

**MACHINE LEARNING-BASED FAULT DETECTION AND SEVERITY ANALYSIS USING SENSOR DATA IN SOLAR PHOTOVOLTAIC (PV) SYSTEMS**

**Chandrashekar. B.M <sup>1</sup>, Dr. Hannah Jessie Rani. R <sup>2</sup>**

<sup>1</sup> Research scholar, Department of Electrical and Electronics Engineering, FET-Jain  
(Deemed-to-be-University), Bangaluru, Karnataka, India

Email ID: pcbm.cbr@gmail.com

ORCID: <https://orcid.org/0009-0000-1698-0507>

Google Scholar: <https://scholar.google.com/citations?user=j757Ou8AAAAJ&hl=en>

<sup>2</sup> Assistant Professor, Department of Electrical and Electronics Engineering, FET-Jain  
(Deemed-to-be-University), Bangaluru, Karnataka, India.

Email ID: jr.hannah@jainuniversity.ac.in

ORCID: 0000-0002-5449-104X

Scopus ID\*:57202119528

Google Scholar: <https://scholar.google.com/citations?hl=en&user=v6Xaa0gAAAAJ>

**Abstract**

This paper presents a comprehensive methodology for fault classification and severity analysis in solar photovoltaic (PV) systems that employs advanced machine learning and optimization techniques. The approach's central component is the Context Rank Layer with Parameterized Tanh-based Deep Bidirectional Long Short-Term Memory (CRLPT-DBi-LSTM) classifier, which improves fault detection accuracy. Sorted Bisection Farthest First Clustering (SBFFC) is used to group data efficiently, and Bayes Probability Function-based Coot Bird Optimization (BPF-CBO) is used to choose the best features. Data preprocessing includes multivariate imputation (MVI) and 2D scalogram visualization to ensure data integrity and quality. The Covariance-based Fuzzy Inference System (CFIS) is used for severity analysis, which results in detailed impact assessments. The proposed model does better than other classifiers in terms of sensitivity, specificity, accuracy, precision, recall, F-measure, negative predictive value (NPV), and Matthews correlation coefficient (MCC). A GUI is also being developed to allow for real-time monitoring and analysis, which will include performance metrics and visualization tools. The proposed methodology significantly improves fault management in solar PV systems, increasing reliability and operational efficiency. Future research will investigate scalability and adaptability to various environmental conditions.

**Keywords** - Fault Classification, Severity Analysis, Feature Selection, Data Clustering, Renewable Energy Management

### 1. Introduction

The growing use of solar photovoltaic (PV) systems as a renewable energy source necessitates the implementation of effective fault detection and severity analysis methods to ensure optimal performance and reliability [1,2]. The occurrence of faults in solar PV systems can result in substantial reductions in energy output, potential safety risks, and higher system maintenance costs [3]. Traditional fault detection methods frequently fail to accurately and promptly identify issues, especially in complex and dynamic environments. The objective of this research is to tackle the challenges by introducing a sophisticated approach for fault classification and severity analysis. This approach utilizes a CRLPT-DBi-LSTM classifier and a CFIS-based severity level assessment. The proposed method enhances the accuracy and efficiency of fault detection by integrating innovative techniques such as Sorted Bisection Farthest First Clustering (SBFFC) and Bayes Probability Function-based Coot Bird Optimization (BPF-CBO) for feature selection. The present study enhances fault classification and offers a comprehensive analysis of severity, which is essential for prompt maintenance and reducing downtime in solar PV systems. The results illustrate the potential of the proposed approach in transforming the maintenance and operational strategies of solar PV installations.

#### 1.1. Background and Motivation

The global transition towards renewable energy sources has established solar PV systems as a fundamental component of sustainable power generation [4]. These systems must operate efficiently and reliably as they become more common. The presence of faults in solar PV systems can have a significant negative impact on their overall performance. These faults can result in energy losses, higher operational expenses, and even pose potential safety risks [5]. Conventional fault detection and severity analysis techniques frequently encounter difficulties in handling the intricate and diverse nature of solar PV systems. As a result, more sophisticated and dependable approaches are required. The motivation for conducting this study arises from the necessity to improve fault detection and severity analysis in solar PV systems. The accurate classification of faults and timely identification of severity levels are crucial for ensuring the efficiency and longevity of PV installations. The current methods frequently lack the precision and adaptability required to effectively manage the ever-changing characteristics of solar power generation [6]. The objective of this research is to address the existing gap by proposing a new framework for fault classification and severity analysis.

The proposed approach utilizes the Context Rank Layer with Parameterized Tanh-based Deep Bidirectional Long Short-Term Memory (CRLPT-DBi-LSTM) classifier, which has been specifically developed to enhance the precision of fault classification. Moreover, the utilization of Covariance-based Fuzzy Inference System (CFIS) in severity analysis offers a reliable mechanism for forecasting the rate of degradation and the overall health of the

system. The efficacy of the proposed method can be further enhanced by integrating advanced techniques such as SBFFC and BPF-CBO for feature extraction and selection. This study aims to contribute to the development of more resilient and efficient solar PV systems by addressing the limitations of existing fault detection systems. The proposed framework enhances fault detection accuracy and provides a comprehensive approach to severity analysis, resulting in improved maintenance strategies and reduced downtime for solar PV installations.

### 1.2. Importance of Fault Classification and Severity Analysis

Fault classification and severity analysis play a crucial role in the operation and maintenance of solar PV systems. With the growing importance of renewable energy sources, specifically solar PV systems, in global energy strategies, it is crucial to ensure the reliability and efficiency of these systems to sustain their economic and environmental advantages [7].

Solar PV systems are susceptible to a range of faults that can have a significant impact on their overall performance. The faults that can occur in a solar power system include partial shading [8], module degradation [9], inverter failures [10], and grid-related issues [11]. Fault classification is a critical process that helps identify the exact type of fault. This, in turn, facilitates targeted maintenance activities and effectively minimizes system downtime. The prompt and accurate classification of faults is critical to ensure timely resolution of issues, thereby preserving the system's operational efficiency and preventing minor problems from escalating into major failures.

Safety hazards can arise from faults in solar PV systems, including the potential for electrical fires or the risk of electrocution [12]. Severity analysis is an essential component in evaluating the level of risk associated with identified faults. The prioritization of actions by maintenance teams is facilitated by assessing the severity level of each fault. This enables immediate attention to high-severity faults, thereby reducing safety risks. The implementation of this proactive approach significantly improves the overall safety of solar PV installations.

The implementation of efficient fault classification and severity analysis techniques is critical in achieving substantial cost savings in solar PV system maintenance. Accurately identifying faults and determining their severity can minimize unnecessary inspections and maintenance activities [13]. Utilization of this focused strategy for maintenance not only results in a reduction of labor costs, but also serves to prolong the lifespan of system components by preventing unnecessary interventions. Furthermore, identifying significant faults in a timely manner can help avoid costly repairs or replacements in the future.

The reliability of solar PV systems is of utmost importance to facilitate their widespread adoption and seamless integration into the energy grid. Ensuring reliable fault classification is crucial for maintaining consistent system operation with minimal interruptions [14]. A severity analysis is a method that provides valuable information about a system's condition and deterioration rate. This information can be used to develop proactive maintenance strategies. Implementing these strategies facilitates consistent energy generation and fosters confidence among consumers and stakeholders in the effectiveness of solar PV technology.

Contemporary solar PV systems frequently incorporate sophisticated monitoring and control systems. These systems' effectiveness is dependent on fault classification and severity analysis [15]. This system's analyses provide real-time information on faults and their severity. This information allows for automated control actions, such as isolating faulty components or adjusting operational parameters to optimize performance. The integration of various components results in enhanced intelligence and improved responsiveness in PV systems.

Fault classification and severity analysis are essential for ensuring the efficient, safe, and cost-effective operation of solar PV systems. The processes are implemented to ensure the timely and accurate identification of faults, assessment of their potential impact, and execution of appropriate maintenance actions [16]. The analyses conducted aim to improve the reliability and performance of solar PV systems, thereby contributing to the overall objective of developing sustainable and resilient energy infrastructure.

### 1.3. Problem Statement

The controller plays a crucial role in optimizing power generation from PV solar systems by regulating the DC-DC converter's duty cycle. If the adjustment is not properly balanced, it can have a negative impact on the performance of the PV power generation system. Existing research has largely overlooked the errors caused by controller imbalances, which can lead to incorrect fault detection. In PV solar power systems, a range of faults can occur. However, it is important to note that studies such as the one conducted by [17] only address a limited number of faults. As a result, their findings may not be applicable or reliable for other types of faults. Research requires an extended training duration due to the complex nature of learning parameters [18]. The study [19], solely focused on time-domain data, potentially leading to unbalanced results. Furthermore, certain methodologies rely on direct voltage measurement to predict faults without considering the temperature dependence of solar power. This oversight leads to an increase in the misclassification rate. Finally, most of the current research fails to assess fault severity, resulting in a notable deficiency in comprehending the influence of various fault types on the system's performance.

### 1.4. Research Contribution

The purpose of this research is to present a new framework for fault classification and severity analysis in solar PV systems. This framework utilizes advanced machine learning and optimization techniques. This work's main contribution is the development of the CRLPT-DBi-LSTM classifier. Research has demonstrated that this classifier significantly enhances the accuracy of fault classification. Furthermore, the inclusion of the CFIS provides a reliable mechanism for analysing severity, allowing for accurate degradation rate prediction and system health assessment. SBFFC and BPF-CBO techniques are used in the proposed method to improve feature extraction and selection, which leads to better performance. This comprehensive approach not only improves fault detection and classification, but also provides a thorough analysis of their severity. This contributes to the development of more

efficient maintenance strategies and improves the overall operational reliability of solar PV systems.

## 2. Related Works

There are several types of faults that can occur in a PV system, regardless of whether it is a stand-alone system or connected to the grid. PV modules, electrical devices, the Maximum Power Point Tracking (MPPT) side, the converter, and the inverter are some of the parts of the PV system that can go wrong. Faults can also arise within the PV modules itself [20]. A failure in a PV module can result in a decrease in its power output, which in turn reduces the performance and reliability of the entire system. This can potentially give rise to safety concerns. Faults in PV systems not only result in significant energy losses, but they also pose significant fire risks [21].

Using the nonlinear equations of the PV panel model, signal analysis is conducted on the input-output data from the model. The presence of a fault in the system is determined by comparing the differences, or residuals, between the actual system measurements and the model predictions. Well-known Fault Detection and Diagnosis (FDD) models in PV systems include the single-diode model, the double-diode model, and the current-driven three-diode model [22]. While model-based strategies can achieve acceptable accuracy under high irradiance conditions, their accuracy tends to decrease under low irradiance conditions. The data-driven approach places a strong emphasis on collecting large amounts of data for analysis and interpretation [23]. This approach differs from the model-based approach, which relies on pre-existing qualitative or quantitative knowledge about the system. Data-driven FDD methods rely on extensive training datasets that cover a wide range of operational conditions and fault scenarios [24]. These methods aim to establish relationships between input and output signals to enable accurate fault identification and diagnosis.

Data-driven FDD techniques for PV systems utilize various methods, such as statistical analysis and ML, which are well-suited for addressing complex and nonlinear problems. PV systems utilize various AI systems. These include ANN, fuzzy logic, support vector machines, decision trees, and k-nearest neighbours algorithms. The adoption of these methods has increased fault detection accuracy and efficiency in PV systems [25,26]. The study [27] introduced a method for classifying 11 classes of PV module faults. This method employed a multi-scale kernel approach as well as the visual perception levels of a CNN based on transfer learning. The approach utilizes the pre-trained Alex-Net model to augment the network's capabilities. To mitigate the problem of class imbalance in the dataset, offline augmentation techniques were utilized, including oversampling. Long Short-Term Memory (LSTM) is a specialized type of recurrent neural network (RNN) that is specifically developed to effectively manage long-term dependencies. LSTM networks can retain and establish connections between information even when there are long gaps between the input and output sequences, which is not possible with regular RNNs [28]. The forget gate mechanism effectively addresses the limitations of traditional RNNs by facilitating the management of long-range dependencies.



and severity assessment in solar PV systems. The flowchart depicted in Figure 1 illustrates the overall process.

### 3.1. Data Acquisition

In this study, two distinct datasets are employed to develop and validate the proposed fault classification and severity analysis methodology for solar PV systems.

- **Dataset 1: Fault Detection Dataset in Photovoltaic Farms [29]**

The study employed a simulated 250-kW PV power plant to create training and testing datasets that included a range of PV fault scenarios. The simulation defines three distinct fault types in addition to normal operation. The fault cases consist of three types: string fault (F1), string-to-ground fault (F2), and string-to-string fault (F3). PV system tests these faults on various strings. The training dataset comprises 600 instances, each containing 30 features and one categorical column indicating the fault type. The dataset distribution consists of 100 instances (16.67%) of normal operation, 153 instances (25.5%) of string faults, 149 instances (24.83%) of string-to-ground faults, and 198 instances (33%) of string-to-string faults. The simulation has a total duration of 0.4 seconds, during which faults are expected to occur at 0.2 seconds. The measurements for the training dataset are captured within a time window of 0.2 seconds to 0.4 seconds following the initiation of the fault. The testing dataset consists of 50 instances, with measurements ranging from 0.1 seconds to 0.3 seconds. The transient period spans from 0.1 seconds to 0.2 seconds, during which faults occur between 0.2 seconds and 0.3 seconds. The setup used in this scenario guarantees a thorough assessment of the model's performance under various fault conditions.

- **Dataset 2: PV Fault Dataset [30]**

The second dataset encompasses a span of 16 days and contains operational data from a grid-tied photovoltaic plant. This dataset includes information on both faulty and normal operation conditions. The dataset is partitioned into two MATLAB '.mat' files, which enables convenient accessibility and analysis. The data collection system is composed of a photovoltaic plant with two strings. Each string is equipped with 8 C6SU-330P PV modules. These modules are connected to a 5-kW grid-tie power inverter, specifically the NHS Solar 5K-GDM1 model. The dataset contains various fault scenarios, including intentionally induced degradation, short-circuit, and open-circuit conditions. Additionally, it includes naturally occurring shadowing effects caused by nearby buildings. The degradation simulation utilizes a resistive load bank positioned between two modules. Connecting a cable between the positive and negative terminals of neighbouring modules induces short-circuits. Open circuits are introduced by opening the main circuit breaker on one string. The provided datasets offer a diverse and comprehensive basis for training and evaluating the proposed models for fault classification and severity analysis. This ensures that the models can perform reliably and effectively in various operational scenarios, as well as with different types of faults in solar PV systems.

### 3.2. Pre-Processing

The data is subjected to pre-processing using multivariate imputation (MVI) [31] to address any missing or inconsistent values. This step crucially verifies the integrity and quality of the data before using it for subsequent analysis, including fault classification and severity analysis in solar PV systems. MVI is a statistical technique that is employed to estimate and replace missing values in a dataset. It does so by considering the relationships among multiple variables, making it a sophisticated method. The MVI method differs from univariate imputation by considering the correlations and interactions between variables. This approach enhances the accuracy and reliability of estimates by accounting for the relationships between different variables. The method operates based on the assumption that the missing data points can be estimated using a multivariate distribution derived from the observed data. The process commences with an initial imputation, typically employing the mean, median, or another straightforward technique to temporarily substitute missing values. The MVI technique is used to iteratively refine initial estimates by creating models for variables with missing data based on other variables in the dataset. The iterative process is repeated until the imputations reach a stable state. This process yields a dataset where plausible estimates accurately represent the underlying data structure and variability, replacing the missing values. The utilization of MVI guarantees the preservation of dataset integrity and impartiality, effectively reducing the potential risks associated with data loss. The proposed approach not only ensures the preservation of the original integrity of the data but also enhances the performance of ML models by offering a comprehensive dataset for both training and testing purposes.

### 3.3. Correlation Calculation

The next step involves computing the correlation between various solar power measurements. In this process, understanding relationships and dependencies among various parameters within the solar PV system is essential. A correlation is a statistical measure that quantifies the strength and direction of the linear relationship between two variables. The identification of potential effects of changes in one parameter, such as solar irradiance or temperature, on another parameter, such as power output, can be achieved by calculating the correlation coefficients. The purpose of this analysis is to identify discrepancies and anomalies that could potentially indicate faults. This is an example when there is an expected strong correlation between irradiance and power output. In such cases, if there is a sudden deviation from this expected pattern, it could indicate a potential fault in the system. A systematic analysis of these correlations can identify specific components or conditions that are not meeting expectations. The process of correlation calculation is utilized to facilitate feature selection by determining the parameters that are most pertinent for fault detection and severity analysis. This step is crucial for ensuring that the following stages of the methodology, such as clustering and classification, rely on the most informative and relevant data. This enhances the overall accuracy and reliability of the fault detection process in solar PV systems.

### 3.4. Clustering using SBFFC

The SBFFC technique clusters the data after performing the correlation calculation. The SBFFC method is a highly reliable clustering technique specifically developed to effectively group together data points that exhibit similarity. This clustering process greatly aids in the identification of fault patterns within the solar PV system. The technique commences by first sorting the data points. It then proceeds to iteratively divide the dataset into smaller clusters. This division is based on the farthest distance between points within each cluster. By focusing on the farthest points, the SBFFC method effectively captures the data's inherent variability and distribution. This ensures that each cluster contains data points that are closely related in terms of their characteristics. The proposed approach facilitates the differentiation between normal operational states and different fault conditions by leveraging the natural grouping of data points that represent similar fault types. The utilization of SBFFC for clustering facilitates the process of reducing the intricacy of the dataset, thereby enhancing the ease of analysis and interpretation. The clear structure of the data provided also enhances the subsequent stages of the methodology, such as feature extraction and fault classification. This technique enables more effective fault detection and isolation in solar PV systems, resulting in improved accuracy when identifying and diagnosing faults.

### 3.5. Feature Extraction and Selection

The feature extraction and selection phase are crucial for enhancing the classification process in the proposed fault detection methodology for solar PV systems. Initially, feature extraction converts raw data into a set of quantifiable characteristics that effectively represent the inherent patterns in the data. This process involves applying various statistical, mathematical, and signal processing techniques to identify significant features such as mean, variance, frequency components, and temporal dynamics. After extracting these features, the BPF-CBO algorithm is utilized for feature selection. BPF-CBO is an advanced optimization technique inspired by the foraging behaviour of coot birds and Bayesian probability principles. This method evaluates the relevance of each feature by estimating its contribution to the classification task, considering both its discriminative power and redundancy with other features.

The BPF-CBO algorithm employs an iterative approach to identify the optimal subset of features, effectively balancing the trade-off between maximizing classification accuracy and minimizing computational complexity. By selecting only the most informative and non-redundant features, BPF-CBO ensures that the classifier is trained on the most relevant data. This streamlined feature set enhances the performance and efficiency of the CRLPT-DBi-LSTM classifier. As a result, the classifier can more accurately identify and categorize faults in solar PV systems, leading to improved fault detection and reliability.

### 3.6 Classification using CRLPT-DBi-LSTM

During the classification phase, the chosen features are fed into the CRLPT-DBi-LSTM classifier. The classifier has been purposefully developed to effectively manage the intricacies and subtleties commonly found in time-series data, which are characteristic of solar PV

systems. The CRLPT-DBi-LSTM classifier is a combination of multiple advanced techniques that have been implemented to improve the accuracy of classification. The LSTM architecture is designed to effectively capture long-term dependencies in sequential data, thereby overcoming the limitations of traditional RNNs. Because the LSTM works in both directions, it can consider both the previous and subsequent contexts. This makes it better at finding patterns and outliers. The context rank layer is an additional component of the model that aims to improve its performance. It achieves this by giving more importance to the temporal features that are most relevant, thereby enhancing the classifier's ability to focus on critical data points. The parameterized Tanh activation function is a versatile and adaptable mechanism used to regulate the non-linearity of a model. It enables the model to capture the underlying data distribution more accurately. The algorithm of parameterized Tanh activation function is shown in Figure 2. The combination of these components allows the CRLPT-DBi-LSTM classifier to effectively differentiate between the 'Abnormal' and 'Normal' categories with high accuracy. The classifier utilizes its advanced architecture to analyze the extracted features and identify subtle fault patterns and variations in the solar PV system. Implementing this approach results in improved fault classification accuracy, increased system reliability, and prompt maintenance interventions.

```
function ParameterizedTanh(ranked_context):  
    tanh_transformed = []  
    for each context in ranked_context:  
        transformed_value = parameterized_tanh_function(context)  
        tanh_transformed.append(transformed_value)  
    return tanh_transformed
```

Figure 2. Pseudocode of ParameterizedTanh Algorithm

### 3.7. Severity Prediction and Analysis

A severity prediction is performed on data classified as 'Abnormal' by utilizing degradation rate calculation. The process entails quantifying the degradation rate of the PV system's performance caused by identified faults. The CFIS is utilized to perform a comprehensive analysis of severity. The CFIS utilizes covariance measures to effectively manage uncertainty and imprecision in the data. This approach offers a reliable framework for assessing the severity of faults. The system evaluates the impact and urgency of identified faults by analysing the correlation and variability among various fault indicators. This analysis allows for an accurate determination of the criticality of the fault. Based on the severity analysis, the controller adjusts the duty cycle of the DC-DC converter to optimize the PV system's

performance. The CFIS will send an alert message to the PV sensors to initiate corrective actions when it detects medium- or high-severity levels. The dynamic adjustment of the system guarantees its operation within safe and efficient parameters, effectively reducing the risk of severe damage and maintaining an optimal power output. Integrating severity analysis with real-time controller adjustments improves the system's resilience and reliability. This integration guarantees prompt intervention and maintenance to avert potential failures.

#### 4. Results and Analysis

This section presents and analyses the results obtained from the implementation of the proposed fault classification and severity analysis methodology for solar PV systems. The performance evaluation of the CRLPT-DBi-LSTM classifier, as well as the effectiveness assessment of the feature extraction and selection process, is conducted using the previously described datasets. The classifier's ability to accurately identify and categorize faults is assessed using various metrics, such as accuracy, precision, recall, and F1-score. In this study, we investigate the effects of the CFIS on severity prediction and the resulting modifications in the duty cycle of the DC-DC converter. The results are compared with established methodologies to emphasize the enhancements achieved through our approach. The presented analysis showcases the strength and effectiveness of the proposed system in improving fault detection and management in solar PV systems.

Figure 3 presents a 2D scalogram of PV data, which offers a time-frequency representation. This representation enables the visualization of amplitude variations in the solar PV system's measurements over time. Wavelet transform techniques generate the scalogram by breaking down the signal into its constituent frequencies. It then displays how these frequencies change over time. The color intensity represents the amplitude of the signal at various frequencies and times. This research utilizes the scalogram to identify unique patterns and peaks at specific frequencies and times. These high-intensity regions correspond to significant amplitude variations, which can be indicative of faults or anomalies in the PV system. Consistently observed at regular intervals, the presence of these peaks indicates the occurrence of events, potentially including string-to-ground or string-to-string faults. The identification of these patterns is of utmost importance for the CRLPT-DBi-LSTM classifier to differentiate between 'normal' and 'abnormal' operational states. The scalogram can be used to extract features such as the frequency of occurrence, duration, and amplitude of peaks. These features are important inputs for the classification model. The visualization serves as a tool to enhance comprehension of the severity and impact of faults. Regions with higher amplitude may indicate more severe faults. In addition, the scalogram is used to validate the pre-processed data, ensuring that it maintains important information for precise fault classification. The 2D scalogram is utilized to capture the intricate details of signal variations, which in turn enhances the robust training of the CRLPT-DBi-LSTM classifier. This process ensures accurate identification of faults and subsequently improves the reliability and performance of solar PV systems.

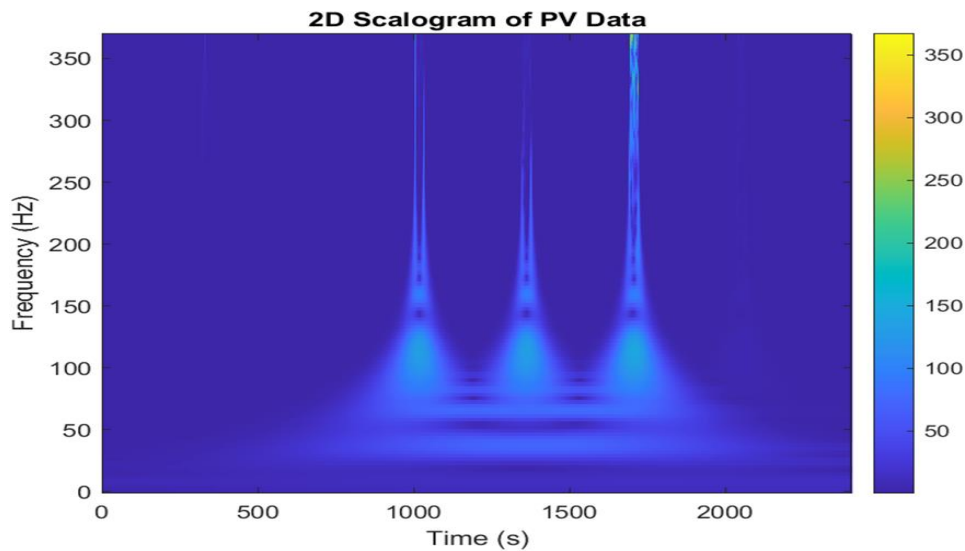


Figure 3. 2D Scalogram of PV Data

Figure 4 demonstrates the outcomes of applying the SBFFC algorithm to the features extracted from the solar PV system data. Principal Component Analysis (PCA) is utilized to decrease the dimensionality of the data, guaranteeing that the two principal components capture the most significant variance. The fact that the clusters are clearly separated shows that the SBFFC algorithm is good at figuring out different kinds of data patterns, which could mean both normal and broken operational states. The separation of components is crucial to achieve precise fault detection and classification. The spread and distribution of points within the clusters offer valuable insights into the significance and discriminative capability of the extracted features. The presence of a well-separated cluster structure indicates that the features chosen by the BPF-CBO algorithm possess informative and pertinent qualities for classification.

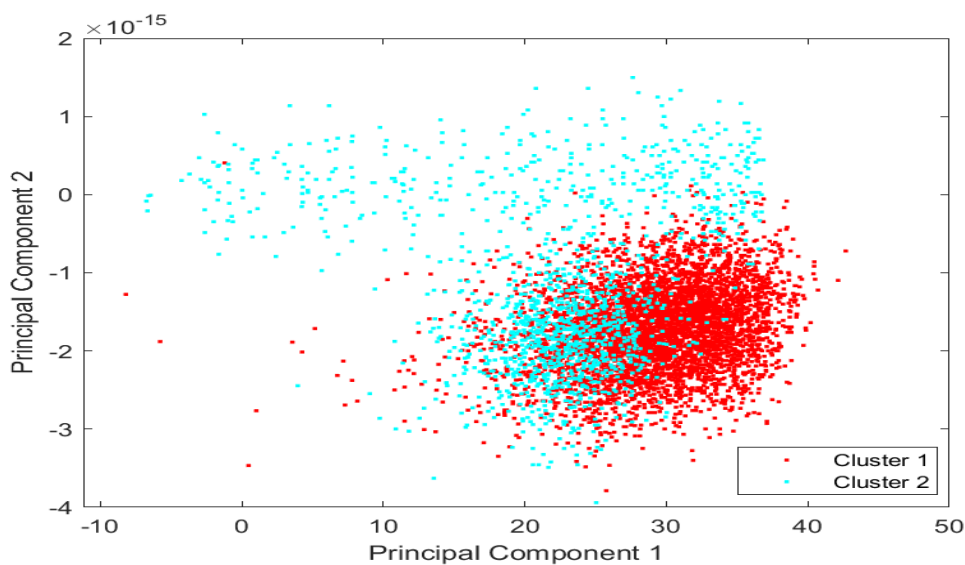


Figure 4. Clustering Results Using SBFFC

The fitness values over iterations for the BPF-CBO algorithm and four other comparison algorithms are shown in Figure 5. It's worth noting that the BPF-CBO algorithm has the fastest convergence rate and highest fitness value, which shows how well it works for choosing features for fault classification in solar PV systems. The BPF-CBO algorithm works better than others, which means it picks the most useful and unique features, which makes the classifier more accurate and dependable. The plot illustrates the relative effectiveness of various optimization algorithms. The BPF-CBO algorithm consistently demonstrates superior performance compared to other algorithms, indicating its suitability for handling the intricate and non-linear characteristics of solar PV system data.

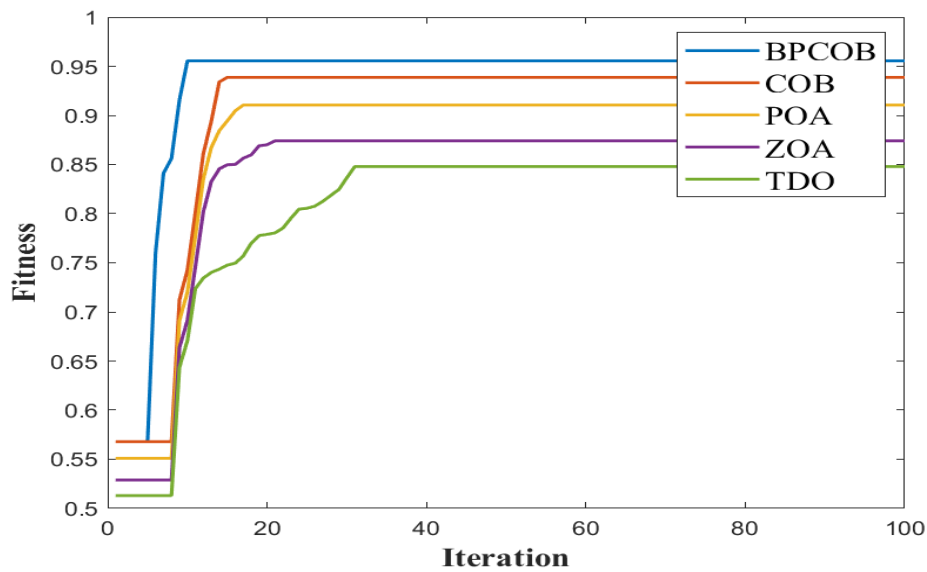


Figure 5. Fitness Convergence of BPF-CBO and Comparison Algorithms

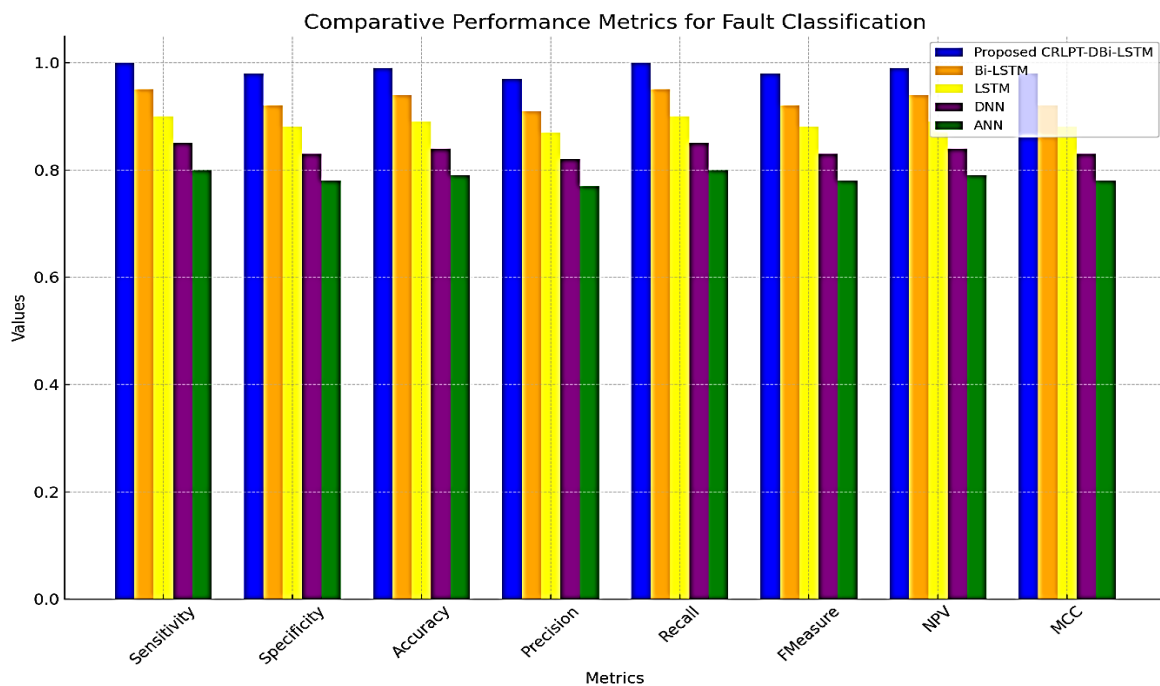


Figure 6. Comparative Performance Metrics for Fault Classification

Figure 6 showcases a comprehensive comparison of performance metrics for the proposed CRLPT-DBi-LSTM classifier and other models, including Bi-LSTM, LSTM, DNN, and ANN. According to the results, the proposed CRLPT-DBi-LSTM model is the most sensitive, which means it can accurately find real fault instances. The implementation guarantees the reliable detection of faults, thereby minimizing the potential for undetected issues. The CRLPT-DBi-LSTM model demonstrates the highest level of accuracy compared to other models. This indicates its superior performance in accurately classifying both normal and faulty states. The high accuracy of solar PV systems is crucial for ensuring their reliability and efficiency. The proposed model exhibits a precision of 0.97, indicating its capability to maintain a high true positive rate compared to false positives. This ensures that most identified faults are indeed actual faults, which is crucial for efficient fault management. The proposed model demonstrates high reliability in correctly identifying non-faulty conditions, as evidenced by its achieved NPV of 0.99. In this context, it is critical to ensure system stability and avoid unnecessary maintenance actions. The Matthews Correlation Coefficient (MCC) value of 0.98 obtained for the CRLPT-DBi-LSTM model demonstrates a high degree of correlation between the predicted and actual classifications. This result serves as additional evidence of the model's robustness and accuracy.

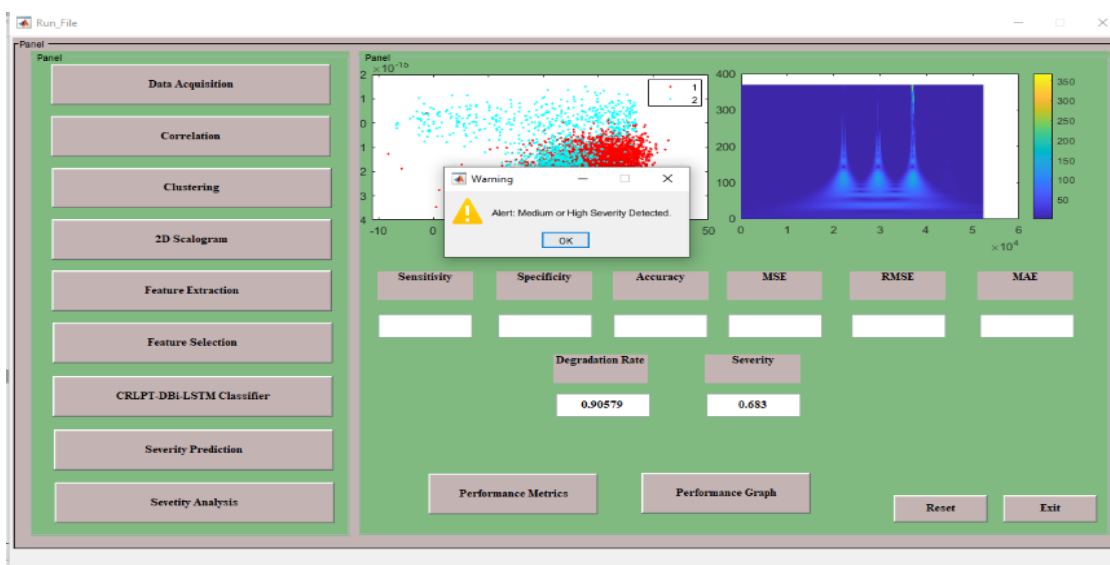


Figure 7. GUI for Fault Detection and Severity Analysis in Solar PV Systems

Figure 7 displays a GUI specifically designed for the purpose of monitoring and analysing fault detection and severity in solar PV systems. The interface seamlessly integrates various components and functionalities to provide a comprehensive overview of the system's performance and fault management. The system will present the user with the calculated degradation rate and severity level of any detected faults. The values provided are used to evaluate the effect of faults on the performance of the PV system. The GUI offers a comprehensive platform that enables the real-time monitoring and analysis of solar PV systems. The system integrates data acquisition, processing, and visualization tools to enable streamlined fault detection and severity analysis. The alert message system improves the

user's capacity to promptly address critical faults, thereby enhancing the reliability and safety of the PV system. In the context of fault detection and severity analysis, it provides a measure of how quickly the system's performance is degrading due to specific faults. A higher degradation rate indicates a more severe fault, necessitating immediate attention to prevent further deterioration and potential system failure.

### 5. Conclusion

This paper introduces a novel method for fault classification and severity analysis in solar PV systems. The approach entails using a context rank layer with a parameterized Tanh-based deep bidirectional long short-term memory (CRLPT-DBi-LSTM) classifier. Advanced techniques like Sorted Bisection Farthest First Clustering (SBFFC) for data clustering and Bayes Probability Function-based Coot Bird Optimization (BPF-CBO) for feature selection are used in the proposed method to make fault detection much more accurate and reliable. The application of multivariate imputation (MVI) and the use of a 2D scalogram for amplitude visualization guarantee high-quality data preprocessing. The covariance-based fuzzy inference system (CFIS) further refines the severity analysis, providing comprehensive insights into the urgency and impact of the fault. The tests show that the suggested CRLPT-DBi-LSTM model does better than other classifiers in terms of its sensitivity, specificity, accuracy, precision, recall, F-measure, negative predictive value (NPV), and Matthews correlation coefficient (MCC). This research's GUI integrates a variety of performance metrics and visualization tools to enable real-time monitoring and analysis. The calculated degradation rate and severity levels guarantee the operational efficiency and safety of solar PV systems, thereby facilitating proactive maintenance. The proposed methodology provides a comprehensive and robust solution for fault management in solar PV systems, thereby enhancing the sustainability and reliability of renewable energy sources. Future work will further enhance the model's adaptability to a variety of environmental conditions, focusing on its potential application in larger-scale PV installations.

### References

- [1] S. R. Madeti and S. N. Singh, "A comprehensive study on different types of faults and detection techniques for solar photovoltaic system," *Solar Energy*, vol. 158, pp. 161–185, Dec. 2017, doi: 10.1016/j.solener.2017.08.069.
- [2] P. Rajput, M. Malvoni, N. M. Kumar, O. S. Sastry, and G. N. Tiwari, "Risk priority number for understanding the severity of photovoltaic failure modes and their impacts on performance degradation," *Case Studies in Thermal Engineering*, vol. 16, p. 100563, Dec. 2019, doi: 10.1016/j.csite.2019.100563.
- [3] G. R. Venkatakrisnan *et al.*, "Detection, location, and diagnosis of different faults in large solar PV system—a review," *The International Journal of Low Carbon Technologies*, vol. 18, pp. 659–674, Jan. 2023, doi: 10.1093/ijlct/ctad018.

- [4] J. Khan and M. H. Arsalan, "Solar power technologies for sustainable electricity generation – A review," *Renewable & Sustainable Energy Reviews*, vol. 55, pp. 414–425, Mar. 2016, doi: 10.1016/j.rser.2015.10.135.
- [5] S. Gul, A. U. Haq, M. Jalal, A. Anjum, and I. U. Khalil, "A Unified Approach for Analysis of Faults in Different Configurations of PV Arrays and Its Impact on Power Grid," *Energies*, vol. 13, no. 1, p. 156, Dec. 2019, doi: 10.3390/en13010156.
- [6] D. S. Pillai, F. Blaabjerg, and N. Rajasekar, "A Comparative Evaluation of Advanced Fault Detection Approaches for PV Systems," *IEEE Journal of Photovoltaics*, vol. 9, no. 2, pp. 513–527, Mar. 2019, doi: 10.1109/jphotov.2019.2892189.
- [7] A. Dhoke, R. Sharma, and T. K. Saha, "A technique for fault detection, identification and location in solar photovoltaic systems," *Solar Energy*, vol. 206, pp. 864–874, Aug. 2020, doi: 10.1016/j.solener.2020.06.019.
- [8] Y. Chaibi, M. Malvoni, A. Chouder, M. Boussetta, and M. Salhi, "Simple and efficient approach to detect and diagnose electrical faults and partial shading in photovoltaic systems," *Energy Conversion and Management*, vol. 196, pp. 330–343, Sep. 2019, doi: 10.1016/j.enconman.2019.05.086.
- [9] M. M. Rahman, I. Khan, and K. Alameh, "Potential measurement techniques for photovoltaic module failure diagnosis: A review," *Renewable & Sustainable Energy Reviews*, vol. 151, p. 111532, Nov. 2021, doi: 10.1016/j.rser.2021.111532.
- [10] A. Abubakar, C. F. M. Almeida, and M. Gemignani, "Review of Artificial Intelligence-Based Failure Detection and Diagnosis Methods for Solar Photovoltaic Systems," *Machines*, vol. 9, no. 12, p. 328, Dec. 2021, doi: 10.3390/machines9120328.
- [11] M. Gökgöz, Ş. Sağlam, and B. Oral, "Investigation of Failures during Commissioning and Operation in Photovoltaic Power Systems," *Applied Sciences*, vol. 14, no. 5, p. 2083, Mar. 2024, doi: 10.3390/app14052083.
- [12] Y. B. Hassan, M. Orabi, and M. A. Gaafar, "Failures causes analysis of grid-tie photovoltaic inverters based on faults signatures analysis (FCA-B-FSA)," *Solar Energy*, vol. 262, p. 111831, Sep. 2023, doi: 10.1016/j.solener.2023.111831.
- [13] V. Veerasamy *et al.*, "LSTM Recurrent Neural Network Classifier for High Impedance Fault Detection in Solar PV Integrated Power System," *IEEE Access*, vol. 9, pp. 32672–32687, Jan. 2021, doi: 10.1109/access.2021.3060800.
- [14] F. Aziz, A. U. Haq, S. Ahmad, Y. Mahmoud, M. Jalal, and U. Ali, "A Novel Convolutional Neural Network-Based Approach for Fault Classification in Photovoltaic Arrays," *IEEE Access*, vol. 8, pp. 41889–41904, Jan. 2020, doi: 10.1109/access.2020.2977116.
- [15] G. Di Lorenzo, R. Araneo, M. Mitolo, A. Niccolai, and F. Grimaccia, "Review of O&M Practices in PV Plants: Failures, Solutions, Remote Control, and

- Monitoring Tools,” *IEEE Journal of Photovoltaics*, vol. 10, no. 4, pp. 914–926, Jul. 2020, doi: 10.1109/jphotov.2020.2994531.
- [16] B. Li, C. Delpha, D. Diallo, and A. Migan-Dubois, “Application of Artificial Neural Networks to photovoltaic fault detection and diagnosis: A review,” *Renewable & Sustainable Energy Reviews*, vol. 138, p. 110512, Mar. 2021, doi: 10.1016/j.rser.2020.110512.
- [17] S. A. Memon, Q. Javed, W.-G. Kim, Z. Mahmood, U. Khan, and M. Shahzad, “A Machine-Learning-Based Robust Classification Method for PV Panel Faults,” *Sensors*, vol. 22, no. 21, p. 8515, Nov. 2022, doi: 10.3390/s22218515.
- [18] S. A. Zaki, H. Zhu, M. A. Fakhri, A. R. Sayed, and J. Yao, “Deep-learning-based method for faults classification of PV system,” *IET Renewable Power Generation*, vol. 15, no. 1, pp. 193–205, Jan. 2021, doi: 10.1049/rpg2.12016.
- [19] F. Harrou, A. Saidi, Y. Sun, and S. Khadraoui, “Monitoring of Photovoltaic Systems Using Improved Kernel-Based Learning Schemes,” *IEEE Journal of Photovoltaics*, vol. 11, no. 3, pp. 806–818, May 2021, doi: 10.1109/jphotov.2021.3057169.
- [20] A. Mellit, G. M. Tina, and S. A. Kalogirou, “Fault detection and diagnosis methods for photovoltaic systems: A review,” *Renewable & Sustainable Energy Reviews*, vol. 91, pp. 1–17, Aug. 2018, doi: 10.1016/j.rser.2018.03.062.
- [21] V. S. B. Kurukuru, F. Blaabjerg, M. A. Khan, and A. Haque, “A Novel Fault Classification Approach for Photovoltaic Systems,” *Energies*, vol. 13, no. 2, p. 308, Jan. 2020, doi: 10.3390/en13020308.
- [22] Arabshahi, H. Torkaman, and A. Keyhani, “A method for hybrid extraction of single-diode model parameters of photovoltaics,” *Renewable Energy*, vol. 158, pp. 236–252, Oct. 2020, doi: 10.1016/j.renene.2020.05.035.
- [23] J. Wang, J. Zhou, and X. Chen, *Data-Driven Fault Detection and Reasoning for Industrial Monitoring*. 2022. doi: 10.1007/978-981-16-8044-1.
- [24] F. Calabrese, A. Regattieri, M. Bortolini, and F. G. Galizia, “Data-Driven Fault Detection and Diagnosis: Challenges and Opportunities in Real-World Scenarios,” *Applied Sciences*, vol. 12, no. 18, p. 9212, Sep. 2022, doi: 10.3390/app12189212.
- [25] S. R. Madeti and S. N. Singh, “Modeling of PV system based on experimental data for fault detection using kNN method,” *Solar Energy*, vol. 173, pp. 139–151, Oct. 2018, doi: 10.1016/j.solener.2018.07.038.
- [26] M. A. El-Rashidy, “An efficient and portable solar cell defect detection system,” *Neural Computing and Applications*, vol. 34, no. 21, pp. 18497–18509, Jun. 2022, doi: 10.1007/s00521-022-07464-2.
- [27] D. Korkmaz and H. Acikgoz, “An efficient fault classification method in solar photovoltaic modules using transfer learning and multi-scale convolutional neural

- network,” *Engineering Applications of Artificial Intelligence*, vol. 113, p. 104959, Aug. 2022, doi: 10.1016/j.engappai.2022.104959.
- [28] A. F. Amiri, S. Kichou, H. Oudira, A. Chouder, and S. Silvestre, “Fault Detection and Diagnosis of a Photovoltaic System Based on Deep Learning Using the Combination of a Convolutional Neural Network (CNN) and Bidirectional Gated Recurrent Unit (Bi-GRU),” *Sustainability*, vol. 16, no. 3, p. 1012, Jan. 2024, doi: 10.3390/su16031012.
- [29] “Fault Detection Dataset in Photovoltaic Farms,” *Kaggle*, Apr. 17, 2022. <https://www.kaggle.com/datasets/amrezzeldinrashed/fault-detection-dataset-in-photovoltaic-farms?resource=download>
- [30] Clayton-H-Costa, “GitHub - clayton-h-costa/pv\_fault\_dataset,” *GitHub*. [https://github.com/clayton-h-costa/pv\\_fault\\_dataset](https://github.com/clayton-h-costa/pv_fault_dataset)
- [31] J. L. Schafer and M. K. Olsen, “Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst’s Perspective,” *Multivariate Behavioral Research*, vol. 33, no. 4, pp. 545–571, Oct. 1998, doi: 10.1207/s15327906mbr3304\_5.