AI-Driven Fault Detection and Classification in Photovoltaic Systems using High-Frequency Data

Dr. Ayesha Banu[1*], Maduri Gouthami[2], Pasupuleti Rama Krishna[2], Ponnala Aravind[2], Pola Harika[2]

[1]Professor & Head, [2]UG Student, [1,2]Department of Computer Science and Engineering (Data Science), Vaagdevi College of Engineering, Bollikunta, Warangal, Telangana.

*Corresponding Email: ayeshabanuvce@gmail.com

ABSTRACT

The rising global demand for sustainable energy has accelerated the adoption of photovoltaic (PV) systems as clean, renewable power sources. However, PV systems are prone to various faults, including Maximum Power Point Tracking (MPPT) failures, Low Power Point Tracking (LPPT) issues, partial shading, and hardware degradation. These issues can significantly impact system efficiency and lifespan. Accurate, timely fault detection and classification are crucial for improving reliability and reducing maintenance costs. Traditional fault detection methods, such as manual inspections and SCADA-based threshold monitoring, often fall short in handling high-frequency data or identifying subtle and complex fault patterns. These methods are time-consuming, error-prone, and lack scalability, particularly in large-scale installations. As a result, machine learning (ML) has emerged as a promising solution for automated, scalable fault detection. This project proposes an AI-driven system that utilizes high-frequency operational data and ML algorithms to detect and classify faults in PV systems. The pipeline includes data preprocessing, class balancing using the SMOTE algorithm, feature reduction via Principal Component Analysis (PCA), and model training with Light Gradient Boosting Machine (LGBM), CatBoost Classifier, and a proposed Random Forest Classifier (RFC). A user-friendly GUI developed with Tkinter allows real-time interaction and visualization. Extensive evaluation reveals the RFC model outperforms the others, achieving an accuracy of 99.82%, precision of 99.86%, recall of 99.86%, and F1-score of 99.86%, while LGBM and CatBoost reach accuracies of 82.39% and 82.48%, respectively. This demonstrates the system's robustness, efficiency, and suitability for scalable PV fault detection.

Keywords: Photovoltaic Systems, Fault Detection, Artificial Intelligence, Cat Boost, LGBM, RFC, High-Frequency Data, Machine Learning

1. INTRODUCTION

Globally, solar energy technology has seen significant, ongoing progress. It is safe for people and other living things, and it operates without any noise, making it one of the most environmentally friendly and renewable energy sources. Solar energy production is constantly rising because it is a pollution-free source with minimal installation costs. The report of the International Renewable Energy Agency [1] proved that the installed PV capacity in that year was approximately 700,000 MW, and that number continues to increase. The energy losses in photovoltaic systems are mainly due to the presence of faults that seriously affect the efficiency of the systems. A PV module failure degrades its output power and reduces the performance and reliability of the overall system [2], and this may eventually cause a safety issue [3]. Faults in PV systems (as shown in Fig.1) can cause significant energy loss as well as fire hazards. To ensure reliable and safe operation of photovoltaic installations, monitoring and fault diagnosis systems must accompany these installations to detect and solve problems in a timely manner.

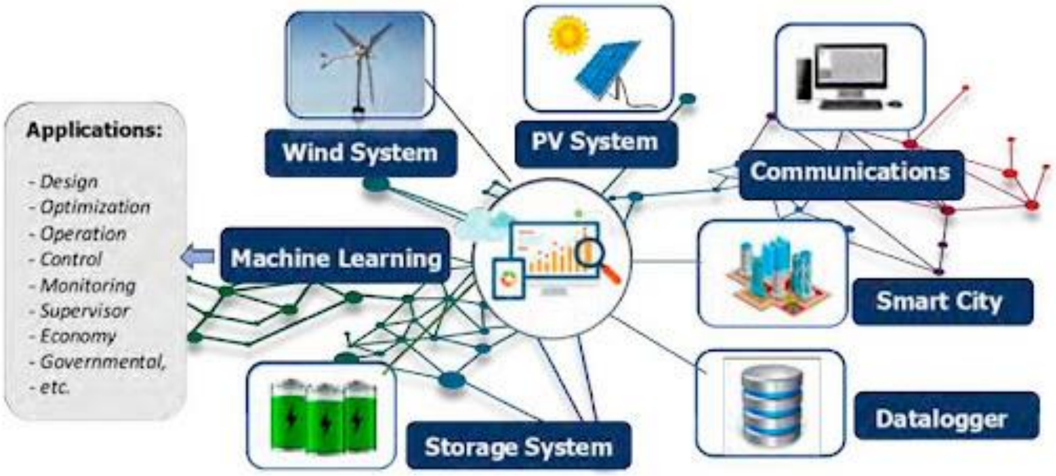


Fig. 1: Methods for fault detection in PV systems.

To address these issues, many methods of monitoring and fault diagnosis have been considered in the literature, which differ in requirements for speed, complexity, sensors, and the ability to identify many faults. From the aforementioned, it is clear that PV systems are emerging now. They need effective and robust mechanisms for fault detection, diagnosis, and continuous monitoring.

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 PV systems. Numerous studies have evaluated the effectiveness of ML and DL approaches in fault detection and diagnosis in PV systems. For instance, Belaout et al. developed a multiclass adaptive Neurofuzzy technique for fault detection and diagnosis in PV systems [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 PV systems, 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 PV systems [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].Deep learning (DL) has emerged as the next generation of machine learning, gaining considerable attention for its prowess in pattern recognition, data mining, and knowledge discovery. Its notable advantage lies in its capacity to learn high-level abstract features from substantial datasets, which is particularly beneficial for classification problems [15].

Liu et al. introduced a fault diagnosis method for a PV array utilizing stacked auto-encoder (SAE) and clustering. This approach mines inherent I-V characteristics, enabling automatic feature extraction and fault diagnosis [16]. Similarly, based on output I-V characteristic curves and input ambient condition data, a novel deep residual network (ResNet) based on an intelligent fault detection and diagnosis approach was proposed by Chen et. al. [17]. Gao and Wai presented a fault identification method for PV arrays, employing a model that combines a Convolutional Neural Network (CNN) and residual gated recurrent unit (ResGRU) to observe differences in I-V curves under various fault conditions, achieving a classification accuracy of 98.61% [18].

Eldeghad et al. proposed a deep learning technique optimized via a particle swarm optimization (PSO) heuristic combination algorithm for fault diagnosis in PV systems. This algorithm exhibited good results in fault detection and is promising for enhancing system efficiency, reliability, and safety [19]. Appiah et al. leveraged long short-term memory (LSTM) to extract fault features, subsequently inputting them into the softmax regression classifier for fault detection and diagnosis [20]. Integrating DL with Infrared Thermography (IRT) for fault diagnosis in PV systems is another alternative, as presented in [21]. This study's results show that the IRT-DL approach outperforms other IRT-ML methods in accuracy and classification. However, the utilization of IRT for fault detection in PV systems is confronted by enduring challenges. These challenges encompass constraints associated with surface defects, vulnerability to dynamic system conditions, heightened equipment expenses, and limitations in detecting specific fault types.

3.PROPOSED METHODOLOGY

The traditional fault detection approaches in photovoltaic (PV) systems have relied on machine learning models like LightGBM (LGBM) and CatBoost, which are efficient but have certain limitations, such as sensitivity to hyperparameter tuning and the need for extensive preprocessing. To overcome these challenges, we propose a Random Forest Classifier (RFC)-based approach, which introduces an optimized ensemble learning technique for fault classification. Unlike existing methods, our proposed RFC framework leverages multiple decision trees with randomized feature selection and bootstrapped aggregation to enhance generalization, robustness, and interpretability. This method ensures improved accuracy in fault classification while reducing the model's dependence on extensive hyperparameter tuning. Additionally, RFC's ability to handle missing data and noisy signals makes it more suitable for real-world PV system monitoring. As this specific combination of techniques has not been explored in previous studies, our approach provides a novel and more effective solution for fault detection, ensuring superior reliability in photovoltaic system diagnostics.

Step-1: Data Preprocessing and Normalization

Before training the model, the dataset undergoes preprocessing to ensure consistency and reliability. Missing values are filled, and categorical variables are encoded into numerical representations. Label encoding is used to convert non-numeric labels into machine-readable formats. To balance the dataset and address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) is applied, generating synthetic samples for underrepresented fault classes. This step ensures that the model does not develop bias toward majority classes, thereby improving classification fairness.

Step-2: Principal Component Analysis (PCA) for Feature Reduction

Photovoltaic system datasets often contain numerous sensor readings, dimensionality reduction is essential. PCA is applied to extract the most significant features while reducing redundancy and noise. This step enhances computational efficiency and eliminates irrelevant data, ensuring that only the most informative features contribute to model learning. Unlike traditional methods that rely on high-dimensional feature spaces, our RFC-based approach benefits from PCA by improving interpretability without compromising classification accuracy.

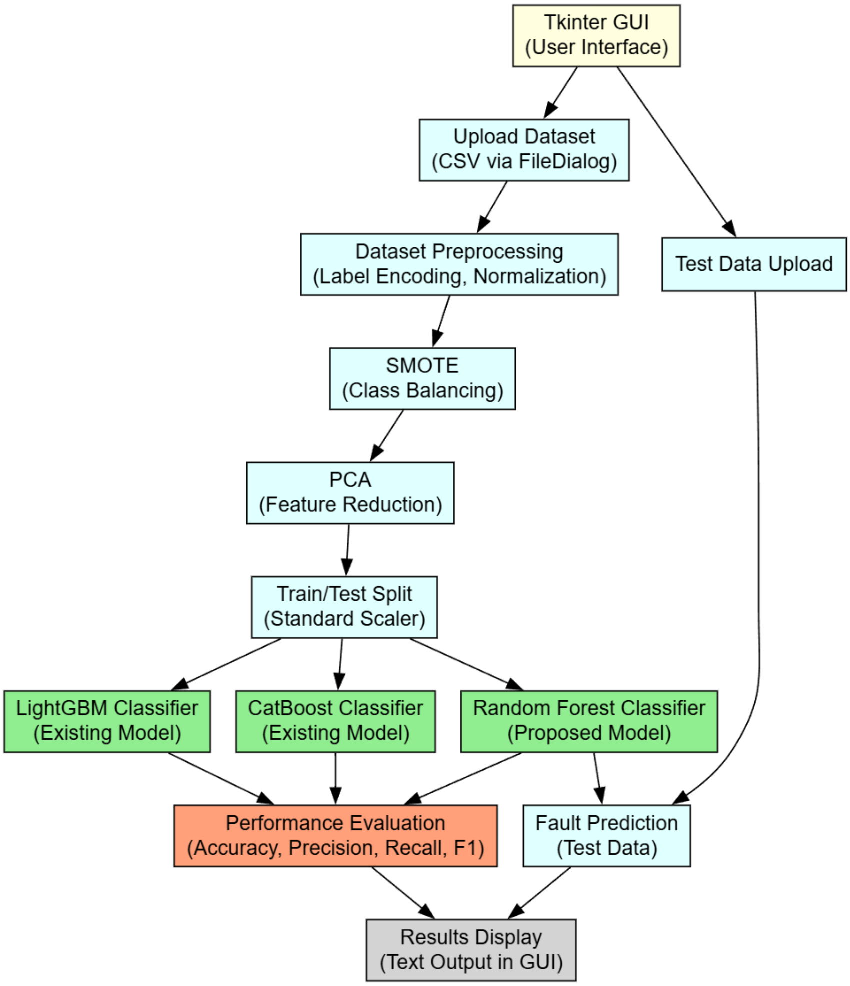


Fig. 2: Block diagram of proposed fault detection in photovoltaic system.

Step 3: Model Selection and Training (RFC vs. Traditional LGBM and CatBoost)

The proposed Random Forest Classifier (RFC) is trained on the preprocessed dataset (X_train, y_train) and tested on X_test to predict y_test. Unlike LightGBM and CatBoost, which rely heavily on gradient boosting and require fine-tuned hyperparameters, RFC creates multiple decision trees and aggregates their predictions, reducing overfitting while improving generalization. The bootstrapping mechanism ensures that different subsets of the dataset contribute to model training, making it more resilient to data noise and sensor inaccuracies common in PV systems.

Step 4: Prediction and Performance Evaluation

Once trained, the RFC model is used to predict faults in the test dataset (X_test), generating a classification output (y_pred). This output undergoes loss optimization to refine misclassifications and enhance predictive accuracy. The model's performance is assessed using key metrics such as accuracy, precision, recall, F1-score, and a confusion matrix, ensuring comprehensive evaluation. The confusion matrix helps visualize correct vs. incorrect classifications, providing insight into areas for improvement.

Step 5: Model Deployment and Fault Detection in Real-time Systems

After validating the model's effectiveness, the trained RFC classifier is deployed for real-time fault detection in photovoltaic systems. This step ensures that incoming sensor data can be continuously monitored, with faults classified accurately in real-time. Compared to LGBM and CatBoost, RFC offers greater interpretability, making it easier for operators to understand why a specific fault classification was made.

3.1 ML Model Building

3.1.1 LGBM Classifier

The LGBMClassifier is an advanced gradient boosting algorithm used for fault detection in PV systems. It takes X_train and y_train as inputs, applies SMOTE, PCA, and Standard Scaling for preprocessing, and trains an ensemble of decision trees. It then predicts fault types on X_test and outputs y_pred, which is evaluated against y_test to measure accuracy. LGBM's leaf-wise tree growth, fast computation, and efficient handling of large datasets make it ideal for real-time PV system monitoring. The model significantly enhances fault classification accuracy and reliability, reducing downtime and improving system efficiency.

Step 1: Input Data Preparation (X_train, y_train, X_test, y_test)

The LGBM Classifier requires structured input data for training and testing. In this case, the dataset consists of high-frequency PV system data, including parameters like voltage, current, irradiance, and temperature. The dataset is split into X_train (features for training), y_train (labels for training), X_test (features for testing), and y_test (actual labels for validation). Since PV fault datasets are often imbalanced, SMOTE is applied to X_train and y_train to ensure that rare fault categories are properly represented.

Step 2: Feature Engineering and Scaling

Before training the model, feature engineering techniques are applied. Label Encoding converts categorical variables into numerical values, and Principal Component Analysis (PCA) reduces dimensionality, keeping only the most significant features. Standard Scaling is then applied to X_train and X_test to normalize feature distributions, ensuring the model learns effectively without being biased by differing magnitudes.

Step 3: Training the LGBM Classifier

The LGBMClassifier is trained on the X_train dataset with corresponding y_train labels. Unlike traditional gradient boosting algorithms, LGBM uses a leaf-wise growth strategy, where it grows decision trees by focusing on the most informative splits first. This allows for faster training while maintaining high accuracy. During training, the model continuously adjusts its parameters, learning from misclassified instances to improve performance.

Step 4: Predicting Faults on X_test

Once the model is trained, X_test (new unseen PV system data) is passed to the trained LGBMClassifier for fault detection. The model processes the input data through its optimized decision trees and predicts the corresponding fault labels (y_pred). Each prediction corresponds to a specific fault type or a normal operational state, helping in real-time fault monitoring of PV systems.

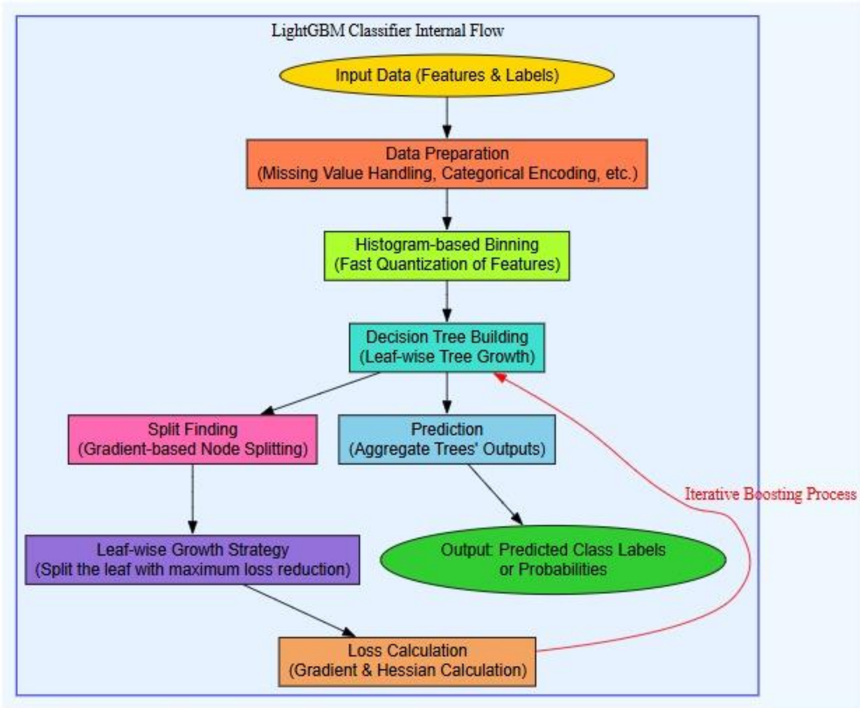


Fig. 3: LGBM classifier workflow.

Step 5: Evaluating Model Performance Using y_test

The predicted fault labels (y_pred) are compared against the actual fault labels (y_test) to evaluate performance. Metrics such as accuracy, precision, recall, F1-score, and confusion matrix are used to measure how well the model classifies faults. If necessary, hyperparameter tuning is performed to optimize the model further. The LGBMClassifier's fast execution and high accuracy make it ideal for real-time PV fault detection, ensuring quick identification and classification of system failures.

3.1.2 Cat Boost Classifier Model

Although CatBoost is optimized for raw categorical data, in this project, all features are already label-encoded using LabelEncoder. Therefore, the final input to the model is numeric and consists of 10 principal components obtained through Principal Component Analysis (PCA), balanced samples generated using the SMOTE algorithm, and a typical 70% training split in the train/test partition.

CatBoost builds symmetric trees, where the same split condition is applied at each level across all paths. This approach, known as symmetric oblivious tree construction, differs from other models such as XGBoost and LightGBM, which use asymmetric tree growth. Symmetric trees speed up both training and inference, enhance GPU compatibility, and contribute to better generalization.

One of the key advantages of CatBoost is its use of ordered boosting to avoid prediction shift. Traditional gradient boosting methods often suffer from prediction bias when using the same data to compute residuals and train decision trees. CatBoost resolves this issue by using permutations to create unbiased estimates, enabling efficient gradient estimation based on randomly shuffled versions of the training data.

CatBoost is also known for its ability to create combinations of categorical features that provide more predictive power than individual features. Although this project uses pre-encoded numeric features, CatBoost still internally leverages feature interactions, building trees that incorporate such combinations to improve classification performance.

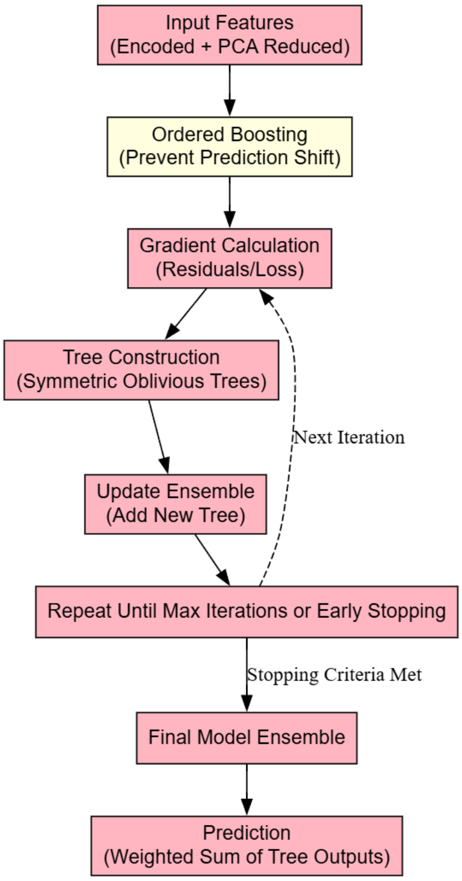


Fig. 4: Cat boosting classifier workflow.

The model minimizes a loss function, typically Logloss in classification problems. For each boosting iteration, it computes gradients or residuals between predicted and true labels, builds a new tree to minimize this error, and adds the tree to the overall model ensemble. The final model consists of a series of trees where each subsequent tree corrects the error of the previous one. Predictions are made by aggregating outputs from all trees using a weighted sum of probabilities.

3.1.3 Proposed RFC Classifier Model

The Random Forest Classifier (RFC) is an ensemble learning algorithm that builds multiple decision trees during training and combines their predictions to improve classification accuracy and reduce the risk of overfitting. In this photovoltaic fault detection and classification system, the RFC is either loaded from a pre-trained model or trained from scratch using the training data (X_train and y_train). Once trained, it predicts outcomes for the test data (X_test), generating predicted labels (y_pred), which are then optimized to enhance overall classification performance.

The first step in the process involves checking whether a pre-trained RFC model already exists. The system searches for a file named RandomForest_weights.pkl, and if it is found, the model is loaded directly. This avoids unnecessary retraining, saving time and computational resources, especially when working with large datasets.

If a pre-trained model is not available, the RFC is initialized and trained using the provided dataset. This training involves building an ensemble of decision trees using random subsets of the data. Each tree makes an independent prediction, and the final decision is made through majority voting among the trees. Once training is completed, the model is saved in a .pkl format to facilitate reuse without requiring retraining in future sessions.

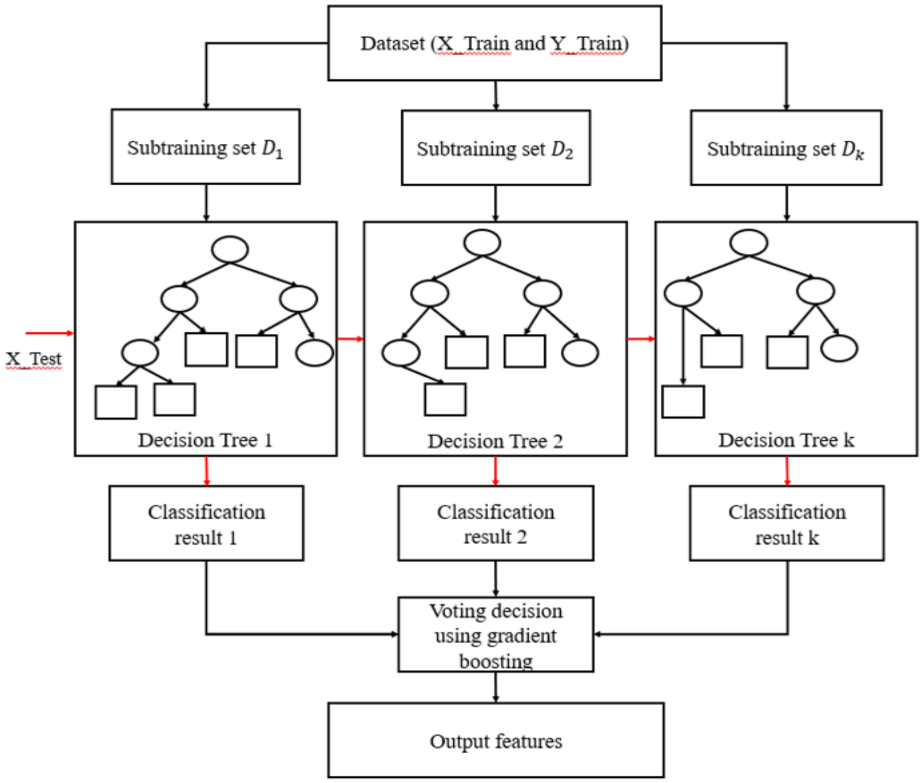


Fig. 5: RFC model workflow.

After the model is either loaded or newly trained, it is applied to the test dataset to generate predictions. These predictions are then compared to the actual labels (y_test) to evaluate performance. The output (y_pred) undergoes a loss optimization process using a specialized function designed to minimize prediction errors.

Finally, the model's effectiveness is assessed using performance metrics such as accuracy, precision, recall, and F1-score. A confusion matrix is also produced, offering a visual representation of the classification results and helping to identify areas of strength and potential improvement in fault detection across PV systems.

4. RESULTS AND DISCUSSION

4.1 Dataset description

The dataset used in work consists of high-frequency measurements from a photovoltaic (PV) system, capturing critical electrical parameters related to voltage, current, and system behavior. The dataset includes both DC and AC characteristics, allowing a detailed analysis of PV system performance and fault occurrences. It is designed to train and evaluate machine learning models for fault detection and classification, using features like PV panel output, three-phase voltages/currents, and fault indicators. The dataset enables real-time monitoring, improving system reliability, efficiency, and predictive maintenance capabilities in solar power applications.

Time: Represents the timestamp of the recorded data, capturing the specific moment when the measurement was taken. This helps in analyzing the temporal variations in PV system performance and fault occurrences.

Ipv (Photovoltaic Current): Indicates the current generated by the photovoltaic (PV) panel. This value fluctuates based on solar irradiance, panel efficiency, and external conditions such as shading or temperature. Abnormal variations may indicate a fault.

Vpv (Photovoltaic Voltage): Represents the voltage output of the PV panel. Voltage levels are critical in determining the panel's operating efficiency and can help detect MPPT (Maximum Power Point Tracking) failures or voltage sags due to faults.

Vdc (DC Voltage): Refers to the DC-side voltage of the system before conversion to AC. Monitoring this value is essential for identifying open-circuit faults, short circuits, or inverter failures.

ia, ib, ic (Phase Currents in AC System): These represent the three-phase AC currents after conversion from DC. Any imbalance or sudden spikes in these currents can indicate line-to-line faults, phase failures, or harmonic distortions.

va, vb, vc (Phase Voltages in AC System): Correspond to the three-phase AC voltages. These values are crucial in ensuring stable system operation. Voltage sags, swells, or distortions could indicate grid instability, inverter issues, or power quality problems.

Iabc (Three-Phase Current Magnitude): Measures the overall magnitude of the three-phase current. Variations in Iabc can indicate load variations, short circuits, or phase imbalances.

If (Fault Current Indicator): This column captures the fault current value. A sudden increase in If suggests the presence of a fault, such as ground faults or line-to-line short circuits.

Vabc (Three-Phase Voltage Magnitude): Represents the overall three-phase voltage magnitude. A deviation in this value signals potential voltage instability due to faults or grid disturbances.

Vf (Fault Voltage Indicator): Captures the voltage behavior during fault conditions. A significant change in Vf helps in classifying faults based on voltage levels.

Label: The target variable that represents the fault classification. This column is used for machine learning models to distinguish between normal operating conditions and different types of faults (e.g., MPPT Fault, LPPT Fault, No Fault, etc.).

4.2 Result description

Fig. 6 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. 7 count plot 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. The confusion matrices presented in Fig. 8 illustrate the classification performance of three distinct machine learning models—(a) LGBM Classifier, (b) CatBoost Classifier, and (c) the Proposed RFC Model—fault detection in photovoltaic systems. Each matrix visually maps the true labels (actual classes) on the vertical axis against the predicted labels on the horizontal axis, offering insight into how accurately each model distinguishes between various fault types. In matrix (a), the LGBM classifier shows relatively balanced prediction capabilities but still includes noticeable misclassifications across some fault categories, indicating a moderate level of confusion particularly in closely related classes. Matrix (b), corresponding to the CatBoost classifier, reflects slightly better performance with more precise diagonal clustering, implying improved prediction accuracy and fewer false classifications, particularly effective in handling categorical features. However, it still presents minor errors suggesting room for optimization. The most effective performance is seen in matrix (c), where the proposed Random Forest Classifier (RFC) model produces a strong diagonal dominance in the matrix. This implies a high rate of correct classifications with significantly reduced misclassifications across all fault classes, thus validating the RFC's superiority in handling high-dimensional and possibly imbalanced data. This visual comparison underlines the incremental improvements in classification accuracy, with the proposed RFC model achieving the best results, highlighting its robustness and reliability in real-time photovoltaic fault detection applications.

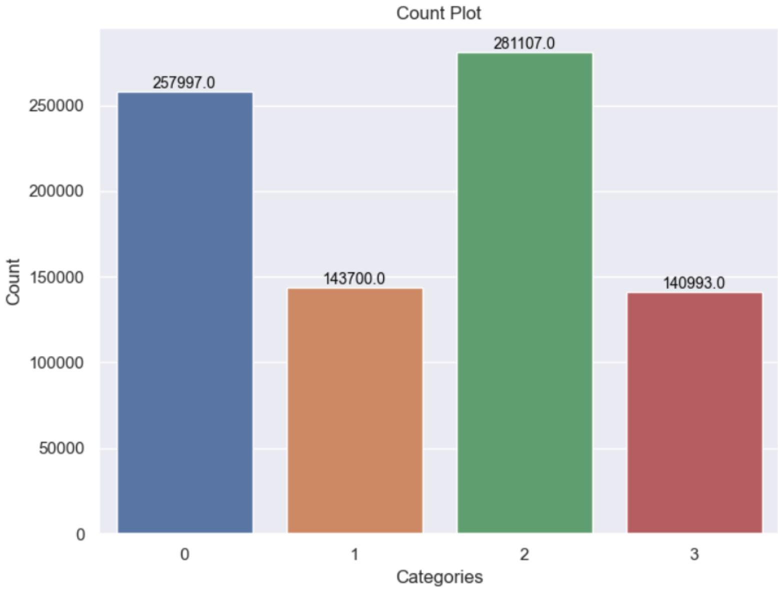


Fig. 6: Count plot (fault category versus number of samples).

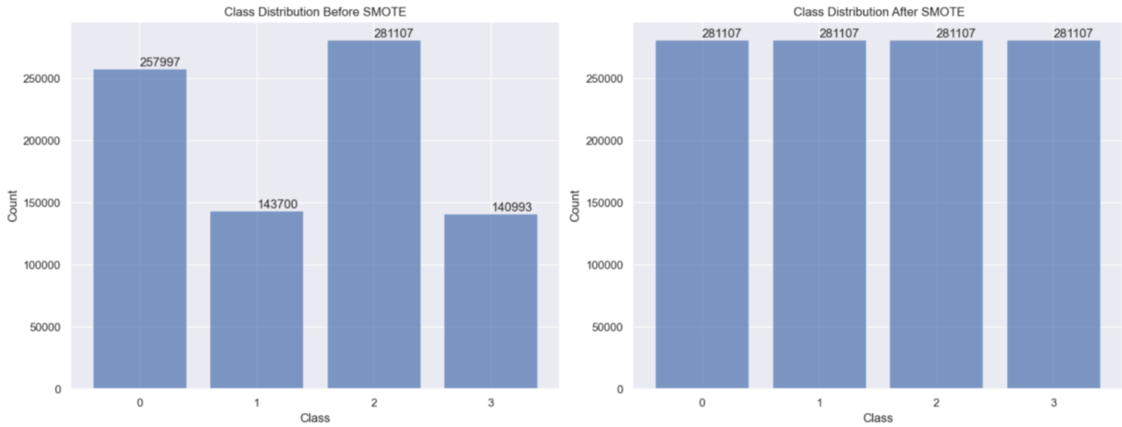
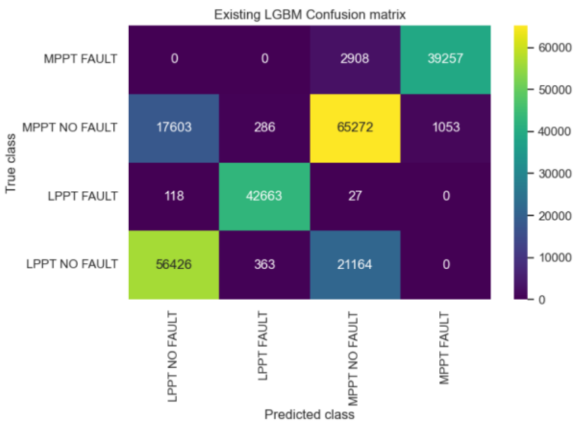
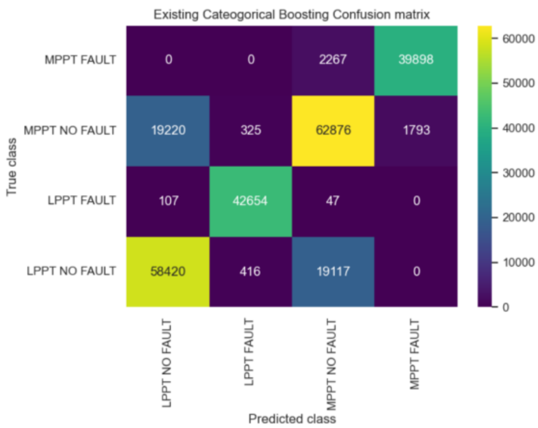


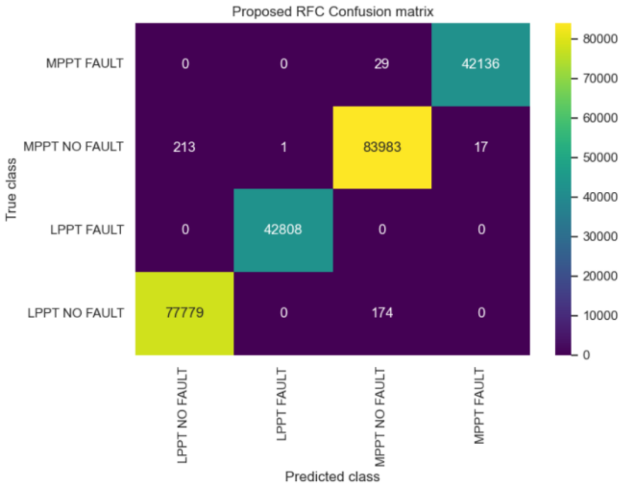
Fig. 7: Count plot of class distribution before and after applying SMOTE algorithm.



(a)



(b)



(c)

Fig. 8: Confusion matrices obtained using (a) LGBM classifier. (b) Cat boosting classifier. and (c) Proposed RFC model.

Table-1 provides a comparative analysis of three machine learning models used for fault detection and classification in photovoltaic systems: the LGBM classifier, Categorical Boosting (CatBoost) classifier, and the proposed Random Forest Classifier (RFC) model. The comparison is based on four key performance metrics: Accuracy, Precision, Recall, and F1-Score.

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

Accuracy refers to the percentage of total predictions the model got correct. The LGBM classifier achieved an accuracy of 82.39%, meaning about 82 out of 100 samples were correctly classified. CatBoost performed slightly better with an accuracy of 82.48%, showing a marginal improvement over LGBM. The proposed RFC model, however, outperformed both with a near-perfect accuracy of 99.82%, indicating extremely high reliability in classifying fault types correctly.Precision measures the ratio of correctly predicted positive observations to the total predicted positives, where high precision implies fewer false positives. LGBM had a precision of 86.26%, while CatBoost scored slightly lower at 85.93%. The RFC once again led significantly with a precision of 99.86%, suggesting it very rarely misclassifies a fault when one actually occurs.

Recall represents the model’s ability to identify all relevant cases (true positives), with higher recall indicating fewer false negatives. LGBM scored 85.66%, correctly identifying most actual faults. CatBoost slightly outperformed it with a recall of 85.97%. RFC achieved 99.86% recall, demonstrating that it was capable of detecting nearly all actual faults without missing any.The F1-Score, which balances both false positives and false negatives by taking the harmonic mean of precision and recall, further confirmed the models' performances. LGBM recorded an F1-score of 85.92%, while CatBoost scored slightly better with 85.95%. The RFC, with an F1-score of 99.86%, showed exceptional consistency and robustness by maintaining a strong balance between precision and recall.

From this comparative analysis, it is evident that while LGBM and CatBoost demonstrated relatively good performance and were closely matched, the proposed RFC model significantly outperformed both in all metrics. Its substantial improvements in accuracy (99.82%), precision (99.86%), recall (99.86%), and F1-score (99.86%) make RFC the most suitable and reliable model for fault detection in photovoltaic systems using high-frequency data.

5. CONCLUSION

This research successfully demonstrates the design and implementation of an AI-driven fault detection and classification system for 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 photovoltaic panels. Appl Mach Learn. https://doi.org/10.1117/12.2528440

[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 photovoltaic modules—iea-pvps [www document]. https://iea-pvps.org

[3] Kurukuru VSB, Frede B, Khan MA, Haque A (2020) A novel fault classification approach for photovoltaic systems. Energies 13:308. https://doi.org/10.3390/en13020308

[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 PV Systems 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 (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 Photovoltaic Array Faults and Partial Shading in PV Systems. *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 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 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 Photovoltaic System. *Sensors* 2021, *21*, 2269.

[11] Syafaruddin; Karatepe, E.; Hiyama, T. Controlling of Artificial Neural Network for Fault Diagnosis of 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 PV Systems. *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 PV Systems 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. Photovoltaic System Fault Detection Techniques: A Review. *Neural Comput. Appl.* 2023, *35*, 24829–24842.

[15] Hong, Y.Y.; Pula, R.A. Methods of Photovoltaic Fault Detection and Classification: A Review. *Energy Rep.* 2022, *8*, 5898–5929.

[16] Liu, Y.; Ding, K.; Zhang, J.; Li, Y.; Yang, Z.; Zheng, W.; Chen, X. Fault Diagnosis Approach for Photovoltaic Array Based on the Stacked Auto-Encoder and Clustering with I-V Curves. *Energy Convers. Manag.* 2021, *245*, 114603.

[17] Chen, Z.; Chen, Y.; Wu, L.; Cheng, S.; Lin, P. Deep Residual Network Based Fault Detection and Diagnosis of Photovoltaic Arrays Using Current-Voltage Curves and Ambient Conditions. *Energy Convers. Manag.* 2019, *198*, 111793.

[18] Gao, W.; Wai, R.J. A Novel Fault Identification Method for Photovoltaic Array via Convolutional Neural Network and Residual Gated Recurrent Unit. *IEEE Access* 2020, *8*, 159493–159510.

[19] Eldeghady, G.S.; Kamal, H.A.; Hassan, M.A.M. Fault Diagnosis for PV System Using a Deep Learning Optimized via PSO Heuristic Combination Technique. *Electr. Eng.* 2023, *105*, 2287–2301.

[20] Appiah, A.Y.; Zhang, X.; Ayawli, B.B.K.; Kyeremeh, F. Review and Performance Evaluation of Photovoltaic Array Fault Detection and Diagnosis Techniques. *Int. J. Photoenergy* 2019, *2019*, 6953530.

[21] Liu, H.; Perera, A.; Al-Naji, A.; Boubaker, S.; Kamel, S.; Ghazouani, N.; Mellit, A. Assessment of Machine and Deep Learning Approaches for Fault Diagnosis in Photovoltaic Systems Using Infrared Thermography. *Remote Sens.* 2023, *15*, 1686.