International Journal of Innovative Research in Technology and Science
ISSN: 2321-1156 www.ijirts.org Volume 13 Issue 1, January 2025

A Convolutional Neural Network-Based Approach for High-Accuracy Fault Diagnosis in Photovoltaic Systems

Pratham Tiwari, Aditya Singh, Ankit Raj, Shahnawaz Ahmad
School of Computer Science Engineering and Technology, Bennett University, Greater Noida, India

Abstract- The increasing demand for renewable energy has led to the widespread adoption of photovoltaic (PV) systems. However, the efficiency and reliability of these systems are often compromised due to various faults, which can significantly impact their performance. This paper presents a machine learning-based fault diagnosis system for solar panels, focusing on the detection and classification of faults in PV systems. We evaluate several machines learning models, including Random Forest, Decision Tree, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Logistic Regression, and Naive Bayes, using a dataset of grid-connected PV system faults. Our results demonstrate that the Convolutional Neural Network (CNN) model outperforms other models, achieving a validation accuracy of 99.8%. The proposed system offers a robust solution for maintaining the efficiency and reliability of PV systems, contributing to the broader goal of sustainable energy development.

Keywords: Photovoltaic (PV) Systems, Fault Diagnosis, Machine Learning, Convolutional Neural Network (CNN), Renewable Energy, Solar Panel Fault Detection

1. INTRODUCTION

Energy consumption is increasing daily due to factors such as industrialisation and population growth. Sustaining equilibrium between supply and demand necessitates an augmentation in energy generation. Consequently, renewable energy is assuming a progressively more significant role on a global scale. The convenient accessibility of renewable energy sources in a local context, coupled with their plentiful supply, diminishes reliance on foreign resources for fulfilling energy requirements, consequently enhancing energy security for nations. An additional rationale for selecting renewable energy sources lies in their eco-friendliness. Due to their lower emissions of greenhouse gases compared to fossil fuels, the heightened adoption of renewable energy leads to a reduction in environmental pollution. Solar energy does not contribute to global warming or harm ecosystems, thereby assisting in the maintenance of ecological balance.

In summary, solar energy is a valuable tool in mitigating climate change, which is essential for the well-being of all life forms. To address climate change, the use of environmentally friendly renewable energy has become a mandatory requirement, as outlined in the Paris Climate Agreement, which was signed at the end of the 2015 United Nations Climate Change Conference [6]. Hence, renewable energy plays a crucial role in attaining sustainable development objectives and securing a stable and sustainable future. Within the realm of renewable energy options, photovoltaic (PV) systems are gaining increasing traction. PV panels, constructed from semiconductor materials, employ a technology that transforms solar radiation into electricity. As sunlight strikes the panel, free electrons within the semiconductor material become mobile, generating an electrical current. PV modules are formed by integrating solar cells in series and parallel connections. These modules are combined to form PV panels, and PV arrays are created by connecting the panels in series and parallel arrangements [1]. Solar power plants are developed through the combination of these arrays. The appeal of PV systems lies in their simplicity of energy conversion, resilience, ease of upkeep, and the universal accessibility of sunlight. As a result of the growing interest in PV systems, the field has seen a surge in research. Literature-based investigations have two main goals: improving the utilisation of PV systems by identifying their strengths and ensuring the inclusion of performance-affecting parameters in PV system designs.

Numerous research efforts focus on harnessing solar energy effectively across diverse geographical regions [2, 18]. However, one of the major challenges remains the efficient fault detection and classification in PV systems, which can impact their performance. Recent research has focused on enhancing solar panel fault detection using machine learning (ML) and deep learning (DL) models. In some papers, the integration of Generative Adversarial Networks (GANs) and Social Spider Optimization (SSO) has shown promise in improving the precision and recall rates in detecting and classifying faults in solar panels. The results indicate superior accuracy and the potential for real-time applications despite some limitations, such as model sensitivity to initial populations and the computational resources required[3,4,19]. In other cases, techniques like InceptionV3 combined with U-Net architectures have achieved high precision (94%) in detecting faults, thus contributing to the overall efficiency of fault diagnosis [4]. The performance of solar panels is highly dependent on various environmental factors. Temperature is a key variable, with higher temperatures leading to increased short-circuit current but decreased voltage, which impacts the overall output. The amount of sunlight reaching the panels positively influences system performance.

Additionally, factors such as panel cleanliness and shading can significantly affect the efficiency of PV systems. Dirty panels or shading can reduce the amount of sunlight hitting the solar cells, decreasing energy production [5]. IoT-based models are playing an increasingly important role in monitoring PV system performance. With sensors capturing real-time data on power generation and weather conditions, these IoT-enhanced frameworks, combined with advanced ML models such as Decision Trees (DT), can predict faults and optimise maintenance schedules for solar power plants. This approach helps in reducing manual intervention while maintaining high system accuracy [5]. The main contribution of the paper is given as follows:

- This study develops a machine learning-based fault diagnosis system for solar panels capable of detecting and classifying various types of faults in PV systems.
- We evaluate and compare the performance of multiple machines learning models, including Random Forest, Decision Tree, SVM, KNN, Logistic Regression, and Naive Bayes, for fault classification in PV systems.
- The proposed Convolutional Neural Network (CNN) model achieves a validation accuracy of 99.8%, outperforming other state-of-the-art models in fault classification tasks.
- The study utilises the GPVS-Faults dataset, which includes various fault scenarios in grid-connected PV systems, to validate the effectiveness of the proposed models.
- The high accuracy and robustness of the CNN model suggest its potential for real-time fault detection and classification in PV systems, enhancing maintenance efficiency and reducing operational costs.

The organisation of the paper is structured as follows: Section 2 reviews related work on fault detection and classification in PV systems, covering advancements in ML, DL, and IoT-based monitoring, along with existing limitations. Section 3 outlines the research methodology, including dataset description, preprocessing, and model implementation, with pseudocode for the CNN model. Section 4 presents results, comparing the proposed CNN model's performance with state-of-the-art methods. Section 5 concludes the paper, summarising findings, discussing implications for renewable energy, and suggesting future research directions like real-time fault detection.

2. Related Work

Fault analysis in photovoltaic (PV) systems has been extensively researched due to the increasing importance of maintaining the efficiency and reliability of solar energy generation. Various methods have been developed to enhance fault detection and classification in solar panels. One study proposed a fault diagnosis algorithm that utilises measures such as open-circuit voltage and load voltage to detect faults in solar cell modules. By comparing stored feature curves and using simulation tools, the algorithm could accurately diagnose module faults. Furthermore, a solar simulator was designed to replicate real-world environmental conditions, facilitating fault data collection and enabling rapid maintenance response by teams [8].

Another study focused on evaluating fault detection strategies, emphasising the identification of failure-specific patterns to improve the performance of solar panels. This research underscored the impact of solar panel efficiency on power generation and highlighted factors such as cell composition, electrical arrangement, and panel ratings. The study also examined projected future developments in solar panel deployment, stressing the need for better performance to meet climate targets [9].

Deep learning techniques have also been explored for fault detection. One notable study used infrared images of solar modules to train a pre-trained deep learning model, which demonstrated high accuracy in fault detection. This reinforced the importance of precise fault identification for system maintenance and efficiency [10]. Another study provided a comprehensive evaluation of various PV fault classification and detection strategies, including condition-based if-then rules, decision trees, statistical techniques, and machine learning methods [11].

Amaral et al. [12] introduced a machine learning-based fault diagnosis approach for PV trackers, utilising image processing techniques and principal component analysis for fault detection. Abubakar et al. [13] developed a unique fault detection method that leverages the Elman Neural Network (ENN) and Boosted Tree Algorithms (BTA) for detecting PV array and inverter faults. Further advancements in deep learning-based fault detection were highlighted by Kellil et al. [14], who proposed a technique for classifying faults in PV modules using deep neural networks and infrared images. Memon et al. [15] developed an intelligent model based on a Convolutional Neural Network (CNN) for robust classification of PV panel faults using historical data.

Another modern approach introduced a sequential fault detection algorithm based on autoregressive models and Generalized Likelihood Ratio (GLLR) tests for fast and adaptive fault detection in PV systems [16]. Furthermore, Ramirez et al. [17] presented an efficient, low-cost condition monitoring system for PV panels using radiometric sensors and image processing techniques. Their study proposed a two-layer solution for detecting problematic areas from images acquired via ortho-tile-based georeferenced spatial heat maps.

In more recent work, a hybrid method combining Generative Adversarial Networks (GANs) and the Social Spider Optimization method was proposed for fault detection and classification in solar-based distribution systems. The study leveraged the digital twin concept to enhance real-time data analysis and system monitoring [1]. Another study explored the use of dimensionality reduction techniques and machine learning for fault diagnosis in PV panels, highlighting the importance of feature extraction and filtering in improving detection accuracy [2]. Lastly, a study proposed a hybrid deep learning model combining handcrafted features and automatic feature extraction for fault detection in PV images. Their method significantly improved fault classification accuracy compared to traditional deep learning models [3].

Table 1 summarises various approaches to fault detection and classification in PV systems. Methods like GAN-SSO and DDRS-CART focus on improving precision and feature extraction but face limitations such as high computational costs and limited generalizability. Deep learning models like VGG-16, RestNet-50 + R-CNN, and InceptionV3-Net achieve high accuracy but are computationally intensive and may struggle with low-magnitude faults. IoT-based models with Decision Trees enable real-time monitoring but depend on sensor data quality. The proposed CNN-based fault classification model outperforms these methods, achieving 99.8% accuracy, though further validation on diverse datasets is needed.

Table 1. Related Previous Work

Reference	Proposed Method	Key Contribution	Limitations
[1]	GAN-SSO	Improved precision and recall rates using GANs and SSO	High computational cost, sensitive to initial populations
[2]	DDRS-CART	Feature extraction/reduction with machine learning	Limited fault types, poor generalizability
[3]	VGG-16	Hybrid deep learning model for fault detection	Requires large amounts of labeled training data
[4]	ResNet-50 + R-CNN	High precision (99%) using deep learning architectures	Computationally expensive, not suitable for real-time
[5]	InceptionV3-Net	94.5% accuracy using deep learning models	Limited fault types, poor performance on low-magnitude faults
[6]	IoT-based models with Decision Trees	Real-time monitoring and fault prediction using IoT sensors and machine learning	Reliant on sensor data quality and reliability
[7]	CNN-based fault classification	99.8% accuracy in fault classification	Needs validation on larger and more diverse datasets

3. Research Methodology

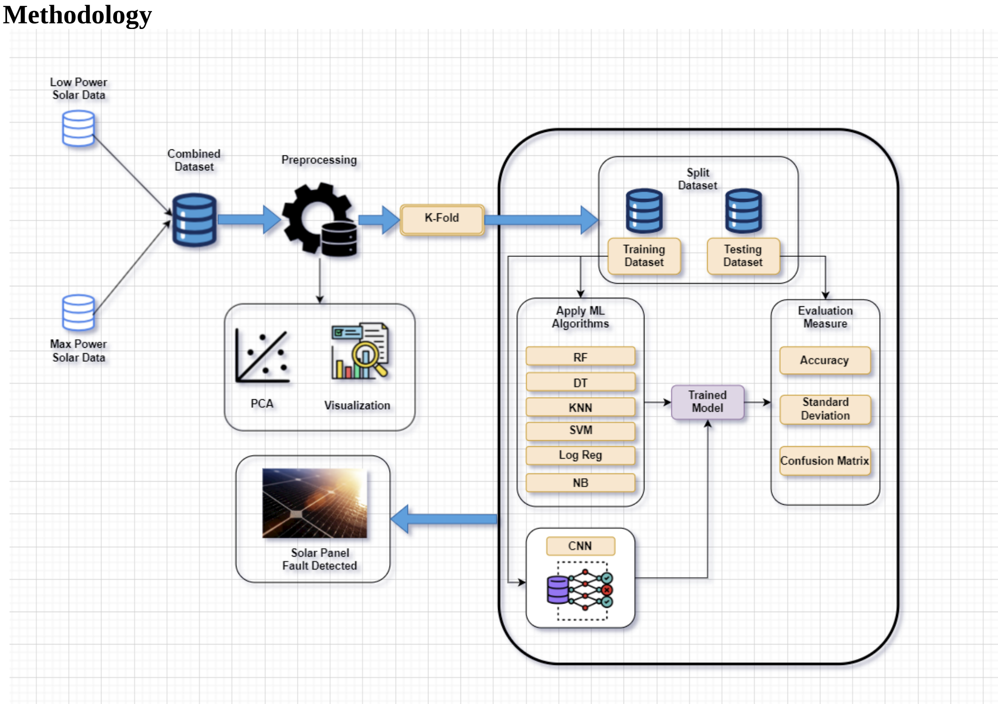

Figure 1. Research Methodology

4. Result and Discussion

4.1. Dataset Description

In this paper, we have used the dataset described in [7]. The Grid-connected PV System Faults (GPVS-Faults) dataset was collected from laboratory experiments on faults in a PV microgrid system and comprises 16 data files in both .mat and .csv formats, each corresponding to a specific experimental scenario. These scenarios include photovoltaic array faults, inverter faults, grid anomalies, feedback sensor faults, and MPPT controller faults of varying severity. The dataset is valuable for designing, validating, and comparing algorithms for fault detection, diagnosis, and classification, thereby supporting PV system protection and reactive maintenance. During the experiments, faults were introduced manually at the midpoint of each experiment. The dataset contains high-frequency measurements that are inherently noisy, with disturbances and fluctuations in temperature and insolation occurring both during and between experiments. Additionally, the MPPT/IPPT operating modes can negatively impact the detection of low-magnitude faults, and following critical faults, the system's operation may be interrupted, leading to potential shutdowns. Therefore, the main challenge is to detect faults before a complete system failure occurs. The structure of the dataset is as follows: the data files are labelled as "Fxy", where x ∈ {0, 1, …, 7} represents the fault scenario (with 0 indicating a fault-free experiment, while 1 through 7 corresponds to seven types of faults), and y ∈ {'L', 'M'} denotes the operation mode ('L' for Limited Power Mode (IPPT) and 'M' for Maximum Power Mode (MPPT)). For example, "F4M" refers to fault F4 in MPPT mode, while "F1L" represents fault F1 in IPPT mode. Each data file includes columns for time (in seconds, with an average sampling interval $T_s = 9.9989$ µs), PV array current measurement (Ipv), PV array voltage measurement (Vpv), DC voltage measurement (Vdc), three-phase current measurements (ia, ib, ic), three-phase voltage measurements (va, vb, vc), current magnitude (Iabc), current frequency (If), voltage magnitude (Vabc), and voltage frequency (Vf).

4.2. Evaluation

In this study, we implemented and compared several machine learning classifiers [15, 18, 19, 20, 21], including Random Forest, Decision Tree, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Logistic Regression, and Naive Bayes. These models were evaluated using 5-fold cross-validation with a test size of 10%, and their performance was assessed based on accuracy metrics. The classifier's accuracy acc) measures how frequently it makes accurate predictions. By dividing the total number of forecasts by the percentage of accurate predictions, accuracy is calculated. The method of determining accuracy is shown in equation eq(1).

$$acc = TP+TN/TP+FP+TN+FN \qquad (1)$$

The results showed that K-Nearest Neighbors achieved the highest accuracy across the models, with a cross-validation accuracy of 97.90%, followed closely by the Random Forest classifier with 97.86%. Both models exhibited low standard deviations, indicating consistent performance across different data splits. In contrast, models like Naive Bayes and Logistic Regression performed moderately, with accuracies around 85-90%. The Support Vector Machine (SVM) model also demonstrated competitive performance with an accuracy of 96.50%, although it required more computational resources during training. A visual comparison of the models is depicted in Fig. 1, showing that KNN and Random Forest consistently outperform other models in this dataset. To perform a more nuanced classification task, we developed a Convolutional Neural Network (CNN) model. The model's performance was evaluated using a confusion matrix, as shown in Figure 2, which provided detailed insights into how well the model distinguished between different categories.

4.3 CNN-Based Fault Diagnosis Algorithm

CNN-Based Fault Diagnosis Algorithm

Start

1. Load Dataset:
- Load Low Power Dataset (df1) and Max Power Dataset (df2)

2. Preprocessing:
- Concatenate df1 and df2
- Standardise features using StandardScaler

3. Sliding Window Technique:
- Set window_length and stride
- Create overlapping windows for each label
- Append features (X) and labels (Y)

4. Label Encoding and Reshaping:
- Encode labels using LabelEncoder
- Convert labels to one-hot encoded vectors
- Reshape X to match the CNN input shape

5. Train-Test Split:
- Split data into training (80%) and testing (20%) sets

6. CNN Model Definition:
- Add Conv2D layers with ReLU activation
- Add MaxPooling2D layers
- Flatten the output
- Add Dense layers with ReLU and softmax activation

7. Model Compilation and Training:
- Compile model with categorical cross-entropy loss, Adam optimiser, and accuracy metric
- Train model with specified batch_size and epochs
- Save trained model

8. Model Evaluation:
- Reload saved model
- Predict labels for the test set
- Inverse-transform predictions to original labels
- Compute confusion matrix
- Visualise the confusion matrix using the heatmap

End

The CNN achieved strong classification performance, with most predictions falling along the diagonal of the confusion matrix, indicating accurate classifications. However, there were some misclassifications, particularly between closely related classes, which can be attributed to the inherent complexity of the data. This suggests that while the CNN model performed well overall, further optimisation could be beneficial to reduce misclassification rates. The validation accuracy for the CNN model was found to be 0.998, indicating excellent performance on unseen data. Fig. 2 provides a comparison of our proposed model with other state-of-the-art models. The results demonstrate that our model outperforms

ISSN: 2321-1156 www.ijirts.org Volume 13 Issue 1, January 2025

existing methods, achieving the highest accuracy of 99.5%. This highlights the effectiveness of our approach in addressing the challenges of fault classification in photovoltaic systems. Table 1 provides a comparison of our proposed model with other state-of-the-art models. The results demonstrate that our model outperforms existing methods, achieving the highest accuracy of 99.5%. This highlights the effectiveness of our approach in addressing the challenges of fault classification in photovoltaic systems.

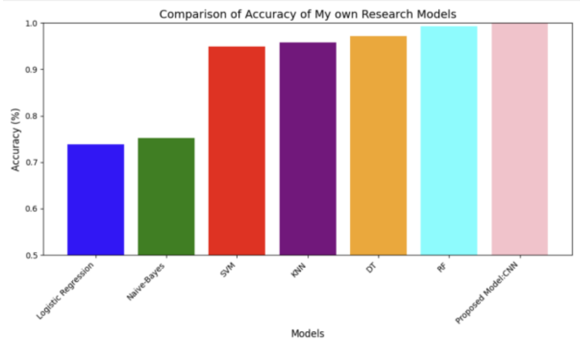

Figure 1: Comparison of Accuracy of My Research Models

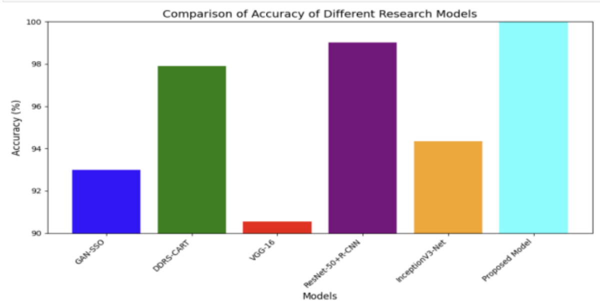

Figure 2: Comparison of Accuracy of Different Research Models

Table 1. Comparative Analysis

Research Papers	**Proposed Model**	**Accuracy**
[1]	GAN-SSO	92.5
[2]	DDRS-CART	98.0
[3]	VGG-16	90.5
[4]	RestNet-50+R-CNN	99.0
[5]	InceptionV3-Net	94.5
*	Proposed Model: CNN	99.8

The results of this study demonstrate that the Convolutional Neural Network (CNN) model is the most effective model for fault classification tasks in photovoltaic (PV) systems. The CNN model achieved a validation accuracy of 0.998, outperforming all other models evaluated in this study. However, it's important to note that further research is needed to validate the generalizability of the CNN model on a larger and more diverse dataset.

5. Conclusion

In this study, we developed a machine learning-based fault diagnosis system for solar panels, focusing on the detection and classification of faults in photovoltaic (PV) systems. Several machines learning models, including Random Forest, Decision Tree, SVM, KNN, Logistic Regression, and Naive Bayes, were evaluated using the GPVS-Faults dataset. Among these, the Convolutional Neural Network (CNN) model emerged as the most effective, achieving a validation accuracy of 99.8%. This high level of accuracy demonstrates the model's capability to distinguish between different fault types, making it a robust solution for maintaining the efficiency and reliability of PV systems. The results highlight the importance of machine learning in advancing solar panel fault diagnosis, offering significant improvements over traditional methods. Future research should focus on detecting low-magnitude faults, real-time implementation, and integrating IoT technologies for continuous monitoring and proactive maintenance. By enhancing the reliability and performance of solar panels, this study contributes to the broader goal of sustainable energy development and aligns with global efforts to combat climate change.

References

[1]. Hanhua Cao, Huanping Zhang, Changle Gu, Yuhua Zhou, Xiu He, Fault detection and classification in solar-based distribution systems in the presence of deep learning and social spider method, Solar Energy, Volume 262,2023,111868, ISSN 0038-092X,

[2]. Bassel Chokr, Nizar Chatti, Abderafi Charki, Thierry Lemenand, Mohammad Hammoud, Feature extraction-reduction and machine learning for fault diagnosis in PV panels, Solar Energy, Volume 262,2023,111918, ISSN 0038-092X.

[3]. Hayder Yousif, Zahraa Al-Milaji, Fault detection from PV images using hybrid deep learning model, Solar Energy, Volume 267,2024,112207, ISSN 0038-092X.

[4]. Sujata P. Pathak, Dr Sonali Patil, Shailee Patel, Solar panel hotspot localisation and fault classification using deep learning approach, Procedia Computer Science, Volume 204,2022, Pages 698-705, ISSN 1877-0509.

[5]. Rifat Al Mamun Rudro, Kamruddin Nur, Md. Faruk Abdullah Al Sohan, M.F. Mridha, Sultan Alfarhood, Mejdl Safran, Karthick Kanagarathinam, SPF-Net: Solar panel fault detection using U-Net based deep learning image classification, Energy Reports, Volume 12,2024, Pages 1580-1594, ISSN 2352-4847.

[6]. COP 21 Paris France Sustainable Innovation Forum 2015 Working with UNEP.

[7]. Bakdi, Azzeddine; Guichi, Amar; Mekhilef, Saad; Bounoua, Wahiba (2020), "GPVS-Faults: Experimental Data for fault scenarios in grid-connected PV systems under MPPT and IPPT modes", Mendeley Data, V1.

[8]. Hwang, Hye & Kim, Berm & Xu, Rui & Lee, In. (2017). Development of A Fault Diagnosis

ISSN: 2321-1156 www.ijirts.org Volume 13 Issue 1, January 2025

Algorithm for Solar Panel. 10.2991/ceie-16.2017.55.

[9]. Tamilselvi Selvaraj, Ramasubbu Rengaraj, GiriRajanbabu Venkatakrishnan, SoundhariyaGanesan Soundararajan, Karuppiah Natarajan, Praveen Kumar Balachandran, PrinceWinston David, Shitharth Selvarajan, "Environmental Fault Diagnosis of Solar Panels Using Solar.

[10]. Duranay, Z.B. Fault Detection in Solar Energy Systems: A Deep Learning Approach. Electronics 2023, 12, 4397.

[11]. Amaral, T.G., Pires, V.F. and Pires, A.J. Fault detection in PV tracking systems using an image processing algorithm based on PCA. Energies 2021, 14, 7278.

[12]. Kellil, N., Aissat, A. and Mellit, A. Fault diagnosis of photovoltaic modules using deep neural networks and infrared images under Algerian climatic conditions. Energy 2023, 263, 125902.

[13]. Ansari, Mohd Aquib, Shahnawaz Ahmad, and Arvind Mewada. "Mitigating risk in medical AI: balancing X-ray datasets for reliable detection." Life Cycle Reliability and Safety Engineering (2025): 1-14.

[14]. Chen, L., Li, S. and Wang, X. Quickest fault detection in photovoltaic systems. IEEE Trans. Smart Grid 2016, 9, 1835–1847.

[15]. Segovia Ramirez, I., Das, B. and Garcia Marquez, F.P. Fault detection and diagnosis in photovoltaic panels by radiometric sensors embedded in unmanned aerial vehicles. Prog. Photovolt. Res. Appl. 2022, 30, 240–256.

[16]. Bakhtiyar, Adeeba, et al. "From Pixels to People: Deep Learning Breakthroughs in Human Detection." 2024 IEEE 16th International Conference on Computational Intelligence and Communication Networks (CICN). IEEE, 2024.

[17]. Ansari, Mohd Aquib, and Dushyant Kumar Singh. "Human detection techniques for real-time surveillance: a comprehensive survey." Multimedia tools and applications 80.6 (2021): 8759-8808.

[18]. Mewada, Arvind, et al. "Deceptive Opinion Detection Using Stacking-Based Deep Ensemble Learning." 2025 3rd International Conference on Disruptive Technologies (ICDT). IEEE, 2025.

[19]. Mewada, Arvind, Sushil Kumar Maurya, and Mohd Aquib Ansari. "Comparative Study of Artificial Intelligence Approaches in Deceptive Opinion Detection." 2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN). IEEE, 2025.