Article

AC Fault Detection in On-Grid Photovoltaic Systems by Machine Learning Techniques

Muhammet Tahir Guneser [1], Sakir Kuzey [2,*] and Bayram Kose [3]

[1] Department of Electronics and Communication Engineering, Istanbul Technical University, Istanbul 34475, Turkey; guneserm@itu.edu.tr
[2] Department of Electronics and Automation, Kocaeli University, İzmit 41500, Turkey
[3] Department of Electrical and Electronics Engineering, Bakircay University, Izmir 35665, Turkey; bayram.kose@bakircay.edu.tr
* Correspondence: sakir.kuzey@kocaeli.edu.tr

Abstract

The increasing integration of solar energy into the power grid necessitates robust fault detection and diagnosis (FDD) guidelines to ensure energy continuity and optimize the performance of grid-connected photovoltaic (GCPV) systems. This research addresses a gap in the literature by systematically evaluating machine learning (ML) algorithms for the detection and classification of AC-side faults (inverter and grid faults) in GCPV systems. We utilized three commonly employed algorithms, namely K-Nearest Neighbors (KNN), Logistic Regression (LR), and Artificial Neural Networks (ANNs), to develop fault detection models. These models were trained using a monthly electrical dataset obtained from the AYCEM-GES-GCPV power plant in Giresun, Turkiye, and their performance was rigorously evaluated using classification accuracy, Area Under the Curve (AUC), and Receiver Operating Characteristic (ROC) analyses. The results demonstrate that the algorithms are highly effective in fault detection, with AUC values consistently exceeding the critical threshold. The obtained accuracies for KNN, LR, and ANN were 0.9826, 0.782, and 0.7096, respectively. These findings emphasize the high effectiveness of ML algorithms, with KNN exhibiting the best performance, for identifying AC-side faults in GCPV installations. While the study focused on AC-side fault detection, subsequent work developed a smart card module to identify complex DC side electrical faults and built a PV array for experimental testing.

Keywords: AC-side faults; ANN; fault detection; KNN; logistic regression; machine learning

Check for updates

Academic Editors: Shruti Vyas and Mahmoud Dhimish

Received: 4 November 2025
Revised: 25 December 2025
Accepted: 15 January 2026
Published: 30 January 2026

1. Introduction

The share of renewable energy sources (RES) among all energy sources is steadily increasing, due to factors such as increasing cost, safety risks, and carbon emissions of fossil fuels. Solar energy stands out due to its sustainable potential and flexibility of setup in many parts of the world. The rapid increase in photovoltaic (PV) power plants in recent years has brought with it several problems. Researchers are developing fault detection and diagnosis (FDD) guidelines and proposing solutions to ensure energy continuity and improve system performance. Faults occurring in PV power plants interrupt energy production, damage equipment, and pose a fire risk. A PV system must operate at optimum efficiency. Therefore, monitoring systems for grid-connected photovoltaic (GCPV) power plants are vital for ensuring a rapid response in the event of a fault and minimizing power loss. However, because GCPV power plants cover a large area of land, and assuming a large number of

https://doi.org/10.3390/solar6010006

PV strings in the system, monitoring each PV module in each string is costly and adds to system complexity. Therefore, array-based monitoring is typically provided through solar inverters. In this case, a fault can only be detected at the PV string level. Furthermore, imaging systems can aid in determining the frequency of periodic maintenance for PV power plants. FDD in PV energy systems is performed using electrical, thermal, and visual methods. In the last decade, researchers have made significant progress in the field of FDD. In recent years, the use of artificial neural networks (ANNs) and statistics-based algorithms in GCPV power systems has been rapidly increasing among FDD methods. A review of FDD-centered scientific studies on GCPV energy systems reveals that methods vary depending on the type of fault. Faults resulting from structural changes, such as yellowing, cracking, discoloration, and cell fracture, are detected visually, while electrical faults that halt system operation or reduce efficiency despite no structural change can be detected using statistical or machine learning (ML)-based techniques.

Related Literature

PV systems are highly dependent on nonlinear, time-varying properties and environmental factors [1]. This complicates parameter extraction, modeling, power estimation, and FDD analysis in PV systems. The unique properties of the K-Nearest Neighbors Method (KNN) make it preferred in parameter extraction, power estimation, and FDD analysis studies in PV energy systems. KNN is widely used among ML techniques due to its simplicity, flexibility, and lack of prior assumptions about the data distribution [2–4]. The KNN method is a nonparametric, lazy learning-based algorithm [3,5]. It does not require any training before making new predictions [3,6]. This feature increases the applicability of the KNN method in different fields and also makes it a simple method that is easy to implement. Since KNN is a nonparametric technique, it can be used in the analysis of complex and nonlinear properties. Due to the high dependence of the output power of PV systems on climatic conditions, it is difficult to determine predefined thresholds for protection devices. A learning technique such as KNN can determine its own threshold values based on observed data [7]. KNN bases the classification rule on a specified similarity function between the training and test samples [8]. In a study where the KNN algorithm was proposed, it was used to detect open-circuit, line-to-line, and partial shading faults, and achieved an average classification accuracy of 98.70% [7]. A modified version of it, called Fuzzy KNN (fKNN), increased the model accuracy from 96.1% to 98.4% and the classification accuracy from 91.8% to 99.4% [5]. In another study, a KNN model developed for grid-connected systems achieved an F1 score of 99.999995% [1]. KNN is well-suited for working with small data samples, especially in cases where the probability of failure is low and therefore the amount of available fault data is small, because it offers high computational speed and accuracy [9].

Logistic Regression (LR) is preferred due to its simplicity, fast response time, and linear classification capabilities [10]. Furthermore, the low complexity of LR has been emphasized in addressing the complexity of systems based on traditional neural network principles [11,12]. While traditionally suitable for interpreting linear relationships, in the context of multi-class classification, its simplicity and efficiency offer a practical and effective option for interpreting the contribution of each variable to the probability of a specific fault occurrence [13]. The LR classifier′s greater robustness in detecting relationships between data samples compared to statistical models allows for the ability to easily incorporate new training data into the model and adjust the classification threshold to find confidence classes [11]. The LR method is particularly suitable for detecting arc faults on the DC side of PV energy systems, as arc fault diagnosis falls into only two classes. Furthermore, LR must provide the probability of a fault occurring, as additional actions can be decided based on the magnitude of this probability [14]. LR is computationally efficient and fast. It has also been stated that LR is a lightweight model requiring low memory and can be calculated very quickly [15]. Despite being a simple linear model, LR has been observed to perform excellently in the fault detection (line-to-line-LL faults vs. no fault) scenario with 99.9649% accuracy, proving that the data has some degree of linear separability [12].

ANN is an effective data-driven and adaptive method for analyzing complex and nonlinear problems [16–18]. It learns from experience throughout the process and can generalize previously observed behavior to adapt to the current new situation. This method is suitable for solving problems where explicit knowledge is difficult to define, but large amounts of data are available [19]. It simplifies the complex processes of previous methods through a simple implementation based on training models. ANN has flexibility and reliability in fault detection and classification [20]. The rapid decision-making ability of ANNs with a well-trained model enables high-precision real-time health monitoring [21]. Indeed, many studies have shown that ANNs provide effective results in fault detection and classification. In a study conducted in this field, the ANN method achieved 100% accuracy in tests on both simulated and real systems [22]. A study presenting a multilayer neural network approach classified PV faults with 99.6% accuracy with a fast computation time of 0.08 s [23]. The ANN method was used to detect short-circuit, open-circuit, and bypass faults, providing an average sensitivity of 92.6% and accuracy of 96.4% [20]. In one study, the average diagnostic accuracy for different PV faults ranged from 95% to 98% [24].

ML algorithms have their own advantages and disadvantages. To suppress these disadvantages, other algorithms are incorporated into the process to create a hybrid algorithm configuration (ensemble learning—EL). Here, the goal is to achieve better performance on the observed parameters. For example, ANN can be used to extract features from the data and reduce the dimensionality, while another algorithm, such as SVM, is used for classification [19–21,25]. In a study to detect PV energy system faults, a Random Subspace (RS) ensemble classifier based on the KNN algorithm was proposed [26]. Analysis under different scenarios was used to predict the classification label using Random Tree (RT), KNN, and LR classifiers. An EL algorithm combining Support Vector Machine (SVM), LR, and KNN algorithms was used to classify open-circuit (OC) and LL faults [27]. Only the PV string's current and voltage were used to extract the features of the analyzed faults. When the studies are reviewed, different EL-based methods, such as KNN, LR, and ANN, have been used in PV system FDD, prediction, and classification, or EL methods performed by combining different algorithms and have been compared with ML-based methods such as KNN, LR, and ANN [3,25,28,29]. In an electrical fault detection study using a sixteen-day dataset from a real GCPV energy system, Decision Tree (DT), Naive Bayes (NB), KNN, and LR supervised learning techniques have been applied to detect mismatch, shading, OC, and short-circuit (SC) faults [30]. The best fault prediction performance in the study was obtained with the KNN method, with 99.2% accuracy and 99.7% area under the curve–receiver operating characteristic (AUC-ROC). In a remarkable study using only the solar radiation and output power data of two different PV systems as input parameters of the ANN algorithm, a comparative analysis of the effect of partial shading fault on the PV system has been performed [19]. In an article presented in the field of prediction and classification of PV system faults, a deep learning-based transformer model based on the rate of change of solar cell parameters was proposed. Unlike other prediction techniques, this model does not rely on previous trends and uses a proactive process [31]. The prediction and classification performance of the proposed deep learning method was compared with ML-based classification techniques such as KNN, ANN, and SVM. In another study, the results of a new ML classifier using an extra-tree cluster ensemble algorithm were compared with six classifiers, including LR and KNN algorithms, to detect faults on the DC side of the PV system [20]. In a study conducted to detect inverter faults in the GCPV power system, dimensionality reduction and ML techniques were used [15]. Principal component analysis (PCA) and autoencoder-based inference techniques were used in the dimensionality reduction process. The performances of four different ML classifiers, including KNN and LR, were compared with respect to their speed and computational requirements.

In recent years, the application of ML algorithms in FDD analysis of GCPV energy systems has increased the success rate in fault detection and significantly improved performance prediction and system reliability. This has brought ML techniques into focus in the FDD field. Such approaches are typically extensible models that require a systematic workflow, starting with data collection, followed by preprocessing, training, feature extraction, model definition, training, and evaluation. These models are evaluated based on their output using standard performance indices such as F1 score, recall, precision, and accuracy. Any serious fault on the AC side of GCPV energy systems can be isolated using power fuses, overcurrent protection, and earth leakage protection devices [32]. However, isolation serves the purpose of protecting the system rather than improving system performance. Rapid detection of AC-side faults will prevent long-term power outages, enhance system reliability, and ensure the sustainable transmission of energy.

This study was conducted to contribute to the field of PV system FDD and forecasting by subjecting both electrical and meteorological data from a real GCPV power plant to a specific training process. This 30-day dataset consists of data available in 5 min periods from a real GCPV power plant. The meteorological data from the PV plant were obtained from the solar inverters currently in use. The temperature data obtained from the solar inverters do not reflect ambient temperature. Due to the lack of a weather station at the GCPV power plant, we requested comprehensive meteorological data from the Giresun Provincial Directorate of Meteorology. Since these meteorological data consist of one-hour periods, they do not align with the five-digit periodic dataset obtained from the solar inverters. Therefore, a Python program was developed to generate temperature data from one-hour temperature data in five-minute system periods. The training process to complement the Python program reduced the GCPV power system's cellular outputs and meteorological impact. This study, which aims to detect and predict faults in the AC side of a PV system, presents an ML-based methodology that comparatively analyzes the fault detection and prediction results of KNN, LR, and ANN methods. The procedural steps of the ML procedures used in the study are illustrated in the distribution updates presented in Figure 1. The methodology follows the results of a study that began with the details and preprocessing of electrical data to eliminate outliers and ensure data integrity. This study, which provides an ML-focused comparative framework, employed six different performance indices, namely accuracy, precision, recall, F1 score, AUC, and optimum threshold, allowing us to analyze the accuracy and reliability of AC-side fault detection and prediction results. Furthermore, this study will be expanded in the future, and the infrastructure for DC-side fault detection has been prepared. Therefore, an electronic smart module was developed to detect DC-side faults at the PV module level. This smart module will measure electrical data at the PV module and array levels, along with temperature and humidity data from both the ambient and smart modules, and store the data wirelessly via an SD card. To test the smart module, a 140 Wp PV array was constructed, consisting of fourteen 10 Wp PV panels.

A summary of the core contributions of this study is provided below.

- AC-side faults in PV power systems were classified/regressed using only current, voltage, and power signals without the need for additional mapping or transformation to extract discriminative features.

- Our research revealed a lack of research on AC-side faults in GCPV systems. The literature contributes to this issue.
- Electrical fault detection on the grid side of the GCPV power system was successfully achieved.
- The comparative effectiveness of KNN, LR, and ANN methods in PV fault detection was systematically evaluated.
- The data used in the study were obtained from a real GCPV power plant that had not been studied previously.
- A smart module was made at the PV panel level.
- Potential economic benefits can be realized through increased system efficiency.

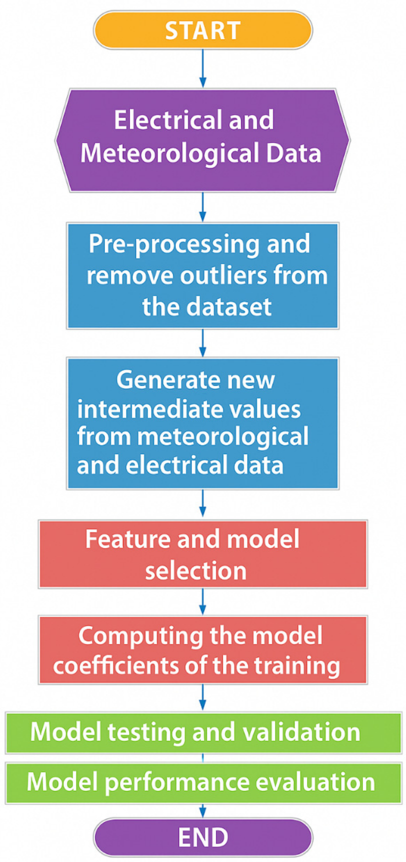
START
Electrical and Meteorological Data
Pre-processing and remove outliers from the dataset
Generate new intermediate values from meteorological and electrical data
Feature and model selection
Computing the model coefficients of the training
Model testing and validation
Model performance evaluation
END

Figure 1. Flowchart of the research progress.

2. Method Analysis in Bibliometrics

In this section, a narrow bibliometric review of AC side FDD methods of PV energy systems is presented. A bibliometric review provides a macro-level statistical analysis of scientific articles published in the field of PV power systems. The bibliometric review conducted in a field creates a dataset about sources, keywords, authors, publishers, and countries related to the field from databases such as Scopus and Web of Science (WoS). Here, a dataset was created based on keywords such as "photovoltaic" and "fault" and keywords including PV AC-side faults, estimation, and FDD methods between 2014 and 2024. The figures in this section were organized based on the Biblioshiny R 4.5.0 package by merging BibTeX files taken separately from the Scopus and WoS library databases.

When literature data were analyzed, we found that studies in the field of PV system FDD analysis have increased rapidly in recent years. The keyword analysis includes the title, keywords, and abstract related to the field of PV energy systems. Figure 2 shows the frequency of the words over time between 2014 and 2024. The frequency of use of the words “fault detection” and “machine learning” is striking in the figure. The abstract word cloud corresponding to the annual growth of the words is shown in Figure 3.

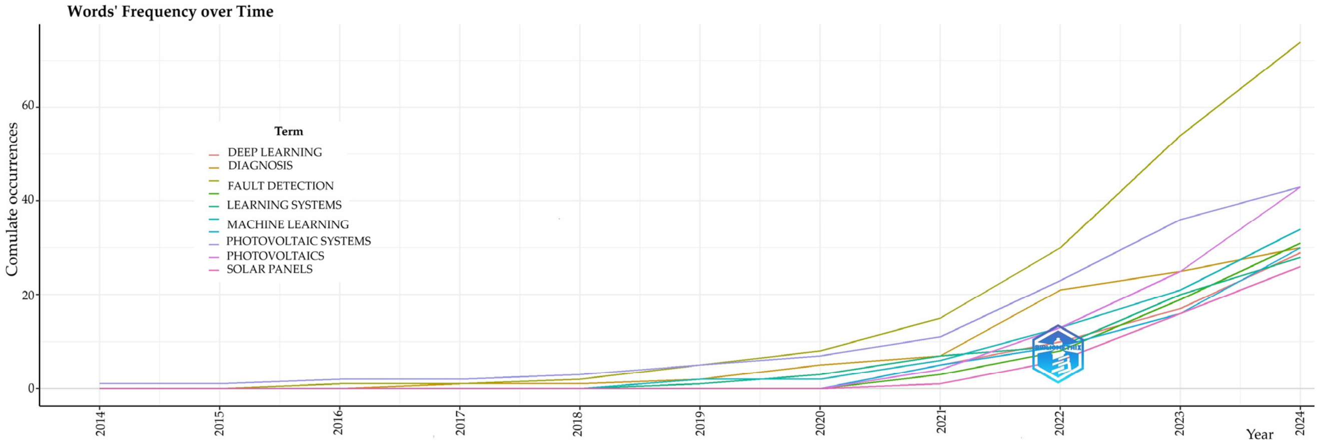

Figure 2. Frequency of keywords by year.

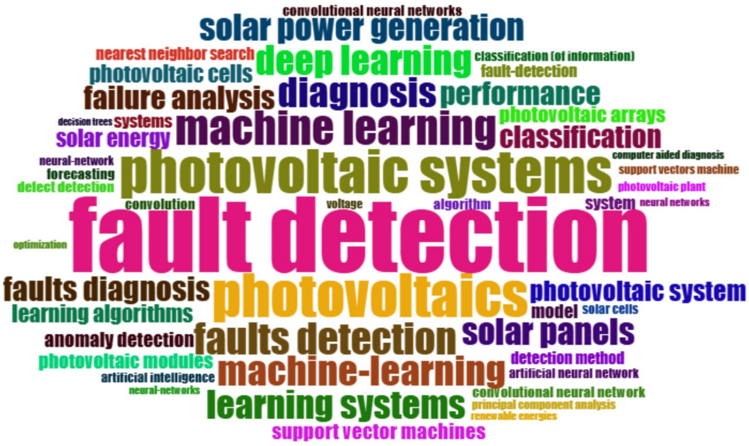

Figure 3. The most frequent words.

When the bibliometric dataset was examined, it was concluded that the focus of the topic of fault detection and detection algorithm structures in PV energy systems in the keyword map was ML and statistical approaches, which had a significant proportion in the literature.

3. Classifications of PV Faults

Faults can occur temporarily or permanently on both the AC and DC sides of the GCPV energy system. Permanent faults occurring in the strings on the DC side of the PV system have a devastating effect on the system’s lifespan and efficiency [33]. In PV power systems, the part from the PV modules where DC power is produced to the inverter constitutes the DC side, while the part starting from the inverter to the distribution grid constitutes the AC side. Faults occurring on the distribution side of a grid-connected or standalone PV system only affect the AC side. It is expected that the PV modules in the DC energy generation stage will operate under conditions that maximize their power output. If the fault is located on the DC side of the PV system, analytical approaches using microscopic analysis may be necessary to determine the cause. SEM, ATIR, and X-ray techniques can be cited as among the most important and newest microanalysis methods in the literature [34]. Fault detection and classification techniques can be categorized according to the signal structure obtained [35,36]. Fault detection techniques used in PV energy systems can be listed as visual detection (deletion, yellowing, cracking, discoloration, oxidation, and cell breakage), infrared thermography analysis (connector corrosion, hot spots, snail trails, and microcracks), electrical characterization (electroluminescence, photoluminescence, and UV fluorescence), and STD (signal transmission device). In recent years, UAV applications and smart fault diagnosis methods based on intelligent monitoring have been increasingly employed. An examination of the literature shows that many fault classifications have been made based on different characteristics of the PV system and are presented schematically [33,37–39].

3.1. DC-Side Faults

DC-side PV arrays consist of DC-DC converters where a certain duty-cycle value is applied depending on MPPT. The DC side terminates at the inverter, where DC-to-AC energy conversion takes place. While some faults directly affect PV cells, others directly impact the PV system. These faults, which cause performance loss, also pose a risk of fire and loss of life. DC-side faults consist of open-circuit faults, short-circuit faults, grounding faults, mismatch faults, junction box faults, arc faults, and MPPT faults. Since our study focuses on the AC-side faults, DC-side faults are not explained in detail.

3.2. AC-Side Faults

The energy production process and components, starting from the inverter to the distribution network, constitute the AC stage. Faults occurring on the distribution side of a PV system affect only the AC side and can be easily detected and diagnosed by employing the protection standards specified in [40–42]. Therefore, they can be detected, diagnosed, and isolated more easily than DC-side faults. AC-side faults consist of inverter faults and network abnormalities.

A solar inverter is a high-efficiency device that converts DC energy into AC energy. Since many strings are connected to the inverter in a PV power system, a fault in the inverter can cause significant power loss. The MPPT inputs of the solar inverter are grouped according to their current-carrying capacity. The strings should be connected to the inverter in these groups. Technical personnel should be prepared for various situations, such as inverter disconnection and fault events.

There can be two types of faults on the grid side of the GCPV system: grid outage and total power outage. In both cases, a significant decrease in power production or even a halt in production is expected.

4. Methodologies

4.1. K-Nearest Neighbors Method

The KNN approach is a statistical classification technique that determines the class label of a new instance by analyzing its distances to existing training samples in the feature space and assigning it according to the majority class of its nearest neighbors [7]. KNN is one of the easiest-to-implement and most widely used supervised learning algorithms. Although it is used in solving both classification and regression problems, it is generally preferred for classification problems [43]. KNN is a nonparametric algorithm; therefore, it does not make any explicit assumptions about a specific functional form that defines the relationship with the data. In addition, it is directly based on the training data. The model predicts outcomes as a function of the classes or values identified from the nearest neighbors using distance-based similarity. This method makes decisions based on the similarity between examples, and classification is usually performed using distance measurements. The model assigns a class label by evaluating the similarity among the training instances [44]. The KNN algorithm works based on a set of examples whose class labels are known and calculates the distance of a data point to the existing examples to determine its class. The class label of the new data is determined based on many of the class labels of its nearest k neighbors. These distances are usually calculated with Euclidean, Manhattan, or Minkowski distance functions; the relevant formulas are given in Equations (1)–(3) [45].

$$Euclidean\ Distance = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)}\ i \neq j : 1, 2, \ldots, n \tag{1}$$

$$Manhattan\ Distance = \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|\ i \neq j : 1, 2, \ldots, n \tag{2}$$

$$Minkowski\ Distance = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)^q\right)^{1/q}\ i \neq j : 1, 2, \ldots, n \tag{3}$$

The algorithm consists of five basic steps:

- First, the value of k is determined.
- The distances between the target object and other objects are calculated; Euclidean distance is usually used, but Manhattan or Minkowski distances can also be preferred.
- The calculated distances are sorted, and the nearest neighbors are selected according to the smallest distances.

The categories of the selected nearest neighbors are combined.

4.2. Logistic Regression Analysis

LR analysis is a descriptive, inferential statistical method that establishes the relationship between the independent variables and the target variable, predicts the possible outcomes of the target variable, and achieves optimal fit with the fewest number of variables [46–48]. LR is relatively simple to understand and implement compared to other ML methods. In LR, the relationship between the independent variables and the target variable is assumed to be linear, but this may not always be the case. LR can be sensitive to outliers, which can affect the predicted probabilities and class labels [49]. LR performs poorly when there are multiple or nonlinear decision boundaries. Naturally, they are not flexible enough to capture more complex relationships [50]. LR, a simple supervised ML technique, is generally used to solve binary classification problems. In this study, the LR classification technique with multiple nominal features is used to solve multiple classification problems. As can be seen from the S-shaped curve shown in Figure 4, estimates may have values that are not within the limits between 0 and 1. In Figure 4, x_0 is the midpoint of the x input value, and L is the maximum value that can be taken at the output. Since the limit values are linear, the classes can be separated linearly. The mathematical expression used in LR analysis is given in Equations (4) and (5) [51].

$$\hat{y_i} = 1/(1 + \exp(-z)) \tag{4}$$

$$z = b + w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_px_p \tag{5}$$

In this equation, the only input expression is x, while the estimated output is y. When the result of the linear regression equation z is written to the sigmoid function, the sigmoid function produces a probability value between 0 and 1, and a class assignment is made as 1 or 0 according to this probability value. In the classification problem, the probability of belonging to class 1 is always calculated. The problem is solved by finding the weights that can minimize the *Log Loss* value given in Equation (6) [52] regarding the differences between the real value and the estimated values.

$$Log\ Loss = \frac{1}{m}\left(\sum_{i=0}^{m} -y_i \log\left(p\left(\hat{y_i}\right)\right) - (1 - y_i)\log\left(1 - p\left(\hat{y_i}\right)\right)\right) \tag{6}$$

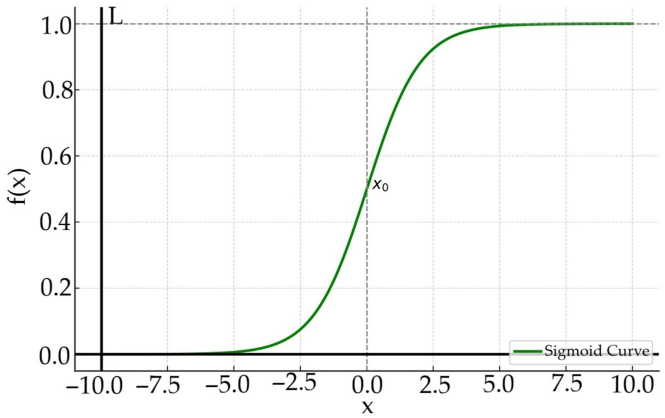

Figure 4. The Sigmoid function for LR.

In classical regression methods, when the dependent variable consists of 0 and 1, the obtained values can't be between 0 and 1 in every business problem. The sigmoid function comes into play exactly at this stage and maps the values of the z linear form to values between 0 and 1. In summary, the values obtained from the mapping using the Sigmoid function, based on the independent variables xi, correspond to the probability of the first class of the dependent variable occurring. In the method, values lower than the threshold value are accepted as 0. This transformation process in LR forms the basis of artificial and deep learning classification problems [51]. The model creation strategy can be adapted depending on the field of study. It is important to determine the variables that should be included in the model during the model creation process. Variable selection becomes important in cases where there are three or more independent variables [53]. There may be more than one model that shows good compatibility in a field of study where LR analysis is applied. Whether the model fits the data within the framework of scientific values should be determined and interpreted with various statistical methods. LR combined with cross-validation has achieved 97.11% accuracy in detecting faults in the DC components of PV systems, including OC, SC, and mismatch faults [52]. While commonly applied to electrical feature analysis, LR is less utilized in thermal imaging. Its suitability for binary and multiclass classification provides an advantage over other AI methods in processing image-based PV fault data. This study used the LR algorithm to detect and predict AC-side faults in a GCPV power plant.

4.3. *Artificial Neural Network Method*

ANNs are a computational approach that can generate meaningful results in the output layer by processing trained input data through interconnected layers of neurons. They model the neural processes of the human brain to recognize and analyze patterns, generalize knowledge, and derive new meaning from objective and quantitative data through comparison with prior knowledge. Their adaptability over time, ability to process complex and large datasets with high accuracy, and dynamic nature enable them to successfully perform functions such as classification and prediction. The model trained with the preprocessing of past data is a statistical ML-based structure used to detect, classify, and predict failures [54]. It can adapt to many work areas. However, the high computational requirements for large-scale datasets and their closed-loop nature, which make them less interpretable than other methods, limit the ANN approach. The effectiveness of the proposed approach is directly influenced by the quality of the dataset employed during training. The critical diagnostic accuracy rate assumed for ANN approaches to be successful in FDD analysis is over 90% [16]. In GCPV energy systems, electrical parameters at the input and output are recorded with internal or external imaging systems in 5 or 10 min intervals.

Solar inverter data in a PV power system is typically used as a reference. Depending on the electrical parameters to be controlled, they can also be recorded with a specialized imaging system. The short data acquisition period means that a large-scale dataset will be used in the input layer of the ANN model to be created. The algorithms processing this data are trained to produce output parameters derived from the input data for the purpose of detecting potential faults in the GCPV energy system. The proposed ANN model processes data in a closed-loop manner through its hidden layers, generating outputs for FDD, classification, and prediction. The neural network employed to identify AC faults in the GCPV energy system consists of a multilayer feedforward architecture, as shown in Figure 5. The structure consists of six neurons, consisting of the currents and voltages obtained from the solar inverter over the course of a month. The hidden layer, consisting of an interconnected processing network that defines the relationship between input and output, has a structure of 10 neurons and consists of an output layer. After eliminating outlier data in the input layer, the trained dataset was used. Training was terminated when the mean square error (MSE) reached 1×10^{-4} or lower during neural network training.

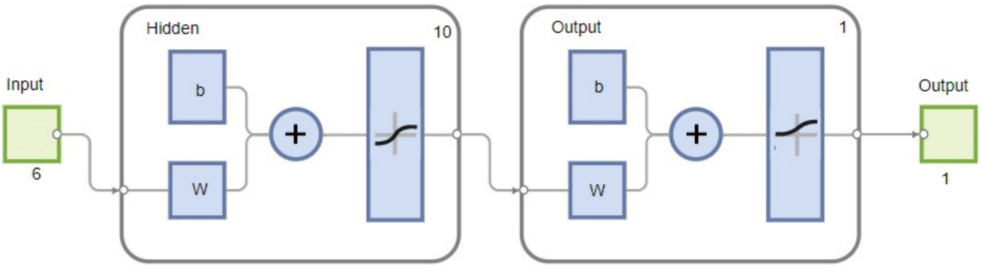

Figure 5. Proposed ANN training.

As shown in the flowchart in Figure 6, the model training process began with processing electrical parameters obtained from PV systems, independent of meteorological inputs. In the data preprocessing step, erroneous, incomplete, or outlier records were removed to create a suitable structure for analysis. Feature extraction was performed on this structure, and then classification models were trained using ML methods, including KNN, LR, and ANN. Model performance was assessed through the evaluation phase outlined in the flowchart, and the methods were compared at the end of the process.

In our study, all data preprocessing, application, training, cross-validation, and testing of ML algorithms (KNN, LR, and ANN) were performed using statistical and ML libraries related to Microsoft Excel and the Python programming language. A Python program was developed to align hourly meteorological data (temperature) with 5 min electrical data, which is one of the critical preprocessing steps of the article. This program provided inferences compatible with the data obtained in 5 min periods. The obtained 5 min temperature data were then made ready for DC-side fault diagnosis. Analyses for the training and comparative evaluation of the ML algorithms were conducted using real field electrical data obtained from the AYCEM-GES-GCPV power plant in Giresun province, Turkiye. A dataset of 5566 rows recorded in 5 min periods from a solar inverter (GoodWe GW50K-MT model 65 kW) for May 2021 was used. Independent variables: The independent variables of the ML algorithms are the phase voltages and phase currents taken from the solar inverter. The dependent variable is the mode status. This state is designated as Mode 0 for normal/standby and Mode 1 for fault conditions. The dataset for experimental ML methods is 70% for training to capture the relationships between the input parameters and the output power. 15% for cross-validation to prevent overfitting and optimize hyperparameters. 15% for testing to evaluate the generalization ability of the network on the test data. The proposed ANN model has a multilayer feedforward architecture. The model consists of 6 input neurons, a hidden layer of 10 neurons, and an output layer. Training was terminated when the mean square error (MSE) reached 1×10^{-4} or less.

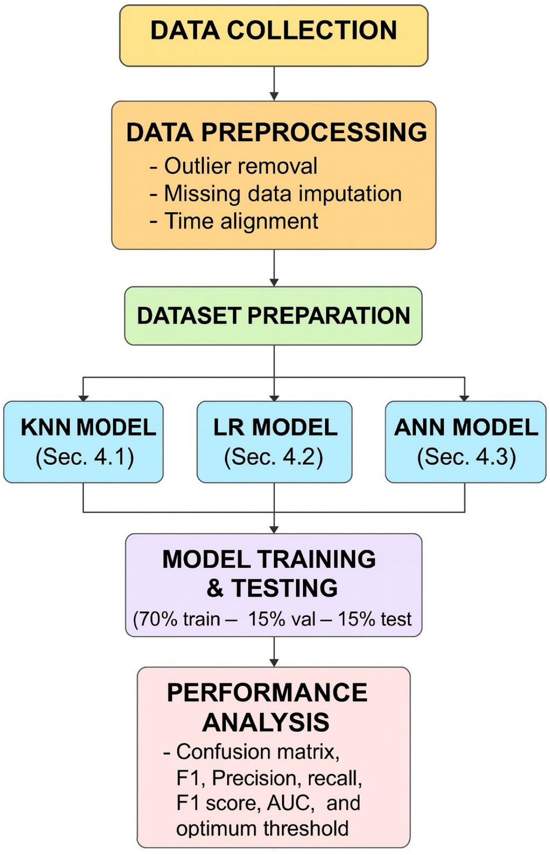

Figure 6. Methodology flowchart.

5. Comparison of ML Approaches

When analyzing ML algorithms, each algorithm may have superior characteristics depending on its application, but it may also have inadequate characteristics. Therefore, it is important to use ML algorithms correctly in a suitable field. Table 1 categorizes the strengths and weaknesses of ML methods based on specific characteristics, according to references. Because the ANN algorithm works by experimenting with the input data, it requires a very large amount of training data. Despite the high training time, it operates quickly after training. The ANN algorithm is inherently more complex and susceptible to noise. However, it is a frequently preferred method due to its very high accuracy rate. The low data training requirement of the LR algorithm also shortens the training time. The LR algorithm is very fast, has a very simple structure, and is less sensitive to noise. Except for binary classification, its accuracy is not very high. The KNN algorithm, with its lazy learning feature, requires very little data training. This reduces training time. The KNN algorithm is particularly slow on large datasets. Its accuracy rate is high on small datasets. It has a very simple structure.

Table 1. Feature comparison of ML algorithms.

Method	Ref.	Data Training Requirement	Duration of Training	Working Speed	Accuracy Rate	Simplicity	Sensitivity to Noise
ETC	[20]	Middle-High	Very Low	Fast	Very High	Middle	Low
RF	[3,15,20,30,55,56]	Middle-High	Middle-High	Middle-Fast	Very High	Middle	Low
DT	[3,12,15,20,57]	Low-Middle	Very Low	Very Fast	Middle-High	Very High	High
AdaBoost	[5,12]	Middle-High	Middle-High	Middle-Low	Low-Middle	Middle	Middle/High
SVM	[3,13,18,20,56,58]	Middle	Middle-High	Middle	High	Middle	Low-Middle
KNN	[1,3,9,15,20,56,59–61]	Low (lazy learning)	No	Low-Middle (slow if data is large)	High (small data)	Very High	Low-Middle
LR	[10,11,15,20,55,59]	Low	Very Low	Very Fast	Middle-High (dual classes)	Very High	Low
ANN	[13,16,18,39,59,62]	Very High	High	Fast (after training)	Vey High	Low	Middle/High

Despite the many strengths of ANN compared to other ML methods, literature analysis also highlights some challenges and limitations encountered in practical applications of this method. Developing ANN models requires more computational resources, especially when using backpropagation (BP) algorithms and when hyperparameter optimization is required [1,59]. Determining many aspects, such as the model structure (number of layers, number of neurons), learning algorithm, loss function, and activation function, is a difficult and tedious task [16]. ANN models are generally data-driven and require a large amount of data to perform effectively. MLP-type networks require high-quality labeled data that describes the process very well [63]. Models such as ANN and ELM are typically considered "black-box" models because they lack interpretability [59]. While ANN generally performs well, in some cases, EL methods such as Enhanced Trees (ETC), RF, KNN, or XGBoost can achieve higher accuracy [20,57]. In summary, ANN is a preferred tool in PV FDD systems due to its robust ability to model nonlinear system behavior and its high accuracy results. However, the complex training process, high computational requirements, and the need for large, high-quality data are key limitations that must be considered during implementation. In this study, the data we obtained from a real GCPV power plant are unique. The large size of the data we will analyze and the resulting need for the highest accuracy in classification led us to choose the ANN method.

The KNN method is less sensitive to outliers than other methods [2]. This feature facilitates classification with fewer outliers, especially in cases where the illumination differs from the sensor readings due to environmental effects such as shading [7]. The KNN method is capable of distinguishing normal and abnormal features [4]. Despite the simplicity and flexibility of the KNN algorithm, the literature indicates serious limitations that should be taken into account in PV FDD applications [58]. One of the biggest challenges of the KNN algorithm is choosing the right value for *k*. It is highly sensitive to the setting of the *k* parameter. The KNN algorithm does not work well with large and high-dimensional datasets [6,25]. Although there is no learning time (lazy learning), when a large training dataset is available, the computation time increases because the distance to all training samples must be calculated for each new test sample [6,8,32]. In conclusion, KNN is a preferred ML technique for FDD in PV systems, particularly due to its simple structure, nonparametric nature, and ability to process small datasets quickly.

In general, the LR method has shown inferior metrics compared to more complex models such as RF, SVM, or Ensemble methods [15,55]. LR's simple structure and linear assumptions can lead to drawbacks, especially in more complex or multi-class PV fault diagnosis scenarios. LR assumes that the input features are linear, but this was not the case in the analyzed datasets, and it was found to fail to effectively capture the input feature pattern [12,30]. This can lead to an increase in false positives (FP) and false negatives (FN) compared to more complex data. While LR is an ideal tool for FDD in PV systems due to its speed and simplicity, this feature is also its biggest drawback.

6. Experimental System

Smart card designs created for real-time monitoring and FDD at the module or string level in PV energy systems usually provide data streams every 5–10 min. This improves measurement accuracy and allows for quicker PV system interventions. These designs have become a key area of research in recent years. STM32-based embedded platforms, with high-sampling-rate ADC architectures, enable instant and reliable monitoring of PV power quality, making them a valuable addition to the hardware setup of smart cards [64]. Studies on wireless data transfer at the PV module level have shown the benefits of LoRa-based long-range communication modules, which feature low power consumption and facilitate remote monitoring of temperature, humidity, current, and voltage behaviors of PV modules [65]. However, research focused on PV module-level performance analysis has shown how processing local sensor data helps identify fault types and analyze PV module behavior [66]. These studies clearly indicate that STM32-based high-accuracy measurements at the PV module level, combined with proper sensor selection and low-power, long-range communication setups, greatly improve data quality and diagnostic efficiency in fault detection within PV energy systems. When reviewing these studies on smart card module designs, our module stands out because it also monitors weather data, such as internal and ambient temperature and humidity. This module records and transmits electrical parameters, including PV module-level current and voltage, via wired and wireless means. In this study, a smart electronic module was developed to perform system data acquisition (DAQ) for detecting, diagnosing, and predicting DC-side faults in the GCPV system. Operating at the PV module level, this smart module can measure environmental meteorological parameters like ambient temperature, humidity, module temperature, and interior temperature, which may affect the operation of the smart card module shown in Figure 7. It can also measure electrical data such as PV module current, voltage, and power. The maximum PV module voltage is 48 V using a voltage divider, but this can be increased to enable DAQ at the PV module level, extending from the PV string level. Since PV modules in the string are usually connected in series in GCPV systems, the smart card module's current sensor (20–30 A) is suitable for measuring string current.

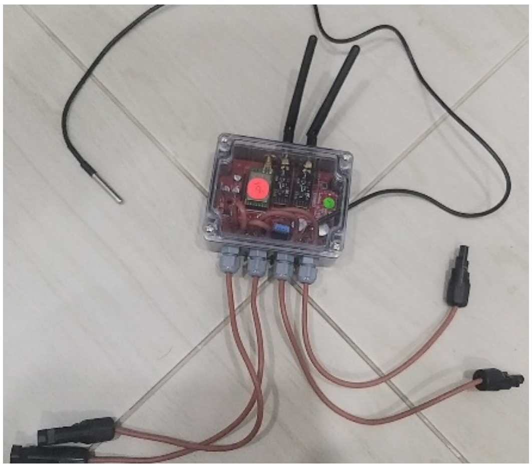

Figure 7. The designed smart module.

The smart module integrates an STM32F401RB microcontroller (STMicroelectronics, Geneva, Switzerland) and an E32-433T20DC wireless serial port module based on Semtech's SX1278 RF chip (Semtech Corporation, Camarillo, CA, USA), supporting Low-Power Wide Area Network (LPWAN) communication. A microSD card module was also added to provide uninterrupted DAQ in the event of an unfavorable situation, such as an internet outage. Figure 8 shows the MCU configuration of the smart module.

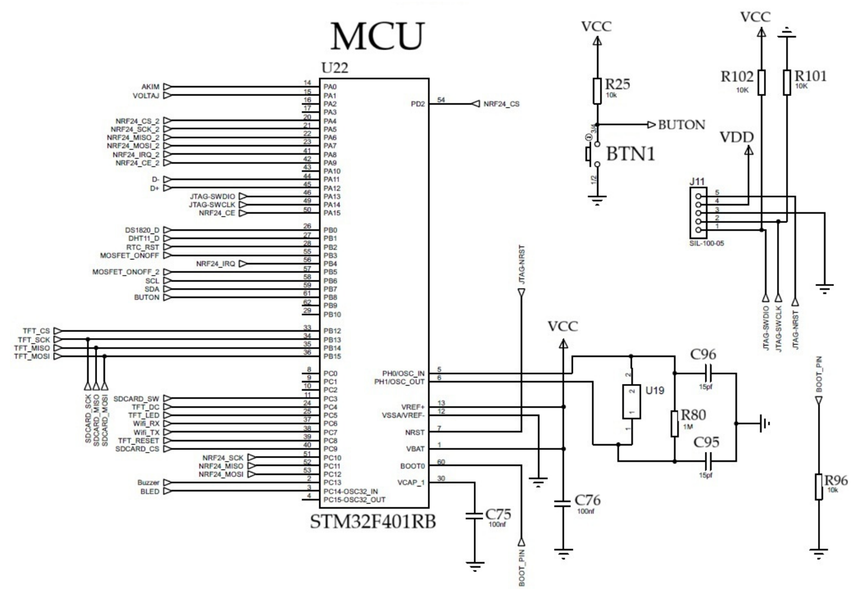

Figure 8. Smart module's MCU configuration.

The low-power NRF24L01 transceiver module (Nordic Semiconductor, Trondheim, Norway) was used to wirelessly transfer PV system data to the central computer. The NRF24L01 module operates at 3.3 V. However, the SPI pins on the module require 5 V. It has a maximum data rate of 2 Mbps. It is a 2.4 GHz radio transceiver. When it starts transmitting data, the current increases to 150 mAh. The data transmission distance in an open area is around 100 m. In fact, this distance can cause limitations in data transfer in a utility-scale PV power plant, especially at the PV module level. This limitation can be overcome by using master nodes. Figure 9 shows the electronic circuit diagram of the NRF24L01, which enables the transfer of PV module current, voltage, power, module temperature, and ambient temperature data.

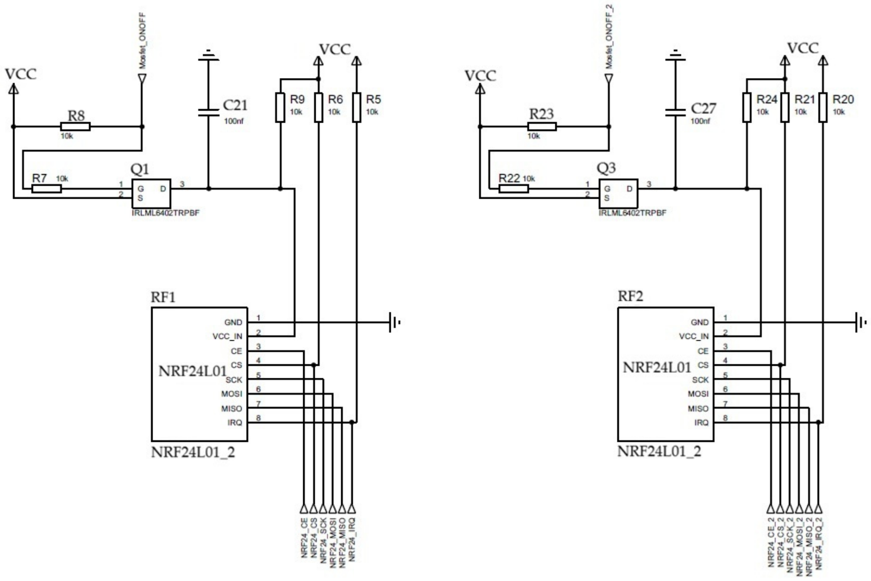

Figure 9. NRF24L01 receiver and transmitter.

The designed smart circuit module uses the E32-433T20DC wireless serial port module based on the SX1278 RF chip produced by Semtech, which has a low-power wide area network (LPWAN) infrastructure. It has multiple LoRa transmission modes operating in spread spectrum technology between 410 MHz and 441 MHz. It supports the globally license-free ISM 433 MHz frequency band. The tested communication distance of the LoRa module with a maximum transmission power of 100 mW is up to 3 km. It features data compression and encryption. Data compression provides shorter data transmission time and a lower interference rate, improving security and transmission efficiency. The module's encryption–decryption algorithm, which has very strong data hiding capabilities, protects against data interference from unwanted external environments. The module, which has a transmission power of 20 dBm, uses industrial-grade crystal oscillators to ensure stability and consistency, and its sensitivity is lower than the commonly used 10 ppm. The LoRa module's input power supply supports a voltage range of 3.3 V to 5.5 V. However, it performs best above 5.0 V. A voltage above 5.2 V can cause permanent damage to the module. The risk of burnout is lower when used at short distances. The LoRa module can operate between −40 °C and 85 °C [67]. Figure 10 shows the connection circuit between the LoRa module and the microcontroller.

Prior to deployment in a GCPV plant, laboratory tests were conducted using an off-grid PV array comprising fourteen modules with a total capacity of 140 Wp, as illustrated in Figure 11.

The proposed method was experimentally implemented using electrical data obtained from the grid-connected AYCEM GES PV plant, with an 800-kWp capacity, located in Giresun, in the Eastern Black Sea Region of Turkiye, as shown in Figure 12, along with meteorological measurement data provided by the Giresun Directorate of Meteorology. In the PV plant, twenty AS-M60 model 295 Wp PV modules were connected in series, and the outputs were individually fed into the maximum power point (MPP) inputs of nine series-connected solar inverters. The plant is equipped with a total of sixteen solar inverters. The technical specifications of the PV modules used in the system are provided in Table 2.

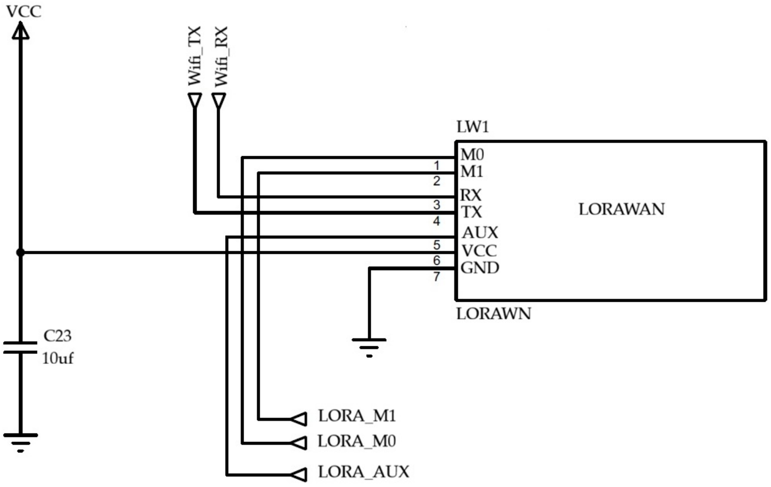

Figure 10. MCU-LoRa module connection structure.

Table 2. PV module technical data.

Electrical Characteristics	AS-M60-295 W
Maximum power (P_{max})	295 Wp
Maximum power voltage (V_{mp})	31.3 V
Maximum power current (I_{mp})	9.42 A
Open-circuit voltage (V_{oc})	39.3 V
Short-circuit current (I_{sc})	9.87 A
Temperature coefficient ($P_{m)}$	−0.41%/°C
Temperature coefficient (V_{oc})	−0.31%/°C
Temperature coefficient I_{sc}	0.05%/°C
Power Tolerance	±3%

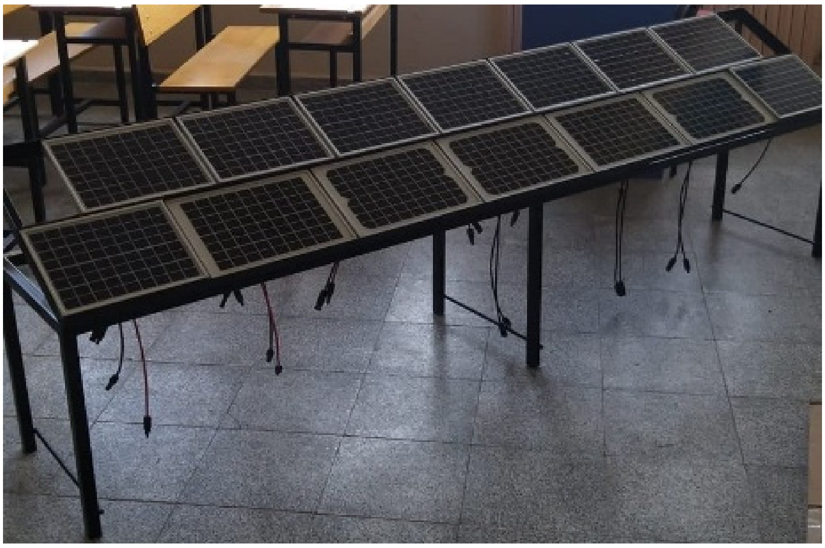

Figure 11. The experimental PV array for DC faults.

Figure 12. AYCEM GES PV plant.

Three years of electrical data from the PV power plant, covering the period between 2020 and 2022, were collected. A GoodWe (Suzhou, China) brand GW50K-MT model 65 kW solar inverter was used in the system. This solar inverter has a total of 10 MPP input terminals, of which nine were utilized in the plant. The technical data of the solar inverter used in the system are given in Table 3. An examination of the technical parameters of the solar inverter presented in Table 3 indicates that the voltage at the MPP inputs must exceed 200 volts. Therefore, the inverter enters standby mode from sunset until the first minutes of the next day when solar radiation resumes. In the applied ML, phase voltages and phase currents are independent variables, and mode status is the dependent variable.

When the electrical data obtained from the solar inverters of a real GCPV power plant were evaluated, the data were recorded in 5 min periods from sunrise to sunset. The electrical data obtained from the solar inverter included PV array currents, MPP voltages, MPP currents, phase currents, and voltages measured at the solar inverter output, as well as DC and AC power values. In addition, the solar inverter temperature data were included among the obtained data. It can be inferred that a very large dataset is generated because the data are recorded every 5 min. For this reason, only the data of one solar inverter were used in the proposed method. LR analysis was performed on the data of the first 5 arrays connected to the 16th solar inverter of the AYCEM-GES power plant for May 2021. These arrays consist of modules positioned adjacent to each other and at the same tilt angle. Data exceeding the study limits and meaningless data were eliminated from the 5566-line data group obtained from the GCPV power plant for May 2021. Daily and hourly humidity, wind, precipitation, and temperature weather forecast data for the relevant dates were obtained from the Giresun Meteorology Directorate. These data were associated with the grid-connected PV power plant data. Since the weather forecast data are provided hourly, inferences compatible with the 5 min electrical data were generated using the estimation algorithm implemented in Python. ML can be used in PV power generation estimation or in the process of determining and predicting PV failures.

Table 3. Solar inverter's technical data.

Electrical Characteristics	**GW50K-MT**
Maximum PV power (W)	65,000
Maximum DC input voltage (V)	1000
MPPT range (V)	200~850
Starting voltage (V)	200
Nominal DC input voltage (V)	620
Maximum input current (A)	30/30/20/20
Maximum short current (A)	38/38/25/25
No. of MPP trackers	4
No. of input strings per tracker	3/3/2/2
AC Output Data	**GW50K-MT**
Nominal output power (W)	50,000
Maximum output power (W)	55,000; 57,500 @415 Vac
Nominal output frequency (Hz)	50/60
Maximum output current (A)	80
Output THDİ (@ nominal output)	<3%
European efficiency	98.7%

In ML analysis, in the process of estimating the power to be produced in the next period by processing the trained data, the generated power is the dependent variable. In contrast, the electrical properties of the system and weather data can be considered independent variables. Using a tolerance value, the estimated probability is converted to a binary result (0 or 1). The evaluation criteria used in the ML analysis for AC fault detection and prediction in the GCPV power system are given in Table 4. In the table, the prediction is based on the phase voltages and phase currents from the electrical data obtained from the solar inverter. The state with phase voltages and phase currents is the normal operating state (mode 0). The state with phase voltages and no phase currents is the standby state for the inverter (mode 0). The state with both phase voltages and no phase currents is considered a fault state. Since the standard and standby states are normal states that do not cause a fault, mode 0 is active in the ML analysis, while mode 1 is active in the event of a fault.

This study was conducted using 5566 rows of real empirical electrical data from May 2021, obtained from the AYCEM-GES GCPV power plant in Giresun, and we independently applied KNN, LR, and ANN algorithms for AC-side fault detection. In the data preprocessing phase, the ".xlsx" dataset from the PV power plant was converted to Python 3.9.13 compatibility (comma-dot conversion, filtering of NaN values) to ensure data integrity. The dataset exhibits a highly skewed class distribution, containing only 6 rare fault instances (Mode 1) compared to normal (Mode 0) cases. To correct class imbalance and improve generalization capability, 70% of the dataset was allocated for training, 15% for cross-validation, and 15% for testing using stratified splitting, and the SMOTE adjustment technique was applied to the training data. Normal Class (Negative) samples constitute approximately 99% of the total data, while Fault Class (Positive) samples account for less than 1%. This imbalance resulted in approximately 5–10 positive samples in the test set. The data covers the period from May 2021; however, the total duration of each class depends on the sampling frequency in the data, which is fixed. The models' noise robustness and the effect of the SMOTE equalization technique were observed. The results demonstrate that all models exhibit high efficiency in detecting inverter and grid-related AC-side faults, with high AUC values exceeding the critical threshold (KNN: 0.9917, LR: 0.8709, ANN: 0.8555) and correctly classifying all six fault samples (Mode 1) in the test set with zero False Negatives (FN). The best overall performance was achieved by KNN (k = 10) with an AUC of 0.9917, highlighting the fault-tolerant detection capability of these models. The study contributes to the literature by focusing on AC-side errors, the methodology, the GoodWe GW50K-MT inverter data used, and the details of the Python 3.9.13-based computational environment, which provides transparency for full reproducibility.

Table 4. ML methods' evaluation modes.

Phase Voltages ($\mathbf{U_{a,b,c}}$)	Phase Currents ($\mathbf{I_{a,b,c}}$)	Case	Mode	Mode Value
✓	✓	Normal	0	0
✓	⊗	Wait	2	0
⊗	⊗	Fault	1	1

7. Evaluation of the Model

A dataset from a real GCPV power plant was used for model training, while 30% was used as test data for prediction. True positives (TP) are the values in which the model correctly classifies a positive sample in the test dataset. FN are the values in which the model incorrectly classifies a negative sample. FP are the values in which the true value is negative, but the model incorrectly predicts a positive value. True negatives (TN) are the values in which the true value is negative, and the model correctly predicts a negative sample. The confusion matrix evaluates the performance of the model used in fault detection and prediction across these four primary classification outputs, enabling the calculation of correct and incorrect classification rates. The accuracy score is calculated as the ratio of the number of samples correctly classified by the model to the total number of predictions. This performance metric is calculated by dividing the sum of true positives and true negatives by the total number of predictions made by the model. The accuracy score is a reliable performance metric when the class distribution in the dataset is balanced. However, its use in datasets with unbalanced class distributions may give misleading results.

$$Accuracy = (TP + TN)/(TP + FP + FN + TN) \tag{7}$$

Precision score is the proportion of correctly predicted positive examples to the total number of examples that the model predicted to be positive. The precision score is utilized in determining how much of the data was correctly labelled as positive by the model.

$$Precision = TP/(TP + FP) \tag{8}$$

The recall measure is defined as the proportion of correctly classified positive cases to the number of positive cases in the real class. The recall measure indicates the proportion of actual data classified by the model as positive.

$$Recall = TP/(TP + FN) \tag{9}$$

https://doi.org/10.3390/solar6010006

The F1 score can be defined as the harmonic formulation of precision and recall rates. It helps us estimate the model's overall performance, taking into account the FP and FN. The F1 score offers a better and more meaningful measure of performance relative to the accuracy score, particularly for instances when there isn't an even distribution between classes.

$$F1 = (2 \times precision \times recall)/(precision + recall) \tag{10}$$

Model performance assessment in this paper takes inspiration from a set of widely used classification metrics: precision, recall, F1 score, and accuracy. These metrics capture various aspects of ML models' prediction performance. Range is 0.00 to 1.00, where 1.00 represents ideal classification with zero error, and 0.00 represents the worst, where there is total failure in classification. Accuracy determines the total accuracy of the model by determining the number of instances that are correctly classified for all predictions. Precision determines the precision of positive predictions by determining the number of true positives in comparison to all instances that have been predicted as positive, with perfect marking showing that there are no false positives. Recall measures the model's performance at detecting all the actual true positive instances, and 1.00 means all true positives are detected with no FN. F1 Score is the harmonic mean of precision and recall, and a unidimensional measure that finds a compromise between both; precision and recall both achieve their best value results in the F1 score also achieving its best value, and that is an ideal value of 1.00. In comparative performance research of DT, KNN, and SVM algorithms for PV array fault detection, Bayesian optimization was used for allocating optimal hyperparameters to fault classifiers. The SVM-based model provided the best classification performance [68]. In a study comparing the performance of different ML algorithms in fault detection in solar power systems, DT, KNN, Random Forest, and Extra Trees classifiers achieved an F1 score of 1.00, whereas the LR algorithm achieved an F1 score of 0.993 [69]. However, the findings of the research can be made more valid by using a dataset with a wider diversity of fault modes and more performance evaluation methods of the model.

8. Results and Discussion

This study uses one month of operating data from the AYCEM-GES grid-connected photovoltaic power plant, and measurements were obtained from a 65 kWp GW50K-MT model solar inverter manufactured by GoodWe. Electrical values are read in this inverter using the Modbus protocol via an RS-485-based serial communication interface, providing reliable access to string-based current and voltage values, AC/DC power components, module temperature, instantaneous production information, and system status parameters. Because the RS-485 line requires a physical connection to a data acquisition device, a SCADA unit, or an external monitoring module, wired data transfer was initially implemented at the plant. In subsequent stages, a DAQ-based gateway/data logger device was integrated via RS-485 or the built-in WiFi module, enabling the inverter's monitoring interface and enabling wireless transmission of measurement data to the manufacturer's portal. This architecture increased both operational flexibility and the accuracy of continuous, reliable data collection in the PV field. The data obtained were preprocessed to eliminate outliers. Additionally, temperature data were compared with hourly data obtained from the Giresun Meteorological Directorate. Following the comparison, a temperature extraction algorithm was created using Python software, yielding hourly data with a 5 min period. These data will be used to identify DC-side faults, which will be analyzed in the future. A smart module was also designed to collect PV module-based data for diagnosing DC-side faults.

The meteorological dataset used in this study has a time resolution where measurements are reported as hourly averages. In contrast, inverter and electrical performance variables were recorded at 5 min intervals. To enable the analysis of data sources with different sampling frequencies, the time axis needs to be harmonized. Therefore, the linear interpolation method was preferred for converting hourly average temperature values to a 5 min resolution. Linear interpolation is based on the assumption that values between two known points change in a linear relationship and is a widely used scale adjustment technique in time series. In this method, transitions between hourly temperature measurements are estimated along a linear line without requiring any additional assumptions. Thus, both data continuity is ensured, and high-frequency electrical data and meteorological variables can be compared at the same time resolution. The basic formulation of linear interpolation is expressed as follows [70]:

$$T(t) = T(t_1) + [T(t_2) \; - \; T(t_1)] \; \times (t - \; t_1 \;) / \; (t_2 - \; t_1) \tag{11}$$

Here, $T(t_1)$ and $T(t_2)$ represent the previous and next temperature measurements, respectively, and t represents the intermediate (5 min) time point. This equation allows the temperature value corresponding to any intermediate time point between t_1 and t_2 to be calculated based on the linear rate of change between the two values. Since meteorological temperature data are a relatively slowly changing atmospheric parameter, the linear interpolation method was considered a suitable approach in terms of both estimation accuracy and computational cost. The resulting 5 min temperature series was synchronized with all the variables used in inverter performance analysis.

A confusion matrix is one of the most widely used methods to evaluate the performance of models created in ML [71]. This matrix presents the classification results of the models on predefined target datasets as positive and negative. The use of a confusion matrix facilitates the analysis of the obtained results and provides benefits in terms of evaluating the accuracy and errors of the models. Figure 13 presents the general structure of the confusion matrix. As shown in the figure, the actual data indicate that the number of failures is low. This is clearly illustrated in Figure 13. The experimental ML methods were trained, validated, and tested using historical electrical data from a real GCPV energy plant. 70% of the dataset was allocated for training to capture the relationships between input parameters and output power, while 15% was used for cross-validation to optimize hyperparameters and prevent overfitting. The remaining 15% of unseen data evaluated the network's generalization capability under novel conditions. The use of real empirical data enabled the neural model to effectively capture complex nonlinear relationships, thereby ensuring reliable and accurate predictive performance [72].

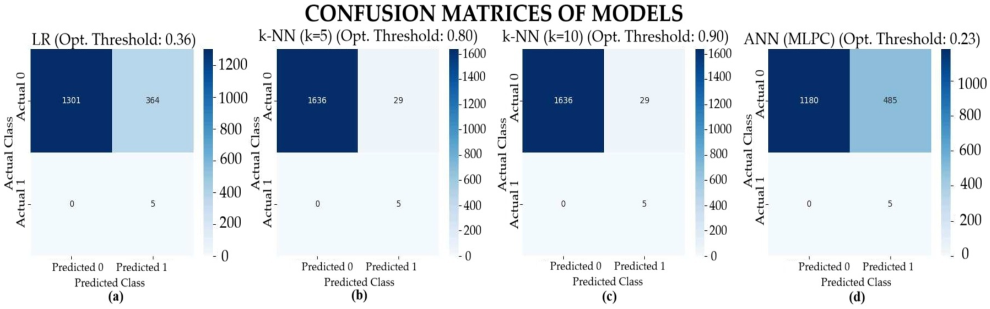

Figure 13. Confusion matrices: (**a**) confusion matrix for LR; (**b**) confusion matrix for KNN (k = 5); (**c**) confusion matrix for KNN (k = 10); (**d**) confusion matrix for ANN.

The performances of KNN, LR, and ANN were compared according to the performance indices of accuracy, precision, recall, and F1 score, and the obtained values are presented in Table 5. The accuracy and AUC index values of all three machine learning methods ranged from a minimum of 0.7096 to a maximum of 0.9917. This shows that machine learning models can classify and predict faults with high accuracy in all performance indices. Among the machine learning methods, the accuracy and AUC values of KNN are 0.9826 and 0.9917, respectively, and high-performance index values were obtained. The LR algorithm showed high performance with accuracy and AUC values of 0.782 and 0.8709, respectively. Finally, the ANN algorithm showed the lowest performance with accuracy and AUC values of 0.7096 and 0.8555, respectively.

Table 5. Test set's performance summary table.

Model	Accuracy	Precision	Recall	F1 Score	AUC	Opt. Threshold
KNN (k = 5)	0.9826	0.1471	1	0.2564	0.9911	0.8
KNN (k = 10)	0.9826	0.1471	1	0.2564	0.9917	0.9
LR	0.782	0.0136	1	0.0267	0.8709	0.3605
ANN (MLPC)	0.7096	0.0102	1	0.0202	0.8555	0.2306

When the confusion matrices in Figure 13 are evaluated, it is seen that the performance of the KNN, LR, and ANN ML algorithms applied in the study is high despite the high class imbalance with 6 fault samples compared to 1663 normal samples in the test set. The most critical finding is that all models successfully detected all 6 fault samples in the test set, thus achieving a zero false negative (FN = 0) result. This result proves that the models exhibited excellence (1.00) in the recall metric and eliminated the risk of missing critical errors. However, the high false positive (FP) values observed, especially in the LR (364 false positives) and ANN (485 false positives) models, indicate that the models tend to overestimate the positive class and therefore their precision scores remain low. In GCPV systems, the operational cost of failing to detect a fault is more critical than the cost of a false alarm, so this performance, supported by high recall and F1 Score, strongly confirms that the models have achieved their fundamental fault detection objective.

To ensure the reproducibility and transparency of the study, the critical hyperparameters of the KNN, LR, and ANN models in AC-side fault detection in GCPV systems using ML techniques are detailed. To minimize the effect of the dataset with high class imbalance (1663 normal and 6 fault samples in the test set), the LR model uses class_weight = 'balanced' and max_iter = 1000 iterations. The KNN algorithm was tested with $k = 5$ and $k = 10$ neighbors to determine the optimal performance. The ANN (MLPC) model is structured with two hidden layers (50 and 20 neurons), the ReLU activation function, the Adam optimization algorithm, and an alpha = 0.001 L2 regularization with an early_stopping = True mechanism to optimize its ability to model complex relationships and control overfitting. Training, cross-validation, and test set separation (70%-15%-15%) were performed using stratified splitting, and consistency across all procedures was fixed at random_state = 42. These extensive hyperparameter adjustments ensured the scientific reproducibility and methodological robustness of the obtained experimental results (particularly, for KNN (k = 10), AUC = 0.9917).

Phase voltages and phase currents from the GCPV power plant dataset serve as independent variables for the ML algorithm models. The results for the mode value, the dependent variable, were analyzed using KNN, LR, and ANN models. There are three modes: wait, normal, and fault. The correlation matrix shown in Figure 14 illustrates the relationship between phase voltages and phase currents in the PV power plant. The correlation coefficient, which indicates the relationship between the two groups of variables, ranges from -1 to 1. A value of -1 signifies a perfect negative linear relationship. A value of +1 signifies a perfect positive linear relationship. A value of 0 indicates no linear relationship between the variables. The independent variables in the correlation matrix are grouped into two categories, as shown in the first and second columns of Table 4. The correlation between the independent variables, phase voltages, and phase currents is strongly positive, ranging from 0.8426 to 1. Since the standby mode, where phase voltages are present but phase currents are absent, is considered the normal mode, the correlation between the first three columns and the last row of the matrix, as well as between the first three rows and the last three columns, indicates weak positive correlations in the range of 0.17 to 0.1947. When evaluating the correlation matrix in Figure 14, yellow indicates positive correlation, blue indicates negative correlation, and green indicates weak correlation. The correlation between row 1 and columns 2, 3, and 4 is negative. The correlation between row 1 and columns 5, 6, and 7 shows a very weak correlation. Columns 2, 3, and 4 are highly positively correlated (0.90–0.98), as are columns 5, 6, and 7. These results suggest that the data are linear and have high predictive value. Consequently, the presence or absence of phase currents accounts for a linear relationship aligned with the operating conditions of the solar inverter in the PV power system, while the absence of both phase voltages and phase currents indicates a fault on the AC side.

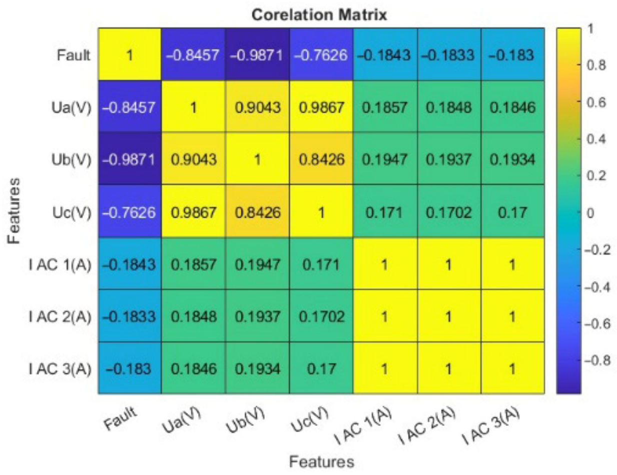

Figure 14. The correlation matrix—fault status relationship.

Figure 15 shows the classification accuracy on the ROC curve of the AC-side fault detection model for a PV energy system developed using three different machine learning techniques, namely KNN, LR, and ANN algorithms. The ROC curve is a widely used method for evaluating the accuracy performance of algorithm models. The ROC curve presented in Figure 15 facilitates the comparative analysis of the machine learning techniques used in the study. The figure shows the true positive rate (TPR) and false positive rate (FPR) values under varying decision thresholds for the KNN, LR, and ANN models trained on the GCPV power plant dataset. The probability values estimated by the machine learning algorithms were converted into binary classes using thresholds between 0.5 and 0.8. Accordingly, the TPR and FPR rates were obtained at different threshold values. The critical threshold values that determine the success of the models are shown as dashed lines in Figure 15. For the evaluated methods to be considered successful, the ROC curves are expected to remain above this dashed reference line, indicating satisfactory model performance. In addition, AUC is considered an important metric representing the overall accuracy of the models. AUC is a summary metric ranging from 0 to 1; AUC values of 0.5–0.7, 0.7–0.8, 0.8–0.9, and >0.9 indicate poor, good, excellent, and excellent predictive accuracy, respectively [73]. In this study, as seen in Figure 15, values higher than the reference AUC value (0.5) were obtained. The AUC values obtained for the KNN, LR, and ANN algorithms were found to be 0.9917, 0.8709, and 0.8555, respectively. The critical AUC level was exceeded for all three ML algorithms, confirming successful fault detection. In the comparative analysis of ML algorithms used in AC-side fault detection, the ANN method successfully performed the classification with lower accuracy performance compared to the other methods. Considering the strong correlation threshold of 0.80, the accuracy values for the KNN, LR, and ANN methods reached 0.9826, 0.782, and 0.7096, respectively. Among other machine learning methods, the KNN algorithm stands out with the highest accuracy value.

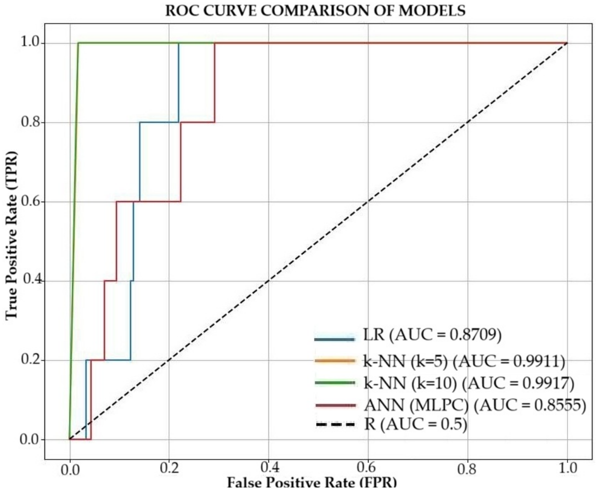

Figure 15. ROC curve comparison of models.

9. Conclusions

This study makes a significant contribution by focusing on fault detection and classification on the AC side of GCPV systems; our research has not found any similar studies focusing on this topic in the literature. Three different machine learning methods—KNN, LR, and ANN—were applied using a unique electrical dataset from a real GCPV power plant. All applied machine learning methods successfully detected faults originating from the solar inverter and the grid. When evaluating the overall performance of the models, the KNN algorithm performed best with the highest accuracy rate of 98.26%. Among the other methods, LR achieved an accuracy rate of 78.20%, while the ANN algorithm performed relatively poorly compared to other machine learning algorithms, with an accuracy rate of 70.96%. The AUC values of all models exceeded the critical threshold (KNN: 0.9917, LR: 0.8709, ANN: 0.8555), confirming their success in fault detection.

The proposed approach has some limitations and potential drawbacks. The study used only one month of real field data from the AYCEM-GES-GCPV power plant in Giresun. This limited timeframe could potentially limit the model's ability to generalize under different seasonal conditions or to less common fault types (Mode 1). Furthermore, the test dataset (Mode 1) contained only six rare fault examples, resulting in a significant class imbalance. Therefore, additional metrics, such as the F1 score, were used to prevent metrics such as accuracy from being misleading.

The study intentionally focused only on AC-side faults. This approach cannot directly detect complex DC-side faults (arcing, SC, OC), which pose a fire risk and are more destructive in PV systems. This limited the scope of the FDD study to the AC side.

These limitations identified the focus areas for future research: An electronic smart card module was designed to read and record data at the PV module level to detect DC-side faults. In addition, a test PV array consisting of 14 10Wp PV modules was created for experimental applications of the smart card module. With these improvements, future studies will focus more on detailed research and development using ML methods.

Author Contributions: Conceptualization, M.T.G.; methodology, S.K. and B.K.; software, S.K. and B.K.; validation, S.K. and B.K.; investigation, M.T.G. and S.K.; resources, S.K.; data curation, M.T.G. and S.K.; writing—original draft preparation, S.K.; writing—review and editing, S.K.; visualization, S.K. and B.K.; supervision, M.T.G.; funding acquisition, M.T.G. and B.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset will be shared with you upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Singh, V.; Beniwal, R. Automated model for fault detection in grid-connected solar systems. *J. Eng. Appl. Sci.* **2025**, *72*, 32. [CrossRef]
2. Varaprasad, K.S.; Mukherjee, A.; Basak, R.; Thakur, A.K.; Sanyal, S. K-Nearest Neighbour (KNN) Algorithm approach in Machine Learning to Power output prediction of solar photovoltaic Renewal Energy. In *Optimization and Artificial Intelligent Strategies for Engineering and Management*; BS Publications: Hyderabad, India, 2025; pp. 211–222. Available online: https://bspublications.net/9781978522200/26.pdf (accessed on 14 January 2026).
3. Dhibi, K.; Mansouri, M.; Bouzrara, K.; Nounou, H.; Nounou, M. An Enhanced Ensemble Learning-Based Fault Detection and Diagnosis for Grid-Connected PV Systems. *IEEE Access* **2021**, *9*, 155622–155633. [CrossRef]
4. Harrou, F.; Taghezouit, B.; Sun, Y. Improved *k*NN-Based Monitoring Schemes for Detecting Faults in PV Systems. *IEEE J. Photovolt.* **2019**, *9*, 811–821. [CrossRef]
5. Qin, J.; Wang, L.; Huang, R. Research on Fault Diagnosis Method of Spacecraft Solar Array Based on f-KNN Algorithm. In Proceedings of the 2017 Prognostics and System Health Management Conference (PHM-Harbin), Harbin, China, 9–12 July 2017; pp. 1–4. [CrossRef]
6. Iheanetu, K.; Obileke, K. Short-Term Forecasting of Photovoltaic Power Using Multilayer Perceptron Neural Network. Convolutional Neural Network, and k-Nearest Neighbors' Algorithms. *Optics* **2024**, *5*, 293–309. [CrossRef]
7. 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. [CrossRef]
8. Ağır, T.T. Prediction of Losses Due to Dust in PV Using Hybrid LSTM-KNN Algorithm: The Case of Saruhanlı. *Sustainability* **2024**, *16*, 3581. [CrossRef]
9. Chen, Q.; Li, Q.; Wu, J.; He, J.; Mao, C.; Li, Z.; Yang, B. State Monitoring and Fault Diagnosis of HVDC System via KNN Algorithm with Knowledge Graph: A Practical China Power Grid Case. *Sustainability* **2023**, *15*, 3717. [CrossRef]
10. Eskandari, A.; Nedaei, A.; Milimonfared, J.; Aghaei, M. A multilayer integrative approach for diagnosis, classification and severity detection of electrical faults in photovoltaic arrays. *Expert Syst. Appl.* **2024**, *252*, 124111. [CrossRef]
11. Al-Rousan, N.; Isa, N.A.M.; Desa, M.K.M.; Al-Najjar, H. Integration of logistic regression and multilayer perceptron for intelligent single and dual axis solar tracking systems. *Int. J. Intell. Syst.* **2021**, *25*, 5605–5669. [CrossRef]
12. Prashanth, M.V.; Bharath, K.N.; Fathima, N.; Latha, B.M.; Sathisha, M.S.; Ramji, B.R.; Yathiraj, G.R. Machine Learning Approaches for Solar PV Fault Identification. *SN Comput. Sci.* **2025**, *6*, 839. [CrossRef]
13. Zwirtes, J.; Libano, F.B.; Silva, L.A.L.; Freitas, E.P. Fault Detection in Photovoltaic Systems Using a Machine Learning Approach. *IEEE Access* **2025**, *13*, 41406–41421. [CrossRef]
14. Fan, J.; Liwen, L.; Shiyue, G.; Jian, Y. Logistic Regression Based Arc Fault Detection in Photovoltaic Systems Under Different Conditions. *J. Shanghai Jiao Tong Univ.* **2019**, *24*, 459–470. [CrossRef]
15. Tufail, S.; Sarwat, A.I. A Comparative Study of Dimensionality Reduction Methods for Accurate and Efficient Inverter Fault Detection in Grid-Connected Solar Photovoltaic Systems. *Electronics* **2025**, *14*, 2916. [CrossRef]
16. Li, B.; Delpha, C.; Diallo, D.; Migan-Dubois, A. Application of Artificial Neural Networks to photovoltaic fault detection and diagnosis: A review. *Renew. Sustain. Energy Rev.* **2020**, *138*, 110512. [CrossRef]
17. El-Banby, G.M.; Moawad, N.M.; Abouzalm, B.A.; Abouzaid, W.F.; Ramadan, E.A. Photovoltaic system detection techniques: A review. *Neural Comput. Appl.* **2023**, *35*, 24829–24842. [CrossRef]
18. Mouleloued, Y.; Kara, K.; Chouder, A. A Developed Algorithm Inspired from the Classical KNN for Fault Detection and Diagnosis PV Systems. *J. Control Autom. Electr. Syst.* **2023**, *34*, 1013–1027. [CrossRef]
19. Hussain, M.; Dhimish, M.; Titarenko, S.; Mather, P. Artificial neural network based photovoltaic fault detection algorithm integrating two bi-directional input parameters. *Renew. Energy* **2020**, *155*, 1272–1292. [CrossRef]
20. Tchio, G.M.T.; Kenfack, J.; Voufo, J.; Mindzie, Y.A.; Njoya, B.F.; Ouro-Djobo, S.S. Diagnosing faults in a photovoltaic system using the Extra Trees ensemble algorithm. *AIMS Energy* **2024**, *12*, 727–750. [CrossRef]

21. Pandian, P.S.; Denosha, T.G.; Kumar, R.S.; Sivaiah, V. Real-Time Fault Detection in Solar PV Systems Using Hybrid ANN-SVM Machine Learning Algorithm. *Int. J. Intell. Syst. Appl. Eng.* **2024**, *12*, 1367–1374.
22. Islam, M.; Rashel, M.R.; Ahmed, M.T.; Islam, A.K.M.K.; Tlemçani, M. Artificial Intelligence in Photovoltaic Fault Identification and Diagnosis: A Systematic Review. *Energies* **2023**, *16*, 7417. [CrossRef]
23. Ul-Haq, A.; Sındı, H.F.; Gul, S.; Jalal, M. Modeling and Fault Categorization in Thin-Film and Crystalline PV Arrays Through Multilayer Neural Network Algortihm. *IEEE Access* **2020**, *8*, 102235–102255. [CrossRef]
24. Zamzeer, A.S.; Farhan, M.S.; AlRikabi, H.T.H.S. Fault Detection System of Photovoltaic Based on Artificial Neural Network. *Wasit J. Eng. Sci.* **2023**, *11*, 93–104. [CrossRef]
25. Emamian, M.; Eskandari, A.; Aghaei, M.; Nedaei, A.; Sizkouhi, A.M.; Milimonfared, J. Cloud Computing and IoT Based Intelligent Monitoring System for Photovoltaic Plants Using Machine Learning Techniques. *Energies* **2022**, *15*, 3014. [CrossRef]
26. Swarna, K.S.V.; Vinayagam, A.; Ananth, M.B.J.; Kumar, P.V.; Veerasamy, V.; Radhakrishnan, P. A KNN based random subspace ensemble classifier for detection and discrimination of high impedance fault in PV integrated power network. *Measurement* **2022**, *187*, 110333. [CrossRef]
27. Eskandari, A.; Aghaei, M.; Milimonfared, J.; Nedaei, A. A weighted ensemble learning-based autonomous fault diagnosis method for photovoltaic systems using genetic algorithm. *Int. J. Electr. Power Energy Syst.* **2023**, *144*, 108591. [CrossRef]
28. Amiri, A.F.; Oudira, H.; Chouder, A.; Kichou, S. Faults detection and diagnosis of PV systems based on machine learning approach using random forest classifier. *Energy Convers. Manag.* **2024**, *301*, 118076. [CrossRef]
29. Hassan, M.S.; Chin, V.J.; Gopal, L. Accurate diagnosis of concurrent faults in photovoltaic systems using CONMI-based feature selection and Support vector machines. *Energy Convers. Manag.* **2025**, *344*, 120293. [CrossRef]
30. Eldeghady, G.S.; Kamal, H.A.; Hassan, M.A.M. Comparative analysis of the performance of supervised learning algorithms for photovoltaic system fault diagnosis. *Sci. Technol. Energy Transit.* **2024**, *79*, 27. [CrossRef]
31. Khalil, I.U.; Haq, A.U.; Islam, N.U. A deep learning-based transformer model for photovoltaic fault forecasting and classification. *Electr. Power Syst. Res.* **2024**, *228*, 110063. [CrossRef]
32. Chen, J.L.; Kuo, C.L.; Chen, S.J.; Kuo, C.C.; Zhan, T.S.; Lin, C.H.; Chen, Y.S. DC side fault detection for photovoltaic energy conversion system using fractional-order dynamic-error-based fuzzy Petri net integrated with intelligent meters. *IET Renew. Power Gener.* **2016**, *10*, 1318–1327. [CrossRef]
33. Zhao, Y.; Lehman, B.; Palma, J.F.; Mosesian, J.; Lyons, R. Fault analysis in solar PV arrays under low irradiance conditions and reverse connections. In Proceedings of the 37th IEEE PVSC, Seattle, WA, USA, 19–24 June 2011. [CrossRef]
34. AbdulMawjood, K.; Refaat, S.S.; Morsi, W.G. Detection and prediction of faults in photovoltaic arrays: A review. In *Proceedings of the 2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018), Doha, Qatar, 10–12 April 2018*; IEEE: New York, NY, USA, 2018; pp. 1–8. [CrossRef]
35. Hong, Y.Y.; Pula, R.A. Methods of photovoltaic fault detection and classification: A review. *Energy Rep.* **2022**, *8*, 5898–5929. [CrossRef]
36. Et-taleby, A.; Chaibi, Y.; Benslimane, M.; Boussetta, M. Applications of Machine Learning Algorithms for Photovoltaic Fault Detection: A Review. *Stat. Optim. Inf. Comput.* **2023**, *11*, 168–177. [CrossRef]
37. Ghaffarzadeh, N.; Azadian, A. A comprehensive review and performance evaluation in solar (PV) systems fault classification and fault detection techniques. *J. Sol. Energy Res.* **2019**, *4*, 252–272. [CrossRef]
38. Duman, S.; Li, J.; Wu, L.; Yorukeren, N. Symbiotic organisms search algorithm-based security-constrained AC-DC OPF regarding uncertainty of wind, PV, and PEV systems. *Soft Comput.* **2021**, *25*, 9389–9426. [CrossRef]
39. Pillai, D.S.; Rajasekar, N. A comprehensive review on protection challenges and fault diagnosis in PV systems. *Renew. Sustain. Energy Rev.* **2018**, *91*, 18–40. [CrossRef]
40. *IEC Standard 60363-7-712*; Electrical Installations of Buildings-Part 7-712: Requirements for Special Installations or Locations-Solar Photovoltaic (PV) Power Supply Systems. International Electrotechnical Commission: Geneva, Switzerland, 2020. Available online: https://standards.der-lab.net/standard/electrical-installations-of-buildings-part-7-712-requirements-for-special-installations-or-locations-solar-photovoltaic-pv-power-supply-systems/ (accessed on 14 January 2026).
41. *IEC Standard 62548*; Installation and Safety Requirements for PV Generators. International Electrotechnical Commission: Geneva, Switzerland, 2023. Available online: https://webstore.iec.ch/en/publication/64171 (accessed on 14 January 2026).
42. Article 690-Solar Photovoltaic Systems of National Electrical Code, NFPA70. 2011. Available online: https://enkonnsolar.com/wp-content/uploads/2023/08/Article-690-Photovoltaic-PV-System.pdf (accessed on 14 January 2026).
43. Gaaloul, Y.; Kechiche, O.B.H.B.; Oudira, H.; Chouder, A.; Hamouda, M.; Silvestre, S.; Kichou, S. Faults Detection and Diagnosis of a Large-Scale PV System by Analyzing Power Losses and Electric Indicators Computed Using Random Forest and KNN-Based Prediction Models. *Energies* **2025**, *18*, 2482. [CrossRef]
44. Sarikh, S.; Raoufi, M.; Bennouna, A.; Benlarabi, A.; Ikken, B. Fault diagnosis in a photovoltaic system through I-V characteristics analysis. In *Proceedings of the 9th International Renewable Energy Congress (IREC), Hammamet, Tunisia, 26–28 March 2018*; IEEE: New York, NY, USA, 2018; pp. 1–6. [CrossRef]

45. Bower, W.; Wiles, J. Investigation of ground-fault protection devices for photovoltaic power system applications. In *Proceedings of the Twenty-Eighth IEEE Photovoltaic Specialists Conference, Anchorage, AK, USA, 15–22 September 2000*; IEEE: New York, NY, USA, 2000; pp. 1378–1383. [CrossRef]
46. Zeb, K.; Islam, S.U.; Khan, I.; Uddin, W.; Ishfaq, M.; Busarello, T.D.C.; Muyeen, S.M.; Ahmad, I.; Kim, H.J. Faults and fault ride through strategies for grid-connected photovoltaic system: A comprehensive review. *Renew Sustain Energy Rev.* **2022**, *158*, 112125. [CrossRef]
47. Triki-Lahani, A.; Abdelghani, A.B.B.; Slama-Belkhodja, I. Fault detection and monitoring systems for photovoltaic installations: A review. *Renew. Sustain. Energy Rev.* **2018**, *82*, 2680–2692. [CrossRef]
48. Garoudja, E.; Harrou, F.; Sun, Y.; Kara, K.; Chouder, A.; Silvestre, S. Statistical fault detection in photovoltaic systems. *Sol. Energy* **2017**, *150*, 485–499. [CrossRef]
49. Pei, T.; Hao, X. A fault detection method for photovoltaic systems based on voltage and current observation and evaluation. *Energies* **2019**, *12*, 1712. [CrossRef]
50. Aziz, F.; Ul-Haq, A.; Ahmad, S.; Mahmoud, Y.; Jalal, M.; Ali, U. A novel convolutional neural network-based approach for fault classification in photovoltaic arrays. *IEEE Access* **2020**, *8*, 41889–41904. [CrossRef]
51. Al-sheikh, H.; Moubayed, N. Fault detection and diagnosis of renewable energy systems: An overview. In Proceedings of the 2012 International Conference on Renewable Energies for Developing Countries (REDEC), Beirut, Lebanon, 28–29 November 2012; IEEE: New York, NY, USA, 2012; pp. 1–7. [CrossRef]
52. Thakfan, A.; Salamah, Y.B. Artificial-Intelligence-Based Detection of Defects and Faults in Photovoltaic Systems: A Survey. *Energies* **2024**, *17*, 4807. [CrossRef]
53. Hariharan, R.; Çakkarapani, M.; Ilango, G.S.; Nagamani, C. A method to detect photovoltaic array faults and partial shading in PV systems. *IEEE J. Photovolt.* **2016**, *6*, 1278–1285. [CrossRef]
54. Basnest, B.; Chun, H.; Bang, J. An Intelligent Fault Detection Model for Fault Detection in Photovoltaic Systems. *J. Sens.* **2020**, *2020*, 6960328. [CrossRef]
55. Abdelsattar, M.; AbdelMoety, A.; Emad-Eldeen, A. Advanced machine learning techniques for predicting power generation and fault detection in solar photovoltaic systems. *Neural Comput. Appl.* **2025**, *37*, 8825–8844. [CrossRef]
56. VenkateshS, N.; Sripada, D.; Sugumaran, V.; Aghaei, M. Detection of visual faults in photovoltaic modules using a stacking ensemble approach. *Heliyon* **2024**, *10*, e27894. [CrossRef]
57. Nassreddine, G.; Arid, A.E.; Nassereddine, M.; Khatib, O.A. Fault Detection and Classification for Photovoltaic Panel System Using Machine Learning Techniques. *Appl. AI Lett.* **2025**, *6*, e115. [CrossRef]
58. Saxena, N.; Kumar, R.; Rao, Y.K.S.S.; Mondloe, D.S.; Dhapekar, N.K.; Sharma, A.; Yadav, A.S. Hybrid KNN-SVM machine learning approach for solar power forecasting. *Environ. Chall.* **2024**, *14*, 100838. [CrossRef]
59. Al-Dahidi, S.; Hammad, B.; Alrbai, M.; Al-Abed, M. A novel dynamic/adaptive K-nearest neighbor model for the prediction of solar photovoltaic systems' performance. *Results Eng.* **2024**, *22*, 102141. [CrossRef]
60. Veerasamy, V.; Wahab, N.I.A.; Othman, M.L.; Padmanaban, S.; Ramachandran, R.; Vinayagam, A.; Islam, M.Z. LSTM Recurrent Neural Network Classifier for High Impedance Fault Detection in Solar PV Integrated Power System. *IEEE Access* **2021**, *9*, 32672–32687. [CrossRef]
61. Wang, F.; Zhen, Z.; Wang, B.; Mi, Z. Comparative Study on KNN and SVM Based Weather Classification Models for Day Ahead Short Term Solar PV Power Forecasting. *Appl. Sci.* **2018**, *8*, 28. [CrossRef]
62. Reddy, O.Y.; Chatterjee, S.; Chakraborty, A.K. Bilayered fault detection classification scheme for low-voltage DCmicrogrid with weighted KNN and decision tree. *Int. J. Green Energy* **2021**, *19*, 1149–1159. [CrossRef]
63. 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. [CrossRef]
64. Nie, Y.; Huang, Y.; Luo, W.; Zhou, W. Design of On-line Monitoring System for Photovoltaic Power Generation Power Quality Based on Stm32. *J. Phys. Conf. Ser.* **2023**, *2418*, 012021. [CrossRef]
65. Kim, M.; Kim, D.; Kim, H.; Prabakar, K. A Novel Strategy for Monitoring a PV Junction Box Based on LoRa in a 3 kW Residential PV System. *Electronics* **2022**, *11*, 709. [CrossRef]
66. Feng, L.; Amin, N.; Zhang, J.; Ding, K.; Hamelmann, F.U. Module-Level Performance Evaluation for a Smart PV System Based on Field Conditions. *Appl. Sci.* **2023**, *13*, 1448. [CrossRef]
67. E32-433T20DC User Manual SX1278 433MHz 100Mw dıp Wireless Module. Available online: https://mikroshop.ch/pdf/E32-433T20DC.pdf (accessed on 14 January 2026).
68. Badr, M.M.; Hamad, M.S.; Abdel-Khalik, A.S.; Hamdy, R.A.; Ahmed, S.; Hamdan, E. Fault Identification of Photovoltaic Array Based on Machine Learning Classifiers. *IEEE Access* **2021**, *9*, 159113–159132. [CrossRef]
69. Abdelsattar, M.; AbdelMoety, A.; Emad-Eldeen, A. Comparative Analysis of Machine Learning Techniques for Fault Detection in Solar Panel Systems. *Int. J. Eng. Sci. Appl.* **2024**, *5*, 140–152. [CrossRef]
70. Chapra, S.C.; Canale, R.P. *Numerical Methods for Engineers*, 8th ed.; McGraw-Hill Education: New York, NY, USA, 2021; p. 497.

71. Hernandez, J.C.; Vidal, P.G.; Jurado, F. Guidelines to requirements for protection against electric shock in PV generators. *IEEE Trans. Energy Convers.* **2009**, *24*, 274–282. [CrossRef]
72. Bouzidi, M.; Nasri, A.; Ouledali, O.; Hamouda, M. Optimising Solar Power Plant Reliability Using Neural Networks for Fault Detection and Diagnosis. *Elektron. Elektrotechnika* **2025**, *31*, 32–39. [CrossRef]
73. Li, Y.; Li, Z.; Wang, J.; Zeng, H. Analyses of driving factors on the spatial variations in regional eco-environmental quality using two types of species distribution models: A case study of Minjiang River Basin, China. *Ecol. Indic.* **2022**, *139*, 108980. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.