International Journal of Engineering Research and Science & Technology

IJERST

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 3 (1) 2025

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 3(1), 2025

Research Paper

INTELLIGENT FAULT DETECTION AND THERMAL MANAGEMENT SYSTEM FOR HIGH-ENERGY DENSITY EV BATTERIES

P. Satish[1], K. Akhila[2], K. Akhila[2], E. Eshwar goud[2], J. Ravi Teja[2]
[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (Data Science)
[1,2]Sree Dattha Group of Institutions, Sheriguda, Ibrahimpatnam, 501510, Telangana.

ABSTRACT

The growing global demand for sustainable energy has accelerated the deployment of thermal photovoltaic (PV) systems as an eco-friendly and renewable power source. These systems depend on uninterrupted operation and peak performance to ensure efficient energy output. However, they are vulnerable to various faults such as Maximum Power Point Tracking (MPPT) failures, Low Power Point Tracking (LPPT) issues, partial shading, and hardware degradation, all of which can significantly diminish efficiency and shorten system lifespan. As a result, accurate and timely fault detection and classification are critical for enhancing reliability and reducing maintenance costs. Traditionally, fault detection in EV batteries has relied on manual inspections and threshold-based monitoring via Supervisory Control and Data Acquisition (SCADA) systems. These conventional methods often fall short in handling large volumes of high-frequency data and identifying subtle or complex fault patterns. They are typically time-consuming, error-prone, and lack scalability—particularly in large-scale or distributed solar setups. To address these limitations, this project introduces an AI-powered fault detection and classification system that utilizes high-frequency operational data from EV batteries in combination with machine learning (ML) techniques to automatically detect and categorize faults. The system features a comprehensive pipeline including data preprocessing, class balancing using the Synthetic Minority Over-sampling Technique (SMOTE), feature reduction via Principal Component Analysis (PCA), and model training using Light Gradient Boosting Machine (LGBM), CatBoost Classifier, and the proposed Random Forest Classifier (RFC). A user-friendly graphical user interface (GUI) developed using Tkinter enables easy interaction with the system and provides real-time visualization of predictions and performance metrics. Experimental results demonstrate that the RFC model significantly outperforms the other models, achieving an accuracy of 99.82%, precision of 99.86%, recall of 99.86%, and F1-score of 99.86%, compared to 82.39% and 82.48% accuracy for LGBM and CatBoost, respectively. These findings validate the effectiveness of the proposed system as a robust, efficient, and scalable solution for fault detection in thermal PV applications.

Keywords: Thermal Photovoltaic Systems, Fault Detection, EV Batteries, Machine Learning, Random Forest Classifier, SMOTE, PCA, Real-Time Monitoring, LGBM, CatBoost

Received: 14-7-2025 Accepted: 21-8-2025 Published: 28-8-2025

1. INTRODUCTION

Solar energy technology has witnessed continuous global growth as a clean, quiet, and cost-effective renewable power source, with installed PV capacity reaching approximately 700,000 MW according to IRENA, yet thermal photovoltaic (PV) systems suffer efficiency losses and safety risks due to faults such as partial shading, inverter failures, and MPPT errors; traditional manual and threshold-based methods are inadequate for subtle or emerging anomalies in real time, especially as EV battery integration expands. To address this, we propose an intelligent, automated machine learning framework—leveraging models like LightGBM and Random Forest Classifier—that preprocesses high-frequency sensor data (handling missing values, encoding, and balancing via SMOTE), reduces dimensionality through PCA, and trains on balanced datasets to accurately detect and classify multiple fault types. By evaluating accuracy, precision, recall, and F1-score—metrics chosen for their sensitivity to class imbalance and operational consequences—our approach enables scalable, real-time fault classification that minimizes human intervention, supports preventive maintenance, and integrates seamlessly with IoT, SCADA, and smart-grid platforms, thereby enhancing reliability, reducing downtime and repair costs, and advancing the transition to sustainable, intelligent solar monitoring systems.

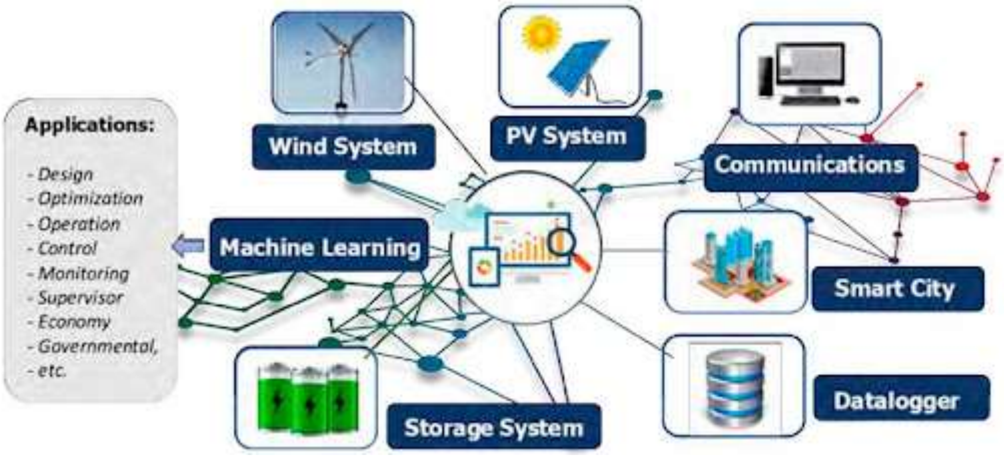


Fig. 1: Methods for fault detection in EV batteries.

2. LITERATURE SURVEY

In recent years, in contrast to standard model-based fault detection procedures [4, 5, 6], which involve simulating the PV installation's performance and comparing the simulated output power with the monitored one, machine learning (ML) and deep learning (DL) techniques have gained popularity and are considered promising solutions for fault detection and diagnosis in EV batteries. Numerous studies have evaluated the effectiveness of ML and DL approaches in fault detection and diagnosis in EV batteries.

For instance, Belaout et al. developed a multiclass adaptive Neurofuzzy technique for fault detection and diagnosis in EV batteries [7]. This algorithm can detect partial shading conditions, increased series resistance, faulty bypass diodes, and PV module short circuit faults. However, this technique cannot detect short circuits or defective strings under varying weather conditions.

Madeti and Singh introduced an algorithm based on k-nearest neighbors (KNN) for real-time fault detection in EV batteries, demonstrating its capability to detect and classify open-circuit faults, line-to-line faults, and partial shading faults [8]. However, it is worth noting that the method, while computationally efficient, is not flawless in terms of accuracy. Chen et al. proposed an intelligent fault detection approach based on I-V characteristics, utilizing an emerging kernel-based extreme learning machine. This method exhibits high accuracy in detecting and classifying faults in PV arrays [9].

Bendary et al. proposed two adaptive neuro-fuzzy inference system (ANFIS)-based controllers to address cleaning, tracking, and faulty issues in EV batteries [10]. This method relies on 1these parameters, accounting for ambient changes in conditions, including irradiation and temperature. Syafaruddin et al. suggested a simple and fast method based on several artificial neural networks (ANNs) capable of independently identifying the short-circuit location of PV modules in one string [11].

Garoudja et al. introduced a fault detection and diagnosis approach based on a Probabilistic Neural Network (PNN), which was tested using noisy and noiseless data [12]. Similarly, Vieira et al. proposed a fault detection method combining PNN with a Multilayer Perceptron algorithm. The process does not necessitate extensive datasets from pre-existing systems and primarily focuses on detecting short-circuited modules and disconnected strings [13]. While the results derived from the use of these machine learning approaches are promising, they also exhibit drawbacks, particularly with respect to extensive databases leading to overfitting. Furthermore, machine learning methods present limitations in representing features of complex high-dimensional data [14].

3. PROPOSED SYSTEM

The proposed approach introduces an optimized Random Forest Classifier (RFC)-based method for fault detection in thermal photovoltaic (PV) systems, addressing limitations in traditional models like LightGBM and CatBoost, which require extensive preprocessing and fine-tuning. RFC leverages ensemble learning through multiple decision trees with randomized feature selection and bootstrapped aggregation, enhancing robustness, generalization, and interpretability. The preprocessing phase begins with handling missing values and encoding categorical variables using Label Encoding. The dataset is then split into features (X) and labels (Y). To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is employed, generating synthetic samples for minority classes to prevent model bias and enhance classification fairness. Following this, Principal Component Analysis (PCA) is applied for feature reduction, retaining the most significant components while eliminating redundancy and noise, thereby improving computational efficiency without sacrificing accuracy. After PCA, the dataset is divided into training and testing sets to ensure the model generalizes well to unseen data. The RFC model is then trained and compared with traditional models, showcasing its superior ability to handle noisy and incomplete data while maintaining high performance.

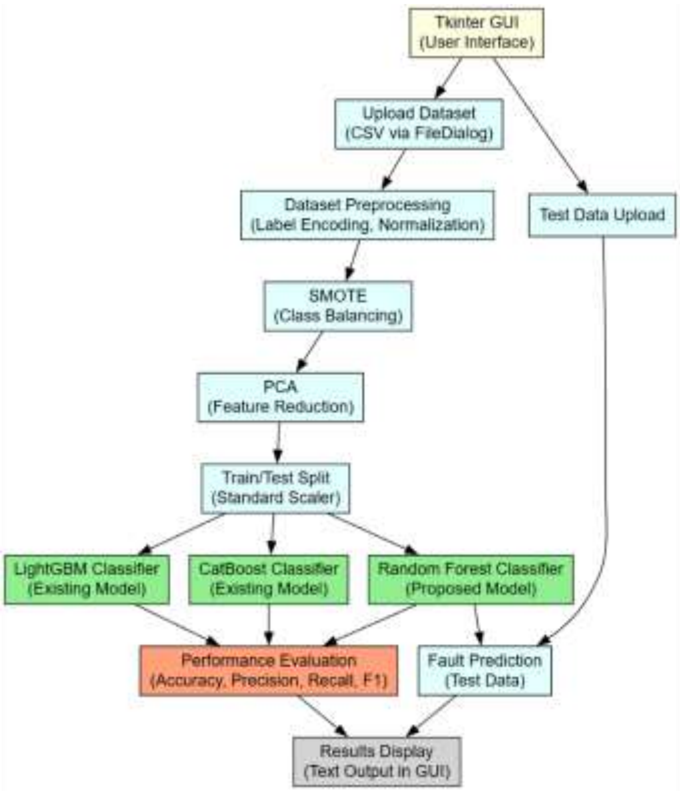


Fig. 2: Block diagram of proposed fault detection in thermal photovoltaic system.

Model evaluation is conducted using accuracy, precision, recall, F1-score, and a confusion matrix to thoroughly assess classification performance. Once validated, the trained RFC model is deployed in real-time PV systems for continuous monitoring and fault detection, offering greater interpretability and resilience compared to gradient-boosting models. This end-to-end pipeline—spanning data preprocessing, class balancing, dimensionality reduction, model training, evaluation, and deployment—forms a comprehensive and reliable solution for thermal PV system diagnostics. By integrating these steps, the proposed framework not only improves fault detection accuracy but also ensures scalability and practicality in real-world applications, making it a robust alternative to existing machine learning approaches in PV fault classification.

3.1 Proposed RFC Classifier Model

The Random Forest Classifier (RFC) is an ensemble machine learning algorithm that constructs multiple decision trees during training and aggregates their predictions to improve accuracy and reduce overfitting.

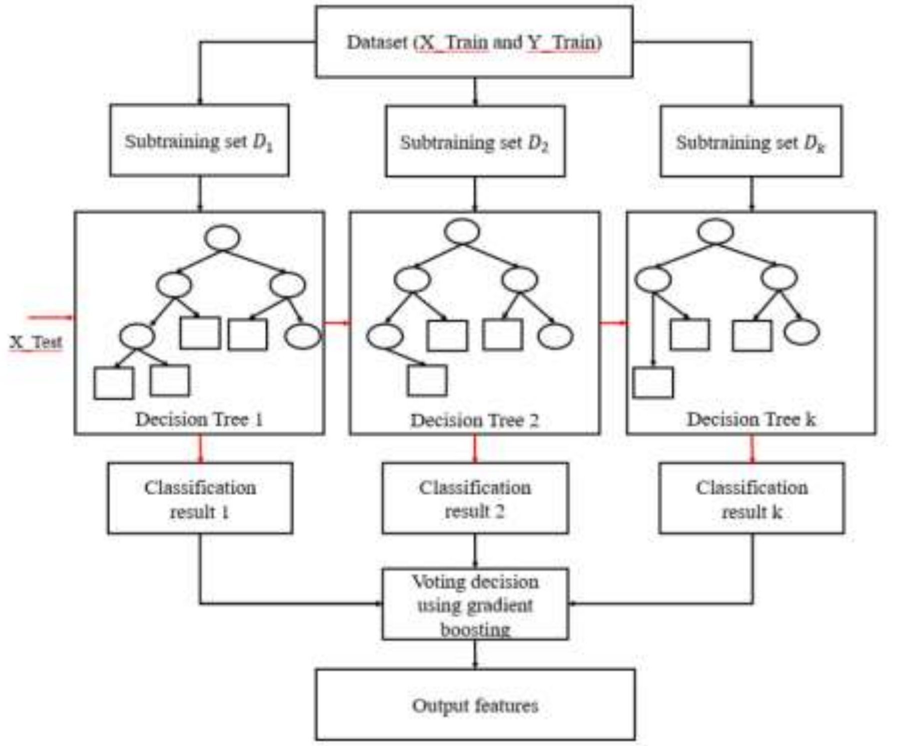


Fig. 3: RFC model workflow.

In the given fault detection and classification system, the RFC is either loaded from a pre-trained model or trained from scratch using X_train and y_train. Once trained, it predicts outcomes for X_test and generates y_pred, which undergoes optimization to enhance classification accuracy. The model's performance is then evaluated using precision, recall, F1-score, and confusion matrix analysis, ensuring reliable fault detection in thermal photovoltaic systems.

Step 1: Model Initialization and Pre-trained Model Check

The first step in the process involves checking whether a pre-trained Random Forest Classifier (RFC) model already exists. The code verifies the presence of a previously trained model stored as RandomForest_weights.pkl. If the model exists, it is loaded directly, allowing the system to bypass retraining and quickly move to making predictions. This approach saves computational resources and time, especially when handling large datasets.

Step 2: Training the Random Forest Classifier

If a pre-trained model is not found, the RFC is initialized and trained using the provided training dataset (X_train and y_train). The model is trained using an ensemble learning technique, where multiple decision trees are built using random subsets of data. Each tree independently predicts the outcome, and the final prediction is made based on majority voting among the trees. Once trained, the model weights are saved as a .pkl file, ensuring the model can be reused without needing retraining every time.

Step 3: Making Predictions on Test Data

Once the model is either loaded or trained, it is used to make predictions on the test dataset (X_test). The classifier processes the test input and generates predicted labels (y_pred). This step is crucial in evaluating the model's ability to detect and classify faults in thermal photovoltaic systems based on unseen data. The predicted output is compared against the actual test labels (y_test) to assess accuracy and overall model performance.

Step 4: Loss Optimization and Performance Metrics Calculation

To enhance the model's performance, the predicted labels (y_pred) undergo loss optimization through a specialized function (loss_optimization). This step ensures that errors are minimized and the predictions align as closely as possible with the true labels. Finally, the performance of the model is evaluated using metrics such as accuracy, precision, recall, and F1-score. A confusion matrix is also generated to visually analyze the classification results, helping in identifying areas where the model excels and where it may need improvement.

Advantages:

The Random Forest Classifier (RFC) offers several advantages that make it a strong candidate for fault detection in thermal photovoltaic systems. Its ensemble learning approach enhances accuracy and robustness by reducing overfitting and improving generalization to unseen data. RFC is capable of handling missing values effectively, minimizing the need for extensive preprocessing. It also provides valuable insights into feature importance, enabling more informed feature selection. With its ability to process large datasets efficiently through parallel tree-building, RFC is well-suited for real-time applications. Additionally, its resilience to noise and outliers ensures stable and reliable predictions. These strengths allow RFC to maintain an optimal balance between predictive performance, interpretability, and computational efficiency, making it highly effective for classifying faults in complex PV system environments.

4. RESULTS

Fig. 4 is a count plot visualizing the number of samples available for each fault category. This provides an overview of class distribution in the raw dataset before applying any balancing technique. Fig. 9.4 is a bar graph comparing class distributions before and after applying the SMOTE (Synthetic Minority Over-sampling Technique) algorithm. This visualization confirms the effectiveness of SMOTE in balancing minority and majority fault classes.

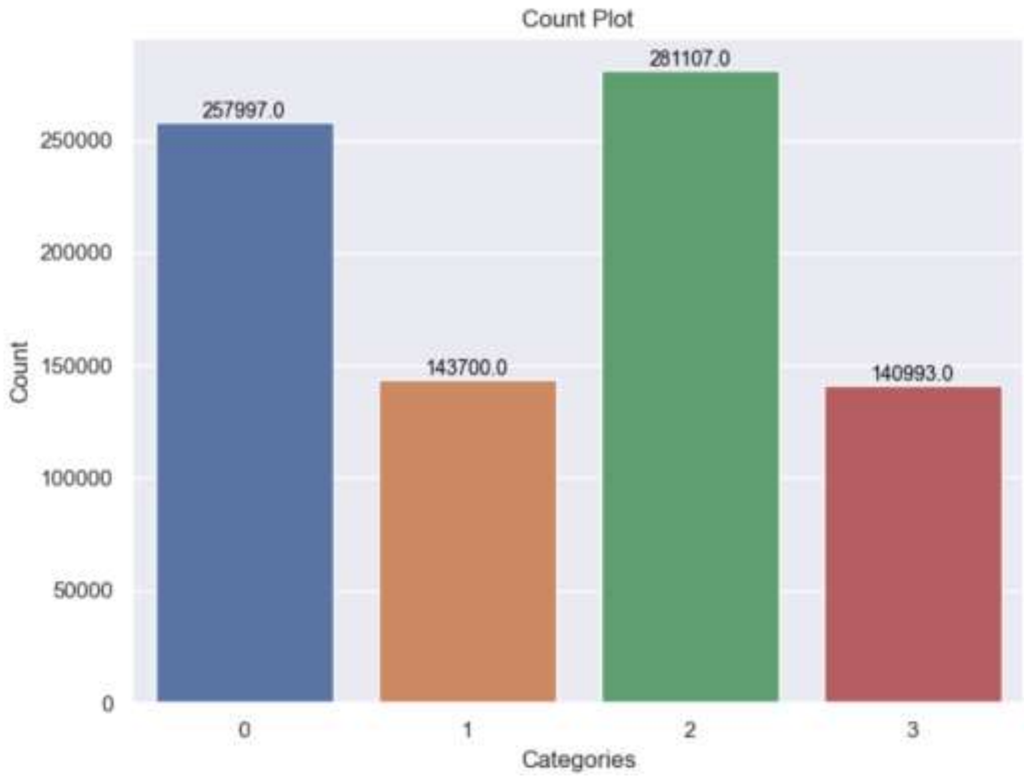


Fig. 4: Count plot (fault category versus number of samples).

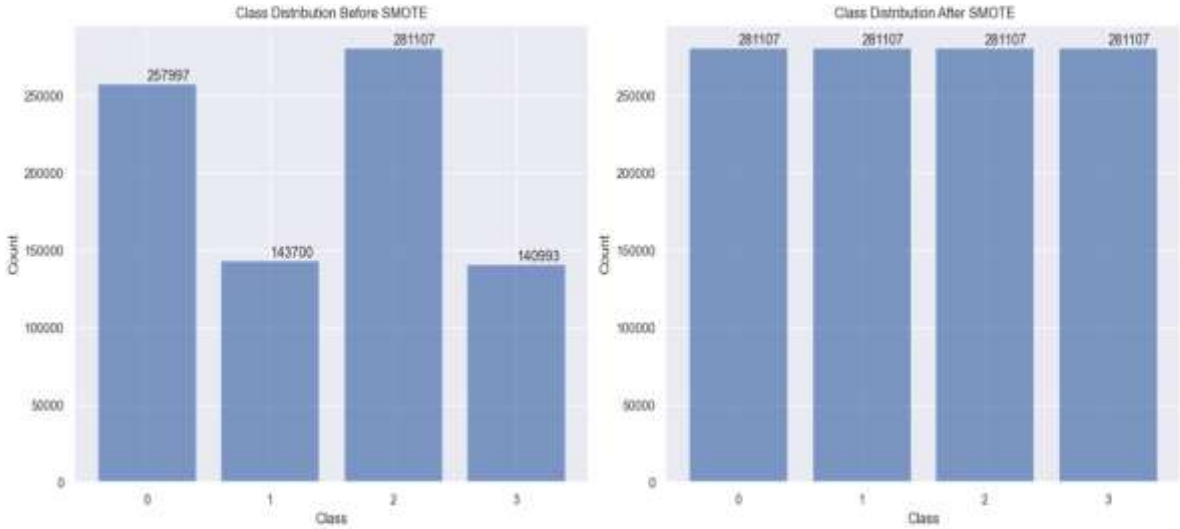


Fig. 5: Bar graph of class distribution before and after applying SMOTE algorithm.

Fig. 6 shows the scatter plots representing (a) the original high-dimensional data and (b) the data after transformation using PCA. These visualizations highlight the data structure and clustering behavior post-feature reduction. Fig. 7 demonstrate the confusion matrices of fault classification results using different machine learning models such as LightGBM classifier, Cat Boost classifier, and proposed RFC models. These matrices compare true vs. predicted classifications, with RFC demonstrating superior accuracy and balanced predictions across all fault categories.

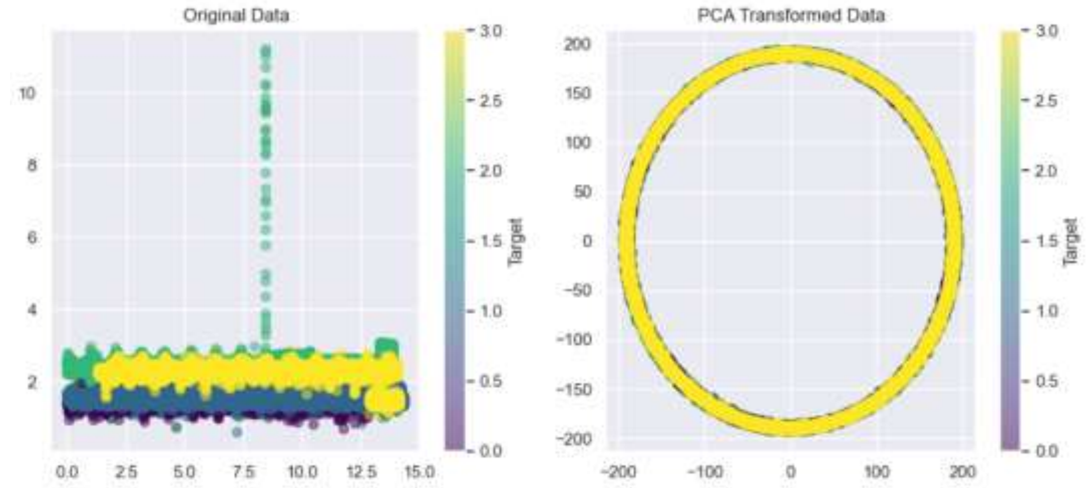
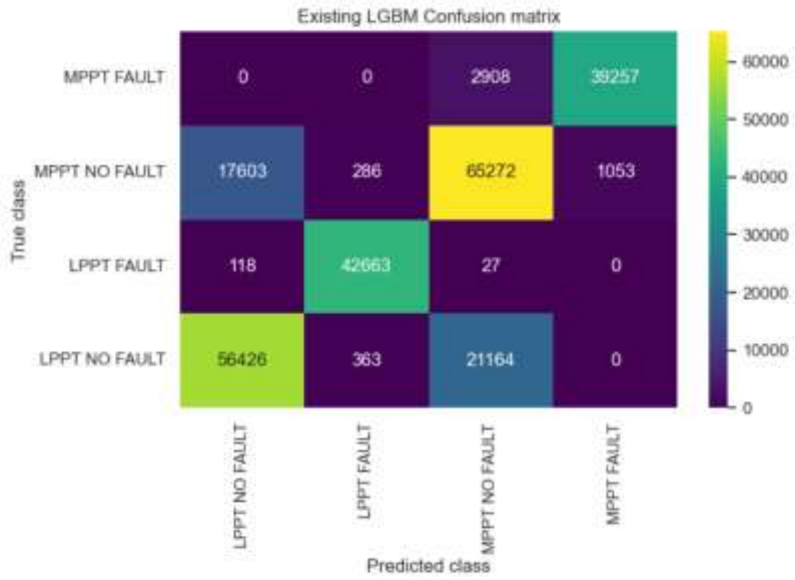
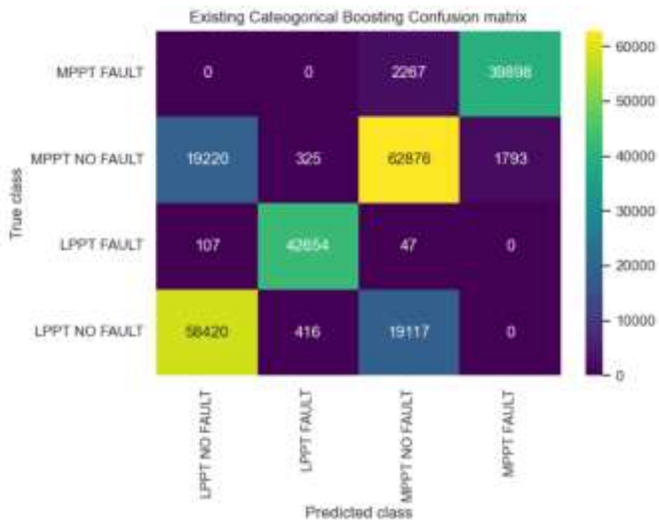


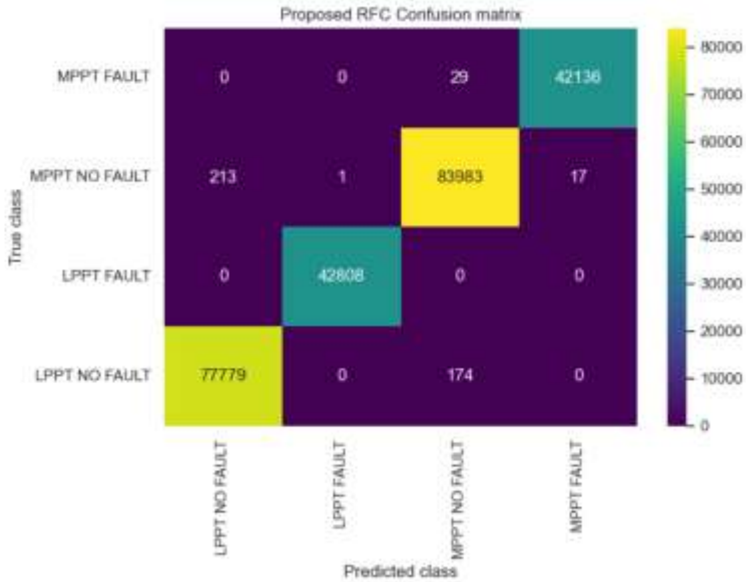
Fig. 6: Original data and PCA transformed data.



(a)



(b)



(c)

Fig. 7: Confusion matrices obtained using (a) LGBM classifier. (b) Cat boosting classifier. and (c) Proposed RFC model.

Table 1: Performance comparison of existing and proposed models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
LGBM classifier	82.39	86.26	85.66	85.92
Cat boost classifier	82.48	85.93	85.97	85.95
Proposed RFC model	**99.82**	**99.86**	**99.86**	**99.86**

Table 1 presents a comprehensive comparison of three machine learning models—LGBM, CatBoost, and the proposed Random Forest Classifier (RFC)—for fault detection and classification in thermal photovoltaic systems, evaluated using four key performance metrics: accuracy, precision, recall, and F1-score. While LGBM and CatBoost demonstrated similar and moderate performance, with accuracies of 82.39% and 82.48% respectively, RFC achieved a near-perfect accuracy of 99.82%, reflecting its superior reliability in classifying fault types. In terms of precision, LGBM and CatBoost scored 86.26% and 85.93%, whereas RFC led with an outstanding 99.86%, indicating minimal false positives. Recall values further emphasized RFC's dominance, scoring 99.86% compared to 85.66% (LGBM) and 85.97% (CatBoost), showcasing its ability to capture nearly all actual faults. The F1-score, which balances precision and recall, was highest for RFC at 99.86%, compared to 85.92% for LGBM and 85.95% for CatBoost. These results highlight that while LGBM and CatBoost offer decent performance, the RFC model significantly outperforms both, demonstrating superior consistency, accuracy, and robustness, making it the most reliable model for real-time fault detection in thermal PV systems using high-frequency sensor data.

5. CONCLUSION

This project successfully demonstrates the design and implementation of an AI-driven fault detection and classification system for thermal photovoltaic (PV) systems using high-frequency sensor data. The system employs a complete machine learning pipeline within a user-friendly Tkinter-based GUI, enabling users to upload data, preprocess it, balance the dataset using SMOTE, apply dimensionality reduction with PCA, and train multiple classifiers. Among the evaluated models—Light Gradient Boosting Machine (LGBM), CatBoost Classifier, and the proposed Random Forest Classifier (RFC)—the RFC model significantly outperformed the others. Specifically, the RFC achieved an accuracy of 99.82%, precision of 99.86%, recall of 99.86%, and F1-score of 99.86%, compared to LGBM's accuracy of 82.39% and CatBoost's 82.48%. These results confirm that the RFC model is highly effective in accurately identifying and classifying faults such as MPPT fault, LPPT fault, and their respective no-fault states. The integration of PCA and SMOTE played a crucial role in enhancing the model's robustness and generalization capabilities. Overall, the project delivers a reliable, scalable, and interpretable solution that contributes to improving the operational efficiency and fault resilience of solar energy systems.

REFERENCES

[1] Mantel C, Villebro F, Alves dos Reis Benatto G, Rajesh Parikh H, Wendlandt S, Hossain K, Poulsen PB, Spataru S, Séra D, Forchhammer S (2019) Machine learning prediction of defect types for electroluminescence images of thermal photovoltaic panels. Appl Mach Learn.

[2] Köntges M, Kurtz S, Packard C, Jahn U, Berger K, Kato K, Friesen T, Liu H, Iseghem MV, Wohlgemuth J, Miller D, Kempe M, Hacke P, Reil F, Bogdansk N, Herrmann W, Buerhop-Lutz C, Razongles G, Friesen G (2014) Review of failures of thermal photovoltaic modules—iea-pvps [www document].

[3] Kurukuru VSB, Frede B, Khan MA, Haque A (2020) A novel fault classification approach for thermal photovoltaic systems. Energies 13:308.

[4] Drews, A.; de Keizer, A.C.; Beyer, H.G.; Lorenz, E.; Betcke, J.; van Sark, W.G.J.H.M.; Heydenreich, W.; Wiemken, E.; Stettler, S.; Toggweiler, P.; et al. Monitoring and Remote Failure Detection of Grid-Connected EV batteries Based on Satellite Observations. Sol. Energy 2007, 81, 548–564.

[5] Silvestre, S.; Kichou, S.; Chouder, A.; Nofuentes, G.; Karatepe, E. Analysis of Current and Voltage Indicators in Grid Connected PV (Thermal photovoltaic) Systems Working in Faulty and Partial Shading Conditions. Energy 2015, 86, 42–50.

[6] Hariharan, R.; Chakkarapani, M.; Saravana Ilango, G.; Nagamani, C.; Member, S. A Method to Detect Thermal photovoltaic Array Faults and Partial Shading in EV batteries. IEEE J. Photovolt. 2016, 6, 1278–1285.

[7] Belaout, A.; Krim, F.; Mellit, A.; Talbi, B.; Arabi, A. Multiclass Adaptive Neuro-Fuzzy Classifier and Feature Selection Techniques for Thermal photovoltaic Array Fault Detection and Classification. Renew. Energy 2018, 127, 548–558.

[8] Madeti, S.R.; Singh, S.N. Modeling of PV System Based on Experimental Data for Fault Detection Using KNN Method. Sol. Energy 2018, 173, 139–151.

[9] Chen, Z.; Wu, L.; Cheng, S.; Lin, P.; Wu, Y.; Lin, W. Intelligent Fault Diagnosis of Thermal photovoltaic Arrays Based on Optimized Kernel Extreme Learning Machine and I-V Characteristics. Appl. Energy 2017, 204, 912–931.

[10] Bendary, A.F.; Abdelaziz, A.Y.; Ismail, M.M.; Mahmoud, K.; Lehtonen, M.; Darwish, M.M.F. Proposed ANFIS Based Approach for Fault Tracking, Detection, Clearing and Rearrangement for Thermal photovoltaic System. Sensors 2021, 21, 2269.

[11] Syafaruddin; Karatepe, E.; Hiyama, T. Controlling of Artificial Neural Network for Fault Diagnosis of Thermal photovoltaic Array. In Proceedings of the 16th International Conference on Intelligent System Applications to Power Systems, Hersonissos, Greece, 25–28 September 2011.

[12] Garoudja, E.; Chouder, A.; Kara, K.; Silvestre, S. An Enhanced Machine Learning Based Approach for Failures Detection and Diagnosis of EV batteries. Energy Convers. Manag. 2017, 151, 496–513.

[13] Vieira, R.G.; Dhimish, M.; de Araújo, F.M.U.; da Silva Guerra, M.I. Comparing Multilayer Perceptron and Probabilistic Neural Network for EV batteries Fault Detection. Expert Syst. Appl. 2022, 201, 117248.

[14] El-Banby, G.M.; Moawad, N.M.; Abouzalm, B.A.; Abouzaid, W.F.; Ramadan, E.A. Thermal photovoltaic System Fault Detection Techniques: A Review. Neural Comput. Appl. 2023, 35, 24829–24842.