detection for anomaly-based intrusion detection. The comparative test study done for the research found that the proposed method, median absolute deviation for anomaly identification in smart metering datasets, produced the most accurate and precise differentiations with the highest accuracy and precision. The median absolute deviation (MAD) algorithm model is trained using test data, and raw predictions are made, before true data are used to derive final test result parameters, precision, recall, and F1 scores. The methodology of the entire study is discussed in this paper, as well as how the MAD algorithm is best suited for anomaly-based intrusion detection, as well as result comparisons of other machine learning algorithms.

1. Introduction

Cyberattacks such as confidential data breaches, malware injection, and other cyberattacks are common in smart grid IoT-based components such as advanced metering infrastructure and data transmission systems. Additional methods have been integrated directly into the database management system to counter the system's vulnerabilities in order to protect the smart grid from such malicious attacks. Intrusion detection systems are one such countermeasure, which can be classified into signature-based and anomaly-based detection. Behaviour is compared to rules in signature-based detection, whereas behaviour is compared to profiles in anomaly-based detection [1, 2]. The three main components of cyber security must be adequately provided: confidentiality, validity, and availability. One of the major difficulties in integrating today's critical infrastructures is cyber security concerns and cyberattacks [3, 4]. Cyber

security issues and cyberattacks are the main obstacles to integrating today's key infrastructures. Smart grids are exceptional at effectively producing, transferring, and delivering energy, which is a critical component of critical infrastructure. Smart grid utilisation is expanding, necessitating the classification, detection, and application of threat mitigation measures. The highest level of information security must be provided in order to, at the very least, protect smart grids from cyberattacks. The research in this area looked at several attack types, smart grid assaults, and significant cyberattacks on the power system. Electricity is a crucial component of crucial infrastructure, and smart grids excel at generating, transmitting, and distributing it efficiently. The increasing use of smart grids demands the categorisation and identification of threats, as well as the execution of countermeasures [5, 6]. Smart grids must be secured from cyberattacks at the very least; thus, the most complete information security must be provided. In the twenty-first century, the smart grid has played a critical role in altering the concept of electrical power engineering. In the past, the generation and consumption had to match in all circumstances, but with the advent of nonconventional grids, everything changed, and customers now had to use the equivalent amount of electricity as the generators were producing. The demand response (DR) strategy is the means to accomplish all of that. Consumer consumption patterns should be more flattened than sharp curves, which result in higher prices because more electricity is generated during peak demand periods. The programmes for demand response are mentioned and explained with references to the documentation. In this regard, the research examined various attack types, smart grid assaults, and serious cyberattacks on the power system [7, 8]. To distinguish abnormal data from standard data values, these profiles still need to define a threshold limit. Intrusion prevention techniques are not the same as these detection techniques. These detection techniques are not the same as preventative measures. On the detection and false-positive ratio, the best parameters for classifying the data as normal or anomalous are determined. Then, the intrusion detection model receives the smart meter dataset aimed to compare various machine learning techniques, including the median absolute deviation technique, the local outlier factor, and deviation-based outlier detection. Because hackers are constantly seeking for vulnerabilities in organisations' security protocols, both large corporations and small- and medium-sized businesses (SMEs) must take proactive measures to fend off attacks. The deduction of optimal parameters for the classification of the data from normal to anomaly is decided based on the dataset's rate of detection and ratio of false positives. Cybersecurity aids in defending against online dangers and computer system assaults. It locates many system flaws and vulnerabilities that hackers and attackers could use, and it automatically fixes all of the flaws with the capacity to enhance performance problems. Attacks on huge servers connected to wide-area networks are a serious problem that cybersecurity helps to address. It upholds the industry standard, strict safety standards for users to abide by cybersecurity precautions in order to secure the devices.

The smart meter dataset is then entered into the intrusion detection model. For this paper, we have aimed to compare multiple machine learning algorithms such as local outlier factor, deviation-based outlier detection, one-class support vector machines, and median absolute deviation method. The optimal parameters for identifying the data as normal or anomalous are calculated based on the dataset's rate of detection and false-positive ratio. The smart meter dataset is then sent to the intrusion detection model. In this study, we examine various machine learning algorithms, including the median absolute deviation approach, the local outlier factor, deviation-based outlier detection, and one-class supported vector machines. The methodology and application of the MAD technique produced the greatest and most accurate findings, according to the examination of precision, recall, and F1 score. After the analysis of precision, recall, and F1 score, it was found that MAD technique has the highest and most accurate results, and the methodology and implementation of the technique have been discussed in the later section of the paper.

The rest of the paper is divided into following sections. Section 2 presents the literature review, Section 3 presents the cyber threats on smart meters, Section 4 presents the frequently used intrusion detection, Section 5 presents the proposed methodology, and Section 6 presents simulation test case. Simulation result of the current study and conclusion presents in Sections 7 and 8, respectively.

2. Literature Survey on Advanced Metering Infrastructure (AMI)

Advanced metering infrastructure (AMI) is the most important component of a smart grid because it connects all of the important subcomponents to form an efficient and smart network. The importance of power application security works with cyber infrastructure security to prevent, mitigate, and tolerate. Electric utilities are required to purchase the electric energy and capacity made available by qualified facilities at a price that reflects the savings, and they will experience by using these sources instead of other sources. Cybersecurity is the safeguarding of networks, hardware, and software against intrusions, harm, illegal access, and denial of service. An AMI is made up of smart meters, a data management system (DMS), automatic meter reading (AMR), a communication, and data transmission network, and it allows for two-way communication between consumers and utilities for the transmission of metered data. AMI is used not only in power systems but also in other utilities such as gas and water metering around the world. In order to maintain attributes of an AMI such as display units, metering sensors, and communication modules, microcontrollers are embedded into the subsystem components [1]. For instance, the general manufactured smart meters are embedded with an AVR RISC-based microcontroller, ATMEL mega 2560 that executes powerful instructions in a single clock cycle and allows to strike a fine balance between metered data and processing speed. On the consumer end, a smart meter is equipped with a wireless network data transmission module that collects real-time data for the

consumers as well as for the host system. Thereupon, the metered data are sent to DMS (data management system) where the data are analysed and sent to the service provider [2]. DMS also manages data storage received from individual smart meters, thus contributing greatly to the infrastructure.

2.1. Smart Meters and Their Components. Smart meters are end-user devices that are equipped with electronic hardware and software for real-time measurements, communication of metered data to the service host, and execution of utility host commands. A consumer can view their consumption details through display units installed in a smart meter and then further analyse and regulate their use as they wish. Many sensors and actuation components are integrated into SMs for data transmission, processing, data storage with time stamping, and so on. A microcontroller for command execution and data processing, communication interface modules, sensors, and an LCD display module make up the hardware [9]. The notion of a "smart grid" enables a two-way information flow from utilities to customers and vice versa, improving power usage efficiency [10, 11]. This will be possible with the implementation of "advanced metering infrastructure." A consumer's power usage data are quickly gathered, and the "smart meter" situated at the user's premises gathers and analyses these data. The analysed data are sent to the utilities through AMI [12, 13]. The AMI's advanced communication system includes home area networks, district network cables, and wide area networks. As a consequence, AMI not only supplies utilities with smart meter data but also communication from utilities to customers about peak demand and energy consumption costs, allowing the customer to transfer peak demand [14, 15].

Accurate measurements of various parameters, data transmission to and from the host, and effective execution of operation commands from the host are just a few of the important characteristics that a smart meter should have. The most important things to consider are data collection and transmission synchronisation, outrage notification, energy theft alert, and power management in the event of a primary supply unit failure.

2.2. Need for Cyber Security in Smart Meter Infrastructure. There are numerous technical considerations to be made when choosing a communication network. For example, DNP3 does not provide adequate security for collaborative operations, so it has been enhanced with data object security and a security layer. Furthermore, most smart meter communication networks have low bandwidth, resulting in high traffic and limiting the amount of data that can be transmitted. Integration of modulation and demodulation devices, as well as additional memory for storing data logs, could raise overall deployment costs.

Smart meters send a large amount of data to the service provider at regular intervals via wireless communication. The information gathered is private and sensitive, and it should not fall into the hands of shady characters, so security rules and regulations must be followed to the letter. To ensure this, all smart meters are given unique identifiers, which can then be used to access information about the data being measured. The communication network must be cost-effective [16] as well as have the necessary transmission range and security functions. RF, Wi-Fi, ZigBee, and other secure communication networks are examples [17, 18]. Because it can be integrated with a home area network (HAN) for data transfer and control signals, ZigBee has become one of the most popular data transmission methods [19].

However, the challenge of cyber security comes with a highly secure and sophisticated wireless communication network. Smart meters without additional security features are vulnerable to attacks that can result in data breaches or manipulations that are both valuable and confidential. Attackers can use the data to launch a malware attack on the data's confidentiality [20, 21]. The frequency of malware attacks such as ransomware is much higher [22].

3. Cyber Threats on Smart Meters

Before developing a security measure, an objective analysis of the smart grid's security and potential cyberattacks is conducted. In a smart grid, smart meters are installed at the user end and communicate with the host directly over the wireless network. Because these smart meters are being deployed in such large numbers, cyber security maintenance will necessitate the service provider upgrading the infrastructure with cyberattack countermeasures, as the basic hardware lacks any specific and effective security features [23]. Data encryption and data theft detection are two of the most basic security features. Data security's basic foundation is encryption. It is the simplest and most crucial approach to guarantee that data on a computer system cannot be taken and read by someone who intends to use it maliciously. To secure user information transmitted between a browser and a server, both small businesses and individual consumers frequently utilise data security encryption. It is possible to create a symmetric encryption that, in theory, can only be cracked with a lot of processing power by using data encryption technology, also referred to as an encryption algorithm or cypher. The act of stealing digital information from computers, servers, or other electronic devices in order to access sensitive information or violate privacy is known as data theft. Malicious actors that intend to sell the information or exploit it for identity theft are the main causes of data theft. If identity thieves have enough data, they can use it to open secure accounts, available credit cards in the victim's name, or perform other unauthorised acts using the victim's identity.

Smart meters collect and store data, which is then used to analyse peak load patterns, variable tariff systems, and other issues. The distinctions that were made most precisely and accurately came from the datasets from smart meters. The primary emphasis is on the advanced metering infrastructure (AMI) technologies. Smart meters use wireless connectivity to provide a lot of data to the service provider on a regular basis. Since the information acquired is confidential and sensitive, strict adherence to security laws and regulations is required to prevent it from falling into the hands of

nefarious persons. As a result, the data extracted must be accurate and complete, representing all aspects of a customer's energy consumption and grid status. These data can be manipulated or altered, posing a risk to the data in the smart grid [24]. An attacker's motivation could include data manipulation, breach of private confidential data, credential extraction, kindling outrages, and so on as normal or anomalous. Then, the intrusion detection model receives the smart meter dataset. Government, business, and academic institutions are rapidly becoming more interested in the smart grid. It is a next-generation electricity network with autonomous meter reading, dynamic, and two-way communications among its components. It is more dependable, efficient, and self-healing. Cyberattacks on smart meters can be classified into four categories based on the methodology and type of attack: availability attacks, integrity attacks, confidentiality attacks, and authenticity attacks. Availability attacks are Dos/DDos attacks that disrupt utility supply and cause services to be unavailable to customers [25]. Interoperability, network communications, demand response, energy storage, and distribution system management are only a few of the obstacles that the development of the smart grid faces. A power network called the "smart grid" that uses digital communications technology must contend with issues like rising load demands, blackouts, overloads, and voltage sags, as well as cyberattacks. Internet-connected smart grid equipment has developed a number of vulnerabilities in recent years. These methods can be used to simulate and analyse cyber vulnerabilities in the smart grid, such as relay protection, power flow control, grid security, and dependability. Radiofrequency jamming and replay attacks are two other types of availability attacks. Integrity attacks: data manipulation by a cyberattack that compromises the data's integrity and causes the grid to lose power.

Unauthorised access to sensitive data, traffic analysis, and MitM attacks can cause grid parameters to change and attack the network, with the goal of influencing energy values, extracting authorisation commands, eavesdropping, and masquerading attacks. It is critical to consider the likelihood of threats in the smart grid when constructing a smart metering infrastructure. Antivirus software and other traditional security tools are vulnerable to anomaly-based cyberattacks [26–28]. As a result, new techniques must be implemented in order to ensure cyber security in the smart grid.

Inspection of cyber security resources in the smart grid includes hardware, software, network parameters, and communication set-up for any vulnerabilities. Simulating a real-time model and entering manipulated data from external sources to mimic an attack can provide a comprehensive analysis of attacks and their aftermath. The proposed method, which can be seen in Section 7, was subjected to the same procedure.

Additional and sophisticated methods are used to fill the void in existing security tools in order to provide necessary cyber security in the smart grid. Firewalls, anti-Dos hardware, incorporating the latest versions of encryption protocols, antimalware installations, and endpoint detection and response systems are just a few of the widely used security practices.

In a smart grid, the network is the most vulnerable against threats and risks; thus, to counter these cyberattacks, some of the proposed/suggested/profound techniques are briefed below [29]. *Malware protection*. The embedded systems of a smart meter are only exposed to running the software and other commands from the host, and such systems are protected using a manufacturing key for validation. But the general purpose systems make use of third-party software. Thus, to ease the risk of using third-party software, other solutions are implemented in a system. Implementing malware protection provides extra layers of protection [30–32]. *Network security*. Adapting VPN provides security measures such as data protection during transmission across networks and encryption of data during transmission, which can be at a risk whilst making use of the public network. *Data Encryption*. Using the Internet to connect to a VPN can directly encrypt your connection. The advanced encryption standard (AES) is a 256-bit cypher that provides the highest encryption standard and is used for data security in the financial and government sectors around the world [33]. Intrusion detection system (IDS) and intrusion prevention system (IPS) are two types of intrusion detection and prevention systems. An intrusion prevention system can prevent an attack and protect the smart grid by controlling network access [34]. The intrusion detection system is primarily concerned with the detection of cyberattacks and their classification, which can be based on signature detection or anomaly detection. The following section discusses the IDS and its methods for detecting an intrusion in the smart grid network.

4. Frequently Used Intrusion Detection

In a smart grid system, methods like IDS and IPS are used to exploit a system's vulnerabilities by detecting abnormal and security violating patterns. Attempts to break into a system, masquerading, virus, Trojan horse, and denial of service are among the violations. By identifying unusual and security-violating patterns, techniques like IDS and IPS are utilised to exploit a system's weaknesses. The infractions include attempting to hack into a system, masquerading, viruses, Trojan horses, and denial of service. An IDS model is a type of rule-based patterns combining technique. Only attacks can be detected by an IDS. Attacks cannot be stopped by it. An IPS, on the other hand, restricts attacks by detecting and stopping them before they reach their target. Any effort to jeopardise accessibility, integrity, or confidentiality is considered an attack. When a potential attack, malicious activity, or an unauthorised user is discovered, an IPS solution has more autonomy and takes action. An IDS model can be thought of as a pattern conjoining method based on rules [35]. In this paper, we want to see which detection technique has the best accuracy, as well as a higher detection rate and fewer false alarms. The first step in preparing the dataset for the algorithm's training was to determine the parameters that were most accurate in discriminating attacks, after which we developed the python code and trained and tested

the algorithm, the results of which are detailed in the paper's later sections. The two main methods for detecting cyberattacks on the system are anomaly-based and signature-based [36].

4.1. Signature-Based Intrusion Detection. For threats that are known to the host or have occurred in the past, signature-based intrusion detection is used. It is based on a preprocessed list of known system vulnerabilities and indicators of compromise, which could include any abnormal system behaviour. Threat signatures include malicious network attack patterns, file hashes, known byte sequences, malicious domains, and known malicious IP addresses. Other preprogrammed patterns help detect and classify the type of intrusion on the system when using a signature-based intrusion detection technique. Because it uses a pre-entered pattern to determine the type of threat, signature-based intrusion detection has a faster processing speed, reducing false alarm rates and allowing the detection technique to detect threats quickly and accurately. Nonetheless, signature-based intrusion detection techniques fail to detect a zero-day exploit attack that specifically targets the weak links of the system that even the host is unaware of. Such an attack cannot be detected by signature-based intrusion detection techniques because only the attacker is aware of such vulnerabilities in a system and thus targets it to compromise confidential data and cause system damage [37].

4.2. Anomaly-Based Intrusion Detection. Anomaly-based detection detects intrusive behaviour that goes beyond the white list or outside the acceptable range. It is capable of detecting unknown suspicious behaviour in the smart grid. In an anomaly-based intrusion detection technique, the model is trained with a normalised baseline against which all activity is compared [38]. When an abnormality crosses or does not align with the normalised baseline, the IDS raises an alarm. These alarms can also be triggered by unusual user logins, new IP addresses attempting to connect to the smart grid network, new devices admitted to the network without permission, and other events. Anomaly-based intrusion detection techniques can have higher false-positive rates because they report even the tiniest unusual activity in the network. Attackers are more likely to try new methods to sabotage the smart grid system, and attacks are more likely to be detected with anomaly-based intrusion detection [39–42].

5. Proposed Methodology

Classifier algorithms in cyber security are constantly learning by analysing data and formulating patterns in order to better detect Malware in the system [20]. The classifier algorithms can recognise patterns and detect threats in large datasets, and by automating the analysis, hosts can effectively detect and isolate compromised situations without disrupting the system's healthy components. Intrusion detection using classifier algorithms monitors network functions in real time for anomalous behaviour and processes data to analyse threats. Such techniques aid in the detection of insider threats, unknown malware, cyberattacks, false data manipulation, policy violations, and malicious Internet activity, as well as the analysis of attack infrastructure for any system threats [43]. Anomaly-based intrusion detection classifier algorithms can even detect malware and threats that have never been seen before; they detect new malicious files or activity based on the characteristics of previously identified attacks. We have detailed the implementation of the mean absolute deviation technique for anomaly-based intrusion detection in the following sections of the paper.

The average distance between the mean of the entire dataset and each individual value determines the mean absolute deviation (MAD) of a dataset; it aids in determining and describing variation in the dataset, as well as plotting the values to determine how dispersed the values are to distinguish anomaly data. The MAD value-based modified z-score of the voltage-time series is proposed as a threshold for clustering normal and anomalous points in the dataset in this study. Anomalies clusters are formed by all voltage data points with decision scores greater than the threshold value. And values below the threshold value are sorted into neat clusters. The classification methodology functions by supposing X_{i} as a datapoint in the voltage-time series and the classification of the datapoint as a normal or an anomaly are based on the formula by calculating the product of a scaling factor and the median of the absolute value of the difference of the data value point and the series median:

$$X_i = X_{\text{anomalous}} \text{when} \frac{\left|X_i - \text{med}_i * X_i\right|}{\text{MAD}_n} > \text{threshold}, \tag{1}$$

where X_i is the point at time "i," $X_{\text{anomalous}}$ is the anomalous point, $\text{med}_i * X_i$ is the middle value in the time series/data threshold is the modified z-score based on MAD value of the series, $\text{MAD}_n = k * \text{med}_i * |X_i - \text{med}_i * X_i|$, which gives the MAD value of the series, and K(=1.4826) is the scaling factor.

$$X_i = \begin{cases} 1 & \text{score} > \text{threshold} \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

By default, threshold value is 3.5 for classification. Based on the decision score, the data point is classified as normal and anomaly. A data point is labelled as 1 for an outlier point and 0 is labelled for an inlier point. The same procedure is applied to all the values in the dataset, and the points are categorised as outlier and inlier.

There are numerous reasons to use the MAD algorithm to detect abnormalities in a univariate dataset; however, in this study, we tested the voltage value dataset. One of the main advantages of using median absolute deviation is that it is unaffected by outliers. The breakdown point for MAD is 0.5, which is the maximum proportion of observations that can be contaminated without causing the estimator to produce a false value. This means it can handle up to 50% of contaminated values in a dataset, which in a time series dataset can be a huge challenge for a hacker to overcome. Also, it is immune to sample size and can perform the operation even on a dataset with billions of values in an affordable time frame as well as with much lesser

computational power, thus making it suitable for anomaly detection in an advanced metering infrastructure.

6. Simulation Test Case

The simulation test set-up is set up to run experiments using the proposed method as well as the other nine methods that use PyOD library algorithms. The dataset preparation, parameter selection, experimental design for the proposed method, and performance evaluation are all part of this set-up.

6.1. Dataset Details and Parameter Selection. The dataset used for obtaining the test results was energy meter readings from National Renewable Energy Lab (NREL). The dataset contains three main parameters, namely, voltage, current, and timestamp. For the training of the proposed intrusion detection technique, anomalies in the dataset were introduced using external means; in order to test the detection of stealth and nonstealth attacks, the anomalies are introduced at a specific percentage range. For stealth-based anomaly dataset, the anomaly percentage varied between ±6% of the average value of the dataset. For nonstealth anomalies based dataset, the abnormalities vary apart from ±6% band.

Table 1 summarises the datasets used for the comparative study.

A MAD model is strictly a univariate input model; thus, we considered only one parameter of the given two. The MAD model for outlier detection utilises voltage data points with anomalies in order to test and train the model.

6.2. Simulation Data Requirements. Data preprocessing is done for each algorithm's analysis goal, and then, all algorithms go through two main processes: clustering and labelling. The main experiments for each detection method are then carried out in the order listed below. Clustering is a partitioning method for generating clean and defective clusters. *Labelling*. Using the numbers "0" and "1," each data point is then labelled as an inlier or an outlier. We used Python libraries to archive the desired output accuracy and precision after writing a Python code to run the detection programme and plot the datapoints. The Sklearn libraries prepare the dataset for the evaluation procedure, and the PyOD libraries use it to run the algorithms. All of the methods used in the comparative study, as well as their data classification techniques, are listed in Table 2.

6.3. Performance Evaluation. For calibrating performance of the algorithm, Sklearn library is used. Accuracy, precision, recall, and F1 score are calculated for all algorithms for the same dataset and are compared.

Accuracy is calculated by the following formula:

$$\frac{Tp + Tn}{Tp + Tn + Fp + Fn}. \tag{3}$$

Table 1: Dataset for the comparative study.

Dataset description	Anomaly type	Number of features	Number of datapoints	Anomaly percentage
Smart meter 1	Stealth	6	500	6.21
	Normal	6	500	7.82
Smart meter 2	Stealth	6	500	6.78
	Normal	6	500	8.91

Table 2: Comparison of few existing methods for data classification techniques.

Technique/methods	Data set	Type
Mean absolute deviation	Real	Probabilistic
Minimum covariance determinant	Real	Linear model
One-class support vector machines	Real + historical	Linear model
Deviation-based outlier detection	Real	Linear model
Connectivity-based outlier factor	Real + historical	Proximity-based
Stochastic outlier selection	Real + historical	Probabilistic
k-Nearest neighbours	Real	Proximity-based
Isolation forest	Real + historical	Outlier ensembles
Lightweight online detector of anomalies	Real + historical	Outlier ensembles

Precision is calculated by the following formula:

$$\frac{Tp}{Tp + Fp}. \tag{4}$$

Here, for a detection model, precision is the intuitive ability to not classify a negative sample (anomalous data point) as a positive one (clean data point).

Recall is calculated by the following formula:

$$\frac{Tp}{Tp + Fn}. \tag{5}$$

Recall is the intuitive ability to identify all the positive samples (clean data points) in the dataset for a detection model.

The F1 score is the weighted harmonic mean (one of the Pythagorean means) of recall and precision multiplied by a factor of one, implying that recall and precision are equally weighted.

The following are the elements in the formula:

Tp–true positive–the number of normal datapoints correctly identified by the model.

Fp–false positive–the number of defective datapoints incorrectly identified as normal by the model.

Tn–true negative–the number of defective datapoints correctly identified by the model.

Fn–false negative–the number of normal datapoints incorrectly identified as defective by the model.

7. Simulation Result

This research is being carried out to see how well the proposed method for detecting anomaly points in a dataset performs. The proposed method's performance is compared to the performance of the other nine methods for detecting anomaly points. As explained in the "Experimental Setup," all the methods are performed on the same dataset. *Cybersecurity Risk Assessment Quantitative Model*. Through the use of qualitative and qualitative criteria, the model can be used to assess the market-place enterprises' readiness for security. We provide a Bayesian network approach that may be utilised to create a cyber security risk score using the security profile and data breach statistics of a company as input. The quantitative model makes it possible to accurately and consistently capture cyber risk. The scoring model aims to establish a standard in the market that could provide incentives for businesses to invest in and advance their security systems. An example of scoring an intrusion detection network finishes this study.

As discussed earlier, the proposed method is applied to two test sets for further validation of the method. Figure 1 shows the graphical representation of the proposed method results, for stealth-type attack on smart meters dataset, below are the results from both the test sets. As seen from the above plots from both the test sets, all the anomaly points are stratified and are confirmed by the classification report generated using respective libraries. Table 3 shows the results for the proposed method. The identification of attack points that are not stealthy in a similar vein and the nonstealth type attack points are plotted and charted after the stealth type attack points. The comparative test study done for the research found that the proposed method, median actual variation for anomaly identification in smart metering datasets, produced the most accurate and precise differentiations with the highest precision. By detecting datapoints above the threshold based on the dataset's MAD value, anomalous datapoints are located.

Figure 2 shows the graphical representation of the proposed method results, for nonstealth type attack on smart meters dataset, and below are the results from both the test sets.

The evaluation result is shown in Table 4. The dataset 1 shows 0.98 precision, whereas datasets 2 and 3 show precision is 1 for test set 1. The F1 score and recall values are 1 for datasets 1, 2, and 3 in test set 1. For the test set 2, the dataset 2 shows 1 precision, whereas dataset 3 also shows precision is 1. The F1 score and recall values are 1 for datasets 1, 2, and 3 in test set 2.

Both test sets are similarly processed from other detection models using the nine methods discussed. The plot below shows the detected anomalous points for the stealth type attack points of test set 1 using all of the listed methods. The comparison chart of the generated result for each detection model follows the plot. The detection of nonstealth type attack points and the results of test set 1 are also shown in the plot and chart below. Similarly, the plots below show the test set 2 results for both the proposed method and the comparison models. The stealth type attack points are plotted and charted first, followed by the nonstealth type attack points.

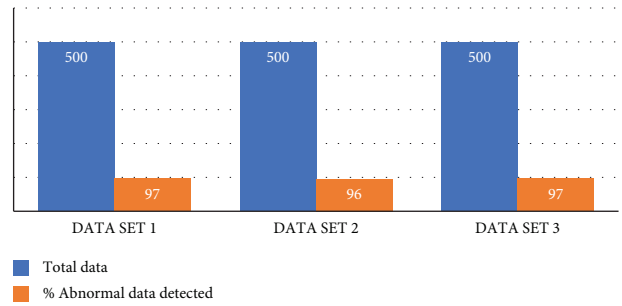

Figure 1: Detection of abnormal data in stealth dataset.

Table 3: Result of the proposed method for stealth type dataset.

Parameters	Test set 1			Test set 2		
	Precision	Recall	F1 score	Precision	Recall	F1 score
Dataset 1	0.93		1	0.96		1
Dataset 2	1	1	1	1	1	1
Dataset 3	1	1	1	1	1	1
No. of anomaly points		25			28	

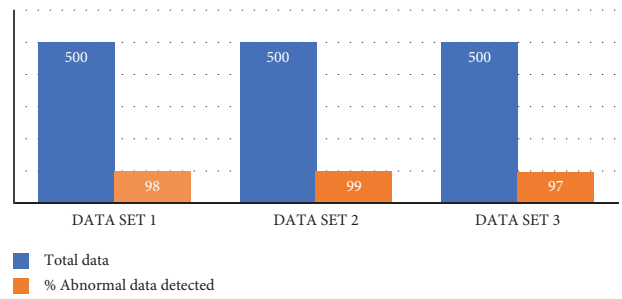

Figure 2: Detection of abnormal data in nonstealth dataset.

Table 4: Result of the proposed method for nonstealth type dataset.

Parameters	Test set 1			Test set 2		
	Precision	Recall	F1 score	Precision	Recall	F1 score
Dataset 1	0.98		1	0.98		1
Dataset 2	1	1	1	1	1	1
Dataset 3	1	1	1	1	1	1
No. of anomaly points		36			24	

8. Conclusion

The proposed method, median absolute deviation for anomaly detection in smart metering datasets, delivered the most accurate and precise differentiations with the highest accuracy and precision, according to the comparative test

study conducted for the paper. Anomaly datapoints are identified by identifying points above the threshold point based on the dataset's MAD value. The research focuses solely on detecting anomaly-based data values in smart metering infrastructure. We can deduce and propose a qualitatively improved anomaly-based intrusion detection system based on the obtained results. Because MAD is a strictly univariant method, it can run algorithms on multiple featured smart meter datasets at once, reducing variance and limiting it to a specific range, lowering the false-positive ratio during anomaly detection. The deployed datasets can be combined with the specified classification of attack type by integrating advanced algorithmic techniques that enable more precise and accurate signature-based intrusion detection system can be modelled to implement it on a larger dataset obtained from an advanced metering infrastructure for further evolution of MAD. In future, we will use smart grid security classification (SGSC), which creates for complex systems like the smart grid. The differences that were made most precisely and accurately came from the datasets from smart meters. The advanced metering infrastructure (AMI) technologies are the main focus.

Data Availability

The data used to support the findings of this study are available from the corresponding author on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] B. O'Gorman, C. Wueest, and D. O'Brien, "Internet security threat report (ISTR) 2019 | Symantec," Broadcom, tech. rep., Symantec, 2019.

[2] C. Castelli, B. Gabriel, J. Yates, and P. Booth, "Strengthening digital society against cyber shocks - key findings from the Global state of information security Survey," tech. rep., PwC, 2018.

[3] K. A. Alissa, B. A. AlDeeb, H. A. Alshehri et al., "Developing a simulated intelligent instrument to measure user behavior toward cybersecurity policies," *International Journal of Communication Networks and Information Security*, vol. 13, no. 1, pp. 82–91, 2021.

[4] R. D. S, L. Shyamala, and S. Saraswathi, "Adaptive learning based whale optimization and convolutional neural network algorithm for distributed denial of service attack detection in software defined network environment," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 10, no. 6, pp. 80–93, 2022.

[5] B. Pabbuleti and J. Somlal, "Implementation of multi-level bidirectional inter allied converter community for global power sharing in hybrid AC/DC microgrids," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 10, no. 6, pp. 52–62, 2022.

[6] C. B. Moorthy and M. K. Deshmukh, "A new approach to optimise placement of wind turbines using particle swarm optimisation," *International Journal of Sustainable Energy*, vol. 34, no. 6, pp. 396–405, 2015.

[7] G. S. Sajja, "Impact of supply Chain management Strategies on business performance," *International Journal of Computer Applications*, vol. 183, no. 38, pp. 45–49, 2021.

[8] M. K. Deshmukh and C. B. Moorthy, "Review on stability analysis of grid connected wind power generating system," *International Journal of Electrical and Electronics Engineering Research and Development (IJEEERD)*, vol. 3, no. 1, 2013.

[9] P. Wood, B. Nahorney, K. Chandrasekar, K. Haley, and S. Wallace, "Internet security threat report," Symantec Corporation, tech. rep, 2016.

[10] F. Guerhardt, T. A. F. Silva, F. M. C. Gamarra et al., "A smart grid system for reducing energy consumption and energy cost in buildings in São Paulo, Brazil," *Energies*, vol. 13, no. 15, p. 3874, 2020.

[11] M. E. El-hawary, "The smart grid-state-of-the-art and future trends," *Electric Power Components and Systems*, vol. 42, no. 3-4, pp. 239–250, 2014.

[12] R. Rashed Mohassel, A. Fung, F. Mohammadi, and K. Raahemifar, "A survey on advanced metering infrastructure," *International Journal of Electrical Power & Energy Systems*, vol. 63, pp. 473–484, 2014.

[13] Hansen A., Staggs J., and Shenoi S., "Security analysis of an advanced metering infrastructure. International Journal of Critical Infrastructure Protection," vol. 18, pp. 3–19, 2017.

[14] M. S. Hamid, N. A. Manap, R. A. Hamzah, and A. F. Kadmin, "Stereo matching algorithm based on hybrid convolutional neural network and directional intensity difference," *International Journal of Emerging Technology and Advanced Engineering*, vol. 11, no. 6, pp. 86–96, 2021.

[15] M. A. Haq, K. Jain, and K. P. R. Menon, "Volumetric Glacier Changes in the Garhwal Himalayas using multitemporal digital elevation model from 2001 to 2010," in *Proceedings of the 35th International Symposium on Remote Sensing of Environment (ISRSE)*, pp. 1–12, IEEE, Beijing, April 2013.

[16] A. P. Morgan, J. A. Cafeo, D. I. Gibbons, R. M. Lesperance, G. H. Sengir, and A. M. Simon, "CBR for Dimensional management in a manufacturing plant," in *Case-Based Reasoning Research and Development*, pp. 597–610, Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.

[17] I. Ghafir and V. Prenosil, "DNS traffic analysis for malicious domains detection," in *Proceedings of the 2nd International Conference on Signal Processing and Integrated Networks*, pp. 613–618, Institute of Electrical and Electronics Engineers Inc, Noida, India, April 2015.

[18] J. J. Davis and A. J. Clark, "Data preprocessing for anomaly-based network intrusion detection: a review," *Computers & Security*, vol. 30, no. 6-7, pp. 353–375, 2011.

[19] P. Casas, J. Mazel, and P. Owezarski, "Unsupervised network intrusion detection systems: detecting the unknown without Knowledge," *Computer Communications*, vol. 35, no. 7, pp. 772–783, 2012.

[20] D. M. Diab, B. AsSadhan, H. Binsalleeh, S. Lambotharan, K. G. Kyriakopoulos, and I. Ghafir, "Anomaly detection using dynamic time Warping," in *Proceedings of the 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, pp. 193–198, IEEE, New York, NY, USA, August 2019.

[21] J. Bhola and S. Soni, "Information theory-based Defense Mechanism against DDOS attacks for WSAN," in *Advances in VLSI, Communication, and Signal Processing*, D. Harvey, H. Kar, and S. Verma, Eds., vol. 683, 2021.

[22] B. AsSadhan, R. AlShaalan, D. M. Diab et al., "A robust anomaly detection method using a constant false alarm rate

approach," *Multimedia Tools and Applications*, vol. 79, no. 17-18, pp. 12727–12750, 2020.

[23] B. AsSadhan, K. Zeb, J. AlMuhtadi, and S. AlShebeili, "Anomaly detection based on LRD behavior analysis of Decomposed control and data Planes network traffic using SOSS and FARIMA models," *IEEE Access*, vol. 5, pp. 13501–13519, 2017.

[24] A. Lakhina, M. Crovella, and C. Diot, "Characterization of network-wide anomalies in traffic flows," in *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*, Association for Computing Machinery, New York, NY, USA, October 2004.

[25] S. Ahmed, Y. Lee, S.-H. Hyun, and I. Koo, "Unsupervised machine learning-based detection of Covert data integrity assault in smart grid networks utilizing isolation forest," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2765–2777, 2019.

[26] M. R. Camana Acosta, S. Ahmed, C. E. Garcia, and I. Koo, "Extremely Randomized Trees-based Scheme for stealthy cyber-attack detection in smart grid networks," *IEEE Access*, vol. 8, pp. 19921–19933, 2020.

[27] M. Mohammadpourfard, Y. Weng, M. Pechenizkiy, M. Tajdinian, and B. Mohammadi-Ivatloo, "Ensuring cybersecurity of smart grid against data integrity attacks under concept Drift," *International Journal of Electrical Power & Energy Systems*, vol. 119, Article ID 105947, 2020.

[28] M. Mohammadpourfard, A. Sami, and Y. Weng, "Identification of false data injection attacks with considering the impact of wind generation and Topology Reconfigurations," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 3, pp. 1349–1364, 2018.

[29] R. Moslemi, A. Mesbahi, J. M. Velni, and A. Fast, "A Fast, Decentralized Covariance selection-based approach to detect cyber attacks in smart grids," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4930–4941, 2018.

[30] B. Li, T. Ding, C. Huang, J. Zhao, Y. Yang, and Y. Chen, "Detecting false data injection attacks against power system state estimation with Fast go-Decomposition approach," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2892–2904, 2019.

[31] Y. Hao, M. Wang, and J. H. Chow, "Likelihood analysis of cyber data attacks to power systems with Markov decision processes," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3191–3202, 2018.

[32] J. Zhao, L. Mili, and M. Wang, "A Generalized false data injection attacks against power system nonlinear state estimator and countermeasures," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4868–4877, 2018.

[33] M. Ostadijafari, R. R. Jha, and A. Dubey, "Conservation voltage Reduction by Coordinating Legacy devices, smart Inverters and Battery," in *Proceedings of the 2019 North American Power Symposium*, IEEE, Wichita, KS, USA, October 2019.

[34] F. Mohammadi, G.-A. Nazri, and M. Saif, "A real-time Cloud-based intelligent car Parking system for smart Cities," in *Proceedings of the 2019 IEEE 2nd International Conference on Information Communication and Signal Processing*, IEEE, Weihai, China, September 2019.

[35] F. Mohammadi, G.-A. Nazri, and M. Saif, "A Bidirectional power Charging control strategy for Plug-in hybrid electric Vehicles," *Sustainability*, vol. 11, no. 16, p. 4317, 2019.

[36] T. Nguyen, S. Wang, M. Alhazmi, M. Nazemi, A. Estebsari, and P. Dehghanian, "Electric power grid Resilience to cyber Adversaries: State of the art," *IEEE Access*, vol. 8, pp. 87592–87608, 2020.

[37] P. K. Gupta, N. K. Singh, and V. Mahajan, "Intrusion detection in cyber-Physical layer of smart grid using intelligent Loop based Artificial neural network technique," *International Journal of Engineering*, vol. 34, no. 5, pp. 1250–1256, 2021.

[38] C. C. Sun, A. Hahn, and C. C. Liu, "Cyber security of a power grid: State-of-the-art," *International Journal of Electrical Power & Energy Systems*, vol. 99, pp. 45–56, 2018.

[39] N. K. Singh and V. Mahajan, "Analysis and evaluation of cyber-attack impact on critical power system infrastructure," *Smart Science*, vol. 9, no. 1, pp. 1–13, 2021.

[40] M. Z. Gunduz and R. Das, "Cyber-security on smart grid: threats and potential solutions," *Computer Networks*, vol. 169, Article ID 107094, 2020.

[41] N. K. Singh, P. K. Gupta, and V. Mahajan, "Intrusion detection in wireless network of smart grid using intelligent trust-weight method," *Smart Science*, vol. 8, no. 3, pp. 152–162, 2020.

[42] Z. E. Mrabet, N. Kaabouch, H. E. Ghazi, and H. E. Ghazi, "Cyber-security in smart grid: Survey and challenges," *Computers & Electrical Engineering*, vol. 67, pp. 469–482, 2018.

[43] K. Kimani, V. Oduol, and K. Langat, "Cyber security challenges for IoT-based smart grid networks," *International Journal of Critical Infrastructure Protection*, vol. 25, pp. 36–49, 2019.