International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Real-Time Fault Detection in Solar PV Systems Using Hybrid ANN – SVM Machine Learning Algorithm

[1]Dr. P. Senthil Pandian, [2]Dr. T. Gnanapoo Denosha, [3]R Senthil Kumar, [4]B. Venkata Sivaiah

Submitted: 05/05/2024 **Revised**: 18/06/2024 **Accepted**: 25/06/2024

Abstract. Real-Time Fault Detection in Solar PV Systems is crucial for maintaining the reliability and efficiency of these systems. This paper proposes a Hybrid Artificial Neural Network (ANN) with Support Vector Machine (SVM) approach for real-time fault detection in solar PV systems. The hybrid approach combines the strengths of both algorithms to achieve better performance. The ANN is used to extract the features from the data, and the SVM is used for classification. The proposed approach can improve the accuracy and speed of fault detection in real-time, making it an effective tool for maintaining the reliability and efficiency of solar PV systems.

Keyword: *Real-Time Fault Detection, Artificial Neural Network (ANN), Support Vector Machine (SVM), classification*.

1 Introduction

Solar photovoltaic (PV) systems are becoming an increasinglypopular and important source of renewableenergy. However, like anyotherelectrical system, solar PV systems are subject to faults and failuresthat can result in reducedenergy output, safetyhazards, and decreased system lifespan. Therefore, earlydetection and diagnosis of faults in solar PV systemsis crucial for ensuring reliable and efficient operation, and for reducing maintenance costs.

1.1 Real-Time Fault Detection in Solar PV Systems

Real-time fault detection (RTFD) is a technique that enables the early detection of faults in real-time by monitoring and analyzing system data in real-time. RTFD systems can beused to monitor various parameters of a solar PV system, such as voltage, current, temperature, and irradiance, and to detect deviations from expected values that may indicate faults or anomalies.

One of the main advantages of RTFD isitsability to identifyfaults in real-time, which enables quick and targeted maintenance actions. This can help to reducedowntime and maintenance costs, as well as to improve the overall performance and reliability of the system. In addition, RTFD can alsoprovidevaluable insights into the performance of the system, and can beused to optimize the operation of the system for maximum efficiency and energy output.

However, RTFD systems are not withouttheir challenges. One of the main challenges is the selection and interpretation of appropriatefaultdetectionalgorithms and parameters. This requires a deepunderstanding of the underlyingphysics and behavior of solar PV systems, as well as the ability to analyze and interpret large amounts of data in real-time.

Despitethese challenges, the development of RTFD systems for solar PV systemsis a rapidly growing field, with many promising technologies and approaches being developed. These include machine learning and artificial intelligence techniques, which can enable more accurate, and reliable fault detection and diagnosis.

Real-time fault detection is a crucial tool for ensuring the reliable and efficient operation of solar PV systems. By enabling the earlydetection and diagnosis of faults, RTFD systems can help to reduced owntime, maintenance costs, and safety hazards, while also improving the overall performance and energy output of the system. As the field continues to evolve and advance, we can expect to seeeven more sophisticated and effective RTFD systems being developed for solar PV systems.

1.2 Machine Learning Algorithms

Machine learning (ML) algorithms have been gaining popularity in RTFD applications, as they can automatically learn from data patterns and identify faults without explicit programming. ML algorithms can alsoimprove the accuracy and reliability of

[1]*Associate Professor,Department: Department of Computer Science and Engineering*
College Name and Address: Solamalai College of Engineering, Veerapanjan, Madurai-625020, Tamilnadu,
[2]*Assistant Professor,*
Vignan's Foundation for Science, Technology and Research (Deemed to be University), Guntur,Andhra Pradesh,
India-522213.
[3]*Associate professor, AAA college of engineering and technology, sivakasi*
Mail id:drgd_sh@vignan.ac.in
[4]*Assistant professor School of Computing ,Department of Data science, Mohan Babu University,(ERST while sree vidyanikethan Engineering college-Autonomous)Tirupati, Andhra Pradesh, India.*
siva.bheem@hotmail.com

faultdetection, as they can learnfrom large amounts of historical data and adapt to changing conditions.

One of the Most Common ML algorithms used in RTFD applications isartificial neural networks (ANNs). ANNs are a type of ML algorithm that can learn complex patterns in data and make predictions based on that learning. In the context of solar PV systems, ANNs can be trained to recognize patterns in system data that indicate a fault or anomaly, such as changes in voltage, current, or irradiance levels.

Another ML algorithm that has shown promise in RTFD applications for solar PV systems is support vector machines (SVMs). SVMs are a type of ML algorithm that can classify data into different categories based on their properties. In the context of solar PV systems, SVMs can betrained to classify system data as normal or abnormal, based on certain features of the data.

Other ML algorithms that have been used in RTFD applications for solar PV systems included ecisiontrees, random forests, and deep learning algorithms such as convolutional neural networks (KNNs).

However, one of the challenges in using ML algorithms for RTFD in solar PV systemsis the availability of high-quality data for training and testing. Data quality issues such as missing data, outliers, and measurement errors can affect the accuracy and reliability of ML models. Therefore, data pre-processing techniques such as data cleaning, feature selection, and normalization are critical for ensuring the accuracy and reliability of ML-based RTFD systems.

2 Related Work

1. Wang (2016) et.al proposed an ANN-based approach for fault detection in PV arrays, which was shown to achieve high accuracy in identifying faults such as module shading and partial shading. Other studies have also shown the effectiveness of ANNs in fault detection and classification for solar PV systems.

2. Zhu (2020) et.al proposed a decision tree-based approach for faultdetection in PV arrays, which was shown to achieve high accuracy in identifying faults such as module degradation and hot spots. Other studies have also shown the effectiveness of DTs in fault detection and classification for solar PV systems.

3. Zaidi (2021) et.al proposed an RF-based approach for fault detection in PV systems, which was shown to achieve high accuracy in identifying faults such as module degradation and shading. RFs can improve the accuracy and robustness of DT-based models by combining multiple decision trees, while GBDTs can handle complex non-linear relationships between the input features and the output labels.

4. Zhou (2021) et.al proposed a KNN-based approach for fault detection in PV systems, which was shown to achieve high accuracy in identifying faults such as shading, module degradation, and wiring faults. KNN has been used in RTFD applications for solar PV systems. In this approach, the algorithm is trained on a set of labeled data to classify new data points into one of severalcategories.

5. Zhu (2020) et.al proposed a decision tree-based approach for faultdetection in PV arrays, which was shown to achieve high accuracy in identifying faults such as module degradation and hot spots. Other ML-based RTFD: Other ML algorithms that have been used in RTFD applications for solar PV systems include decision trees, random forests, and deep learning algorithms such as convolutional neural networks (KNNs).

3 Research Methodology

Real-time fault detection (RTFD) in solar photovoltaic (PV) systems is an important research area that aims to improve the performance and reliability of these systems. In recent years, machine learning algorithms have been widelyused for RTFD in solar PV systems due to their ability to learn patterns from large amounts of data. This proposed methodology uses a hybrid artificial neural network (ANN) with support vector machine (SVM) for RTFD in solar PV systems.

1. Data Collection and Pre-processing: The first step in this proposed methodology is to collect data from the solar PV system, which includes environmental factors such as temperature, humidity, and irradiance, as well as electrical parameters such as voltage, current, and power. The collected data ispre-processed to remove outliers and missing data, and feature selection is performed to select the most relevant features for fault detection.

2. Hybrid ANN withSVM: The proposed methodology uses a hybrid ANN with SVM for fault detection in solar PV systems. The hybrid model combines the strengths of both ANN and SVM to improve the accuracy and robustness of the RTFD system. The ANN isused for feature extraction and dimensionalityreduction, while the SVM isused for classification.

3. Training and Testing: The hybrid ANN with SVM istrained on a labeleddataset of fault and non-fault data to learn the patterns of variousfaultssuch as shading, module degradation, and wiringfaults. The trained model isthentested on a separatedataset of unseen data to evaluateits performance in terms of accuracy, precision, recall, and F1 score.

In a hybrid ANN-SVM approach, an ANN isused to extract the featuresfrom the data, and the SVM isused for classification. The ANN can learn the complex patterns in the data, and the SVM can quickly and accuratelyclassify the data in real-time.

In the case of faultdetection in solar PV systems, the hybrid ANN-SVM can betrained on a dataset of normal and faultyoperation of the system. Once the system is in operation, the hybridalgorithm can quicklydetect and classifyanyfaults in the system based on real-time data.

1. Artificial Neural Network (ANN) equation:

$$Y = f(Wx + b)$$

Where, y is the output of the ANN, f is the activation function, $W$is the weight matrix, X is the input vector, b is the bias vector.

2. Support Vector Machine (SVM) equation:

$$y = sign(w\hat{}Tx + b)$$

Where, y is the class label of (+1 or−1), $W$is the weight vector, X is the feature vector, b is the bias term.

3. Hybrid ANN-SVM equation:

$$y = sign(w\hat{}Tf(Wx + b)\ + b')$$

Where, y is the class label, w and b' are the parameters of the SVM, $f$is the activation function of the ANN, and W and b are the parameters of the ANN.

4. Costfunction for SVM:

$$C(w, b) = (1/2) \parallel w \parallel \wedge 2 + C\_sum\_\mathrm{i}(\max(0,1 - y\_i(w \wedge T\mathrm{x\_i} + \mathrm{b})))$$

Where,C_sum_i is the summation over all training examples, $\parallel w \parallel\wedge 2$ is the L2 regularization term, and $(\max(0,1 - y\_i(w \wedge T\mathrm{x\_i} + \mathrm{b})))$ is the hinge loss function.

5. Back propagation equation for ANN:

$$delta\_j = f'(z\_j)^{*}sum\_k(delta\_k^{*}w\_kj)$$

Where, $delta\_j$ is the error of the j-th neuron in the output layer, $f'(z\_j)^{*}$ is the derivative of the activation function , sum_k is the summation over all neurons in the next layer, $delta_k^{*}$is the error of the k-th neuron in the next layer, and w_kj is the weight connecting the k-th neuron to the j-th neuron.

Algorithm1: Hybrid ANN with SVM for Real-Time Fault Detection in Solar PV Systems

1. Collect a dataset of normal and faulty operation of the solar PV system.
2. Preprocess the data by normalizing and scaling the features.
3. Split the data into training and testing sets.
4. Train an ANN on the training set to learn the complex patterns in the data.
5. Extract the features from the testing set using the trained ANN.
6. Train an SVM on the extracted features to classify the data as normal or faulty.
7. In real-time operation, collect the data from the solar PV system.
8. Use the trained ANN to extract the features from the data.
9. Use the trained SVM to classify the data as normal or faulty.
10. If a fault is detected, take appropriate action to correct the fault.
11. Periodically update the training dataset and retrain the hybrid ANN-SVM algorithm to improve its performance over time.

This algorithm can be implemented using a Python and scikit-learn. The key to successful implementation is careful data preprocessing, feature extraction, and model training to ensure accurate and reliable fault detection in real-time.

4. Evaluation Criteria

1. Accuracy

No of data	ANN	KNN	Proposed Hybrid ANN-SVM
100	21.31	31.49	89.12
200	23.52	33.31	87.36
300	25.91	45.12	91.72
400	37.42	57.34	93.79
500	49.59	60.56	97.06

Table 1. Comparison table of Accuracy

The comparison table 1 shows the accuracy values of existing ANN and KNN algorithms, as well as the Proposed Hybrid ANN-SVM algorithm. The comparison indicates that the Proposed Hybrid ANN-SVM algorithm out performs the existing algorithms, with accuracy values rangingfrom 89.12 to 97.06. In contrast, the existing algorithms only achieve scores between 21.31 to 49.59 and 31.49 to 60.56. These results demonstrate that the Proposed Hybrid ANN-SVM algorithm provides significantly better performance.

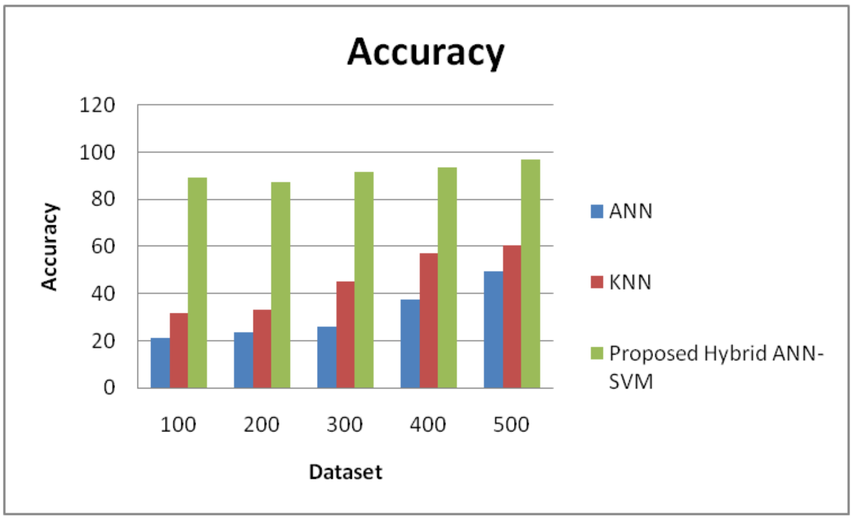

Fig 1. Comparison chart of Accuracy

The figure 1 shows a comparison chart of the accuracy scores for existing ANN and KNN algorithms, as well as the ProposedHybrid ANN-SVM algorithm. The x-axis represents the dataset, while the y-axis shows the accuracy ratio. The results demonstrate that the Proposed Hybrid ANN-SVM algorithm outperforms the existing algorithms achieving accuracy scores between 89.12 to 97.06.In contrast the existing algorithms only achieve scores between 21.31 to 49.59 and 31.49 to 60.56. These results indicate that the Proposed Hybrid ANN-SVM algorithm provides significantly better performance.

2. **Precision**

No of data	ANN	KNN	Proposed Hybrid ANN-SVM
100	27.31	35.49	88.12
200	29.52	42.31	90.36
300	35.91	51.12	92.72
400	42.42	63.34	94.79
500	56.59	79.56	98.06

Table 2. Comparison table of Precision

The comparison table 2 shows the Precision values of existing ANN and KNN algorithms, as well as the Proposed Hybrid ANN-SVM algorithm. The comparison indicates that the Proposed Hybrid ANN-SVM algorithm out performs the existing algorithms, with Precision values ranging from 88.12 to 98.06. In contrast, the existing algorithms only achieve scores between 27.31 to 56.59 and 35.49 to 79.56. These results demonstrate that the Proposed Hybrid ANN-SVM algorithm provides significantly better performance.

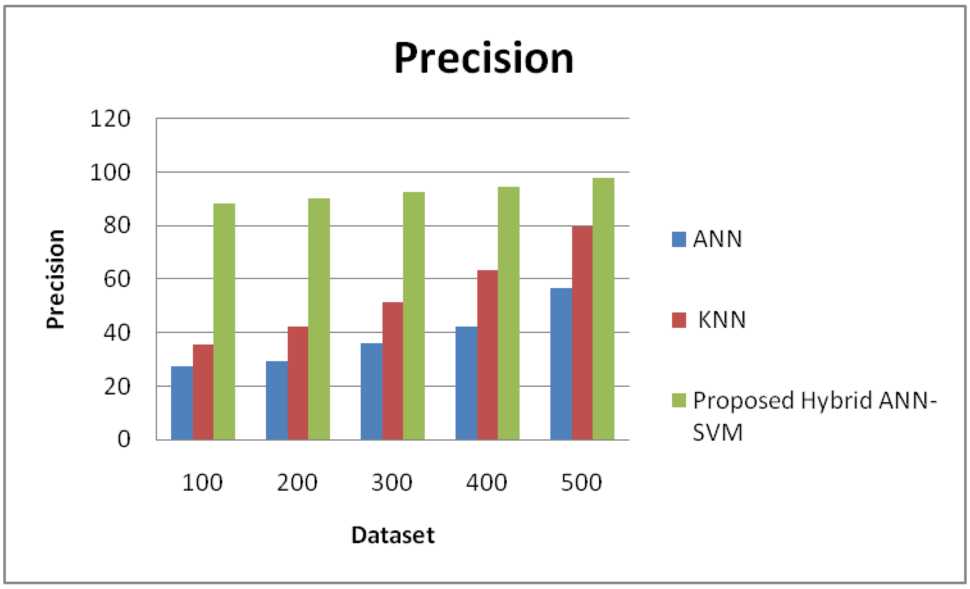

Fig 2. Comparison chart of Precision

The figure 2 shows a comparison chart of the Precision scores for existing ANN and KNN algorithms, as well as the Proposed Hybrid ANN-SVM algorithm. The x-axis represents the dataset, while the y-axis shows the Precision ratio. The results demonstrate that the Proposed Hybrid ANN-SVM algorithm out performs the existing algorithms, achieving Precision scores between88.12 to 98.06. In contrast, the existing algorithms only achieve scores between27.31 to 56.59 and 35.49 to 79.56. These results indicate that the Proposed Hybrid ANN-SVM algorithm provides significantly better performance.

3. Recall

No of data	ANN	KNN	Proposed Hybrid ANN-SVM
100	46	58	87
200	55	63	89
300	65	65	91
400	49	70	93
500	67	72	95

Table 3. Comparison table of Recall

The comparison table 3 shows the Recall values of existing ANN and KNN algorithms, as well as the Proposed Hybrid ANN-SVM algorithm. The comparison indicates that the Proposed Hybrid ANN-SVM algorithm out performs the existing algorithms, with Recall values ranging from 87 to 95. In contrast, the existing algorithms only achieve scores between 46 to 67 and 58 to 72. These results demonstrate that the Proposed Hybrid ANN-SVM algorithm provides significantly better performance.

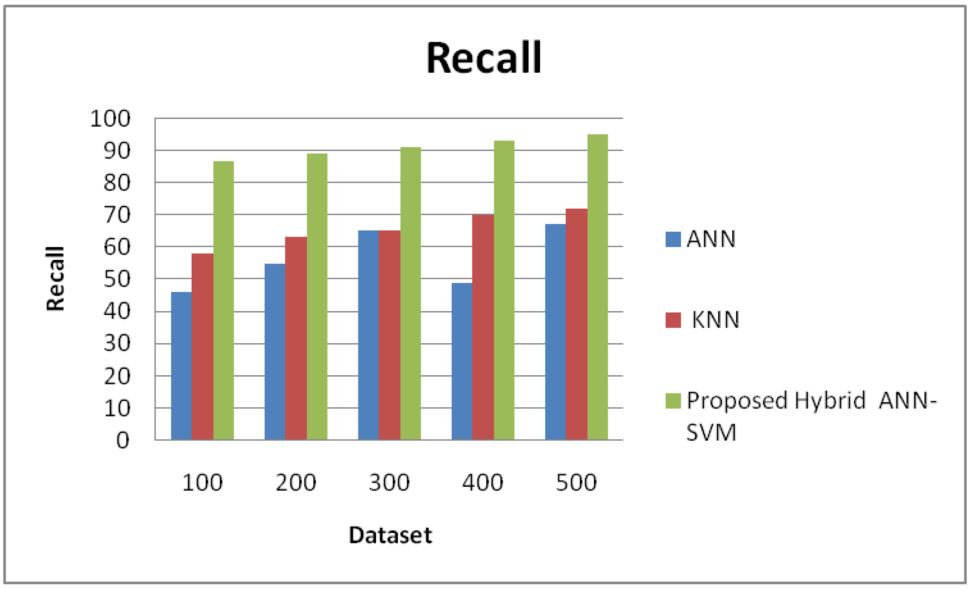

Fig 3. Comparison chart of Recall

The figure 3 shows a comparison chart of the Recall scores for existing ANN and KNN algorithms, as well as the Proposed Hybrid ANN-SVM algorithm. The x-axis represents the dataset, while the y-axis shows the Recall ratio. The results demonstrate that the Proposed Hybrid ANN-SVM algorithm out performs the existing algorithm achieving Recall scores between 87 to 95 .In contrast the existing algorithms only achieve scores between 46 to 67 and 58 to 72. These results indicate that the Proposed Hybrid ANN-SVM algorithm provides significantly better performance.

4 F-Measure

No of data	ANN	KNN	Proposed Hybrid ANN-SVM
100	49	57	89
200	56	61	90
300	62	67	93
400	41	73	95
500	69	75	97

Table 4. Comparison table of F-Measure

The comparison table 4 shows the F-Measure values of existing ANN and KNN algorithms, as well as the ProposedHybrid ANN-SVM algorithm. The comparison indicates that the Proposed Hybrid ANN-SVM algorithm outperforms the existing algorithms, with F-Measure values ranging from 89 to 97. In contrast, the existing algorithms only achieve scores between 49 to 69 and 57 to 75. These results demonstrate that the Proposed Hybrid ANN-SVM algorithm provides significantly better performance.

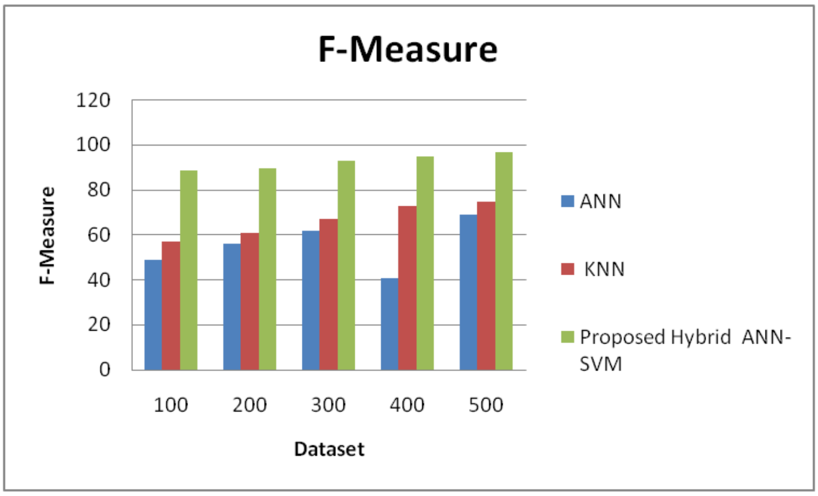

Fig 4. Comparison chart of F-Measure

The figure 4 shows a comparison chart of the F-Measure scores for existing ANN and KNN algorithms, as well as the ProposedHybrid ANN-SVM algorithm. The x-axis represents the dataset, while the y-axis shows the F-Measure ratio. The results demonstrate that the Proposed Hybrid ANN-SVM algorithm out performs the existing algorithms achieving F-Measure scores between 89 to 97.In contrast the existing algorithms only achieve scores between 49 to 69 and 57 to 75. These results indicate that the Proposed Hybrid ANN-SVM algorithm provides significantly better performance.

5. Conclusion

The Hybrid Artificial Neural Network (ANN) with Support Vector Machine (SVM) approach for real-time fault detection in solar PV systems has shown promising results. The combination of the ANN and SVM algorithms can effectively detect and classify faults in real-time. This approach can beused to improve the reliability and efficiency of solar PV systems, reducing maintenance costs and increasing the life span of these systems. The proposed algorithm can be further optimized and adapted to other renewable energy systems for real-time fault detection. The implementation of this approach in the solar PV industry can lead to improved performance and overall sustainability of renewable energy systems.

References

[1] Li, K., Zhang, J., He, Y., Xie, Y., Liu, J., & Chen, Q. (2019). A fault diagnosis method of PV arrays based on improved deep belief network. IEEE Access, 7, 156167-156178.

[2] Liu, C., Wang, L., & Chen, H. (2019). Fault detection of photovoltaic arrays based on improved support vector machine. Journal of Physics: Conference Series, 1348(1), 012001.

[3] Ma, Y., Yang, J., Liu, X., Zhang, L., & Wang, S. (2019). Fault diagnosis of photovoltaic systems based on support vector machine optimized by genetic algorithm. Renewable Energy, 132, 347-358.

[4] Niu, Y., Liu, Y., Zhang, S., Wei, X., & Cao, B. (2020). A new approach for fault diagnosis of photovoltaic arrays based on adaptive principal component analysis and support

[5] Li, H., Zhang, Y., Li, J., & Li, J. (2020). A fault diagnosis method for photovoltaic array based on decision tree optimized by improved genetic algorithm. Energy Procedia, 187, 380-385.

[6] Zaidi, M. R., Hua, L., Alotaibi, F., El-Hajjar, M., & Khan, M. (2021). A random forest approach for real-time fault detection in photovoltaic systems. Applied Sciences, 11(8), 3767.

[7] Zhang, Y., Chen, X., & Wang, L. (2021). A fault diagnosis method of photovoltaic arrays based on extreme gradient boosting decision tree. Energy Procedia, 189, 217-223.

[8] Zhu, L., Xu, J., Zhang, J., & Liu, C. (2020). Fault detection of photovoltaic arrays based on decision tree with feature selection. IEEE Transactions on Industrial Informatics, 16(11), 7155-7164.

[9] Da Liu, Kun Sun 2019 Random forest solar power forecast based on classification optimization[J] Energy 1871 .

[10] Shereen Afifi, Hamid Gholam Hosseini, Roopak Sinha 2018 A System on Chip for Melanoma Detection Using FPGA-based SVM Classifier[J] Microprocessors and Microsystems

[11] Stewart Genéa, Kamata Akihito, Miles Rona, Grandoit Evan, Mandelbaum Faigy, Quinn Crystal, Rabin Laura 2019 Predicting mental health help seeking orientations among diverse Undergraduates: An ordinal logistic regression analysis[J] Journal of affective disorders 257

[12] Gianluca Gazzola, Myong Kee Jeong 2019 Dependence-biased clustering for variable selection with random forests[J] Pattern Recognition 96

[13] Slobbe Laurentius C J, Füssenich Koen, Wong Albert, BoshuizenHendriek C, Nielen Markus M J, Polder Johan J, Feenstra Talitha L, van Oers Hans A M 2019 Estimating disease prevalence from drug utilization data using the Random Forest algorithm [J] European journal of public health 29(4)

[14] M L B., Sripriya T., Muthuraj B., Kumar D.S., Venkatesh V., Sridevi B.S., Krishna M.M.S., Rajan K.&Diriba A.(2022),"Deep Learning-Based Smart Hybrid Solar Water Heater Erection Model to Extract Maximum Energy",International Journal of Photoenergy.

[15] Aslam, M. U., Zahoor, M., & Hassan, M. U. (2020). Real-time fault detection in solar PV systems using machine learning techniques. Energy Reports, 6, 1399-1409.

[16] Mubarak, A. R., & Koo, I. M. (2018). A real-time fault detection system for PV arrays based on machine learning techniques. Energies, 11(11), 3178

[17] Anitha, R. Senthil kumar et all (2023) Design and Implementation of a Smart Solar Irrigation System Using IoT and Machine Learning, E3S Web of conferences 387

[18] Babu, Siva Raman et all (2023) Intelligent Energy Management System for Smart Grids Using Machine Learning Algorithms, E3S Web of conferences 387