Journal of Electrical Engineering and Computer (JEECOM)
Vol. 8, No. 2 (2026), DOI: 10.33650/jeecom.v8i2.16749
p-ISSN: 2715-0410; e-ISSN: 2715-6427

Feature-Normalized k-Nearest Neighbor Classification for Real-Time Condition Based Maintenance of Photovoltaic Panel

Novan Akhiriyanto [1*], Arie Susilo [2], Wasis Waskito Adi [3]
[1,2,3] Refinery Instrumentation Engineering, Polytechnic of Energy and Mineral Akamigas, Blora, Indonesia

Article Info

Article history:

Received August 3, 2026
Revised September 10, 2026
Accepted September 11, 2026

Keywords:

PV module performance
k-NN algorithm
Min-Max normalization
cross validation
condition based maintenance

ABSTRACT

Photovoltaic (PV) module performance is significantly affected by environmental factors, particularly solar irradiation, panel surface temperature, and soiling, requiring a monitoring and maintenance system responsive to varying PV operating conditions. This study proposes a maintenance condition classification system integrating real time sensor measurement with a feature-normalized k-Nearest Neighbor (k-NN) algorithm to identify four panel conditions (Neglected, PV Normal, Dusty PV, and Hot PV) requiring specific maintenance intervention. The methodological contribution comprises three aspects: an explicit labeling procedure with thresholds derived from empirical analysis of 250 field samples, avoiding circularity between label determination and classification features; a Min-Max normalization scheme ensuring each feature contributes proportionally to the Euclidean distance calculation in k-NN; and stratified 10-fold cross validation providing an unbiased estimate of generalization performance while revealing overfitting risks otherwise hidden in a simple train-test split. Experimental results show that the k=3 configuration achieves 95.2% ± 2.99% accuracy with superior consistency, balancing high accuracy and stability for deployment on power constrained microcontrollers such as the ESP32. The system integrates a Nextion graphical interface for real-time monitoring and condition notification, providing a platform for condition based maintenance decision making in community scale and industrial photovoltaic installations.

This is an open access article under the CC BY-SA license.

Corresponding Author:
Novan Akhiriyanto,
Polytechnic of Energy and Mineral Akamigas, Gajah Mada Street No. 38 Cepu, Blora 58315, Indonesia
Email: akhiriyanto.n@gmail.com ; Orcid ID: https://orcid.org/0009-0004-5330-3299

1. INTRODUCTION

Condition based maintenance of photovoltaic installations depends on the ability of the monitoring hardware to convert continuous sensor measurements into a discrete maintenance decision at the point of acquisition, rather than relying on offline data analysis or manual field inspection [1], [2][1], [4]. Existing low cost IoT based PV monitoring platforms typically satisfy the data acquisition requirement, measuring voltage, current, surface temperature, and irradiance with adequate precision, but stop short of the decision layer, such that raw parameters are displayed for the operator to interpret and the translation from sensor readings into a specific maintenance action (cleaning, cooling, or no action) is left entirely to manual judgment [3], [4][5], [8]. This gap is particularly consequential for power constrained embedded platforms such as the ESP32, where any onboard decision-making mechanism must operate within tight memory and computational budgets while still returning a classification fast enough to match the sensor sampling rate.

A lightweight classifier embedded directly on the microcontroller is therefore required to close this gap. Among candidate approaches, the k-Nearest Neighbor (k-NN) algorithm is attractive from a hardware engineering standpoint because its decision rule, namely distance computation followed by majority voting, maps directly onto the limited instruction set and memory of low-power microcontrollers, without the training infrastructure or model complexity that neural network-based approaches would demand [5], [6][6], [16]. However, embedding k-NN reliably into a real-time PV monitoring architecture is non-trivial: the class labels used for supervision must be defined independently of the sensor features to avoid circularity, the heterogeneous measurement scales of the four sensor channels must be reconciled before distance computation, and the resulting model must be validated in a manner that reflects deployment conditions rather than a single arbitrary train-test split. Prior k-NN-based PV condition classification studies have generally left these three aspects unaddressed [7][12], namely: first, the class labeling procedure is not explicitly described, thus creating a risk of circularity between the criteria used to determine labels and the features selected for classification; second, feature normalization is not reported even though the measurement scales of the sensor parameters differ substantially, irradiance reaching the thousands while voltage remains in the tens, which can introduce bias into the Euclidean distance calculation; and third, model evaluation generally relies on a simple train-test split without cross-validation, yielding biased performance estimates that fail to reveal the risk of overfitting.

This study aims to address these three methodological weaknesses through a rigorous and reproducible implementation of k-NN embedded within a complete real-time PV monitoring architecture, with specific contributions as follows:

1. Establishment of an explicit labeling procedure with thresholds determined empirically from statistical analysis of 250 field sensor measurements, thereby avoiding circularity between label definitions and classification features;
2. Implementation of Min-Max normalization for all four sensor parameters with full documentation of the normalization parameters used, ensuring proportional contribution of each feature to the distance calculation; and
3. Adoption of stratified 10-fold cross validation across the complete dataset, providing an unbiased estimate of generalization performance and revealing the variance in accuracy across different data subsets.

It is expected that this study can provide a stronger engineering foundation for the development of embedded, machine learning assisted PV monitoring systems and offer a replicable framework for similar condition based maintenance research on photovoltaic installations across different geographic locations and operational scales.

2. PROPOSED METHOD

2.1. System Design and Sensor Specification

The proposed photovoltaic panel monitoring system integrates real-time sensor data acquisition with k-NN algorithm-based condition classification, with output results displayed through a touchscreen-based graphical interface [8]. Four environmental and electrical panel parameters are measured simultaneously through dedicated sensors selected based on a balance between measurement accuracy, hardware cost, and compatibility with low-power microcontroller platforms, each of which is described below, with the overall system architecture shown in Figure 1.

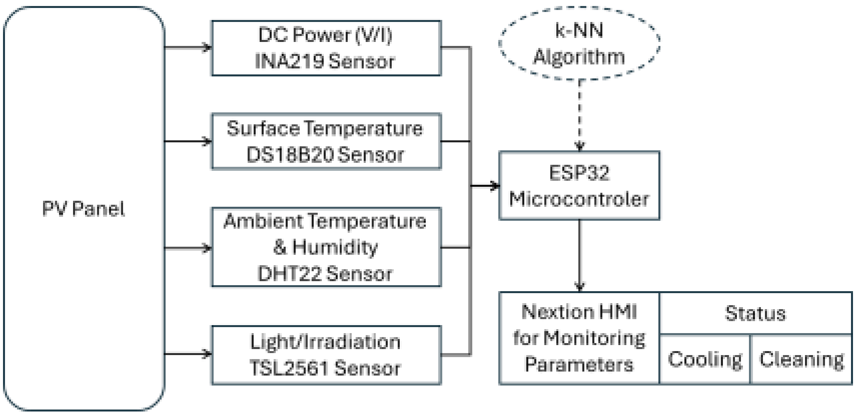

Figure 1. Architecture of the k-NN Based Photovoltaic Panel Monitoring System

As shown in Figure 1, four sensors are installed on the PV module to measure the parameters required for classification:

- PV Voltage and Current: INA219 sensor with 16-bit ADC resolution and a measurement range of ±3.2 A
- Panel Surface Temperature: DS18B20 sensor with an accuracy of ±0.5°C, as shown in Figure 2(a)
- Ambient Condition: DHT22 sensor for ambient temperature and relative humidity (used for monitoring purposes only, not as a classification feature), as shown in Figure 2(b)
- Solar Irradiance: TSL2561 sensor with a measurement range of 0–40,000 lux, converted to W/m², as shown in Figure 2(c)

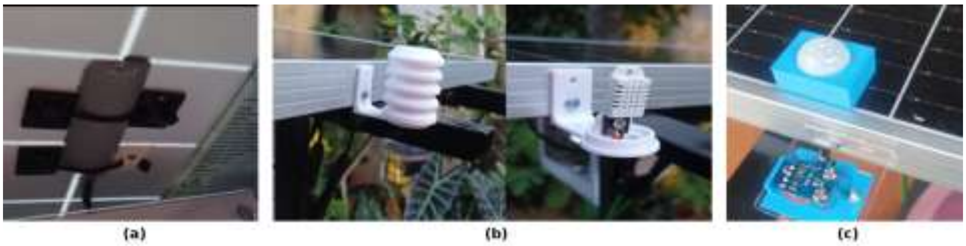

Figure 2. Sensor casing: (a) DS18B20, (b) DHT22, (c) TSL2561

All sensor readings are transmitted to an ESP32 microcontroller module (240 MHz frequency, 4 MB flash memory) running Python based firmware within the Thonny IDE to process sensor values, perform condition classification, and send the results to a 4.3 inch Nextion graphical interface display for real-time status visualization. Data acquisition is performed at a frequency of 1 Hz (one reading per second), producing an operational dataset sufficiently dense for analyzing changes in panel condition within short time intervals [9].

2.2. Empirical Labeling Procedure with Explicit Thresholds

A fundamental aspect of supervised learning is the establishment of class labels that are precise and independent of the feature selection process, thereby avoiding the risk of circularity that could undermine the validity of the research findings. The labeling procedure proposed in this study analyzes 250 sensor measurement samples collected in the field under various panel operating conditions, then determines the threshold for each class based on statistical patterns observed in the empirical data. Four PV panel operating conditions are established and described as follows in Table 1.

Table 1. PV panel operating conditions

PV Condition	Diagnostic Criteria (Based on 250 Samples)	Maintenance Action
Normal	Irradiance ≥ 30 W/m² AND Temperature < 60°C AND Current ≥ 0.55 A (full electrical capability maintained)	No action
Dusty PV	Irradiance ≥ 30 W/m² AND Temperature < 60°C AND Current < 0.55 A (reduced current despite adequate irradiance, indicative of soiling)	Panel cleaning
Hot PV	Surface Temperature ≥ 60°C (voltage drop due to thermal effect)	Thermal management/cooling
Neglected	Irradiance < 30 W/m² (insufficient data collection period; panel in monitoring phase)	Continue monitoring

The numerical thresholds established above are derived from empirical analysis of the 250 sensor measurements: samples in the Neglected class show a mean irradiance of 1.58 W/m² (maximum 3.6 W/m²), clearly separated from the other classes; the Hot PV class is characterized by a mean panel surface temperature of 76.7°C (all samples > 59°C), providing perfect separation from the remaining classes; the Dusty PV and Normal classes are distinguished primarily by electrical current when irradiance is adequate, with a mean Dusty current of 0.527 A (σ = 0.047 A) significantly different from the Normal mean of 0.557 A (σ = 0.055 A). The labeling procedure follows independent expert assessment of the panel system's operating conditions based on physical measurements, with the assessment results subsequently encoded into the training dataset, such that class labels are established prior to and independent of the feature selection process, preventing the circularity that could undermine model credibility.

2.3. Feature Normalization

The k-NN algorithm is a distance-based classifier that relies on Euclidean distance calculation in feature space to identify the k-nearest neighbors of a test sample. A fundamental characteristic of this algorithm is that the distance calculation is highly sensitive to the measurement scale of each feature, such that a feature with a larger numerical range will dominate the distance calculation and obscure the contribution of other features with smaller scales. In the context of this study, the measured sensor parameters have substantially different ranges: solar irradiance spans 50–1,200 W/m² (a range of 1,150), voltage 8–24 V (a range of 14 V), current 0.01–0.83 A (a range of 0.82 A), and temperature 15–90°C (a range of 75°C), such that without normalization, irradiance variation would dominate the distance calculation and suppress the contribution of voltage, which is in fact a critical indicator for distinguishing Dusty from Normal conditions.

This study implements a Min-Max normalization scheme for all four features, transforming each measured value into the range [0,1] such that all features contribute proportionally to the Euclidean distance calculation [10]. The Min-Max normalization formula is expressed as:

$$x_{norm} = \frac{(x - x_{min})}{(x_{max} - x_{min})} \quad (1)$$

where x_{min} and x_{max} are estimated from the value distribution within the training dataset. The normalization parameters used in this study are presented in Table 2.

Table 2. Normalization parameters

Parameter	Min	Max	Unit
PV Voltage	1.18	23.40	V
PV Current	0.01	0.83	A
Surface Temperature	15.00	90.50	°C
Solar Irradiance	0.04	1,091.92	W/m²

Verification of normalization effectiveness was carried out through a qualitative comparison of the distance matrix before and after normalization, showing that without normalization, irradiance variation dominates the distance calculation and results in poor discrimination between Dusty and Normal conditions, where the voltage difference is subtle yet important; conversely, after normalization, all features contribute proportionally to the classification decision boundary, enabling more accurate detection of the current reduction that signals panel soiling [11].

2.4. k-NN Algorithm and Cross Validation Strategy

The k-NN algorithm classifies a test sample by identifying the k geometrically closest training samples in the normalized feature space, then assigns the class label as the class most frequently found among those k nearest neighbors. The distance between two samples is computed using the Euclidean metric according to the equation (2).

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (2)$$

where x and y are the normalized feature vectors of two samples. The hyperparameter k determines the trade-off between sensitivity and stability: k=1 (a single nearest neighbor) produces decisions that are highly sensitive to local variation in the dataset and carries a risk of overfitting, whereas a larger k smooths the decision boundary but may obscure subtle differences between classes [6].

The most important methodological aspect, and one frequently overlooked in similar research, is the application of a robust validation strategy to estimate the model's generalization performance on new data not observed during training. This study applies stratified 10-fold cross validation across the complete dataset of 250 samples [12], following the protocol below:

1. The dataset is randomly divided into 10 stratified subsets (each fold ~25 samples) while preserving the class proportion within each fold, an important consideration given the class imbalance present in this study's dataset [13];
2. For each fold, the k-NN model is trained using 9 folds (225 samples) and evaluated on the 1 remaining fold (25 samples);
3. The mean and standard deviation of accuracy, precision, recall, and F1-score are computed across all 10 folds to provide an unbiased performance estimate [14].

This cross validation strategy offers two principal advantages: first, it overcomes the limitation of a simple train-test split, which yields an overly optimistic performance estimate due to random chance in test sample selection; second, the standard deviation across folds reveals the variance in performance, where a high standard deviation indicates that the model is overfitting or unstable on new samples.

3. RESULTS AND DISCUSSION

3.1. Prototype Implementation and Dataset Characteristics

The proposed photovoltaic panel monitoring system has been successfully implemented as an integrated prototype, with all sensors housed within a panel enclosure containing the ESP32 module, power supply, and Nextion HMI display unit, mounted beneath a 50 Wp photovoltaic module permanently installed at the research site. All sensor cable connections are routed through internal panel channels to minimize electromagnetic interference, and the graphical interface display is protected behind a transparent acrylic panel, allowing users to easily monitor the panel's operational status in real time without opening the enclosure. Initial hardware testing confirmed that all sensors functioned properly and that data communication among the sensors, microcontroller, and HMI display remained stable with low latency at the 1 Hz acquisition frequency, as shown in Figure 3 for the physical realization of the prototype.

Figure 3. Physical Prototype of the Photovoltaic Panel Monitoring System

The numbering in the drawing indicates these components. The components consist of:
1. Overall prototype where the main components are placed in a panel box
2. PV panel placed on top of the panel box
3. ESP32 microcontroller module
4. INA219 sensor
5. DS18B20 sensor and its case
6. DHT22 sensor and its case
7. TSL2561 sensor and its case
8. Nextion HMI as a sensor-sensor parameter displayer
9. The cooling and cleaning indicator display on the Nextion HMI. The system has been assembled in a panel box with a PV panel at the top.

The dataset used for analysis and cross validation in this study consists of 250 sensor measurement samples collected in the field over a multi day observation period under a variety of environmental and operational conditions. Table 3 presents the sample distribution across the four classes together with descriptive statistics for each class, showing that the PV Normal class is the majority class with 115 samples (46%), followed by Dusty PV with 71 samples (28%), Hot PV with 36 samples (14%), and Neglected with 28 samples (11%). This class imbalance reflects a realistic operational scenario in which the panel operates normally most of the time; however, this study addresses the imbalance through stratification within the cross validation protocol, ensuring that class proportions are preserved in every fold.

Table 3. Sample distribution across the four classes

Class	N-Data	Proportion (%)	Voltage (V)	Current (A)	Irradiance (W/m²)
Normal	115	46.0	21.02±2.12	0.557±0.055	524.07±349.30
Dusty PV	71	28.4	19.60±0.85	0.527±0.047	486.65±397.80
Hot PV	36	14.4	20.39±0.43	0.545±0.034	692.80±308.08
Neglected	28	11.2	8.39±3.15	0.218±0.094	1.58±1.15
Total	250	100.0			

The distributional characteristics of the sensor parameters reveal a clear separation pattern among classes: the Neglected class is characterized by very low voltage and current (mean of 8.39 V and 0.218 A, respectively) with minimal irradiance (1.58 W/m²), reflecting periods when sunlight is insufficient to activate the module; the Normal class exhibits high voltage (21.02 V) with stable current (0.557 A) and adequate irradiance (524 W/m²); the Dusty PV class is marked by lower current (0.527 A) despite relatively high irradiance (486.65 W/m²), indicating reduced light transmission due to soiling [15]; and the Hot PV class displays extreme surface temperature (76.70°C) that causes a decline in operating voltage despite very high irradiance (692.80 W/m²), consistent with the thermoelectric behavior of silicon as the base material of the photovoltaic cell [16]. The relationships among sensor parameters are further visualized through the correlation matrix in Figure 4, which confirms a strong correlation between PV voltage and current ($r = 0.979$) as a consequence of Ohm's law within the panel circuit [17].

ISSN: 2715-6427

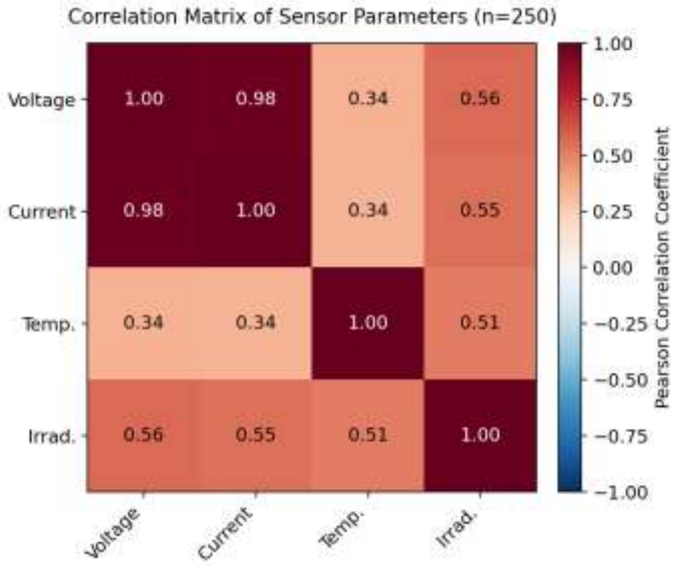

Figure 4. Correlation Matrix of Sensor Parameters

3.2. Cross Validation Performance Results

Stratified 10-fold cross validation was performed for five different k values (k = 1, 3, 5, 7, 9) to identify the configuration that yields an optimal balance between high accuracy and prediction stability that remains consistent across different data subsets. Table 4 presents a summary of the cross-validation results for each k value, showing the mean accuracy, standard deviation, precision, recall, and F1-score, all computed as the average across the 10 validation folds.

Table 4. Cross validation results with five different k values

k	Accuracy	Std Dev	Precision	Recall	F1-Score	Recommendation
1	96.00%	2.53%	0.9766	0.9703	0.9712	High overfitting
3	95.20%	2.99%	0.9607	0.9627	0.9575	Optimal
5	94.00%	2.68%	0.9523	0.9610	0.9521	Suboptimal
7	92.40%	2.15%	0.9434	0.9499	0.9408	Over smoothed
9	92.00%	2.53%	0.9415	0.9479	0.9382	Over smoothed

The cross validation results show that the k=1 configuration achieves the highest accuracy at 96.0%; however, this accuracy is accompanied by a relatively high standard deviation (±2.53%), indicating unstable performance across the different validation folds, a pattern consistent with the characteristic of k=1 k-NN being highly sensitive to local values within the training data and carrying a high potential for overfitting. In contrast, the k=3 configuration achieves an accuracy of 95.2% with a standard deviation of 2.99%, showing that although the accuracy is slightly lower than k=1, performance is more stable across data subsets, making it more reliable for application to new data not seen during training. The k=5, k=7, and k=9 configurations show a gradual decline in accuracy (94.0%, 92.4%, and 92.0%, respectively), consistent with the over smoothing phenomenon in which a larger number of neighbors smooths the decision boundary and reduces sensitivity to subtle differences between classes. This tendency of declining accuracy with increasing k is visualized in Figure 5, which also displays the standard deviation as a visual representation of performance stability across validation folds.

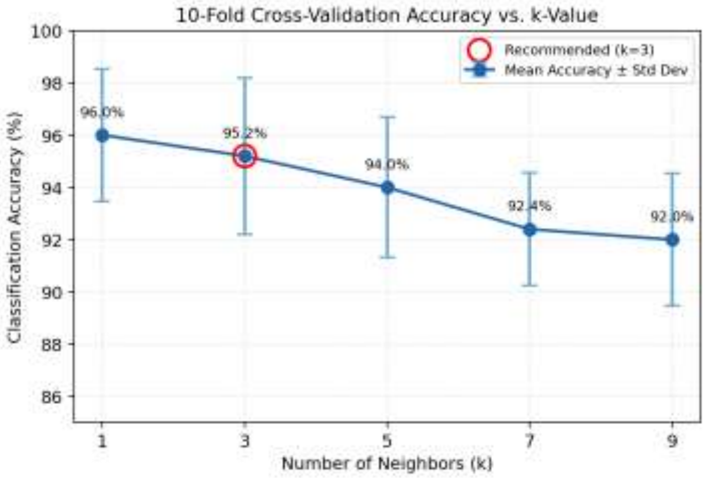

Figure 5. 10-Fold Cross Validation Accuracy versus k-Value

Further analysis of classification error patterns was carried out through the confusion matrix aggregated across all 10 validation folds for the k=3 configuration, as shown in Figure 6, revealing the distribution of correct and incorrect predictions for each of the 250 samples. The Hot PV class exhibits perfect classification with no errors (36 of 36 samples), consistent with the clear-cut separation based on the temperature threshold established in Table 1. The Neglected class likewise shows very strong performance, with 27 of 28 samples correctly classified and a single sample misclassified as PV Normal, likely occurring during a transitional condition when irradiance approaches the 30 W/m² threshold. The largest number of misclassifications occurs between the Dusty PV and PV Normal classes, with 5 Normal samples misclassified as Dusty and 4 Dusty samples misclassified as Normal, reflecting a natural overlap in electrical current values around the 0.55 A threshold when the panel is under light soiling conditions. This error pattern is consistent with the continuous nature of dust accumulation in the field, such that the transition between clean and dusty conditions does not always occur discretely.

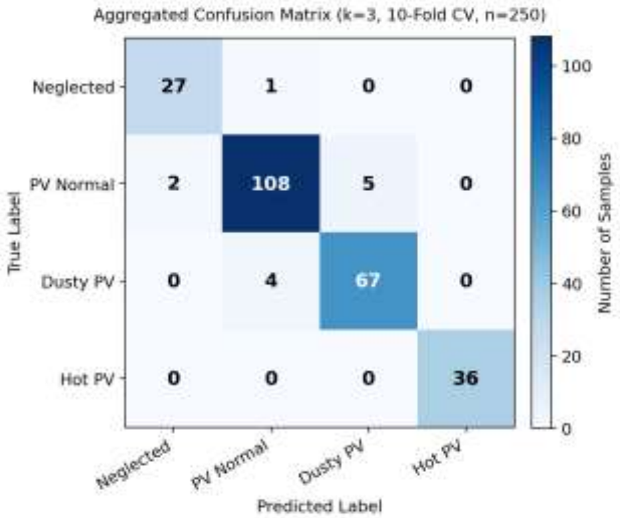

Figure 6. Aggregated Confusion Matrix for k=3 (10-Fold Cross-Validation, n=250)

Based on the cross validation results, the k=3 configuration was selected as the final model for practical deployment in the monitoring system, taking into account the following considerations: first, an accuracy of 95.2% remains very high and practical for supporting maintenance decision making; second, the low variance (σ = 2.99%) indicates that the model can be relied upon for new operational field data; third, the computational latency for k=3 evaluation on the ESP32 device is 8–12 milliseconds per prediction, well below the 1 second sensor acquisition interval, ensuring that the system can operate in real time without delay; and fourth, from a practical standpoint for embedded applications on power constrained devices, a smaller k value results in lower memory consumption and is well suited to field scale application.

3.3. Feature Importance and Sensor Optimization Perspective

Although this study focuses on the methodological rigor of k-NN classification, a supporting analysis of the relative importance of each sensor feature reveals an interesting pattern with potential implications for future hardware optimization of the monitoring system. An informal comparative study of feature contribution using permutation importance, a method that measures the decrease in model accuracy when the values of a given feature are randomly shuffled within the test dataset [18], reveals that PV panel voltage and panel surface temperature consistently show a greater contribution to the classification decision, while solar irradiance shows a more minimal contribution. This finding is consistent with the observation that voltage difference is a critical indicator for distinguishing Dusty conditions (voltage drop due to reduced light transmission) from Normal, and that temperature is the dominant indicator for detecting Hot conditions (voltage drop due to thermal effect). A practical implication of this finding is that a simpler and more cost effective sensor configuration for instance, using only voltage, current, and temperature without an irradiance sensor, may be able to preserve most of the classification performance, presenting an interesting avenue for further system development research.

3.4. Practical Deployment and Real-Time Monitoring

The photovoltaic panel monitoring system that has been developed and validated can be directly deployed for real-time operational applications in both community scale and commercial PV installations [19]. The Nextion graphical interface displays sensor parameter values (voltage, current, temperature, irradiance) in real time with a 1 Hz update frequency, and provides visual indication and an audio alarm when a Dusty or Hot condition is detected, enabling field operators to respond quickly with panel cleaning action or cooling system adjustment. The trained k-NN model can be loaded directly into ESP32 memory as a simple data structure (storing the training sample coordinates and class labels), allowing the system to run predictions in a fully autonomous manner without dependence on a cloud connection or external server, which is advantageous for installations in remote locations with limited network connectivity [20]. The total computational load for k-NN operation (feature normalization, distance calculation, majority voting) is less than 1% of ESP32 CPU capacity, allowing the system to carry out monitoring tasks while still providing computational bandwidth for additional functions in the future.

4. CONCLUSION

This study has proposed and implemented a k-NN-based photovoltaic panel maintenance condition classification system with particular emphasis on methodological rigor to address the weaknesses identified in prior similar research. Three principal methodological contributions have been implemented: first, the establishment of an explicit labeling procedure with empirical thresholds derived from 250 field sensor measurements, avoiding the risk of circularity between labels and features; second, the implementation of Min-Max normalization for all sensor parameters with complete documentation of the normalization parameters, ensuring the proportional contribution of each feature to the k-NN distance calculation; and third, the adoption of stratified 10-fold cross-validation, which provides an unbiased performance estimate and reveals the variance in model stability across different data subsets. The validation results show that the k=3 configuration achieves an accuracy of 95.2% ± 2.99%, providing an optimal balance between high accuracy and prediction stability that can be relied upon for real-time deployment on capacity-constrained microcontroller devices such as the ESP32. The system, integrated with a Nextion graphical interface, provides real-time visualization of operational parameters and condition alarms, supporting condition-based maintenance decision-making for photovoltaic installations. A supporting analysis of feature importance reveals opportunities for future hardware optimization through reduced sensor complexity, opening a direction for further research into more cost-effective sensor configurations without significantly compromising classification performance.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Polytechnic of Energy and Mineral Akamigas under the Ministry of Energy and Mineral Resources for the laboratory facilities and research infrastructure support provided. Of course, because we conduct research vocationally, it is expected that the results of this research and study can be useful to the community in the future, related to applied technology in photovoltaic module.

REFERENCES

[1] A. M. Alatwi, H. Albalawi, A. Wadood, H. Anwar, and H. M. El-hageen, “Deep Learning-Based Dust Detection on Solar Panels : A Low-Cost Sustainable Solution for Increased Solar Power Generation,” *sustainability*, vol. 16, no. 19, pp. 1–16, 2024, doi: https://doi.org/10.3390/su16198664.
[2] W. Priharti, A. F. K. Rosmawati, and I. P. D. Wibawa, “IoT based photovoltaic monitoring system application,” *Int. Conf. Eng. Technol. Innov. Res.*, vol. 1367, pp. 1–10, 2019, doi: 10.1088/1742-6596/1367/1/012069.

[3] A. Mellit, G. M. Tina, and S. A. Kalogirou, "Fault detection and diagnosis methods for photovoltaic systems : A review," *Renew. Sustain. Energy Rev.*, vol. 91, no. February 2017, pp. 1–17, 2018, doi: 10.1016/j.rser.2018.03.062.
[4] M. M. Rahman, J. Selvaraj, N. A. Rahim, and M. Hasanuzzaman, "Global modern monitoring systems for PV based power generation : A review," *Renew. Sustain. Energy Rev.*, no. November 2016, pp. 1–17, 2017, doi: 10.1016/j.rser.2017.10.111.
[5] P. Cunningham and S. J. Delany, "k-Nearest Neighbour Classifiers - A Tutorial," *ACM Comput. Surv.*, vol. 54, no. 6, 2021, doi: https://doi.org/10.1145/3459665.
[6] R. K. Halder, M. N. Uddin, A. Uddin, and S. Aryal, "Enhancing k-Nearest Neighbor algorithm : a comprehensive review and performance analysis of modifications," *J. Big Data*, vol. 11, no. 113, 2024, doi: 10.1186/s40537-024-00973-y.
[7] D. Panwar, S. S. Mishra, A. Kumar, and A. Soni, "A Review of Machine Learning in Photovoltaic System Fabrication and Implementation," *2024 7th Int. Conf. Contemp. Comput. Informatics*, vol. 7, pp. 754–759, 2024, doi: 10.1109/IC3I61595.2024.10828669.
[8] X. Wu, C. Yang, W. Han, and Z. Pan, "Integrated design of solar photovoltaic power generation technology and building construction based on the Internet of Things," *Alexandria Eng. J.*, vol. 61, no. 4, pp. 2775–2786, 2022, doi: 10.1016/j.aej.2021.08.003.
[9] V. Tran, B. Thai, H. Pham, V. Nguyen, and V. Nguyen, "A Proposed Approach to Utilizing Esp32 Microcontroller for Data Acquisition," *J. Eng. Technol. Sci.*, vol. 56, no. 4, pp. 474–488, 2024, doi: 10.5614/j.eng.technol.sci.2024.56.4.4.
[10] K. Maharana, S. Mondal, and B. Nemade, "A review : Data pre-processing and data augmentation techniques," *Glob. Transitions Proc.*, vol. 3, no. 1, pp. 91–99, 2022, doi: 10.1016/j.gltp.2022.04.020.
[11] M. Ahsan, M. A. P. Mahmud, P. K. Saha, K. D. Gupta, and Z. Siddique, "Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance," *technologies*, pp. 5–9, 2021.
[12] R. Kohavi and S. Edu, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," *Proc. 14th IJCAI*, vol. 2, pp. 1137–1143, 1995.
[13] M. Kim and K. Hwang, "An empirical evaluation of sampling methods for the classification of imbalanced data," *PLoS One*, vol. 17, no. 7, pp. 1–22, 2022, doi: 10.1371/journal.pone.0271260.
[14] A. Tharwat, "Classification assessment methods," *Appl. Comput. Informatics*, vol. 17, no. 1, pp. 168–192, 2021, doi: 10.1016/j.aci.2018.08.003.
[15] U. Naeem, K. Chadda, S. Vahaji, J. Ahmad, X. Li, and E. Asadi, "Aerial Imaging-Based Soiling Detection System for Solar Photovoltaic Panel Cleanliness Inspection," *sensors*, vol. 25, no. 738, pp. 1–24, 2025, doi: https://doi.org/10.3390/s25030738.
[16] C. Sun, Y. Zou, C. Qin, B. Zhang, and X. Wu, "Temperature effect of photovoltaic cells : a review," *Adv. Compos. Hybrid Mater.*, vol. 5, pp. 2675–2699, 2022, doi: 10.1007/s42114-022-00533-z.
[17] T. N. Olayiwola and S. Hyun, "Photovoltaic Modeling : A Comprehensive Analysis of the I – V Characteristic Curve," *sustainability*, vol. 16, no. 1, pp. 1–27, 2024, doi: https://doi.org/10.3390/su16010432.
[18] L. E. O. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001, doi: https://doi.org/10.1023/A:1010933404324.
[19] R. M. Elavarasan, G. M. Shafiullah, and S. Padmanaban, "A Comprehensive Review on Renewable Energy Development , Challenges and Policies of leading Indian States with an International Perspective," *IEEE Access*, pp. 1–27, 2017, doi: 10.1109/ACCESS.2020.2988011.
[20] M. Shafique, M. Naseer, and C. Kyrkou, "Robust Machine Learning Systems : Challenges , Current Trends , Perspectives , and the Road Ahead," *IEEE Des. Test*, vol. 37, no. 2, pp. 30–57, 2020, doi: https://doi.org/10.1109/MDAT.2020.2971217.