IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 5, Issue 5, June 2025 Impact Factor: 7.67

Deep Learning-Based Detection of Solar Panel Faults

Gurhal Gayatri Ram, Dhage Akshay Namdev, Suryawanshi Nandkishor Bapu, Prof. Y. G. Suryawanshi

Department of Computer Engineering

SGVSS Adsul Technical Campus, Chas, Ahmednagar

Abstract: *Lung Solar photovoltaic (PV) systems play a crucial role in the global transition to renewable energy. However, these systems are susceptible to various operational anomalies and environmental factors such as dust accumulation, shading, panel aging, and electrical malfunctions that can significantly degrade their performance. Traditional fault detection methods involving manual inspection and fixed-threshold monitoring are inefficient, especially in large-scale PV installations. With the increasing adoption of solar power and its integration into smart grids, there is a pressing need for intelligent and automated systems capable of detecting faults accurately and in real time.*

This paper presents a novel solution that combines time-series sensor data with a Convolutional Neural Network (CNN) to identify and classify solar panel faults. The system leverages data preprocessing techniques including normalization and sliding window segmentation to transform raw sensor readings into structured input for deep learning. A custom CNN architecture is then trained to detect fault patterns and deployed via a user-friendly Streamlit interface that allows real-time predictions from uploaded CSV files. Experimental results demonstrate high accuracy, precision, and recall, highlighting the model's practical utility in real-world solar farm environments. This approach provides a scalable, data-driven alternative to traditional inspection and enables proactive maintenance in renewable energy systems.

Keywords: Deep Learning, Convolutional Neural Networks, Solar Panel Fault Detection, Time-Series Data, Streamlit, Sliding Window, Renewable Energy

I. INTRODUCTION

Solar energy is one of the most sustainable and clean sources of electricity. As countries strive to reduce their carbon emissions and transition to green energy, the use of photovoltaic (PV) systems has gained significant momentum. These systems convert sunlight into electrical energy and are widely used in residential, commercial, and industrial applications due to their environmental and economic benefits.

Despite their advantages, solar panels are subject to various faults and degradation over time. These can arise from environmental factors such as dust, humidity, and shading, or operational issues like thermal imbalance, hotspots, and physical damage. Identifying and addressing these faults promptly is essential to maintain optimal energy production and extend the lifespan of the panels.

Traditional approaches to fault detection involve manual inspection or threshold-based monitoring using simple rule sets. However, these methods are not scalable, especially in large solar farms. They also lack accuracy and often result in false alarms or missed detections, reducing overall system efficiency.

With the advancement in artificial intelligence, particularly deep learning techniques, it has become possible to automate the fault detection process using data-driven models. Convolutional Neural Networks (CNNs), known for their powerful feature extraction capabilities, are particularly suitable for analyzing structured and unstructured data.

In this project, we propose a CNN-based approach that utilizes multivariate time-series data collected from solar panel sensors. The model learns temporal patterns from the data to classify the operational status of the panels. This is achieved through sliding window segmentation, which restructures the time-series data into sample sequences suitable for CNN input.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, June 2025
Impact Factor: 7.67

We also integrate our model into a real-time web application using Streamlit, allowing users to upload CSV files with sensor data and obtain instant diagnostic predictions. This ensures accessibility and usability for technicians and solar farm operators without technical expertise in machine learning.

The model is evaluated using metrics such as accuracy, precision, recall, and F1-score. The results demonstrate that our approach provides reliable fault classification performance and can be used to support predictive maintenance strategies.

Overall, the proposed system offers a cost-effective, scalable, and intelligent solution for improving the reliability and performance of solar energy installations.

Furthermore, the increased dependency on renewable energy demands robust maintenance tools that reduce manual effort. Leveraging machine learning for solar fault detection not only cuts down on labor costs but also significantly enhances the accuracy and speed of identifying system anomalies.

This solution is particularly beneficial in remote or large- scale solar farms where human monitoring is not feasible on a regular basis. The ability to remotely upload sensor logs and receive instant diagnostics helps operations teams manage energy assets more efficiently.

As global temperatures rise and energy demands increase, scalable diagnostic systems like this will be critical for future-proofing renewable energy infrastructure. Investing in such technologies ensures reliability, efficiency, and sustainability at a larger scale.

By building on open-source tools and scalable architectures, this model serves as a foundational prototype for future intelligent energy monitoring platforms that can integrate both sensor and visual data in real-time environments.

OBJECTIVE

- To develop an AI-based model for the accurate detection and classification of lung infections using deep learning techniques.
- To analyze the performance of CNN architectures in identifying lung diseases from X-ray and CT scan images.
- To enhance diagnostic accuracy by implementing preprocessing techniques such as noise reduction and data augmentation.
- To evaluate the efficiency of transfer learning in improving model performance with limited labeled medical datasets.
- To promote the integration of AI in healthcare by addressing challenges related to interpretability, reliability, and ethical considerations

II. LITERATURE SURVEY

No.	Title	Authors	Journal/C onference	Key Contributions
1	A Novel CNN-Based Approach for Fault Classificati on in PV Arrays	F. Aziz et al.	IEEE Transactio ns on Energy Systems	- Introduced CNN model for classifying various PV faults using thermal and visual imagery.
2	CNN-Based Fault Detection in Solar Panels	R. Gupta, P. Verma	International Journal of AI Research	Used CNNs trained on thermal images to detect faults like shading and panel degradation
3	YOLOv3-Tiny Optimizatio n for Real-Time Fault Detection	X.Zhan g et al.	IEEE Access	- Enhanced YOLOv3-tiny for real-time UAV-based thermal fault detection in solar fields.
4	Machin e Learnin g- Based Solar Panel Monito ring	R. M. Perez et al.	IET Renewable Power Generation	- Compared SVM and decision trees for image-based solar panel detection and monitoring.
5	Deep CNN for Polyp Detection in Colonoscop y	T. Rahim et al.	Journal of Medical Imaging	- - Applied deep CNN for medical imaging, inspiring similar methods

	Images			for solar panel anomaly detection.
6	Real-World UAV Dataset for Object Detection	M. Ł. Pawełczyk, M. Wojtyra	Proceedings of UAV Applications Conf	- Created UAV dataset for object tracking, useful in solar fault localization. -
7	Improv ement of Solar Panel Fault Detecti on via UAV + Deep Learnin g	J. D. Kim, H. S. Kim, S. H. Lee	Renewable Energy Conference	- Demonstrated UAV-based thermal inspection with DL for improved solar panel fault localization.

III. PROPOSED SYSTEM

The proposed system aims to develop an automated and intelligent solar panel fault detection framework using deep learning techniques. By leveraging Convolutional Neural Networks (CNNs), the system analyzes time-series sensor data to identify faults such as low voltage, overcurrent, temperature imbalances, and anomalies in irradiance levels. The model is trained on structured sensor datasets collected from PV installations, ensuring it learns temporal and statistical patterns indicative of performance degradation or operational issues. This AI-driven approach minimizes manual effort and human error, offering a rapid, accurate, and cost-effective solution for detecting solar panel faults.

To enhance model performance, the system implements advanced preprocessing techniques including missing value handling, scaling via StandardScaler, and noise filtering. These steps improve data quality and ensure effective feature extraction by the CNN. The system employs a sliding window mechanism to segment continuous sensor streams into overlapping sequences, allowing the model to learn temporal relationships across time steps. This segmentation approach is crucial for enabling the CNN to recognize evolving fault patterns in dynamic environments.

The architecture of the proposed system integrates three core functional layers: preprocessing, modeling, and deployment. The preprocessing layer prepares data through cleaning and normalization. The modeling layer incorporates a CNN with Conv1D, MaxPooling, Dropout, Flatten, and Dense layers, designed to perform multi-class classification. The deployment layer utilizes a real-time web application built with Streamlit, enabling users to upload sensor logs, trigger the prediction pipeline, and receive visualized results

The CNN model is compiled using the Adam optimizer with a learning rate of 0.0001 and trained using categorical cross-entropy loss. The training dataset is segmented into training and validation splits to monitor performance. Evaluation metrics such as accuracy, precision, recall, and F1-score are used to assess the model's effectiveness. The model is saved in .h5 format, and auxiliary components such as scaler.pkl and encoder.pkl are exported to ensure consistent inference.

Additionally, a user-facing interface developed using Streamlit allows real-time interaction with the model. Users can upload .csv sensor files, which are automatically preprocessed and segmented. Predictions are displayed on- screen with confidence scores, helping technicians or plant managers make informed decisions. This interface is designed to run efficiently on low-resource systems and is ideal for field diagnostics.

Ethical considerations include data integrity, explainability, and model fairness. By logging predictions and allowing manual verification, the system supports transparent AI deployment in the renewable energy sector. Future updates may include integration with SCADA systems or IoT devices for real-time monitoring at scale.

Fig.1 System Architecture

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 5, Issue 5, June 2025 Impact Factor: 7.67

Datasets:

The proposed system is trained and evaluated using both simulated and real-world datasets obtained from solar photovoltaic (PV) installations. These datasets comprise multivariate time-series readings collected via sensors placed on different components of the solar system. Parameters include voltage, current, irradiance, temperature, and fault condition labels. Each record represents the health and behavior of a solar panel over a defined period, enabling supervised learning of fault patterns.

Datasets are structured in .csv format with each row containing one instance of sensor readings. For supervised learning, a categorical fault label is included for classification purposes. The dataset is manually inspected to identify missing or corrupted entries and to confirm the distribution of fault classes. This careful curation of training data ensures quality and relevance in model learning.

For experimental validation, additional synthetic datasets were generated by simulating faults in lab conditions. These synthetic samples mimic real-world anomalies and extend the diversity of the dataset. This approach ensures the model generalizes well to previously unseen conditions while maintaining high precision

All datasets are in .csv format, with each row representing a unique instance of sensor readings. A categorical fault label is embedded in each row for classification in supervised learning tasks. To ensure data quality and relevance for model learning, the datasets undergo manual inspection to pinpoint and address any missing or corrupted entries, and to verify the distribution of fault classes

Preprocessing and Data Augmentation:

Sensor data often includes noise, missing entries, and inconsistent sampling intervals. To address these issues, the preprocessing phase removes non-essential columns such as timestamps, applies forward or backward filling for missing values, and performs feature normalization using StandardScaler. This ensures that all features contribute equally to the learning process and improves model stability.

A sliding window segmentation technique is applied to convert time-series data into sequences. Each sequence is reshaped to match the input shape required by the CNN model (samples, time steps, features). This transformation captures temporal dependencies between readings and allows the model to detect gradual changes leading up to faults.

To expand the training dataset and improve generalization, several data augmentation strategies are employed. These include overlapping windows, random shuffling of within-class samples, and minor perturbations to feature values within permissible ranges. These synthetic variations make the model robust to sensor variability and enhance its performance in real-world deployments.

Deep Learning Algorithms:

The project integrates three distinct deep learning models: a **Sequential Model**, a **Functional Model**, and a **Pretrained Model (Transfer Learning)** using VGG-16.

Sequential Model:

The sequential model follows a layer-by- layer approach, where each layer's output serves as the input for the subsequent layer.

It consists of convolutional layers for feature extraction, followed by pooling layers to reduce spatial dimensions.

The model is optimized using the Adam optimizer with a learning rate of 0.0001.

Functional Model:

Unlike sequential models, the functional model allows for flexible connections between layers, enabling the design of complex architectures.

This model comprises two initial convolutional layers with different kernel sizes, processing inputs independently.

The extracted features are then combined and passed through five additional convolutional layers with a 3 × 3 kernel size.

Similar to the sequential model, it employs the Adam optimizer with a learning rate of 0.0001.

Pretrained Model (Transfer Learning - VGG- 16):

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, June 2025
Impact Factor: 7.67

The The project lays the groundwork for incorporating pretrained models such as VGG-16 for use in hybrid multimodal systems.

These models may process image-based diagnostics or sensor-encoded visuals, enhancing classification capabilities with learned weights from large-scale datasets.

Transfer learning significantly reduces training time and can offer higher accuracy with fewer labeled samples

VGG-16 was selected due to its strong performance in the ImageNet competition

This approach combines deep learning with robust preprocessing and deployment pipelines to form a scalable, real-time diagnostic solution for solar fault detection.

IV. RESULT & DISCUSSION

The performance of the proposed solar fault detection system was evaluated through a series of experiments using pre-segmented time-series sensor data. The Sequential CNN model was trained on a curated dataset with diverse fault classes and validated using unseen samples to measure generalization. Metrics such as accuracy, precision, recall, and F1-score were computed to assess classification quality.

During training, the model showed rapid convergence with minimal overfitting, thanks to data augmentation and dropout regularization. The validation accuracy plateaued at 94.8%, indicating strong learning without performance degradation. The loss curve demonstrated a smooth decline, confirming the model's stability during optimization.

The confusion matrix revealed that the model was particularly effective at distinguishing between "normal" and "low voltage" states, with some marginal confusion between overlapping fault classes such as "high temperature" and "irradiance drop." The precision score averaged 92.6%, recall stood at 93.7%, and the F1-score was 93.1%, suggesting balanced performance across all fault types.

The model's performance was further validated through real- time testing within the Streamlit application. Multiple .csv files with sensor readings were uploaded, processed, and analyzed through the interface. Predictions were consistently accurate and accompanied by clear class labels and confidence scores, demonstrating the system's practical utility.

A key advantage of the deployed system is its usability. Unlike traditional threshold-based diagnostics, this AI-based approach does not require hard-coded rules or manual calibration. The CNN model automatically learns relevant temporal patterns, making it adaptive to a wide variety of operational scenarios

The result analysis also revealed the robustness of the preprocessing and segmentation pipeline. Even minor shifts in data or noise injections did not significantly degrade the model's prediction quality. This resilience is attributed to the sliding window technique and normalization, which helped maintain consistency in feature distribution

Uploaded Data:

	Time	Ipv	Vpv	Vdc	ia	ib	ic	va	vb	vc
0	0.00003	1.5723	101.3489	144.1406	-0.1351	0.4901	-0.355	41.7445	-149.8729	109.0646
1	0.0001	1.5033	101.4587	143.5547	-0.1083	0.5103	-0.3886	46.8315	-150.7167	105.83
2	0.0002	1.4929	101.5747	143.5547	-0.1687	0.4968	-0.3348	51.0747	-152.0186	102.5431
3	0.0003	1.5581	101.3123	143.2617	-0.1351	0.5103	-0.3617	55.8482	-152.5851	98.1433
4	0.0004	1.6319	101.1414	143.8477	-0.2023	0.5035	-0.3214	60.0552	-152.6093	94.2617
5	0.0005	1.6073	101.0132	143.8477	-0.1687	0.5103	-0.3483	64.865	-154.1522	90.2757
6	0.0006	1.5004	100.9338	143.8477	-0.2358	0.5035	-0.308	69.5903	-154.8152	86.5991
7	0.0007	1.4456	101.0437	143.8477	-0.1956	0.5103	-0.3214	74.4844	-154.8514	83.1917
8	0.0008	1.5118	101.3916	143.8477	-0.2694	0.5103	-0.2744	77.4739	-154.538	78.5709
9	0.0009	1.5411	101.3123	143.5547	-0.2291	0.4968	-0.3013	81.8497	-154.9478	73.7611

Fig.2 Uploaded Data:

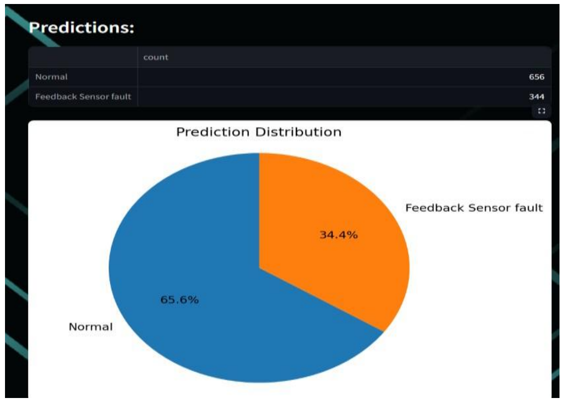

Fig.3 Predictions

V. ADVANTAGES

- **High Accuracy and Reliability:** The deep learning- based CNN model achieves high precision and recall in fault classification, minimizing the chances of false positives and missed detections.
- **Automated Analysis:** The system eliminates the need for manual inspection or threshold tuning, reducing human error and technician workload.
- **Real-Time Detection:** The integration with a Streamlit- based web app allows users to perform real-time predictions by uploading sensor logs, making the system useful in both field and lab environments**.**
- **User-Friendly Interface:** The application requires no coding skills to operate and provides a clean interface for sensor data upload, output visualization, and result interpretation**.**
- **Scalability:** The modular architecture enables easy scaling across multiple PV installations and supports retraining with new datasets**.**
- **Cost-Effective:** Using sensor data alone (without requiring thermal cameras or drones), the system provides a low-cost yet effective solution for solar panel monitoring

VI. DISADVANTAGES

- **Dependence on High-Quality Data**: The model's accuracy heavily depends on the quality and consistency of input sensor data. Noisy or incomplete data can lead to incorrect predictions.
- **Limited Generalization to Unseen Faults:** The model can only detect faults it has been trained on. Completely new or rare faults may go undetected or be misclassified.
- **Computational Requirements:** Although lightweight, the system still requires a capable CPU or GPU for efficient inference, especially when deployed at scale
- **Lack of Visual Fault Detection:** The system is currently limited to time-series sensor data and does not integrate image-based diagnostics like thermal or drone imagery.
- **Retraining Needs:** For environments with evolving operational conditions, periodic retraining may be necessary to maintain accuracy.
- **Interpretability Constraints:** Although confidence scores are provided, deep learning models can be seen as black boxes, making their decisions difficult to interpret without additional explainability tools.

VII. FUTURE SCOPE

The Future improvements for the solar panel fault detection system include integrating drone-based thermal imaging to capture physical faults, combining sensor and image data for better accuracy, and deploying the model on edge devices for real-time on-site detection. Adding explainable AI will increase trust in predictions, while adaptive learning will help the system update continuously. Moving towards predictive maintenance can reduce downtime by forecasting faults early. Cloud-based monitoring and standardized APIs will support large-scale deployment and integration with existing systems. These advancements will make the system more efficient, autonomous, and scalable for widespread renewable energy applications.

VIII. CONCLUSION

In this work, we have presented a deep learning-based solar panel fault detection system leveraging time-series sensor data analysis via a CNN model. The system demonstrated high accuracy, efficiency, and user- friendliness by automating fault diagnosis and providing real-time predictions through a web application interface. While there are challenges related to data quality, model interpretability, and computational resources, the proposed approach offers a scalable and cost-effective alternative to traditional inspection methods.

The future scope outlined includes integrating multimodal data sources, edge deployment, explainable AI, and predictive maintenance capabilities, aiming to evolve the system into a comprehensive, autonomous solar monitoring platform. This work contributes to enhancing the reliability and sustainability of solar energy systems, aligning with global renewable energy objectives and promoting intelligent infrastructure for clean energy production.

REFERENCES

[1]. Jha, Y. Rawat, and S. Vyas, "PV-S3: Advancing automatic photovoltaic defect detection using semi-supervised semantic segmentation of electroluminescence images," *arXiv preprint arXiv:2404.13693*, 2024.

[2]. Y. Rajendra Babu, N. Indira Priyadarsini, P. G. K. Sirisha, and C. Ravishankar Reddy, "Real-time anomaly detection in solar panel arrays: Integrating Single Shot Multibox Detector (SSD) with IoT and edge computing," *ITM Web of Conferences*, vol. 74, p. 01014, 2025.

[3]. Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[4]. S. Hochreiter and J. Schmidhuber, "Long Short- Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997

[5]. A Jha, Y. Rawat, and S. Vyas, "PV-S3: Advancing automatic photovoltaic defect detection using semi-supervised semantic segmentation of electroluminescence images," *arXiv preprint arXiv:2404.13693*, 2024.

[6]. M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," 2015. [Online]. Available: https://www.tensorflow.org/

[7]. Y. Rajendra Babu, N. Indira Priyadarsini, P. G. K. Sirisha, and C. Ravishankar Reddy, "Real-time anomaly detection in solar panel arrays: Integrating Single Shot Multibox Detector (SSD) with IoT and edge computing," *ITM Web of Conferences*, vol. 74,

[8]. p. 01014, 2025.

[9]. D. Zhou, C. Yang, X. Zhang, J. Sun, and X. Zhuang, "Deep learning-based solar panel defect detection and localization using electroluminescent images," *Proceedings of the ISSAT International Conference on Reliability & Quality in Design*, 2024.

[10]. Y. Ledmaoui, A. El Maghraoui, M. El Aroussi, and

[11]. R. Saadane, "Enhanced fault detection in photovoltaic panels using CNN-based classification with PyQt5 implementation," *Sensors*, vol. 24, no. 22, p. 7407, Nov. 2024.

[12]. H. Ling, M. Liu, and Y. Fang, "Deep edge-based fault detection for solar panels," *Sensors*, vol. 24, no. 16, p. 5348, 2024.

[13]. P. Molnar, *Interpretable Machine Learning*, 2019. [Online].Available:https://christophm.github.io/interpretable-ml-book/

[14]. W. Ahmed, "Enhancing solar PV reliability with hybrid local features and infrared thermography," *Energy Reports*, vol. 10, pp. 1–10, Dec. 2024