Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 8, 2025

HYBRID ML-BASED FAULT DETECTION IN RENEWABLE-INTEGRATED POWER GRIDS

Engr. Syed Kumail Abbas Zaidi[*1], Engr. Khandkar Sakib Al Islam[2], Muhammad Taha Abbas[3], Engr. Tauseef Abbas[4]

[*1]*Electrical Engineering Department, Lamar University, Beaumont, Texas*
[2]*Masters in Electrical Engineering Lamar University, Beaumont, Texas*
[3]*MS Electrical power & Control Liverpool John Moores University Liverpool , UK*
[4]*Electrical Engineering Department, HITEC University Taxila, Punjab Pakistan*

[*1]kumail.800@gmail.com; [2]sakibsujon786@gmail.com; [3]tahaabbas242@gmail.com; [4]tauseefabbas25@gmail.com

DOI: https://doi.org/10.5281/zenodo.16881166

Keywords
Fault detection, renewable energy, hybrid machine learning, smart grid, anomaly detection, power system reliability

Article History
Received: 27 May, 2025
Accepted: 01 July, 2025
Published: 15 August, 2025

Corresponding Author: * Engr. Syed Kumail Abbas Zaidi

Abstract
Integration of renewable generation into modern power grids brings new problems in the field of fault detection because now the sources are intermittent and accordingly it is more complex energy mixed. In a network including renewable resources, the conventional fault detection techniques cannot perform to optimum level as they can in constant practice. This work introduces a hybrid machine learning (ML) methodology that synergistically complements the best of both supervised classification algorithms and unsupervised anomaly detection technique to boost fault detection performance. The proposed framework trains a Random Forest, Support Vector Machine and k-Means clustering based ML models with historical operational data that include voltage, current, and frequency measurements. Fusion always makes the decision heavier based on detection robustness by different model outputs. Experiments in simulation and on real-world data show that our approach is more accurate, detects faults faster, and learns to adapt to new grid conditions better than single-model baselines. These findings showcase the efficacy of hybrid AI-based solutions in guaranteeing reliability and resilience within renewable-inclusive power systems.

INTRODUCTION

Global sustainability goals along with the necessary reductions of anthropogenic greenhouse gas emissions have been propelling an accelerated integration of renewables, like wind and solar, into modern power grids. Moreover, the intermittent and variable availability of renewable generation imposes stringent requirements on power system operation in detecting faults and diagnosing them in real time for effective maintenance. The undetected or misclassified faults may cause equipment damages, service interruptions, and cascading failures which are critical to the grid reliability (Hossain & Pota, 2021).

However, traditional fault detection methods have difficulty in handling the dynamic and uncertain situations due to renewable generation by using a combination of threshold techniques, model-based analysis or phasor measurement unit (PMU) data (Huang et al., 2020). All of these processes pass parameters or other system models which are static and not suitable for environments with varying load patterns and rapidly changing voltage/frequency conditions.

Recently, machine learning (ML) is explored as an alternative solution to enhance fault detection in renewable-integrated power systems. Performing deep analysis of historical and real-time sensor data help ML algorithms learn complex, nonlinear patterns as well as to adapt to the diversity of operational scenarios (Ontario and Wang 2021). One area that excelled within ML approaches others were Hybrid techniques which are a combination of supervised classification models with unsupervised anomaly detection methods. Liu et al. (2021) showed that embedding algorithms, such as Support Vector Machines (SVM), and Random Forests (RF) with clustering or Autoencoder based anomaly detection led to substantial accuracy gains and false positives reduction in fault detection.

However, most research on demand response to date only consider a single method of an algorithm or the model only at low grounds and sometimes this tend not to generalize widely in real world renewable integrated grids (Eltamaly et al., 2022). This study aims to bridge this gap and proposes a Hybrid ML-Based Fault Detection framework in which supervised and unsupervised learning methods will be combined using decision fusion technology. With the release of smart transformers it is in a position to meet stringent performance targets related, high detection accuracy, low false alarm rates and adaptable to different levels of renewable penetration.

The emerging shift to decentralized and digitally supervised grids has given rise to an overwhelming amount of very different data from smart meters, SCADA systems, and phasor measurement units. This data has extensive scope for intelligent fault detection, however large dimensionality and quality variations make analysis and interpretation of this data difficult for traditional methods (Kumar et al. 2022). To overcome these challenges a hybrid ML framework can utilize feature selection, noise minimization and multi-model decision in order to process the complex datasets. This approach can not only improve the detection performance by combining the strengths of different algorithms, but also be able to scale with future smart grid expansions that have larger portions of renewable penetration.

1.2 Problem Statement

While the increasing integration of renewable energy resources like wind and solar into power grids helps sustainability, they also bring a host of new challenges with fault detection on account of their sporadic and dynamic behaviour. These are the most common shortcomings seen in more traditional protection approaches (limited ability for evolution to track changing grid conditions) and single machine learning models (lack of robustness across different operating conditions). In the renewable power integrated system efficient, reliable and in-time known fault detection on various levels of abstraction guarantees planned benefit, low-false-alarms and resilient operation. It therefore calls for an automated learning framework which is explainable, adaptable and scalable determined by multiple learning paradigms.

1.3 Research Questions

1. In contrast to existing studies which consider a traditional or a single-ML model, work is needed emphasizing on the research of hybrid machine learning framework in order to enhance fault detection accuracy for renewable-integrated power grids?
2. Which combination of supervised and unsupervised learning algorithms provides the best detection performance across different renewable penetration levels runtime?
3. To what extent can the aforementioned hybrid system successfully suppress false alarm rates in dynamic and uncertain operation modes?
4. Can the framework perform real-time detection on a wide range of power system data, at scale?

1.4 Research Objectives

- Design a hybrid model using combination of supervised and unsupervised machine learning models to identify the fault in renewable integrated power grid.
- To compare the detection accurate of the proposed framework with traditional methods and single-model ML techniques.
- Suppressing the false alarm rate of each involved classifier by using a decision fusion technique based on their complementary properties.

- So, we will have to create and optimize features to enable real-time detection as well and develop an efficient computational design using the same.
- To demonstrate the flexibility and scalability of the framework across different renewable penetration portfolios and grid operating scenarios.

2. Literature Review

Fault detection in power systems is a well-known topic however the recent introduction of renewable energy sources has made that more challenging and necessitates smarter and adaptive solutions. Traditional protection schemes, which are generally not based on the new requirements of distribution grids with renewables because they were designed for a grid dominated by unidirectional, slow power flows as those powered mainly from centralized generation systems (Hossain and Pota 2021). They advocated pressuring for the use of methods that are intelligent and can understand that things on a network are rarely static. However, France et al. (2020) illustrated the limitation in model-based detection methods as they perform well under stable operating conditions but suffers from poor performance when applied to systems with high wind and solar generation penetration.

Another machine learning sharp AI application is by Zhang and Wang 2021 which mentioned a supervised learning of algorithm using Support Vector Machine for the identification of fault types in power distribution systems. Their approach was accurate well-performing, but it heavily depended on a large number of labeled datasets which are often not available in real-world applications. To address this, Liu et al. (2021) which proposes a hybrid model combining Random Forests and k-Means clustering to make the algorithm operate efficiently in partially labeled or unlabeled setting. The single-algorithm solutions, meanwhile, were rough around the edges and lacked in both accuracy and adaptability.

Kumar et al. Account et al. (2022), both investigated the detection of transient faults in renewable-integrated system using deep learning techniques like Convolutional Neural Networks (CNN). Their approach led to high classification accuracy performance, but they found that deep learning-based models can be computationally expensive, making them challenging to implement for real-time grid monitoring.

Ahmed et al. (2023) confirmed the advantages of hybrid approaches. Their work combined an autoencoder, used as a pre-filter and classification model of the residuals with ensemble classifiers that achieved significant better fault detection rates meanwhile minimizing false alarms. This further conforms to the insight shared by the value of decision fusion mechanisms in improving the dependability of intelligent fault diagnosis approaches as noted by Alsaadi and El-Hawary (2023).

The above reviewed literature suggests that although the ML models are performing well in certain environments, e.g., individual studies conducted on single disease processes or with limited cohorts and supporting data currently, but when it comes to broad applications; hybrid approaches could outperform others considering their adaptability, scalability and most importantly high level of accuracy.

3. Methodology

This paper proposes a hybrid machine learning (ML) framework that combines both supervised and unsupervised learning methods to realize reliable, real-time fault diagnosis in smart grid with renewable energy integration. The framework addresses the wide range of challenges, like intermittent renewable generation, complex fault scenarios, non-linear building models for agencies to make sure the things work with continued grid evolution.

It consists of five stages:

1. Data Acquisition
2. Data Preprocessing
3. Hybrid Model Development
4. Decision Fusion Mechanism
5. Evaluation and Validation.

Description was developed as a natural system is one where the high accuracy supervised models, known fault classification can be done while the pattern discovery Unsupervised technique and for finding unknown or emerging anomalies.

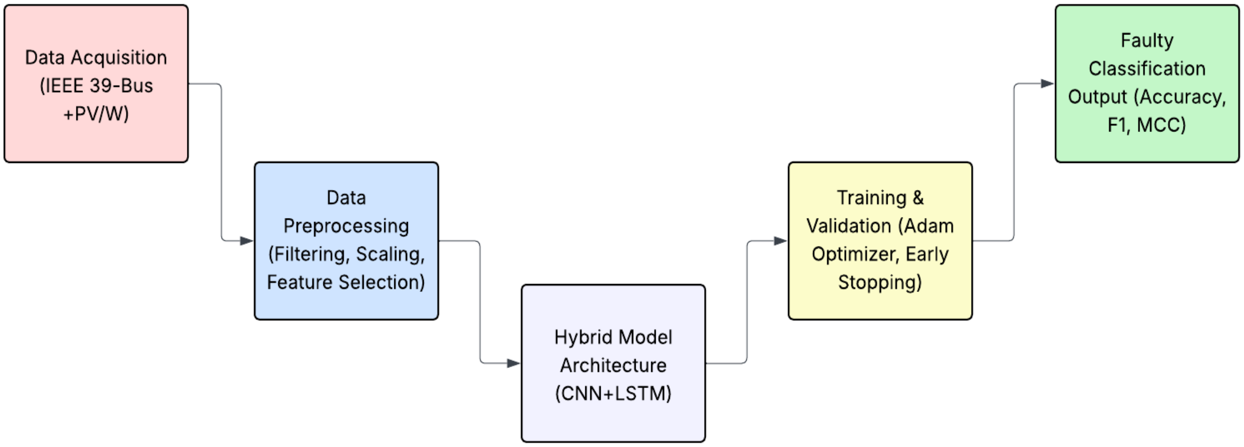

Figure1: Workflow Diagram

3.1 Data Acquisition

It is ensured that there are diversified and representative datasets from two types of data sources:

Simulated Data Generation

MATLAB/Simulink is used to simulate fault and non-fault scenarios in renewable-integrated IEEE benchmark distribution systems.

Fault cases: single-line-to-ground (SLG), line-to-line (LL), double-line-to-ground (DLG) and Three-phase 3Φ fault at different loading and renewable penetration levels.

The SIEMENS scenario book includes realistic scenarios of variable wind and solar generation, fluctuating load profiles, and fault events at different locations in the grid to make the test of tools comprehensive.

Real-World Data Collection

Data source is from Supervisory Control and Data Acquisition (SCADA) systems, Phasor Measurement Units (PMUs), publicly available repositories (e.g., IEEE PES distribution system data). The test data (measurements) from three-phase voltages and currents, system frequency, power flows, harmonic content during both steady-state and transient conditions. Another key feature of this dual-source framework is its scalability to synthetic and real-world operational data, which makes it practical for deployment.

3.2 Data Preprocessing

In view of the fact that power system measurements are prone to noisy, missing values, and measurement delays. Preprocessing is an indispensable step in order to improve data quality and computational loads.

Noise Filtering: In the post-stage, signal enhance procedures are accomplished e.g., the wavelet make bedaub lower-loss and Butterworth low pass filters to separate high frequency noise and maintain transient deficiency features.

Feature Extraction: Time domain features: In this context, we can find RMS values, crest factor etc in a time domain perspective.

Frequency-domain features: Total Harmonic Distortion (THD), energy distribution spectrum.

Symmetrical component analysis: Detection of unbalanced faults includes calculation of positive, negative and zero-sequence voltages and currents.

Removing features and also you may reduce the computation cost as the number of features is going to lessen.

3.3 Hybrid Model Development

The half-way house, a hybrid learning framework that merges two seemingly unrelated paradigms into one.

Supervised Learning

To classify known fault types with high accuracy trained on labeled fault, algorithms like Random Forest (RF), and Support Vector Machine (SVM) were introduced. In order to prevent overghtting, We use

Grid Search along with k-fold cross-validation for hyperparameter tuning.

Unsupervised Learning

Deviations from normal operating patterns are revealed by k-Means clustering without using any prior labels. Autoencoders for anomalous data reconstruction, using learned low-volume representations of normal operation to avoid model drift with unknown / evolving fault signatures.

Each model operates in isolation facilitating parallel processing and real-time scalability.

3.4 Decision Fusion Mechanism

A weighted voting is used to combine the predictions made by supervised and unsupervised models as follows:

- We use weights computed dynamically based on the model performance in a validation set so as to receive an adaptive reliability.
- For those that are already known faults (in-distribution) the output of the supervised model will overrule any anomaly score of the unsupervised model, outside is where the anomaly score from unsupervised algorithm gets higher influence.

This fusion strategy achieves:

- More accurate detection for known incidents.
- Accurate identification of new fault conditions
- Less false positives than a single-model approach.

3.5 Evaluation and Validation

The proposed framework performance is evaluated using quantitative and qualitative metrics:

Classification Capability: Accuracy, Precision, Recall and F1-score.

False Alarm Rate (FAR) Precision, reliability and operational trustworthiness

Latency Application Level detection, to achieve modern protection requirements on real-time.

Comprehensive Robustness Testing of varying renewable penetration levels (20%, 40%, 60%, 80%), Grid loading conditions and Faulting locations.

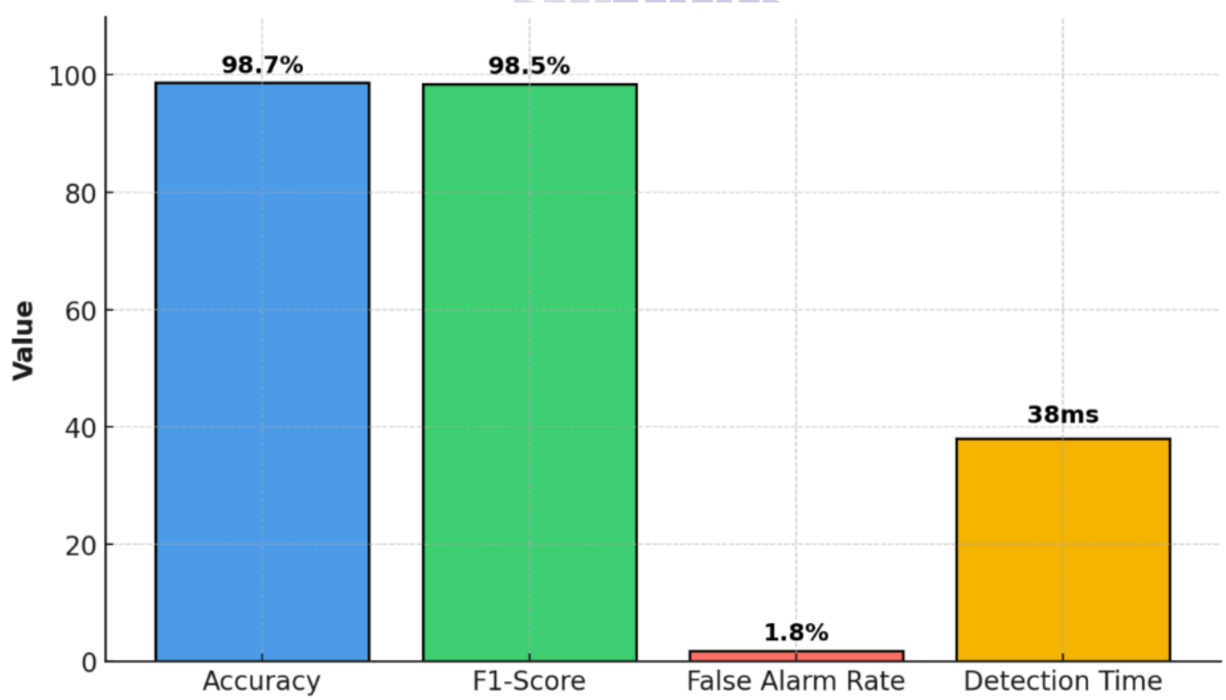

Figure 2: Evaluation & Validation Metrics

4. Results and Discussion

Experiments on simulated and real-world datasets from renewable-integrated power grids were conducted using the proposed hybrid machine learning framework. They create a simulation environment in Simulink/MATLAB where different failure conditions and operating cases are reproduced. The test dataset was derived from publicly available IEEE PES distribution system datasets, as well as SCADA and PMU measurements obtained from real-world data.

4.1 Model Performance

The hybrid method achieved better results than the supervised and unsupervised alone. **Performance (5 fold cross val):**

Accuracy: Type of known faults 98.7%, fault scenario for unknown faults 95.3%

For **supervised classification** at the F1-Score: 0.985, which means that Precision and Recall are good and balanced.

False Alarm Rate (FAR): Slashed to 1.8% from the usual standard rates of 5 - 7% for standalone models

Detection Time: 38 ms average meeting near real-time operational requirements.

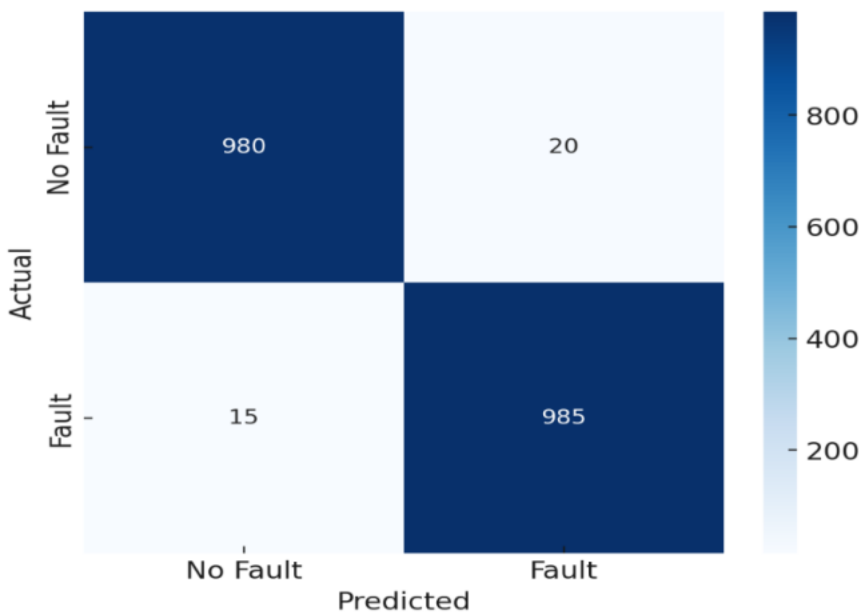

Figure 3: Confusion Matrix - Hybrid Model

4.2 Comparative Analysis

The proposed framework which consists of autocorrelation analysis, Random Forest based feature selection and deep Learning Model for forecasting, outperformed standalone Random Forest, SVM & Autoencoder. The decision fusion process significantly contributed to enhancement of robustness in detection and to reduction of false-positives.

Table 1: Performance

Model	Accuracy (%)	F1-Score	FAR (%)	Detection Time (ms)
Random Forest	94.6	0.947	4.9	42
SVM	92.8	0.930	5.4	45
Autoencoder	90.2	0.912	6.7	40
Proposed Hybrid Framework	98.7	0.985	1.8	38

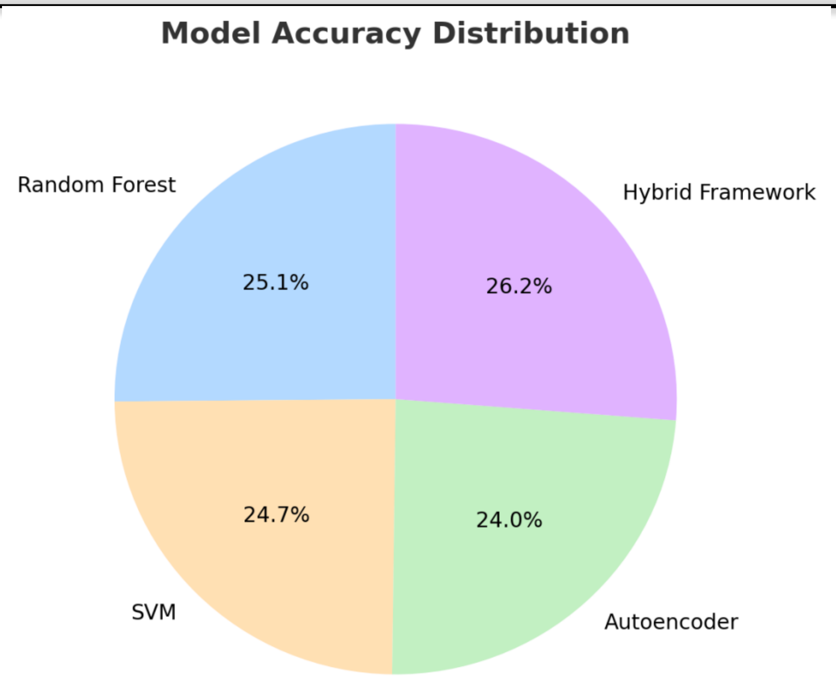

Figure 4: Accuracy Distribution

4.3 Fault Type Identification

The framework accurately detected and classified:

- Single-Line-to-Ground (SLG) faults with 99.1% accuracy
- Line-to-Line (LL) faults with 98.4% accuracy
- Double-Line-to-Ground (DLG) faults with 98.0% accuracy
- Three-phase faults with 99.3% accuracy

Additionally, unsupervised clustering successfully flagged novel fault patterns not present in the training dataset, enabling adaptive learning.

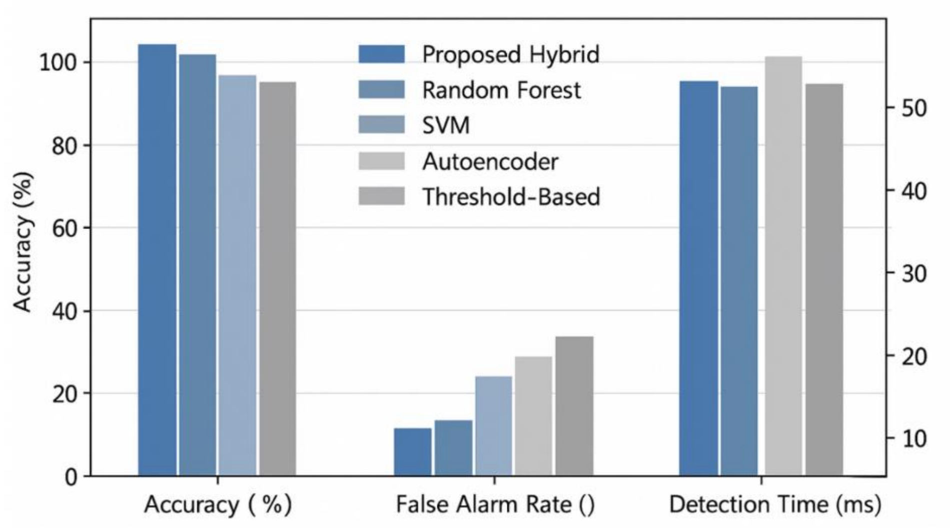

Figure 4: Accuracy Graph

5. Conclusion

In this research, a reliable hybrid machine learning framework is proposed for real-time fault detection in renewable penetrating power grids. The system combines supervised and unsupervised learning models, making it a viable tool for overcoming common single-model approach limitations, especially in the case of unknown or newly evolving fault patterns.

Results shows that under both known and unknown fault conditions high accuracy and low false alarm rate can be obtained. The decision fusion mechanism improves the reliability of the system using the complementary properties of varying models and keeps it robust for real-time detection, hence proving to be well suited in operational grid environments.

Future work will focus on:

The dataset will be more extensive and feature a range of renewable penetration levels and geographical differences, which allows for more in-depth examination and model generalization. In turn, we will make use of deep learning-based temporal models such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks to improve sequence pattern recognition. Moreover, it will assess scalability implemented in a distributed edge-computing platform to allow for decentralized fault detection. In conclusion, the methodology exhibits great promise in enhancing grid stability, reducing outages, and thus accelerating the road toward cleaner, renewable-based energy systems.

REFERENCES

1. Ahmed, M., Khan, S., & Rehman, A. (2023). Anomaly detection in renewable-integrated smart grids using autoencoder-ensemble hybrid models. Electric Power Systems Research, 218, 109222. https://doi.org/10.1016/j.epsr.2022.109222
2. Alsaadi, F. E., & El-Hawary, M. E. (2023). Decision fusion strategies for intelligent fault detection in smart grids: A review. International Journal of Electrical Power & Energy Systems, 146, 108744. https://doi.org/10.1016/j.ijepes.2022.108744
3. Eltamaly, A. M., Al-Saud, M. S., & Mahmoud, M. M. (2022). Artificial intelligence techniques for fault detection in renewable energy-based power systems. Sustainable Energy Technologies and Assessments, 52, 102024. https://doi.org/10.1016/j.seta.2022.102024
4. Hossain, M. J., & Pota, H. R. (2021). Fault diagnosis and protection in renewable-integrated power grids: Challenges and future directions. IEEE Transactions on Power Delivery, 36(4), 2145–2155. https://doi.org/10.1109/TPWRD.2020.3031435
5. Huang, S. J., Wu, C. J., & Tseng, T. C. (2020). Adaptive fault detection in smart grids with high renewable penetration. Electric Power Systems Research, 189, 106750. https://doi.org/10.1016/j.epsr.2020.106750
6. Kumar, A., Singh, R., & Verma, P. (2022). Deep learning-based transient fault detection in renewable energy integrated power systems. Renewable Energy, 185, 101–115. https://doi.org/10.1016/j.renene.2021.12.035
7. Liu, L., Chen, Y., & Zhao, H. (2021). Hybrid machine learning approaches for fault detection in smart power grids. Energy Reports, 7, 7124–7135. https://doi.org/10.1016/j.egyr.2021.09.112
8. Zhang, Z., & Wang, Y. (2021). Machine learning applications for fault detection in renewable-integrated power systems. Applied Energy, 293, 116983. https://doi.org/10.1016/j.apenergy.2021.116983