STM JOURNALS

Trends in
Electrical Engineering

https://journals.stmjournals.com/tee

ISSN: 2249-4774 (Online)
ISSN: 2321-4260 (Print)
Volume 14, Issue 2, 2024
May—August
DOI (Journal): 10.37591/TEE

Research TEE

Enhancing Smart Grid Security: Machine Learning Approaches for Detecting Anomalies

Aadish Jain[1], Abhinav Kumar[1], Gaurav Joshi[1], Surya Pratap Singh[1], Dheeraj Verma[2,*], Praveen Kumar Agrawal[3]

Abstract

The integration of Information and Communication Technology (ICT) with traditional electric grids has led to the development of smart grids. However, this integration has also increased the risk of anomalies, such as cyber-attacks, metering fraud, electricity theft etc. False Data Injection Attacks are a class of cyber-attacks against power grid monitoring systems, where adversaries can inject false data to manipulate the grid's operation. Metering frauds pertain to malicious customers compromising their smart meters to report false readings to achieve financial gains illegally. As a result, there is a growing need for effective detection mechanisms to identify and mitigate anomalies in smart grids. One approach to detecting anomalies in smart grids is through the use of machine learning algorithms. These algorithms leverage the data generated by smart grid systems to identify anomalies that may indicate the presence of false data injection attacks. The detection of false data injection attacks in smart grids is a critical area of research, given the increasing risk of cyber-attacks in these systems. Machine learning algorithms, transfer learning, and deep learning-based methods have been proposed as effective mechanisms for detecting and mitigating anomalies in smart grids, ultimately contributing to the security and reliability of smart grid systems.

Keywords: Anomalies, machine learning, information and communication technology, false data injection attacks

INTRODUCTION

Electricity grids, traditionally reliant on established infrastructure and conventional energy sources, grapple with a myriad of challenges such as sluggish response times and soaring operational expenses. This situation underscores the urgent need for transformational solutions to modernize these grids. The emergence of smart grids, operating as sophisticated cyber-physical systems, represents a significant stride forward in addressing these challenges. By integrating advanced communication networks and embracing renewable energy sources, smart grids offer promising avenues for enhancing efficiency and resilience within the power sector [1]. The integration of communication networks and Internet of Things (IoT) technologies exposes these grids to various cyber threats, including malicious attacks and electricity theft, thereby elevating the risk of grid failure [2] of concern are false data injection (FDI) attacks, a sophisticated form of cyber assault aimed at disrupting the accurate estimation of power system states. These attacks can circumvent conventional detection systems, posing significant challenges to the security and stability of smart grids [3].

***Author for Correspondence**
Dheeraj Verma
E-mail: 2021ree9058@mnit.ac.in

[1]Student, Department of Electrical Engineering, Malaviya National Institute of Technology, Jaipur, Rajasthan, India
[2]Research Scholar, Department of Electrical Engineering, Malaviya National Institute of Technology, Jaipur, Rajasthan, India
[3]Associate Professor, Department of Electrical Engineering, Malaviya National Institute of Technology, Jaipur, Rajasthan, India

Received Date: June 29, 2024
Accepted Date: July 22, 2024
Published Date: July 25, 2024

Citation: Aadish Jain, Abhinav Kumar, Gaurav Joshi, Surya Pratap Singh, Dheeraj Verma, Praveen Kumar Agrawal. Enhancing Smart Grid Security: Machine Learning Approaches for Detecting Anomalies. Trends in Electrical Engineering. 2024; 14(2): 10–19p.

Instances of cyber-attacks on energy infrastructure, such as those in Ukraine and the 2014 attack on South Korean nuclear facilities, underscore the grave vulnerability of the power sector to such threats, necessitating robust cybersecurity measures [4]. These incidents underscore the multifaceted risks posed by cyber-attacks, including economic disruption, national security threats, and erosion of public trust in critical infrastructure [5]. In response to these challenges, the research community has increasingly turned to advanced anomaly detection techniques to fortify the security of smart grids. Anomaly detection encompasses a range of data-driven and model-based algorithms, with machine learning (ML) approaches emerging as particularly promising due to their scalability and real-time capabilities [6]. Techniques such as artificial immune systems (AIS) and support vector machines (SVM) offer distinct advantages over traditional model-based algorithms, enabling efficient detection of anomalies without the need for detailed system models and parameters [7].

Despite the promise of ML-based approaches, several challenges persist. These include handling the high dimensional and computational complexities inherent in smart grid data, as well as addressing the phenomenon of attack sparsity during cyber-attacks. Attack sparsity refers to scenarios where only a small fraction of measurements is compromised, posing a significant hurdle to the accuracy and robustness of detection techniques. Consequently, there is a pressing need for innovative methodologies that can effectively navigate these challenges and enhance the resilience of smart grids against evolving cyber threats [8].

To address these challenges, the study proposes employing various ML models, including k-NN, Isolation Forest, HBOS, CBLOF, and ABOD, for anomaly detection, using load profile data from an IEEE 24-bus RTS system for comparative analysis. The study's key contributions include the development of a robust algorithm capable of handling dimensionality without compromising accuracy, along with techniques for enhancing efficiency and accuracy, ultimately providing a robust model for effectively detecting various anomalies in smart grids.

LITERATURE SURVEY

Smart grids represent a transformative shift in electricity distribution, leveraging digital communication technologies to optimize operations, integrate renewables, and bolster cyber-security, with machine learning enhancing anomaly detection and proactive management. They revolutionize energy management by enhancing efficiency through real-time monitoring and optimized distribution, reducing wastage and costs for utilities and consumers [1].

Integration of renewable energy sources like solar and wind power promotes sustainability while managing variable out-puts, reducing carbon footprints, and supporting environmental goals. Additionally, smart grids ensure heightened reliability with advanced monitoring and fault detection, mitigating downtime. Demand-response initiatives empower consumers to adjust electricity usage, balancing supply and demand for cost savings. Overall, smart grids play a pivotal role in modernizing energy infrastructure, fostering sustainability, and efficiently meeting evolving energy needs. Anomalies in smart grids encompass deviations from normal operation and arise from equipment malfunctions, cyberattacks, environmental factors, human errors, metering fraud, and electricity theft [9]. Early detection of equipment malfunctions prevents disruptions, while robust cybersecurity mitigates cyberattack risks [3]. Proactive monitoring and resilience planning address environmental anomalies, while advanced metering technologies combat fraud and theft. Comprehensive strategies are vital for detecting, addressing, and mitigating anomalies to ensure grid reliability and security.

Cybersecurity is crucial in smart grid infrastructure due to reliance on digital technologies, presenting vulnerabilities that must be addressed for grid reliability, privacy, and security [4]. Safeguarding transmitted data requires encryption and secure communication channels, while protecting control systems prevents unauthorized access [5]. Resilience planning ensures grid functionality in adverse scenarios. Robust cybersecurity measures, including data encryption and threat detection, are

imperative for grid security amid evolving threats. Machine learning focuses on developing algorithms for pattern recognition and prediction, aiding anomaly detection in smart grids [6]. Classification into supervised, unsupervised, and reinforcement learning is based on the learning approach and data characteristics [10]. In smart grids, anomaly detection using machine learning involves employing advanced algorithms to identify deviations from normal behavior by analyzing data from sensors, meters, and control systems in real time [11]. Anomaly detection involves analyzing data from sensors and control systems in real time, with key steps including data collection, preprocessing, and model training. Machine learning enables proactive identification of disruptions and optimization of energy management, highlighting the importance of robust cybersecurity and human factors considerations for grid operation.

Several machine learning algorithms, including k-Nearest Neighbors (k-NN), Isolation Forest, Histogram-based Outlier Score (HBOS), Angle-Based Outlier Detection (ABOD), and Cluster-Based Local Outlier Factor (CBLOF), are employed for anomaly detection in smart grids [12–14]. k-NN identifies outliers by measuring distances between data points and their nearest neighbors, while Isolation Forest isolates anomalies based on the average path length across multiple isolation trees. HBOS constructs histograms to detect anomalies in energy consumption data, ABOD analyzes angles formed by data points to identify deviations, and CBLOF clusters data points to assess local outlier factors. These algorithms enable proactive detection of anomalies such as equipment failures, cyberattacks, and metering fraud, enhancing grid reliability, efficiency, and security.

METHODOLOGY

The flowchart as shown in Figure 1, outlines a structured workflow from gathering load profile data to developing a robust anomaly detection model. It includes preprocessing the data, strategically inserting anomalies, dividing it into training and test sets, and evaluating model performance using various metrics, ensuring a comprehensive analysis and robust model development for power system anomaly detection.

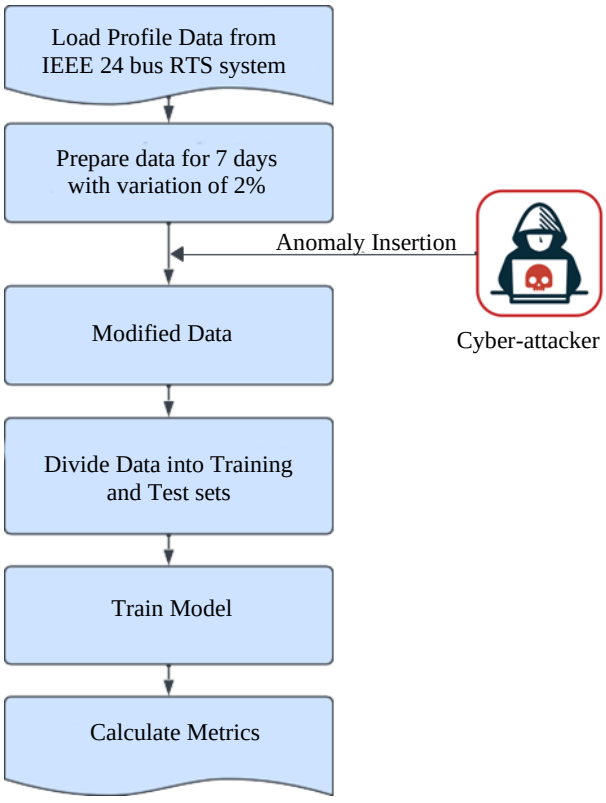

Figure 1. Flowchart representing methodology.

Data Acquisition

The load profile as shown in Figure 2, is obtained from the IEEE 24 RTS bus system which provides a representative data set for studying smart grid behavior [15].

Data Preprocessing

The aggregate system demand per hour and the percentage distribution of load at each node were utilized to calculate the load at each node for every hour of the day. This calculation was performed using standard Excel functions and formulas. To ensure sufficient data for effective ML model training, the available dataset was expanded to cover a period of 7 days. This extension is carried out by replicating the data for each day with a slight tolerance of ±2% to reflect real-world fluctuations [15].

Anomaly Injection

Anomaly injection is the intentional introduction of abnormal data points into a dataset or system. It is used to test the system's ability to detect and handle anomalies, evaluating its robustness and performance.

Mathematics of anomaly injection

Anomalies in this study are created through a stochastic process where individual entities within the dataset are randomly chosen, and a disturbance, a fraction of the data, is added to them.

The subscripts "t" and "*n*" denote the time and node indices, respectively, while uppercase letters signify specific indices within the dataset. And

1. $\tilde{x}$: load vector.
2. $\tilde{a}$: disturbance vector.

The equation for Type 1 anomaly injection is:

$$\tilde{x}_{t,n} \rightarrow \tilde{x}_{t,n} \pm \tilde{a}_{t,n} \tag{1}$$

where at n ∈ $[0.25\tilde{x}_{t,\,n},\ 0.3x_{t,n}]$

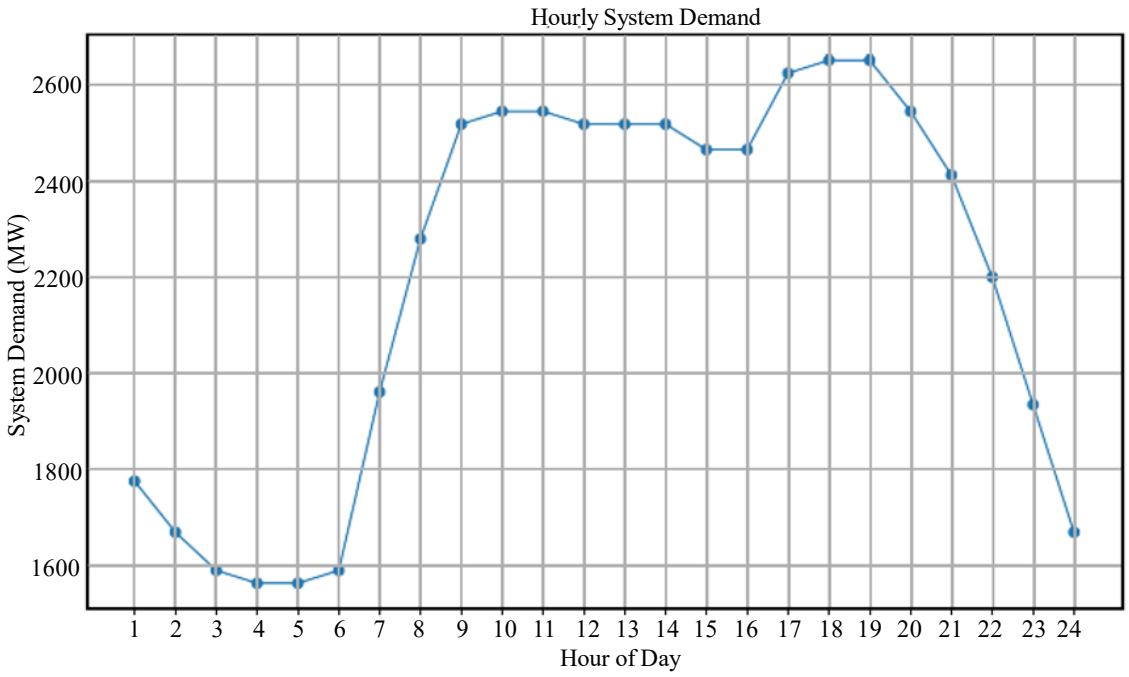

Figure 2. Sample dataset.

The equation for Type 2.1 anomaly injection is:

$$\tilde{x}_{t,N} \rightarrow \tilde{x}_{t,N} \pm \tilde{a}_t \quad (2)$$

where $\alpha_{t,n} \in [0.25x_{t,N}, 0.3x_{t,N}]$

The equation for Type 2.2 anomaly injection is:

$$\tilde{x}_{T,n} \rightarrow \tilde{x}_{T,n} \pm \tilde{a}_n \quad (3)$$

Where, $\alpha_{T,n} \in [0.25x_{T,n}, 0.3x_{T,n}]$

The equation for Type 3 anomaly injection is:

$$\tilde{x}_{t,n} \rightarrow \tilde{x}_{t,n} \pm \tilde{a}_{T,n} \quad (4)$$

where $\Sigma\, \alpha_{t,n}=0$

Methods of Anomaly Injection

After expanding the data to correspond to 7 days, a total of 4032 entries were available for training ML algorithms. Anomalies were introduced in the entire dataset. Anomalies were inserted in the following three ways:

Type 1: injection of Anomalies at Random in the Entire Dataset

Choose a random subset comprising 10% of the indices from the entire dataset for inserting anomalies. Randomly adjust the load values by increasing or decreasing them by 25 to 30% [16]. Set the corresponding 'Anomaly' column value to 1 to mark this row as an anomaly.

Type 2: Anomaly injection at a Particular Node and Time

Type 2.1: injection on a Particular Node at Random Time Instances: Anomalies were inserted into the node with the highest standard deviation. Specifically, anomalies were added at 10% of the time indices for node which had the highest standard deviation. The load values were adjusted by randomly increasing or decreasing them by 25 to 30%. For each modified load value, the corresponding 'anomaly' column was set to 1 to indicate an anomaly.

Type 2.2: injection at a Particular Time (Peak Time) at Random Nodes: Anomalies were inserted at the hour with the highest load. Specifically, anomalies were added at 10% of the nodes for the 18th hour, which had the highest load. The load values were randomly adjusted by increasing or decreasing them by 25 to 30%. Subsequently, the 'Anomaly' column was set to 1 for each modified load value to mark it as an anomaly.

Type 3: injection of Anomalies at Random in the Entire Dataset While Keeping the Total Load Same

Select a random subset representing 10% of the indices from the complete dataset to introduce anomalies. Randomly modify the load values by increasing or decreasing them by 25 to 30% while maintaining the total hourly load constant. Assign a value of 1 to the 'Anomaly' column for each row where an anomaly is introduced, marking it as an anomaly. This ensures that the total load remains unchanged for each hour while allowing variation at individual nodes.

Training and Testing

Initially, a dataset from Kaggle was utilized to familiarize with machine learning training and testing procedures, as well as outlier detection techniques, leveraging the platform's user-friendly interface and sample datasets [17]. Subsequently, after gaining proficiency with these processes and IDE usage, a new dataset sourced from a 24 Bus RTS system, comprising 4032 entries was utilized for

comprehensive algorithm training and testing in a real-world context. Within this dataset, anomalies were deliberately introduced to simulate real-world irregularities, enabling effective assessment of the algorithms' anomaly detection capabilities.

Following the injection of anomalies, 80% of the dataset entries were allocated for algorithm training, ensuring robust training across diverse data points, while the remaining 20% served for independent testing to evaluate performance and generalization capabilities. This systematic approach allowed for rigorous assessment of the algorithms' efficacy in detecting anomalies within the IEEE-24 Bus RTS system dataset [18–24]. After partitioning the data, each algorithm underwent training to learn underlying patterns, followed by the task of identifying anomalies within the testing dataset. Consequently, each machine learning model employed resulted in the generation of a series of graphs as described in the results.

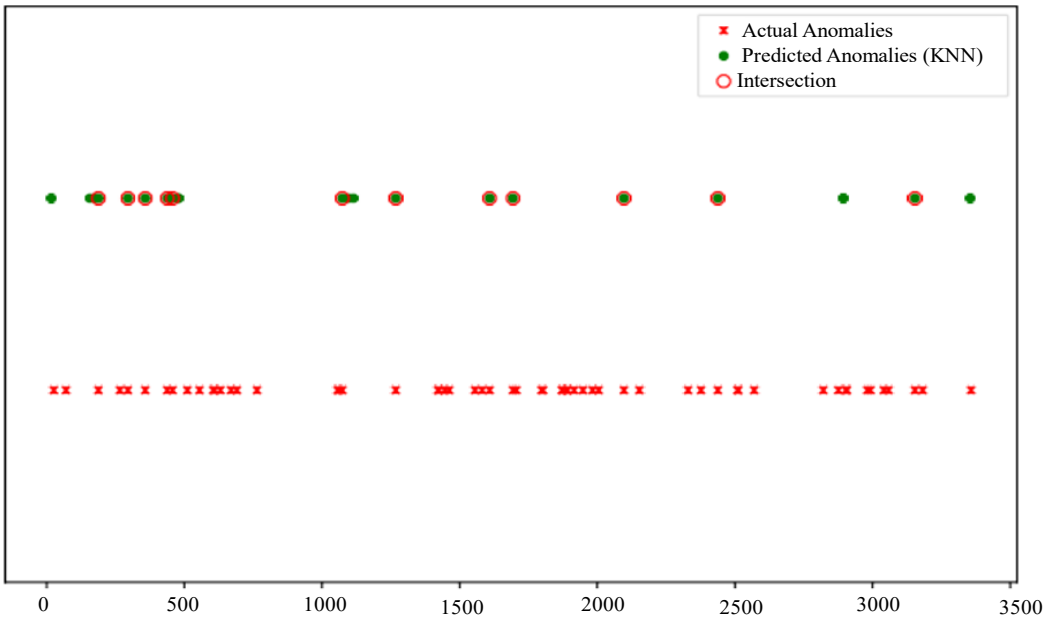

Figure 3. Predicted result by kNN.

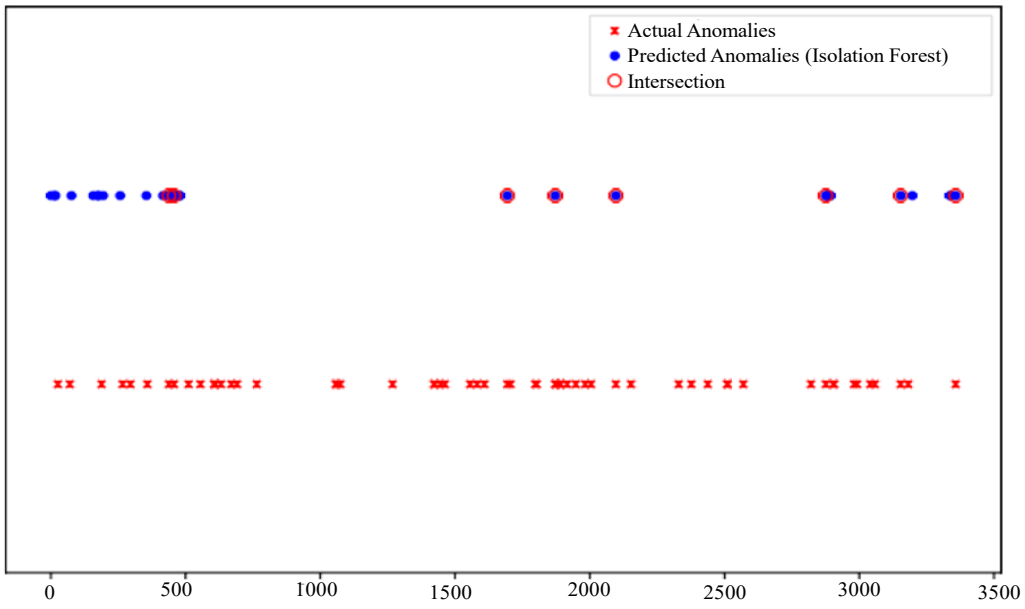

Figure 4. Predicted result by IF.

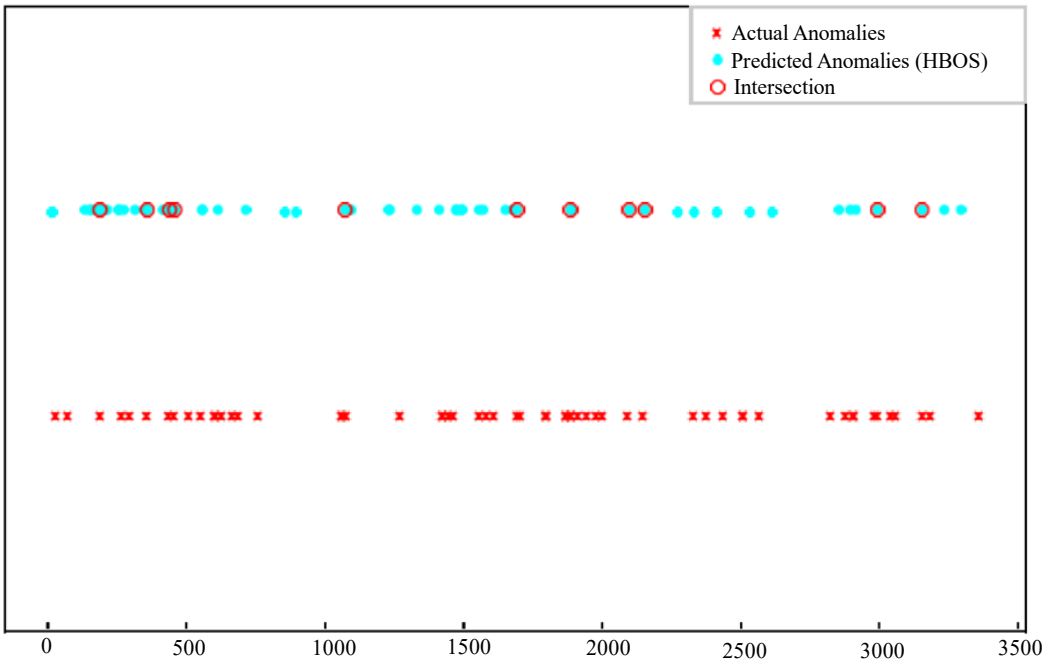

Figure 5. Predicted result by HBOS.

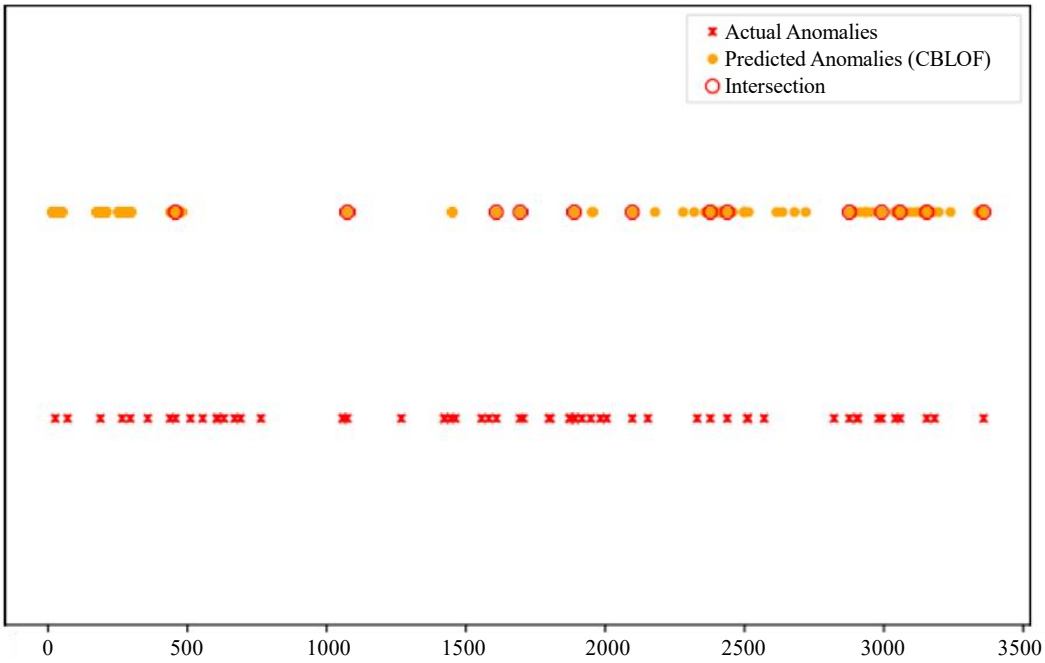

Figure 6. Predicted result by CBLOF.

RESULTS

After introducing anomalies into the dataset through anomaly injection methods discussed above, analysis was done using various machine learning algorithms within the VS Code Integrated Development Environment (IDE). The algorithms examined included k-NN, ABOD, HBOS, Isolation Forest, and CBLOF, aiming to detect anomalies across different types introduced into the dataset. Figures 3–7 illustrate the outcomes of anomaly detection using these algorithms via visualizing indices of predicted anomalies against actual anomalies. These results are in relation to Type 1 anomaly. Anomalies, i.e. disturbance ranging between 25 and 30%, randomly injected at any node or hour are shown in Figure 8. These visuals collectively offer insights into the diverse performance and

capabilities of each algorithm in detecting anomalies. A qualitative performance comparison of methods used is shown in Table 1 by means of accuracy.

Figure 7. Predicted result by ABOD.

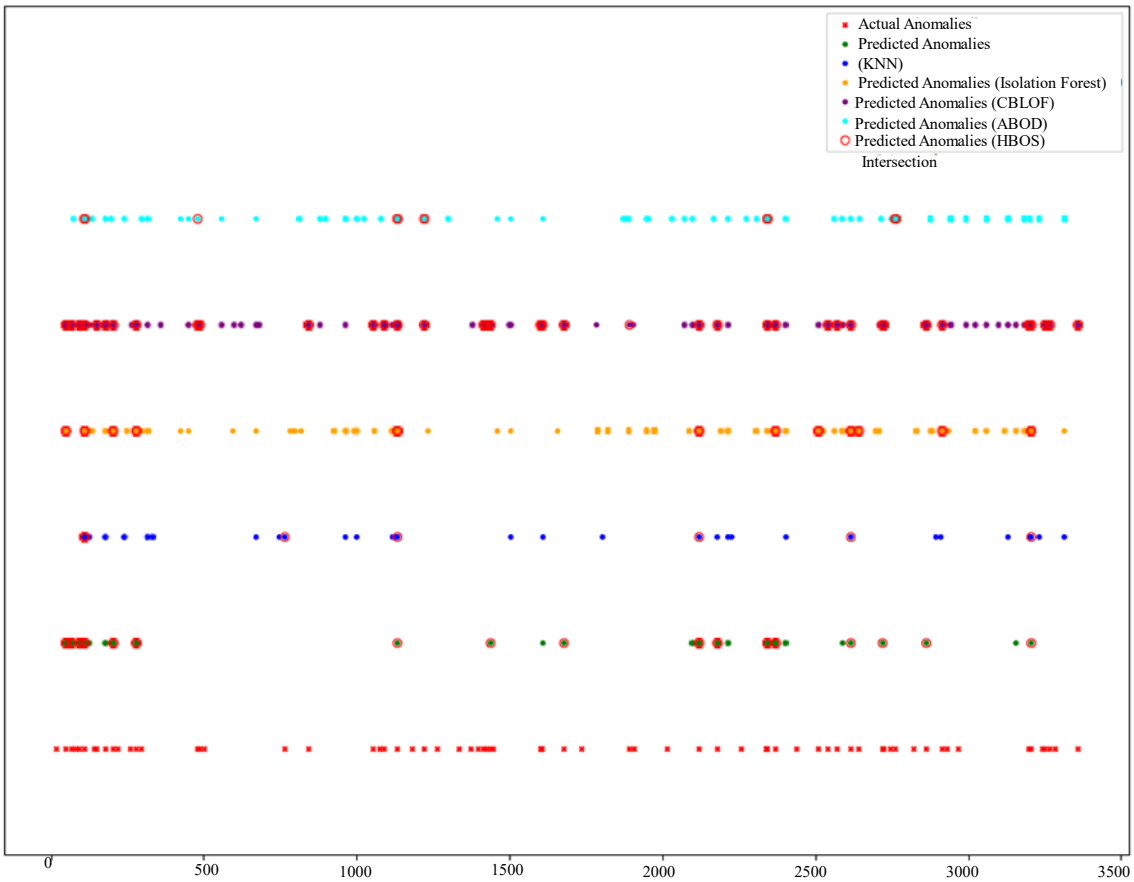

Figure 8. Cumulative Predicted result for Type 3 Anomaly.

Table 1. Accuracy for all types of anomalies.

Method Name	Type of Anomaly			
	Type 1	***Type 2 A***	***Type 2 B***	***Type 3***
k-NN	88.54	94.94	95.83	94.79
HBOS	81.85	89.88	90.47	89.43
ABOD	85.57	89.88	90.62	89.88
CBLOF	81.25	90.02	91.07	88.98
IF	85.56	94.38	94.49	94.34

Table 2. Anomaly Detection Results.

Method	Anomaly 25 to 30%		Anomaly 45 to 50%	
	Accuracy %	***Precision %***	***Accuracy %***	***Precision%***
k-NN	88.54	48.48	91.81	57.14
HBOS	81.84	18.05	86.16	19.29
ABOD	85.56	37.64	89.43	38.59
CBLOF	80.65	12.12	83.18	13.33
IF	86.01	18.18	89.28	26.67

DISCUSSION

This study aimed to assess the efficacy of machine learning (ML) techniques in identifying anomalies within smart grid data, with the objective of determining the most accurate and precise algorithm among the five methods considered. The analysis underscored the effectiveness of ML techniques in anomaly detection, where both supervised and unsupervised learning algorithms facilitated the accurate identification of deviations in load demand compared to the initial load profile before anomaly injection. The effects of depth of disturbance on performance can be seen from Table 2. The successful application of ML methods in anomaly detection suggests their potential to improve the monitoring and maintenance of smart grid systems, enabling utility companies to promptly identify irregular activity and bolster the resilience of electric grid infrastructure. Despite promising outcomes, certain limitations should be acknowledged, including the reliance on historical data for ML model training, which may overlook real-time anomalies, and the potential variability in model effectiveness based on the characteristics of smart grid data and selected features for analysis.

CONCLUSION

Incorporating machine learning for anomaly detection in smart grid data enhances energy management and infrastructure security. Supervised and unsupervised ML algorithms effectively identify deviations, improving reliability and robustness. Historical data patterns enable proactive intervention against potential threats. Continuous monitoring of data streams detects abnormal activity, mitigating cyber threats and ensuring energy system integrity. Future research may develop advanced ML algorithms tailored to smart grid challenges, integrate real-time data streams, and implement distributed anomaly detection systems for improved efficiency and scalability. Such advancements hold promise for fortifying smart grid infrastructure against evolving threats and ensuring reliable energy delivery.

REFERENCES

1. Hentea M. A perspective on research initiatives in cybersecurity engineering for future smart grids. In 2022 IEEE International Conference on Electro Information Technology (eIT). 2022; 352–357.
2. Ganguly P, Nasipuri M, Dutta S. Challenges of the existing security measures deployed in the smart grid framework. In 2019 IEEE 7th International Conference on Smart Energy Grid Engineering (SEGE). 2019; 1–5.
3. Dayaratne T, *et al*. False data injection attack detection for secure distributed demand response in smart grids. In 2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Net- works (DSN). 2022; 367–380.

4. Venkatachary SK, Prasad J, Samikannu R. Cybersecurity and cyber terrorism-in energy sector–a review. J Cyber Secur Technol. 2018; 2(3–4): 111–130.
5. Yadav SA, *et al*. A review of possibilities and solutions of cyber-attacks in smart grids. In 2016 IEEE International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH). 2016; 60–63.
6. Gupta R. A survey on machine learning approaches and its techniques. In 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). 2020; 1–6.
7. Dawoud A, Shahristani S, Raun C. Deep learning for network anomalies detection. In 2018 IEEE International Conference on Machine Learning and Data Engineering (iCMLDE). 2018; 149–153.
8. Dwivedi RK, Rai AK, Kumar R. Outlier detection in wireless sensor networks using machine learning techniques: a survey. In 2020 IEEE International Conference on Electrical and Electronics Engineering (ICE3). 2020; 316–321.
9. Fathnia F, Fathnia F, Javidi DMH. Detection of anomalies in smart meter data: A density-based approach. In 2017 IEEE Smart Grid Conference (SGC). 2017; 1–6.
10. Sharma A, Kaur A, Semwal A. Supervised and unsupervised prediction application of machine learning. In 2022 IEEE International Conference on Cyber Resilience (ICCR). 2022; 1–5.
11. Dalal KR. Analysing the role of supervised and unsupervised machine learning in iot. In 2020 IEEE International Conference on Electronics and Sustainable Communication Systems (ICESC). 2020; 75–79.
12. Wang Z-M, Song G-H, Gao C. An isolation-based distributed outlier detection framework using nearest neighbor ensembles for wire-less sensor networks. IEEE Access. 2019; 7: 96319–96333.
13. Kabir MA, Luo X. Unsupervised learning for network flow based anomaly detection in the era of deep learning. In 2020 IEEE 6th International Conference on Big Data Computing Service and Applications (BigDataService). 2020; 165–168.
14. Chahar RK, Singh AS. Using machine learning techniques for outlier detection application. In 2022 IEEE 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N). 2022; 238–241.
15. Ordoudis C, *et al*. An updated version of the IEEE RTS 24-bus system for electricity market and power system operation studies. Technical University of Denmark, Tech Rep. 13. 2016.
16. Yuan Y, Li Z, Ren K. Modeling load redistribution attacks in power systems. IEEE Trans Smart Grid. 2011; 2(2): 382–390.
17. daython3. Anomaly detection using pyod python library. [Online]. medium.com/ @daython3/ anomaly-detection-using-pyod-85b22ef70c00. Accessed: 5 May 2024.
18. Kumar M, *et al*. A conceptual introduction of machine learning algorithms. In 2023 IEEE 1st International Conference on Intelligent Computing and Research Trends (ICRT). 2023; 1–7.
19. Talati A, *et al.* Cyber-attack detection in smart grids using machine learning approach. In 2023 IEEE 7th International Conference on Computer Applications in Electrical Engineering-Recent Advances (CERA). 2023; 1–6.
20. Prayas Energy. Electricity load patterns. [Online]. energy.prayaspune.org/our-work/article-and-blog/electricity-load-patterns. Accessed: 2024.
21. Jing W. Outlier detection in software system based on machine learning. In 2020 IEEE International Conference on Robots & Intelligent System (ICRIS). 2020; 319–322.
22. Nesa N, Ghosh T, Banerjee I. Non-parametric sequence-based learning approach for outlier detection in iot. Future Gener Comput Syst. 2018; 82: 412–421.
23. Xie M, *et al*. Scalable hypergrid K-NN-based online anomaly detection in wireless sensor networks. IEEE Trans Parallel Distrib Syst. 2012; 24(8): 1661–1670.
24. Mehata S, Linus L, Vinayakvitthal L. Real time data plotting tool using open source platform like raspberry pi and python. In 2019 IEEE Global Conference for Advancement in Technology (GCAT). 2019.

