International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Anomaly Detection in Smart Grid Application

Anuradha V[1], Harshitha R[2], Likith Kumar J S[3], Bharath A[4], Jaya Kumar S K[5]

[1]Asst Professor, Dept. Of CSE, Sambhram Institute of Technology, Bengaluru-560097
[2,3,4,5]Eight Semester, Dept. Of CSE, Sambhram Institute of Technology, Bengaluru-560097, India

---

ABSTRACT

The smart grid, a modernized electrical grid integrating advanced communication and information technologies, enables bidirectional flow of electricity and information, offering numerous benefits such as enhanced reliability, efficiency, and sustainability. However, the complexity and scale of the smart grid introduce new challenges, including the need for effective anomaly detection to ensure its secure and reliable operation. This research paper presents a comprehensive review of state-of-the-art anomaly detection techniques in smart grid applications. We explore various methodologies, including statistical methods, machine learning algorithms, and hybrid approaches, highlighting their strengths and limitations. Furthermore, we discuss the importance of accurate data preprocessing and feature selection to enhance anomaly detection performance. Through a comparative analysis, we demonstrate the effectiveness of different techniques in detecting anomalies in smart grid datasets. Our findings provide valuable insights for researchers and practitioners in developing robust anomaly detection systems for smart grid applications, ultimately contributing to the advancement of smart grid technology.

1. INTRODUCTION

By the help of this ML we can easily predict the concepts of digital transformation, digitalization, and Industry 4.0 are central to leveraging data and technology for enhancing productivity and efficiency. The interconnectedness of data across middleware components and sensors in application systems has led to an abundance of accumulated data. However, the challenge lies in effectively managing this massive volume of data and extracting actionable insights. Preprocessing plays a crucial role in reducing costs and increasing capacity by extracting only the most relevant information. The latest trends in data analytics and machine learning are instrumental in achieving these goals, with anomaly detection standing out as a key application of machine learning [1]. An anomaly, in the context of time series data, represents a deviation from the typical pattern. For instance, consider a scenario where traffic volume is monitored daily. While the first two days may exhibit typical traffic patterns, the third day might show abnormally low traffic volume, indicating an anomaly. Detecting such anomalies in real-time can lead to significant savings in resources and time. Similarly, in the case of machinery or equipment, detecting anomalies in performance metrics can help prevent breakdowns by enabling timely maintenance or replacement. Anomaly detection involves training machine learning models on historical data to identify and categorize distinct patterns. These models can then be used to anticipate future anomalies and take proactive measures. The anomaly score, which quantifies the deviation from the expected pattern, is crucial in defining thresholds for anomaly detection. Various factors, including signal spikes, sudden decreases, or continuous aberrations, can indicate anomalies. It is essential to consider the situational context, analyst's interpretation, and domain expertise when identifying anomalies. Real-time anomaly detection is particularly challenging when dealing with large datasets. However, data visualization techniques, such as plotting data points in two dimensions, can help identify outliers. This paper discusses the importance of anomaly detection in various settings, such as bank fraud, medical issues, and equipment failure. By systematically organizing anomaly detection methods, this paper aims to provide insights into effectively managing anomalies in diverse applications.

2. LITERATURE SURVEY

In 2018, Kabir, Hu, Wang, and Zhuo introduced a novel statistical technique for intrusion detection systems (IDS), termed Optimum Allocation-based Least Square Support Vector Machine (OA-LS-SVM). Their approach was evaluated using the KDD 99 database, a widely accepted benchmark for IDS performance assessment. The proposed method demonstrated realistic performance in terms of accuracy and efficiency, particularly excelling in testing all binary classes. However, a limitation of their approach is its applicability only to static datasets, highlighting a need for further development in dynamic intrusion detection solutions. In 2023, Syariful Ikhwan, Purwanto, and Adian Fatchur Rochim conducted a comparative analysis of intrusion detection in IoT networks using deep learning techniques. Their study employed DNN, CNN, LSTM, and AE algorithms. The results indicated that the DNN algorithm achieved an impressive accuracy of 99.76%. However, it was noted that while the DNN algorithm performed well in detecting anomalies, it was less effective in identifying specific types of attacks. This limitation highlights the need for further research to develop more comprehensive intrusion detection systems for IoT networks. In 2023, Madhuri Telidevara and D. Kothandaraman introduced a novel approach to enhancing intrusion detection in Internet of Things (IoT) networks using feed-forward neural networks. Their study focused on leveraging the UNSW-NB15 dataset for training and evaluation. The proposed technique demonstrated significant advantages, providing a reliable and efficient method for detecting malicious behaviors within IoT networks. However, similar to other studies, the approach primarily focused on detecting anomalies without effectively distinguishing between different types of attacks. While achieving good accuracy in anomaly detection is valuable, the inability to differentiate between various attack types limits the overall effectiveness of the intrusion detection system. This highlights the ongoing challenge in developing intrusion detection systems that not only detect anomalies but also accurately classify and respond to specific types of attacks. Future research efforts should aim to address this limitation and further enhance the capabilities of intrusion detection systems for IoT networks. In 2021, Amer Zaheer, Muhammad Zeeshan Asghar, and Amir Qayyum introduced an innovative framework for intrusion detection and mitigation in Software-

Defined Networking (SDN) controlled IoT networks. Their framework encompasses signature-based, anomaly-based, and machine learning-based methods for intrusion detection. The primary focus of their work is to provide a comprehensive framework for controlling intrusions within IoT networks. While the framework demonstrates good accuracy in detecting anomalies, it falls short in effectively identifying and mitigating different types of attacks. This limitation underscores the need for further research and development to enhance the framework's capability to detect and respond to various attack vectors. Despite this drawback, the framework represents a significant advancement in the field of IoT security, providing a solid foundation for future research in improving the security posture of IoT networks.

3. PROBLEM STATEMENT

Equipment such as pumps, gears, valves, and heat exchangers, whether spinning or non-rotating, consist of various internal components like gearings and sensors. Over time, these machines inevitably reach a point where they do not perform optimally, signalling the onset of potential failure. If such a situation arises unexpectedly, it could lead to prolonged downtime and costly repairs. Hence, it becomes imperative to detect these issues beforehand, allowing for timely maintenance and preventing equipment failure. This proactive approach is known as "condition monitoring," which essentially involves assessing the "health status" of the equipment. Traditionally, condition monitoring has been performed by setting maximum and minimum thresholds for sensor measurements. If a sensor reading falls outside these limits, an alarm is triggered, indicating a potential issue. However, this method is prone to generating false alarms and missing critical alerts. For instance, if one sensor exceeds its maximum value while another falls below its minimum, the equipment may still be in good condition, but an alarm is triggered nonetheless. This inefficiency can lead to wasted resources and, in some cases, actual equipment damage. To address these shortcomings, a more holistic approach is required. Instead of analysing each sensor's data independently, combining multiple sensor data points can provide a more accurate picture of the equipment's overall health. This approach involves the use of supervised machine learning algorithms like KNN, LOF, and SVM for anomaly detection. By leveraging the latest trends in data analytics and machine learning, such as preprocessing techniques, organizations can reduce costs, increase efficiency, and improve equipment reliability. In conclusion, effective condition monitoring relies on the integration of sensor data and advanced analytics to detect anomalies and predict equipment failures before they occur. This proactive approach not only minimizes downtime and maintenance costs but also enhances overall operational efficiency and equipment performance.

4. SYSTEM DESIGN

The system for anomaly detection in smart grid equipment using machine learning applications comprises several key components. First, data collection mechanisms are employed to gather real-time operational data from smart grid devices such as sensors, meters, and other monitoring equipment. This data is then pre-processed to remove noise, handle missing values, and normalize the features. Next, a feature selection process is applied to identify the most relevant features for anomaly detection. Machine learning models, such as decision trees, random forests, or neural networks, are trained on labelled data to classify normal and anomalous behavior. These models are then deployed in a real-time monitoring system that continuously evaluates incoming data streams, triggering alerts or actions when anomalies are detected. Finally, a feedback loop is established to update the models periodically with new labelled data, ensuring their continued effectiveness in detecting anomalies in smart grid equipment.

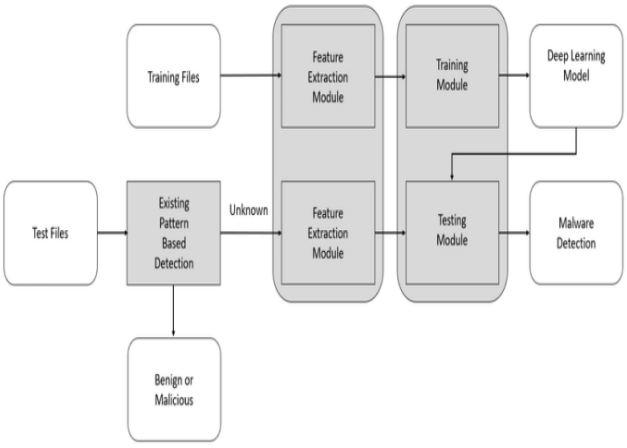


Fig. 1. System design

5. IMPLEMENTATION

5.1 DATASET

- protocol_type
- service
- src_bytes
- dst_bytes
- logged_in
- count
- srv_count
- srv_diff_host_rate
- dst_host_count
- dst_host_srv_count
- dst_host_same_srv_rate
- dst_host_diff_srv_rate
- dst_host_same_src_port_rate
- dst_host_srv_diff_host_rate
- srv_diff_host_rate
- intrusion

5.2 Data Preprocessing

One of the most crucial milestones in the process of putting deep learning models into practice is this one. To make the dataset more appealing and practical for the model training phase, we deliberately applied all data cleaning strategies to our dataset. We removed all the unnecessary and irrelevant data from our dataset throughout the data cleaning process.

- Data cleaning had the following goals in mind-
- Removal of missing data
- Removal of duplicate entries
- Remove rows with Na values

5.3 Exploratory Data Analysis

Exploratory data analysis is a strategy for examining datasets to highlight their key properties, frequently utilizing statistical tools and other techniques for data visualization. It aids we better comprehend our dataset. Executing EDA on our dataset assisted us in:

- Recognize and handle NULL values.
- Recognize and eliminate outliers.
- Identify the underlying relationships and structure.

Additionally, we created a word cloud and several graphs. to learn more about the data.

6. TRAINING

After preprocessing and EDA, we had the final dataset that had been thoroughly cleaned and analyzed. The 80-20 train-test validation method was used, which specifies that 80% of the information is used for planning and 20% is used for testing. The sklearn library is used to divide the data into training and testing portions. Out of 303 samples, 242 examples or instances are chosen and used to create the model . The remaining 61 samples are used as testing data to judge how well the constructed model performs. To implement the model, we proceeded forward. We carried out two distinct sorts of experiments during the implementation. Both the considered deep neural network and the artificial neural network were implemented individually. As a result, we were able to determine how accurately each of these models performed. A Deep neural network has multiple hidden layers. Whereas the Artificial Neural network has one or two hidden layers in it. The activation of neurons is present at the output layer.

$$f(x) = \frac{1}{1+e^{-x}}$$

In the output layer, the sigmoid activation function is applied. The dataset's redundant features are removed using feature selection. Feature extraction and feature selection are different. Finding relevant components from the existing data is called feature extraction. By removing unnecessary features through feature selection, the neural network is fed with pertinent information.

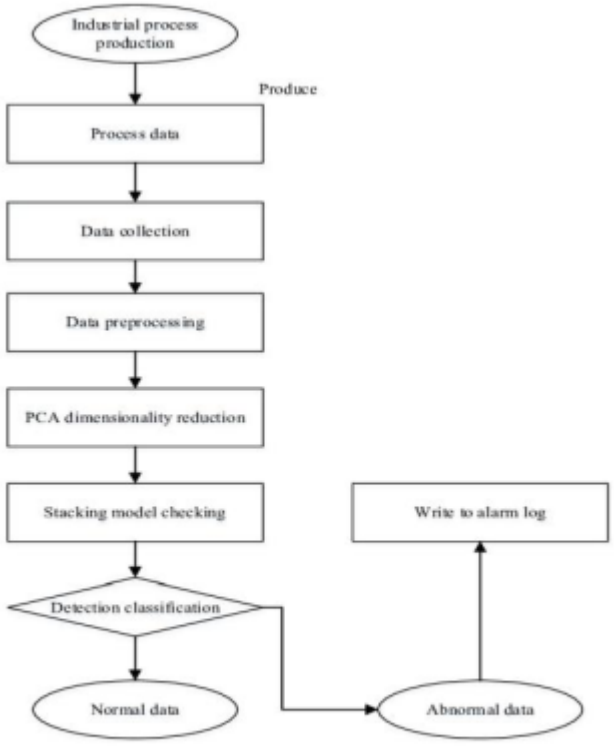
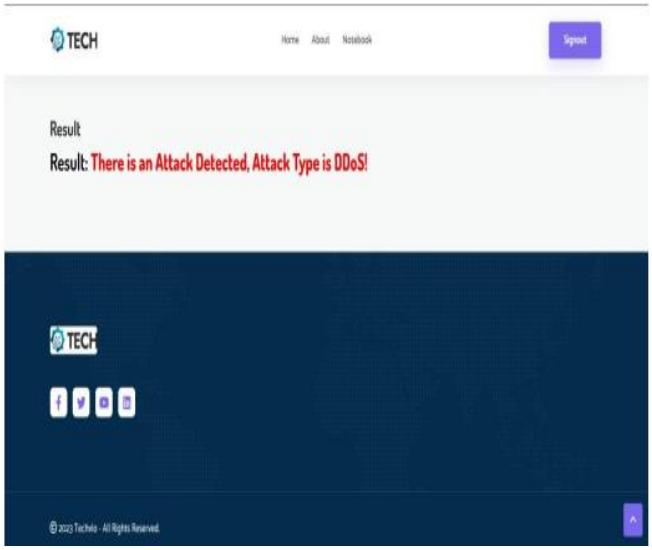
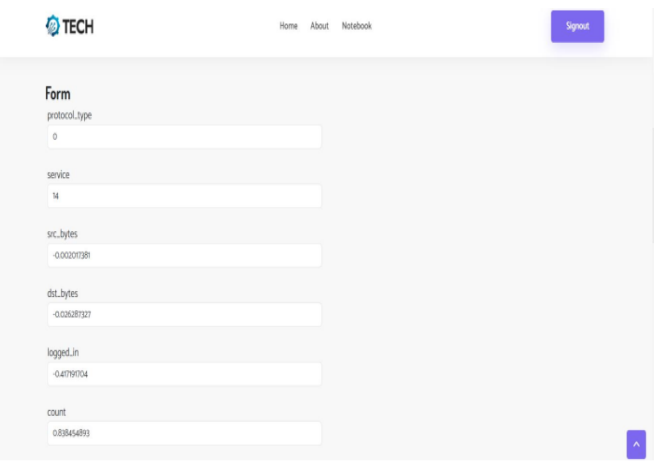


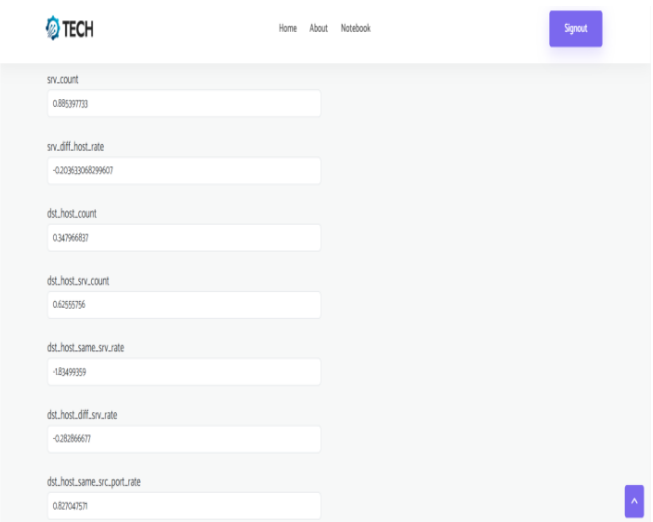
Fig 2 System Architecture

6. RESULTS

The outcomes that we found after testing and putting the suggested algorithms into practice will be covered in this part. The Deep neural network with the more hidden layer performs better than that of Artificial neural network. Accuracy of artificial neural network is 89% and accuracy of deep neural network is 98.5%. So the KNN is more accurate than that of ANN.







7. CONCLUSION AND FUTURE SCOPE

The decreasing cost of sensor data collection and the increasing interconnectedness of devices underscore the importance of extracting actionable insights from data. Machine learning and statistics play a pivotal role in uncovering patterns within vast datasets. By employing two distinct approaches, namely PCA + Mob distance and autoencoders, we successfully preprocess data and apply machine learning to detect anomalies in condition monitoring. These methods have demonstrated the ability to identify anomalies three days prior to actual failures. Looking ahead, the evolution of machine learning holds great promise, potentially enabling even faster anomaly detection. These advancements will be instrumental in predicting future breakdowns, enhancing overall equipment maintenance and operational efficiency.

8. REFERENCES

[1] Karishma Pawar and Vahida Attar, “Deep learning approaches for video-based anomalous activity detection”, World Wide Web, vol. 22.2, pp. 571–601, 2019.
[2] Yuequan Bao et al., “Computer vision and deep learning-based data anomaly detection method for structural health monitoring”, Structural Health Monitoring, vol. 18.2, pp. 401–421, 2019.
[3] Iqbal, A., Rajasekaran, A. S., Nikhil, G. S., and Azees, M. A Secure and Decentralized Blockchain Based EV Energy Trading Model Using Smart Contract in V2G Network. IEEE Access, 9, 75761–75777, 2021.
[4] Rajasekaran, A. S., Azees, M., and Al-Turjman, F. A comprehensive survey on security issues in vehicle-to-grid networks. Journal of Control and Decision, 1–10, 2022.
[5] Leman Akoglu, Hanghang Tong and Danai Koutra, “Graph based anomaly detection and description: a survey”, Data mining and knowledge discovery, vol. 29.3, pp. 626–688, 2015.
[6] Arasan, A., Sadaiyandi, R., Al-Turjman, F., Rajasekaran, A. S., and Selvi Karuppuswamy, K. Computationally efficient and Secure Anonymous Authentication Scheme for Cloud Users. Personal and Ubiquitous Computing, Vol. 25, 2021