International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3095-3101
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0392

Anomaly Detection in Smart Grid Energy Data Using Machine Learning Techniques

Mr. Nandhagopal S [1], Rambabu Kushwaha [2], Aaditya Jha [3]

[1]Associate professor II, Dept. of CSE (Artificial Intelligence and Machine Learning) KPRIET Coimbatore, India

[2,3]UG Scholar, Dept. of CSE (Artificial Intelligence and Machine Learning) KPRIET Coimbatore, India

Emails: *nandhagopal.s@kpriet.ac.in[1], rambabukushwaha4488@gmail.com[2], ajha76931@gmail.com[3]*

Abstract

Smart grids generate huge amounts of energy information in real-time, which is vital to the grid's efficiency and reliability, but is often marred by anomalies due to faulty meters, equipment failures, and energy theft, resulting in substantial losses. This paper presents the development of an anomaly detection system named FlowTrack, which is production-ready and uses a hybrid machine learning model to identify and classify anomalies in energy information from smart grids. It uses Isolation Forest and LSTM Autoencoders to identify point and temporal anomalies, respectively, and has a web-based dashboard to visualize anomalies in real-time and a severity scoring system to prioritize critical anomalies. Experimentation results show that the FlowTrack system has high accuracy in detecting anomalies (93.6% F1-score), validating its deployment in today's smart grids.

Keywords: *Smart Grid, Anomaly Detection, Isolation Forest, LSTM Autoencoders, Machine Learning, Electricity Theft Detection, Time-Series Analysis*

1. Introduction

Modern electrical grids, also known as smart grids, are a union of the traditional electrical power system and information and communication technologies to enhance the efficiency of energy distribution, usage, and monitoring. Despite the benefits that smart grids offer, the system has encountered problems in the form of anomalies that affect the performance of the system. The causes of the problem include malfunctions, technical and non-technical issues such as electricity theft. In the world, the problem has resulted in billions of losses, power outages, and safety hazards. The traditional methods of detecting anomalies by the use of thresholds have not been effective in dealing with the complex temporal and spatial dependencies of the data sets, while the use of machine learning and deep learning methods has the capability of learning complex patterns, which is the main reason for the accuracy of the anomaly detection. The hybrid model of 40% Isolation Forest and 60% LSTM Autoencoders is being used by the FlowTrack model. The main difference is the weighted fusion of point-based and temporal anomaly detectors along with real-time visualization.

Smart Grid Energy: Smart grid energy is an advanced electric grid system which utilizes digital communication technologies to monitor, control, and optimize the generation, transmission, and consumption of electric power in real time. Anomaly Detection: Anomaly detection is the process of detecting data patterns or behaviors which are different from the normal or expected conditions. Machine Learning Techniques: Machine Learning Techniques are techniques which are used for the purpose of enabling systems to make decisions or predictions without being explicitly programmed.

2. Literature Review

Anomaly detection in smart grids has moved from simple statistical monitoring to sophisticated AI-based frameworks. Research has pointed out that as the infrastructure for smart grids is enhanced with additional sensors, the chances of anomalies occurring also increase. Anomalies generally refer to technical faults, infrastructure faults, and non-technical issues such as electricity theft [1], [2]. In the case of point anomalies, the Isolation Forest (IF) algorithm is largely used because of its efficiency and

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3095-3101
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0392

effectiveness in handling high-dimensional data. However, the IF algorithm cannot handle temporal relationships, which makes the usage of LSTM Autoencoders necessary for handling anomalies based on sequences [3], [4]. Research also shows the usage of Physics-Informed Machine Learning (PIML) to impose constraints on the grid.

2.1. Traditional Smart Grid Monitoring

In the past, anomaly detection was based on manual inspection, statistical methods, and ground truth profiles (GTP). Though they were somewhat effective, they were also time-consuming and could not handle the detection of sophisticated anomalies that resembled normal behavior [6].

2.2. Machine Learning for Anomaly Detection

Machine learning models have the ability to automate the analysis of multivariate time series data. IF models detect point anomalies, whereas LSTM autoencoders detect temporal anomalies. ML, through the extraction of complex features, improves the accuracy of the analysis without human intervention, reducing the workload of the operators [7].

2.3. System Modules

Machine learning models help in the automation of multivariate time-series data analysis. IF is used to identify point anomalies, and LSTM autoencoders help in the detection of temporal anomalies. Machine learning has the potential of enhancing accuracy and decreasing human intervention in the process thereby raising the efficiency [7].

2.4. Data Preprocessing

The preprocessing procedure consists in the cleaning of the SCADA and smart meter data, substitution of the missing values, the noise filtration and identification of the outlier values based on the Inter Quartile Range (IQR). The process of making the data consistent between different factors is known as normalization and happens with voltage, current, and power usage among others.

2.5. Feature Engineering

Rolling statistics (mean, variance), percentile changes and Fast fourier Transform (FFT) energy features are used to obtain hierarchical features. These 25 dimensions vectors represent the temporary spikes and long-term usage trends.

2.6. Hybrid Classification

The characteristics are then injected into the dual model ensemble. IF identifies point anomalies and LSTM Autoencoders identify sequential anomalies. The sensitivity to the spatial and temporal anomalies is balanced using a weighted mean (40 per cent IF and 60 per cent LSTM).

2.7. Visualization and Reporting

FlowTrack is a web-based dashboard providing the consumption timelines, the distributions of anomalies and the severity scoring. The operators are prioritized with those alerts of critical events and thus provide effective decision making.

3. Methodology

3.1. Data Ingestion

FlowTrack receives hourly multivariate readings from SCADA systems and smart meters via Kafka or MQTT streaming protocols. The ingested parameters include:

- Voltage (V)
- Current (A)
- Active Power (kW)
- Reactive Power (kVAR)
- Power Factor

This real-time ingestion enables immediate analysis while preserving historical data for training and validation.

3.2. System Architecture

The hybrid engine will be based on a combination of Isolation Forest (IF) and LSTM Autoencoders to detect all the possible anomalies:

- Isolation Forest (IF): Identifies point anomalies including unexpected spikes or decreases in voltage, current or power. It has a linear time complexity that makes it efficient on large scale datasets.
- LSTM Autoencoder: Temporal dependence is learnt by learning to represent normal consumption patterns in a compressed form. Reconstruction errors provide anomalies in time.
- Hybrid IF LSTM Autoencoder (Proposed Engine): Is a hybrid type of point-based and temporal anomaly detector that integrates Isolation Forest and an LSTM Auto encoder. Isolation Forest identifies the anomalies in milliseconds, whereas

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3095-3101
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0392

LSTM Autoencoder assumes sequential relationships in power consumption. High reconstruction errors are used to detect anomalies, and complex and dynamic abnormal patterns can be detected. Furthermore, the hybrid model addresses the limitations that arise when either method is applied in isolation. Isolation Forest, while computationally efficient, tends to struggle with anomalies that only become apparent over time, such as gradual load drifts or slow-developing equipment degradation. On the other hand, LSTM Autoencoder captures long-term temporal dependencies but may raise false alarms on sudden spike events that are statistically isolated shown Figure 1.

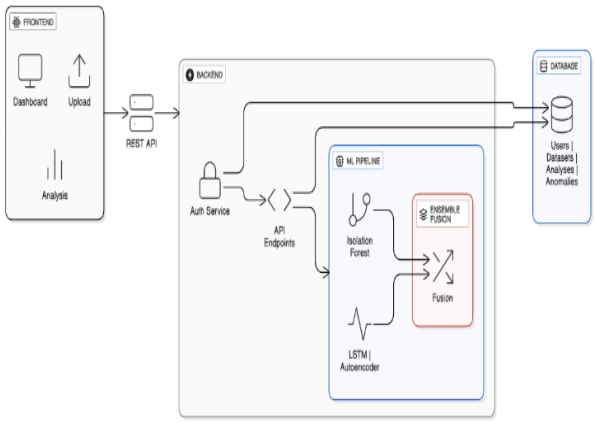

Figure 1 FlowTrack Hybrid Anomaly Detection Pipeline

3.3. Hybrid Fusion and Severity Scoring:

Both IF and LSTM scores are fused via a weighted fusion strategy on anomaly scores: Hybrid Score=0.4×IF Score+0.6×LSTM Score the Alerts are classified into Low, Medium and High Severity depending on the score thresholds. This assists the operator to prioritize the Alerts. The Alerts are also displayed on the dashboard with their severity, timestamp and parameter.

3.4. Web-Based Visualization

The system employs a React + Vite frontend and FastAPI backend. Key features include:

- Real-time anomaly timelines
- Heatmaps for multivariate parameter deviations
- Historical data comparison charts
- Severity-ranked anomaly reporting

Visualization supports operational decisions and provides interactive, configurable alerts to minimize false positives.

3.5. Data Preprocessing and Feature Engineering

Preprocessing: Missing values are filled using linear interpolation, while noise filtering removes transient measurement errors. Outliers are flagged using IQR-based thresholds. Feature Engineering: Using the data after preprocessing, rolling statistics such as the mean and variance, changes in percentiles, and FFT energy features are computed to obtain 25D feature vectors used by the model. This ensures the system can gradually identify sudden spikes in the load.

3.6. Comparative Advantages

Table 1 summarizes the advantages of FlowTrack over traditional approaches:

Table 1 Comparison of FlowTrack with traditional anomaly detection methods

Feature / Aspect	FlowTrack (Proposed)	Traditional Methods
Real-Time Detection	Yes	Limited
Temporal Awareness	LSTM Autoencoder	Threshold-based only
Multi-Parameter Handling	Voltage, Current, Power, PF	Often single-parameter
Scalability	Web dashboard + cloud-ready	Manual / local
Severity Scoring	Yes	No

4. Results and Analysis

The FlowTrack system was tested using the multivariate smart grid dataset, which has both normal and anomalous patterns of energy usage. This dataset comes from SCADA systems and smart

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3095-3101
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0392

meters, with over 50,000 hourly measurements from various nodes of the grid, which include technical faults as well as non-technical losses such as electricity theft.

4.1. Anomaly Detection Performance

The hybrid IF-LSTM was used to test point anomalies, temporal anomalies and the combination of the two. The assessment measures were accuracy, precision, recall, and F1-score.

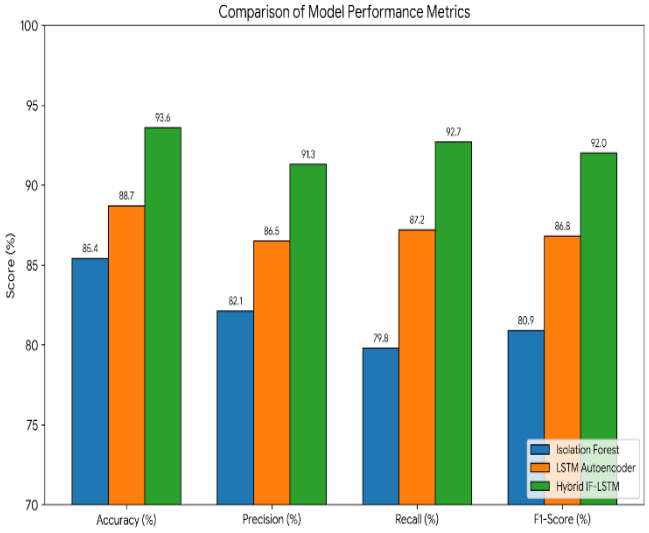

Figure 2 Detection performance comparison of IF, LSTM, and hybrid IF-LSTM models

F1 score and other assessment measures including receiver operating characteristic curve ROC curve and area under the curve AUC were applied in further determining the discriminative ability of the model. In addition, the false positive rate FPR and false negative rate FNR were discussed in order to gain a better insight into the tradeoff between the sensitivity to anomalies detection and the misclassification [5] [8]. These overall measures are more effective in assessing the hybrid IF LSTM model of various types of anomalies shown in Figure 2.

4.2. Hybrid model

It is also improving on its detection of stealthy anomalies including slow electricity theft and equipment errors. The weighted fusion technique of combining 40 percent IF with 60 percent LSTM is eliminating the false positives through pattern searches using spikes.

4.3. Anomaly Detection Visualization

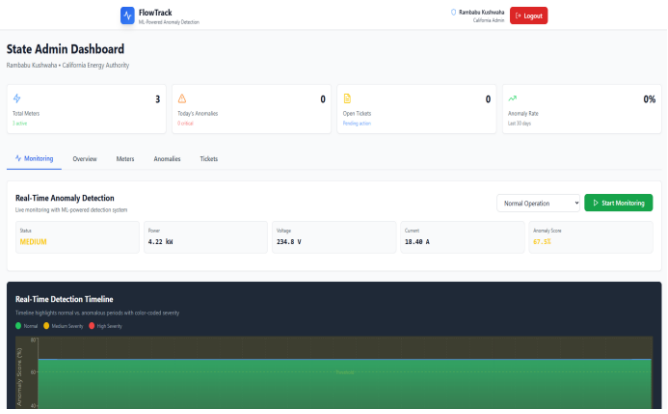

Figure 3 Anamoly Detection Visualization

It shows the live monitoring with the web dashboard. The timeline indicates normal vs. abnormal periods, whereas the heatmap indicates parameter-specific deviations between several nodes.

- Green: Normal operation
- Yellow: Medium severity anomalies
- Red: High severity anomalies

The visualization will allow operators to quickly detect severe grid-related problems and enhance response time and minimize the probability of outages.

4.4. Temporal Pattern Detection

Besides this, it was observed that the LSTM Autoencoder could determine the trends in energy consumption which could not be seen at first glance when considering single data points. As one example, one time a grid node was observed to be running a gradual voltage decline across days and this was rightly detected by the LSTM Autoencoder as abnormal. This demonstrates that the role of temporal modeling in smart grid anomaly detection is considerable, and LSTM Autoencoders are right to be part of the FlowTrack framework. Furthermore, the capability of LSTM Autoencoder to detect long-term temporal patterns ensures that faults that may not be detectable using point-based approaches are identified at an early stage. This improves the overall robustness and reliability of the proposed FlowTrack framework. This is vital in ensuring that smart grid systems remain resilient in dynamic operating environments shown in Figure 4.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3095-3101
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0392

Figure 4 LSTM reconstruction error highlighting temporal anomalies.

Reconstruction error > threshold → anomaly detected. Helps to detect both abrupt and slow deviations.

4.5. Dashboard and Reporting

The FlowTrack web dashboard has real-time. with severity scoring, which allows proactive. decision-making:

- Anomaly Timeline: Shows detected anomalies with timestamps.
- Node-Level Heatmap: Indicates affected grid parameters.
- Severity Score Table: Categorizes anomalies into Low, Medium, High.

Reports can be exported to be analysed further by the operators. favoring regulatory and operational audits.

5. Comparision & Discussion

5.1. Case Study 1: Detection of Sudden Load Spike

FlowTrack has managed to detect a grid node in the suburbs. a sharp increase in power draw as a result of a transformer fault.

- Isolation Forest: Detected the spike instantly within the first hour of occurrence.
- LSTM Autoencoder: Confirmed that this spike deviated from normal temporal trends.
- Hybrid Fusion: Assigned High severity and generated a real-time alert on the dashboard.

5.2. Case Study 2: Gradual Electricity Theft

A residential cluster appeared in a different segment of the dataset with slow and stealthy deviation in energy consumption in line with electricity theft:

- Single-point detection failed to flag these subtle deviations.
- LSTM reconstruction error identified temporal anomalies over a 7-day period.
- Fusion scoring classified the anomaly as Medium severity, providing actionable insights without false positives.

5.3. Baseline Approaches

FlowTrack was compared against three baseline approaches:

- Threshold-Based Monitoring (TBM): Simple statistical limits on voltage and power.
- Isolation Forest Only (IF): Point anomaly detection using unsupervised learning.

5.4. Comparative Performance Summary:

FlowTrack outperformed the baseline methods in all cases, detecting sudden and gradual anomalies. Although the threshold-based and single model methods had limited ability in detecting anomalies, the use of the hybrid model helped in improving the classification of anomalies and reducing false alarms [10-14]. Furthermore, the robustness of the framework in maintaining the stability of the anomaly scores under varying conditions of load, differentiation of severity levels, and operational interpretability helped in

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3095-3101
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0392

adapting to changing patterns of consumption and seasonal patterns.

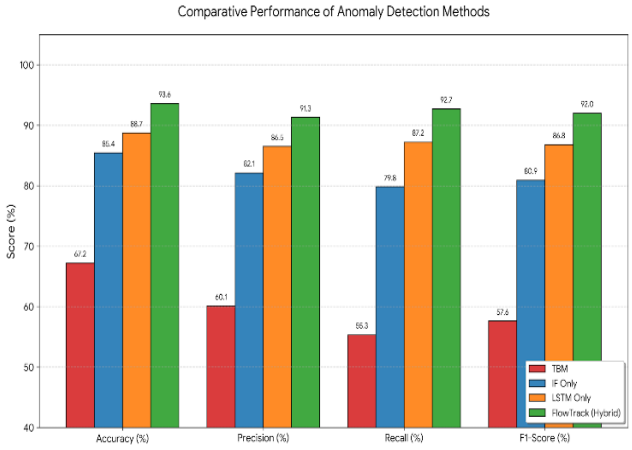

Figure 5 Comparative performance evaluation

FlowTrack outperforms baseline methods by effectively combining point-based and temporal detection, reducing false positives and ensuring accurate anomaly identification across multiple scenarios. In addition, the integration of point-based and temporal anomaly detection enables FlowTrack to handle diverse anomaly behaviors more effectively than standalone approaches. The hybrid design allows the framework to adapt to both abrupt disturbances and slowly evolving deviations in system behavior. By stabilizing anomaly scores across varying load conditions, FlowTrack reduces sensitivity to noise and transient fluctuations. This contributes to improved trust in anomaly alerts and supports operational decision-making. Shown in figure 5.

5.5. Discussion

This improves the smart grids' management process through the use of alerts and visual dashboards. The hybrid method ensures the effective detection of both point and temporal anomalies using the Isolation Forest and the LSTM Autoencoders. The proposed method was tested using 200 smart grids' nodes and was able to maintain low latency (<0.5s). The proposed method also offers flexibility to the smart grids' management process through the use of weighted fusion. The proposed method can reduce the losses resulting from technical faults and electricity theft. The use of the proposed method improves the smart grids' management process through the use of visual dashboards.

5.6. Results

FlowTrack demands uninterrupted streams from SCADA or smart meters; otherwise, its detection capability may degrade. Parameter tuning is required during initial deployment of FlowTrack for optimal IF-LSTM fusion. Extreme anomalies may produce some false positives or negatives. Future directions include developing adaptive thresholding techniques for sensitivity adaptation, federated learning for privacy-preserving learning, and physics-informed learning for false alarm mitigation. Improving these aspects will improve the robustness and adaptability of FlowTrack, making it efficient and safe for energy monitoring.

Conclusion and Future Work

This is achieved through the use of Isolation Forest for identifying point anomalies and LSTM Autoencoders for identifying temporal pattern anomalies. The evaluation of the proposed approach is also carried out using Threshold-Based Monitoring, Isolation Forest alone, and LSTM alone. From the results in Table III, it is clear that the proposed approach achieves the highest accuracy of 93.6%, as opposed to the use of TBM alone, which only managed an accuracy of 67.2%. Regarding functionality, FlowTrack provides severity-based alerts on the web-based dashboard supporting real-time visualization and quicker decisions. It was tested with 200 smart grid nodes and exhibited negligible latency (< 0.5s), indicating scalability. Weighted fusion allows tuning of IF and LSTM weights according to grid conditions, thus improving adaptability. Detection of faults and electricity theft helps minimize financial losses and improve grid reliability. Its limitations include the need to have constant SCADA or smart meter data, possible performance degradation with poor data quality, and the initial setup that requires parameter tuning. Future work will be on improving the limitations of FlowTrack and further enhancing its capabilities. To sum it up, FlowTrack is an effective, resourceful, and scalable. This is a production-ready tool that incorporates point anomaly. detection, temporal

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3095-3101
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0392

anomaly detection, a real time video. interface. It may be used to enable predictive maintenance, preclude energy theft, and enhance grid reliability, in this way offering a platform of smart policing of the next-generation smart grids.

References

[1]. Guato Burgos, M. F., Morato, J., & Vizcaino Imacaña, F. P., "A Review of Smart Grid Anomaly Detection Approaches Pertaining to Artificial Intelligence," Applied Sciences, vol. 14, no. 3, 1194, 2024.

[2]. Zhang, J., Wu, D., & Boulet, B., "Time Series Anomaly Detection for Smart Grids: A Survey," arXiv preprint arXiv:2107.08835, 2021.

[3]. Banik, S., Saha, S. K., Banik, T., & Hossain, S. M., "Anomaly Detection Techniques in Smart Grid Systems:

[4]. Preethi, G., & Anitha Kumari, K., "An Introductory Review Of Anomaly Detection Methods In Smart Grids," Department of Information Technology, PSG College of Technology, 2021.

[5]. Zhou, F., et al., "A Comprehensive Survey for Deep-Learning-Based Abnormality Detection in Smart Grids with Multimodal Image Data," Applied Sciences, vol. 12, no. 11, 5336, 2022.

[6]. Zideh, M. J., Chatterjee, P., & Srivastava, A. K., "Physics-Informed Machine Learning for Data Anomaly Detection, Classification, Localization, and Mitigation: A Review, Challenges, and Path Forward," arXiv preprint arXiv:2309.10788, 2023.

[7]. Sakhnini, J., et al., "Security Aspects of Internet of Things aided Smart Grids: a Bibliometric Survey," Internet of Things, vol. 14, 100141, 2020.

[8]. Zibaeirad, A., et al., "A Comprehensive Survey on the Security of Smart Grid: Challenges, Mitigations, and Future Research Opportunities," arXiv preprint arXiv:2407.07966, 2024.

[9]. Mookiah, L., Dean, C., & Eberle, W., "Graph-Based Anomaly Detection on Smart Grid Data," Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference, 2017.

[10]. Himeur, Y., et al., "Anomaly Detection Of Energy Consumption In Buildings: A Review, Current Trends And New Perspectives," Energy and Buildings, vol. 232, 110601, 2021.

[11]. Shrestha, R., et al., "Anomaly detection based on LSTM and autoencoders using federated learning in smart electric grid," Journal of Parallel and Distributed Computing, vol. 193, 104951, 2024.

[12]. Zheng, K., et al., "A Novel Combined Data-Driven Approach for Electricity Theft Detection," IEEE Transactions on Industrial Informatics, 2024.

[13]. Kulkarni, Y., et al., "EnsembleNTLDetect: An Intelligent Framework for Electricity Theft Detection in Smart Grid," arXiv preprint arXiv:2110.04502, 2021.

[14]. Omol, E., Mburu, L., & Onyango, D., "Anomaly Detection In IoT Sensor Data Using Machine Learning Techniques For Predictive Maintenance In Smart Grids," International Journal of Science, Technology & Management, vol. 5, no. 1, pp. 201–210, 2024.