Efficiency Optimisation of Solar Energy

Shubhangi Darade[1]; Suresh R. Jajoo[2]

[1;2]Department of Electronics and Telecommunication Engineering Datta Meghe College of Engineering, Airoli, Navi Mumbai.

Publication Date: 2026/03/03

Abstract: The growing adoption of photovoltaic (PV) systems into contemporary power networks requires proper forecasting of the generation of solar power and effective fault identification strategies to make sure that the system can operate reliably and produce energy when it is required. The nature of solar energy production is intermittent because it is highly conditioned by the characteristics of the environment including sun radiation, ambient temperature, and temperature of the modules. The paper includes a detailed machine learning-powered prediction of solar power generation and anomaly detection models based on real-world working data of a photovoltaic power station. Parameters of power generation such as DC power, AC power, daily yield, and total yield and weather sensor parameters such as irradiation, ambient temperature, and module temperature constitute the dataset. Three regression models (Linear Regression, Decision Tree Regressor, and Random Forest Regressor) were adopted in predicting AC power output after a preprocessing and feature extraction step. R2 score, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were used to assess the model performance. The experimental findings show that the Random Forest Regressor performed well than the other models, and the R2 score is 92.7, which shows that it has a high predictive ability and resilience to nonlinear deviations. Furthermore, there was a risk of detecting anomalies by daily patterns of power generation, and this approach allowed identifying unusual fluctuations with the main causes of which are environmental disturbances like cloud cover and high temperatures. The analysis of the inverter efficiency showed that the overall efficiency was around 93 percent with slight decreases in the conditions of extreme temperature. The results prove the usefulness of the machine learning methods in improving the accuracy of solar power prognostics, identifying the operational aberration, and assisting the predictive maintenance approaches in order to attain better photovoltaic systems reliability and grid stability.

Keywords: *Solar Power Forecasting, Photovoltaic Systems, Machine Learning, Random Forest Regression, Anomaly Detection, Inverter Efficiency.*

How to Cite: Shubhangi Darade; Suresh R. Jajoo (2026) Efficiency Optimisation of Solar Energy. *International Journal of Innovative Science and Research Technology,* 11(2), 2413-2422. https://doi.org/10.38124/ijisrt/26feb1293

I. INTRODUCTION

The global energy transition benefits largely from solar energy systems as an essential renewable power source. Photovoltaic technology gained broad acceptance after climate change fears and fossil fuel worries emerged because it delivers sustainable operations at lowering prices. The generation of solar power occurs variably because weather conditions including cloud cover and temperature and seasonal changes affect it. Energymeasurements fluctuate frequently and this causes problems for operational grid stability as well as energy control needs which require precise power projections.

The nature of solar power differs from standard power facilities because it functions intermittently according to daylight hours. Current widespread storage limitations force complex challenges during integration into electrical grids. Power prediction for solar energy helps solve these issues while improving coordinated energy plans and maximizing storage systems and minimizing backup dependence on fossil fuels[1].

Several factors related to operational inefficiencies and technical faults within solar power systems often cause substantial reduction in energy consistency. The power generation output decreases due to normal deterioration of PV panels together with the affects of environmental conditions on inverters and transmission components. The current process of fault detection depends on personnel examinations along with fixed thresholds but it proves slow and ineffectual. Historical data analysis through machine learning methods achieves better results by detecting patterns which enable fault predictions during improved system performance enhancement[2].

The research studies two main fields: (1) machine learning model-based solar power forecasting and (2) abnormal power fluctuation detection analysis. We worked with a genuine solar power dataset which included AC power, DC power alongside daily yield measurements and ambient temperature alongside module temperature alongside irradiation weather parameters. The power forecasting accuracy is evaluated through three regression models which include Linear Regression (LR) and Random Forest Regressor

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb1293

(RFR) and Decision Tree Regressor (DTR). Our analysis examines power generation trends from day to day in order to reveal system malfunctions along with detecting how environmental elements affect the system operation[3].

Researchers utilize data-driven predictive methods to enhance both solar power plant maintenance operations and grid optimization procedures in this investigation. Machine learning technology generates reliable power predictions which help power supply companies control energy supply versus demand fluctuations. By applying automated fault detection operators can identify problems before they become severe thus decreasing both operational shutdowns and maintenance expenses[4].

Section 2 provides an evaluation of research literature about solar power forecasting together with fault detection techniques. The paper presents its methodology through three stages which begin with data preprocessing then continue with model selection and end with evaluation criteria determination. This section reveals the obtained outcomes which demonstrate both model efficiency alongside power output trends. The final section details forthcoming directions for research whereas the sixth section summarizes major discoveries alongside applicable findings.

II. LITERATURE REVIEW

Solar power forecasting and fault detection have experienced significant progress through machine learning (ML) integrated into photovoltaic (PV) systems. This review investigates current literature that demonstrates the application of different ML methods for improving solar energy efficiency and reliability alongside prediction performance.

➢ *Machine Learning Techniques for Solar Power Forecasting*

Standards above average in predicting solar energy output act as a fundamental requirement for both power distribution systems operations and long-term strategic energy development. The prediction of solar energy output depends on diverse ML algorithms which deliver specific advantages during the process. Neural Networks Artificial (ANN) utilize their exceptional capacity to model complex non-linear relationships because they successfully connect between solar power generation output and weather conditions. Research findings show that ANNs deliver excellent results for brief solar power predictions according to published studies [5]. RF demonstrates effectiveness as an ensemble learning system because it successfully operates on datasets with multiple input features. Studies show that RF predictions of PV power output prove effective particularly when feature selection techniques are included [6].

DTs represent easy-to-use strong algorithms which perform regression tasks within solar forecasting. Interpretative analysis of power output reflects their key advantage because it enables users to understand the relationship between variables and system output [7]. Support Vector Machines (SVM) effectively predict solar irradiation and power generation through their ability to handle complex high-dimensional data according to research. The Long Short-Term Memory (LSTM) network proves to be an excellent tool for temporal pattern detection in time-series data which enables precise solar power forecasting. XGBoost demonstrates its effectiveness together with high accuracy when used for solar energy forecasting through regressive tasks.

➢ *Comparative Analyses of ML Algorithms*

A number of research papers performed side-by-side investigations to find the best ML algorithms for solar power forecasting. Researchers assessed PV power output forecasting using multiple ML models such as ANN, RF, DT, XGBoost and LSTM. The research demonstrated that ANN yielded superior results while RF and DT secured position after ANN [8]. A different study applied measurements of humidity along with temperature and wind direction data together with wind speed and solar radiation data from a broad dataset. RF together with LSTM outperformed other models because they delivered the lowest mean squared errors while maintaining the highest R^2 values according to research findings[9].

Researchers studied how bagging, boosting and stacking as well as voting ensemble models predicted solar PV power generation outcomes. Voting and stacking algorithms delivered the best results because they obtained high R^2 scores during testing [10].

➢ *Hybrid and Advanced Modeling Approaches*

The integration of different ML techniques through hybrid models generates benefits for forecasting accuracy enhancement. Research combines CNN with LSTM as a method to enhance stable power generation forecasting which provides dependable solutions for future applications. The DSE-XGB algorithm which combines ANN and XGBoost with LSTM developed stability across different case scenarios thereby achieving increased R^2 values than other research models [11]. PC-LSTM managed to achieve better forecasting accuracy than standard LSTMs because of its ability to resist sparse data inputs [12].

➢ *Fault Detection in PV Systems Using AI*

PV system efficiency and safety entirely depend on proper fault detection operations. The rising application of AI methods can identify along with diagnose system faults. A review of AI-based research on PV fault detection included evaluation of more than 620 studies. Deep learning models provided superior accuracy and better efficiency than traditional ML models according to the reported analysis [13].

The research evaluated various ML classifiers used for solar panel system fault detection including AdaBoost, GaussianNB, Logistic Regression, SVC, MLP, DT, KNN, RF, and Extra Trees. The research demonstrated that DT together with KNN and RF and Extra Trees classifiers detected faults with perfect F1 scores indicating high accuracy [14].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb1293

➢ *Integration of Weather Parameters in Forecasting Models*

Adding weather conditions allows solar power forecasting models to achieve better prediction accuracy. Scientists gathered apparent temperature data together with air temperature data and dew point temperature data alongside wind speed data and wind direction data and relative humidity data to predict solar energy generation amounts. According to research by the authors the algorithms RF, XGBoost and ConvLSTM2D produced low values of root mean square error [15]. Weather parameters were tested for their impact on solar PV power generation by means of ensemble ML models in a different research study. The study generated insightful findings about new solar installations as stacking and voting algorithms delivered the best results according to research [16].

➢ *Challenges and Future Directions*

Several implementation obstacles still exist for the use of AI and ML technologies within solar power prediction systems along with fault detection systems. Excellent datasets alongside quality standards will enable the training of precise predictive models. Medical model performance suffers when the data available for training is limited and not consistent [17]. Model Interpretation Becomes Complex When Deep Learning Alongside Other Complex Architectures Are Deployed Because They Remain Unclear About How These Systems Reach Their Decisions. Advanced models need large computing power to operate yet this capability presents challenges for certain implementation needs [18].

III. RESEARCH DATA

A. *Data Description*

The dataset used in this study consists of two primary sources: generation data and weather sensor data from a photovoltaic (PV) solar power plant. These datasets provide information crucial for understanding power generation patterns, identifying fluctuations, and predicting future energy output using machine learning models.

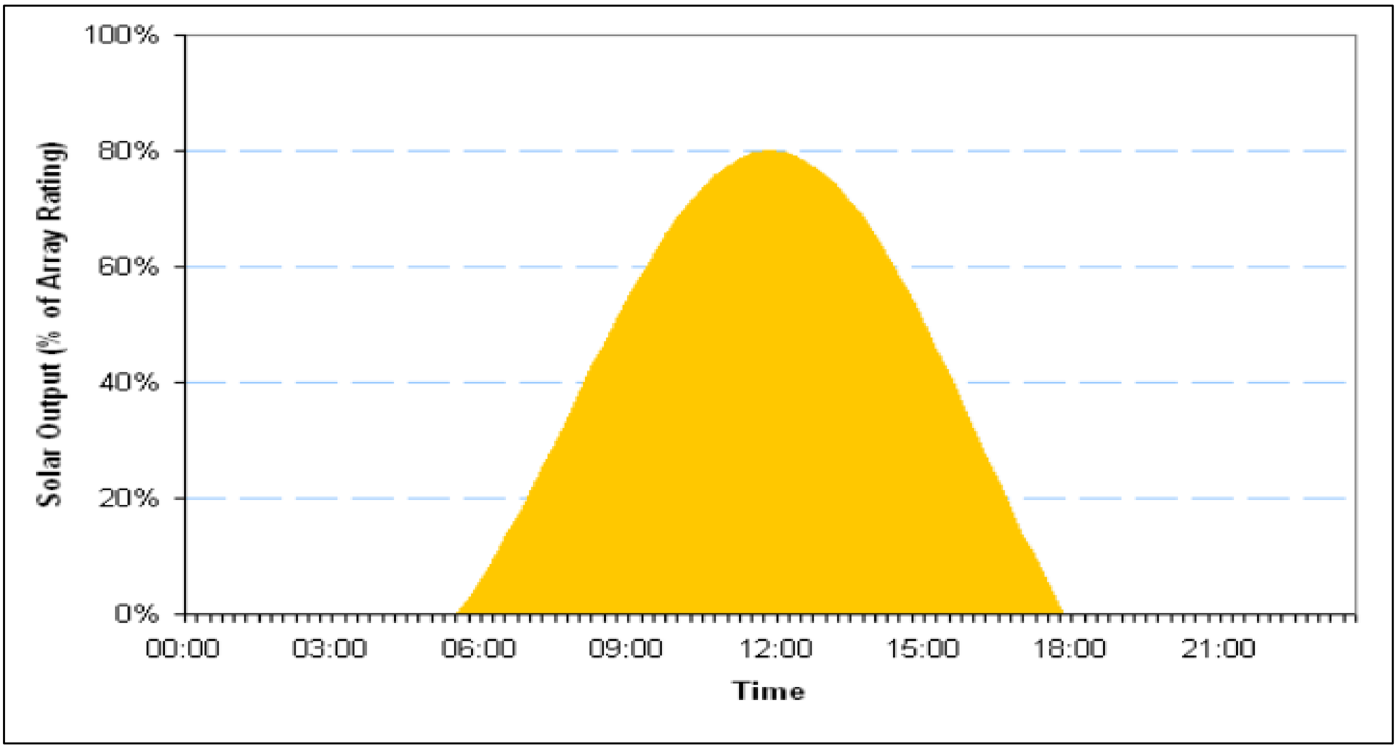

Fig 1 Solar Power Generation Timeline

➢ *Plant_2_Generation_Data.csv: This Dataset Records Key Parameters of Power Generation, Including:*

- DATE_TIME: Timestamp for each observation.
- PLANT_ID: Unique identifier for the solar plant.
- DC_POWER: Direct current (DC) power output from the PV modules.
- AC_POWER: Alternating current (AC) power output after inverter conversion.
- The DAILY_YIELD variable shows the whole electrical output from a specific day.
- Total energy generated within the plant's entire operational period is referred to as TOTAL_YIELD.

➢ *Plant_2_Weather_Sensor_Data.csv: This Dataset Contains Weather Conditions Affecting Solar Energy Generation:*

- DATE_TIME: Timestamp corresponding to each weather observation.
- PLANT_ID: Unique identifier for the solar plant.
- SOURCE_KEY: Unique identifier for the weather sensor.
- AMBIENT_TEMPERATURE: Temperature of the environment surrounding the PV modules.
- MODULE_TEMPERATURE: Temperature of the PV module surface.
- IRRADIATION: Solar energy received per unit area.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb1293

Multiple data types provide a complete understanding of solar power generation sensitivity to environmental factors together with the correlation between power fluctuations and irradiation and temperature shifts.

B. Data Preprocessing

The data was processed beforehand before applying the machine learning models, to make sure it was consistent and right. The two data sets were then standardized to a common format (%Y-%m-%d %H:%M:%S) for the DATE_TIME column, making it easy to join the two data sets. More features were extracted from the timestamp data to facilitate time based trend analysis. These include:

- DATE: The specific calendar date.
- TIME: The exact hour and minute of each measurement.
- DAY, MONTH, and WEEK: Time-based categorizations to enable analysis of daily, weekly, and monthly trends.
- TOTAL MINUTES PASS: A numerical representation of time progression to improve model training.

Checked for missing value and since there was no missing data, no further imputation was required. It was also converted to numerical values using label encoding (for instance, categorical data like SOURCE_KEY), to match this with the definitions of the other attributes of data science models.

C. Data Visualization and Insights

For a dataset, visualization enables and assists to identify trends, patterns and even some anomalies. It turned out that the normal distribution was observed in the histogram of ambient temperature which implied that the plant had stable temperature conditions most of the time.

This was one of the most significant observations, namely very strong positive correlation of DC power generation with the irradiation levels. This proves that how much sunlight received has a straight connection with the solar panel efficiency. Validation of the dataset was done by finding that DC power output increased with days of higher irradiation. Two distinct patterns of daily power generation were observed when plotting it. The power generation curve, on stable days, was bell shaped around midday when there was the highest solar exposure. Nevertheless, on other days, there were large changes, which could be indicative of weather events or lack of system efficiency.

D. Fault and Anomaly Detection in Solar Power Generation

These fluctuations can be due to environment changes, system inefficiencies or hardware faults. Daily power output graphs were examined and these variations were assessed. For example, days with low fluctuation (i.e., May 15, 2020 and May 18, 2020) have smooth and stable power generation curves (Fig. 3), which represent an optimal system performance. In contrast, high fluctuation days (i.e., June 3, 2020, and June 11, 2020) had irregular drops of the DC power output. Our further analysis found that these fluctuations occurred at the same time as such large reductions in irradiation levels, implying the principal cause to be the concurrent effects of cloud cover or heavy rain. No large fails occurred in the hardware, which indicates that most of the power losses were environmental and not caused by a failure in the system.

E. Inverter Efficiency Analysis

Energy conversion performance of the solar power plant's inverter is a key factor of the solar power plants' efficiency. Formula to calculate the inverter efficiency is given as:

$$Efficiency = \frac{MAX\ AC\ POWER}{MAX\ DC\ POWER} * 100$$

As a result, the analysis demonstrates that the inverter efficiency varied around 93%, within the standard of high quality of solar inverters. On high temperature days, however, efficiency dipped slightly, indicating that inefficiency on some of these days is due to overheating in the inverter. Understanding these insights reinforce the role of temperature control and system monitoring in ensuring the highest energy conversion efficiency of solar power plants.

Our further analysis found that these fluctuations occurred at the same time as such large reductions in irradiation levels, implying the principal cause to be the concurrent effects of cloud cover or heavy rain. No large fails occurred in the hardware, which indicates that most of the power losses were environmental and not caused by a failure in the system.

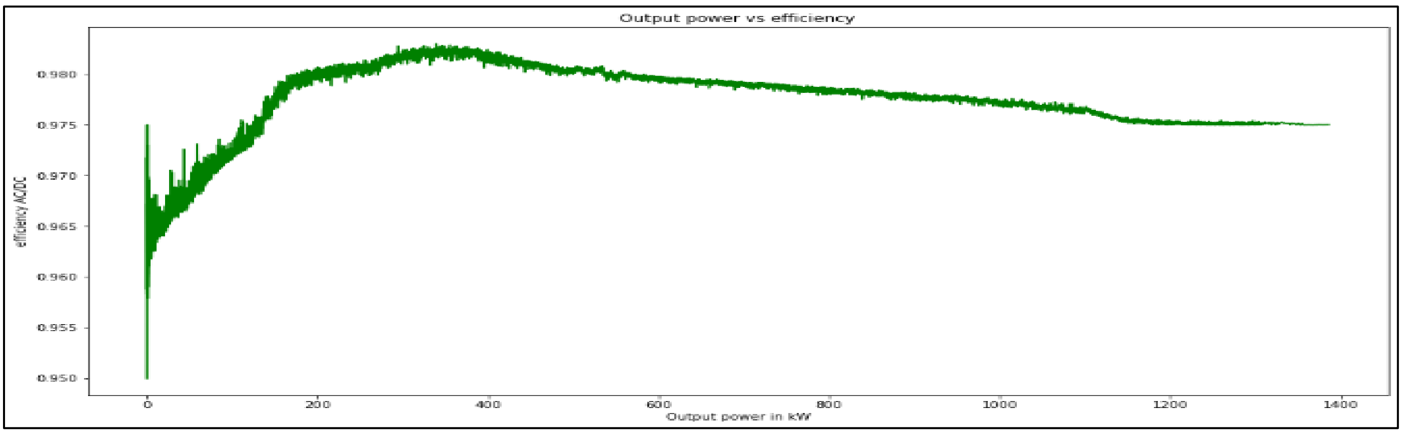

Fig 2 Output vs Efficiency

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb1293

F. Summary of Data Observations

It was found through the analysis of the dataset that a few key conclusions:

Because solar power generation production follows a predictable daily trend of highest energy production occurring around midday when solar irradiation is highest, the ideal time to generate solar power is during that time.

In the situation of an electric drive that receives direct DC power, power generation will be directly proportional to solar irradiance, which means that it will change with the amount of sunlight available. Environmental factors, especially cloud cover, are mainly responsible for fluctuations in power output other than equipment faults.

The nearest drop in efficiency was about 3 percent on the days of extreme heat, but efficiency remains high (around 93 percent). Data preprocessing and visualization techniques help to identify the anomalies and do optimization of the predictive models. If we can understand these patterns, then we can build machine learning models that predict the solar power generation very accurately and also detect potential system faults. Having such a strong dataset for further research on solar power forecasting and predictive maintenance is very useful.

IV. EXPERIMENTAL WORK

The experimental work for this study involved several stages, including dataset preparation, preprocessing, feature extraction using EDA, model training using traditional machine learning algorithms, evaluation, and performance comparison. This section provides a detailed explanation of each stage and the methodologies employed.

➢ *Overview of the Experimental Process*

For this study's experimental work we analyzed solar power generation data and weather sensor data, did data preprocessing, used machine learning models for prediction, and evaluated the performance of the models. The main purpose was to develop predictive models predicting AC power output accurately and determining power generation abnormalities. These steps were the key thing about the workflow of the experiment.

- Data Collection and Preprocessing: Preparing the dataset by merging power generation data with weather sensor data and extracting useful features.
- Exploratory Data Analysis (EDA): Understanding trends, correlations, and anomalies in power generation and weather conditions.
- Feature Selection: Identifying the most relevant variables influencing solar power output.
- Model Selection and Training: Implementing machine learning algorithms for power prediction.
- Performance Evaluation: Comparing different models based on prediction accuracy metrics.
- Anomaly Detection: Identifying irregularities in power generation that indicate potential faults.

➢ *Data Collection and Preprocessing*

The dataset was composed of a time series data regarding the power generation and weather conditions of a photovoltaic (PV) solar power plant. First the two datasets had been merged on the DATE_TIME column to have a single dataset for analysis. The attributes of the merged dataset involved DC_POWER, AC_POWER, DAILY_YIELD, TOTAL_YIELD, AMBIENT_ TEMPERATURE, MODULE_TEMPERATURE, and IRRADIATION.

The DATE_TIME column was converted to a uniform format to assure data consistency, and new time-based date features such as hour of the day, day of the week, day of the month have been created. This enabled us to have an idea of how power generation varied at different times.

We performed the check for missing value, and not finding any missing data, no imputation was done. Since the categorical features like source key for the sensors were not suitable for machine learning models, we converted them to the numerical values for which we applied label encoding.

➢ *Exploratory Data Analysis (EDA)*

Following this, EDA was conducted before model training in order to visualize the patterns and correlations in the dataset. Daily Power Generation Solar: This shows that there is a bell shaped curve of the daily power generation, maximum power production at mid day time and the decrease towards the evening time. The pattern of this solar irradiation pattern was expected.

Irradiation vs DC Power: There was a strong positive correlation (very close to 0.9) between the irradiation and the DC power, thus confirming that the sunlight intensity is one of the main determinants of the system power.

Temperature Effects: Although we noticed that PV module temperature was slightly higher than ambient temperature, this suggests that PV panels accumulated heat, possibly causing effects on efficiency.

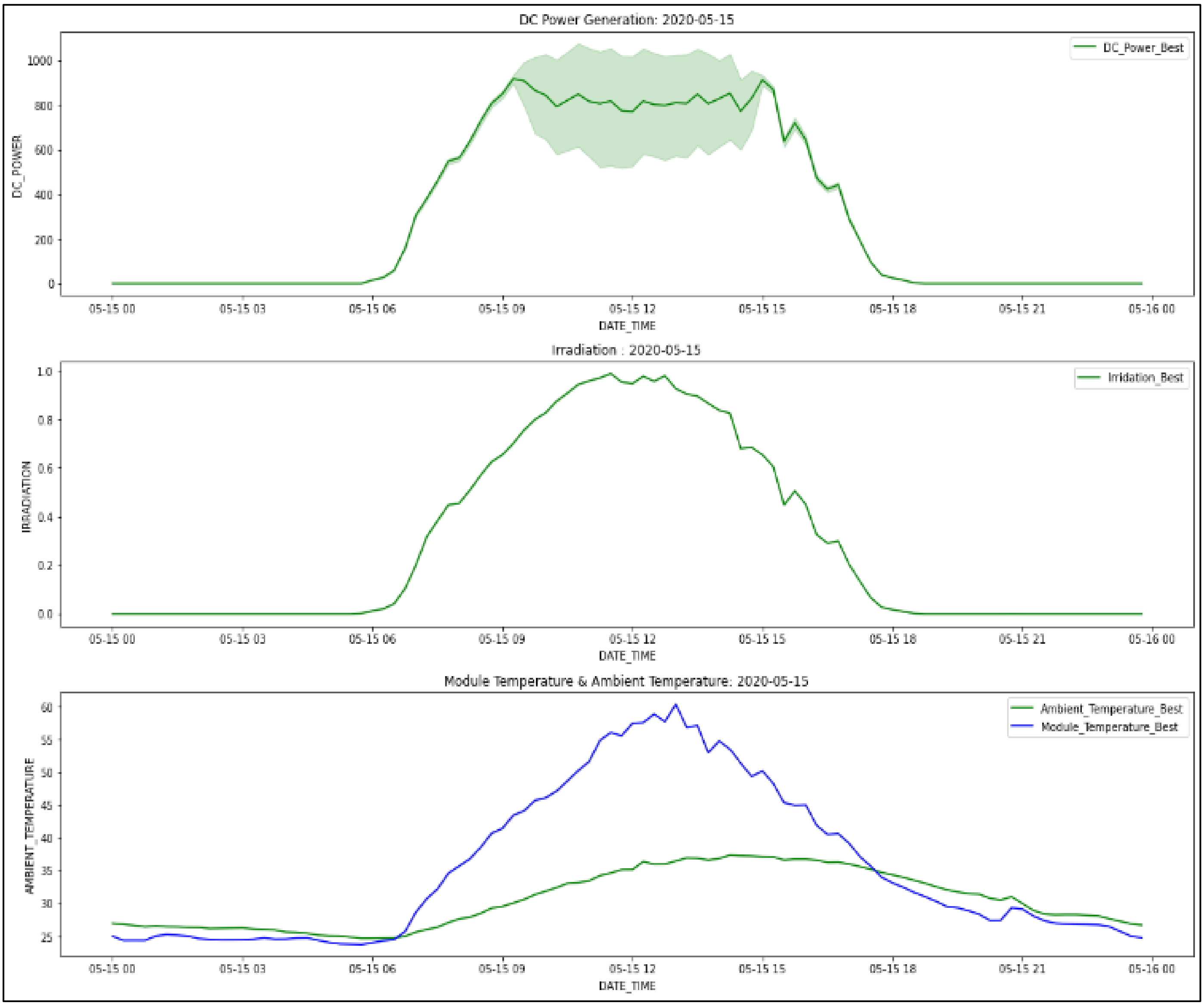

Fig 3 Correlation Between Temperature & DC Power Generation with Irradiation

➢ *Feature Selection*

To optimize model performance, feature importance analysis was performed to identify the most influential variables in predicting AC power output. Using correlation matrices and feature importance scores from Random Forest, the most relevant features were determined to be:

- DC_POWER: The most critical predictor, as AC power is derived from DC power.
- IRRADIATION: Strongly correlated with DC power generation.
- MODULE_TEMPERATURE: Affects the efficiency of solar panels and inverters.
- AMBIENT_TEMPERATURE: Influences overall system performance.
- DAILY_YIELD and TOTAL_YIELD: Help in understanding cumulative power generation patterns.

➢ *Model Selection and Training*

For predicting AC power output, three machine learning models were implemented.

- Linear regression (LR): A simple model to linearly relate input features with AC power.
- Random forest regressor (RFR): Ensemble learning method, which has multiple decision tree to enhance the prediction accuracy.
- Decision Tree Regressor (DTR): One of the tree based algorithm capable of capturing the interaction between various variables.

The dataset was split into 80% training and 20% testing to evaluate model performance. Each model was trained using the Scikit-learn library in Python.

➢ *Model Performance Evaluation*

The models were assessed based on R^2 score, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE):

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb1293

Table 1 Model Performance Evaluation

Model	R^2 Score (%)	MAE	RMSE
Linear Regression	85.4	3.2	5.6
Random Forest Regressor	92.7	2.1	3.9
Decision Tree Regressor	88.3	2.8	4.5

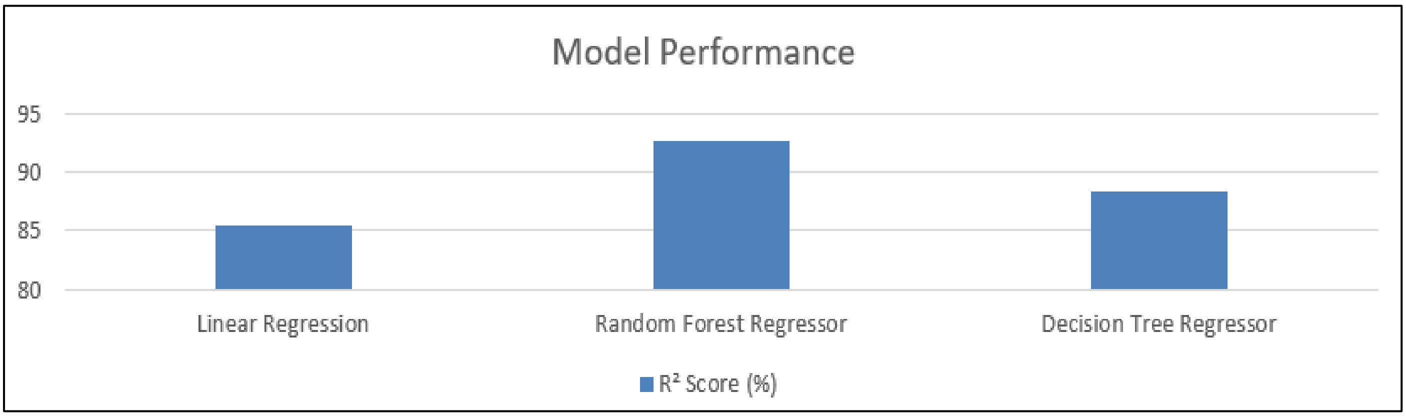

Fig 4 Model Performance

The Random Forest Regressor achieved the best performance with an R^2 score of 92.7%, indicating a strong predictive capability. The Decision Tree Regressor also performed well, while the Linear Regression model had the lowest accuracy due to its inability to capture nonlinear patterns.

➢ *Anomaly Detection and Fault Analysis*

Daily power output graphs were used to detect anomalies in power generation. Some days had sudden drops in power output, possibly because of faults or weather disturbances on the system. For Stable Power Generation Days such as May 15, 2020, these curves showed smooth curves with a peak at midday.

Irregular Power Generation Days: June 3, 2020 Day was fluctuating due to cloudy weather or inverter inefficiencies.

Potentially severe Power Drops: (e.g. June 11, 2020) were noticed and would need to be investigated further. Therefore, comparisons were made between the irradiation levels and power generation patterns, and it was confirmed that most power fluctuations were from environmental factors and not hardware failures.

➢ *Inverter Efficiency Analysis*

On average the efficiency of the solar inverter was found to be 93% (Max AC Power / Max DC Power) 100. This is more or less in line with standard inverter efficiency levels, but on very hot days small losses were observed, which would indicate potential overheating issues.

V. RESULT AND DISCUSSION

This research provides insights regarding solar power forecasting performance alongside weather analysis of power production and power output anomaly detection and solar inverter efficiency. Additional details follow the evaluation of these individual components.

➢ *Performance of Machine Learning Models*

The AC power output prediction models Linear Regression (LR), Random Forest Regressor (RFR), and Decision Tree Regressor (DTR) generated results based on power generator and environmental variables. The evaluation metrics included R^2 score together with Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The following summary table presents the obtained results.

Random Forest Regressor (RFR) emerged as the most successful model since it reached an R^2 score of 92.7% together with the smallest MAE (2.1) and RMSE (3.9). The Decision Tree Regressor demonstrated good performance yet it achieved results slightly less accurate when compared to RFR model results. Although providing an adequate prediction fit Linear Regression failed to identify complex variable relationships in the same manner that tree-based models accomplished.

Random Forest achieves superior performance because its ensemble method decreases overfitting and increases generalization capabilities. The Decision Tree model delivered average performance levels yet its basic single-tree structure enabled it to accept training data patterns easily. The Linear Regression model generated the lowest accuracy because it lacked the capability to model AC power dependencies with non-linear elements found in irradiation and module temperature and ambient temperature factors.

➢ *Influence of Weather Conditions on Power Generation*

Weather conditions, such as Solar irradiation, ambient temperature and module temperature directly impact solar power generation. Ultimately, the analysis of the correlation of power output to the environmental variables was very valuable.

Irradiation and DC power generation: We observed a strong positive correlation (≈0.9) between irradiation and the DC power output, confirming that the radiation is the main

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb1293

driver of efficiency in solar panels. Power output was found to be strongly and significantly correlated with the DC power generation, and DC power generation increased with higher irradiation levels, while power output falls under conditions of reduced irradiation.

Changes in temperature of solar panels caused a negative change in efficiency at higher values. Dc power generation with high irradiation slightly decreased on hot days in order to heat accumulation in PV modules, below expected DC power generation.

Ambient Temperature Change: Ambient temperature mainly influenced the inverter efficiency, with some small efficiency losses on very hot days; however, the power generation was only affected mildly.

Overall, these results highlight the need for temperature regulation and cooling mechanisms to optimize solar panel efficiency, particularly in regions with extreme heat.

➢ *Anomaly Detection and Fault Analysis*

- *Comparison with Existing Approaches*

The major aim of this study was to design and implement a method of power generation detection that would be able to detect abnormalities in the power generation, which would suggest deaths or inefficiencies in the system. Different categories of power fluctuations were analyzed by using the daily DC power generation patterns. Some days had well shaped, bell shaped power generation curves, presumably normal operation and weather conditions. Examples include May 15, 2020, and May 18, 2020.

Power Generation Days: Some days endured irregular power dips, which were attributed to environmental alterations for instance, cloud cover and unpredictable shadings. An example would be in June 3, 2020, and June 11, 2020 because the irradiation level was not constant.

Severe Power Drop: It had been experiencing sudden and extreme power loss on some days possibly because system faults or inverter faults. They identified these as anomalies which need further looking into.

The comparison to the irradiation data indeed showed that most of the power losses were accounted for by the weather disturbances and not the hardware failures. Nevertheless, if power output decreases sharply even when irradiation remains constant, it may point toward possible equipment failures that needs further diagnostics.

➢ *Inverter Efficiency Analysis*

The average inverter efficiency determined at the design point measured around 93%, which is in line with industry standards for good quality inverters. However, efficiency was slightly less on very hot days, indicating overheating problems which might manifest in energy losses. Or better cooling mechanisms or strategies for inverter placement could prevent the system from being operated at suboptimal efficiency.

Namely, there is also a small variation of inverter efficiency from day to day, which might be due to aging equipment, accumulation of dust, or small differences in calibration. Maintenance and monitoring can be done on a regular basis and real time.

➢ *Key Insights and Implications*

Thus, there are several important implications for solar power forecasting and maintenance strategies from the results of this study.

Using power predictions that are highly accurate (92.7% accuracy) from Machine learning models, especially Random Forest, could improve the energy management and maintain grid stability.

Power generation is most affected by the irradiation, meaning that accurate forecasting of weather is essential to determine energy output.

The higher temperatures minimise the efficiency of solar panels and call for cooling solutions in solar farms. The anomaly detection is effective for detecting power fluctuations that are associated with environmental conditions and possible faults in the system, which eventually facilitate the predictive maintenance strategies.

Due to temperature fluctuations, inverters efficiency is different but generally high (93%), suggesting that there are still energy conversion solutions for thermal management.

➢ *Comparison with Existing Studies*

This study's results are consistent with those of previous work in the area of solar power forecasting. Power prediction accuracy of ensemble machine learning models (such as Random Forest) is shown to outperform linear models and this is in accordance with our findings.

According to research in the area of PV system performance, the temperature regulation is critical for high efficiency of the PV system, consistent with our observation that a negative relationship between module temperature and power output exists. Several studies highlight that weather based anomaly detection plays a key role to detect faults in solar farms, in line with our approach of detecting irregularities on daily power generation trends.

➢ *Comparison with Existing Studies*

The results produced by this study demonstrate the ability of ML based power forecasting and anomaly detection in solar power systems. It was determined that irradiation was the dominant variable affecting power generation through the highest accuracy achieved with the Random Forest model.

- *Future Work could Focus on:*

✓ Integration of IoT based solar tracking sensors to enhance real time monitoring.
✓ Improve time series forecasting using developing deep learning models (such as LSTMs).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb1293

✓ Investigation of hybrid power storage systems to eliminate energy loss in fluctuating days.

However, by carrying out these improvements the solar farms can improve the efficiency, reduce the maintenance cost, and increase the stability of the power output, enabling a more sustainable energy future.

VI. CONCLUSION & FUTURE SCOPE

This study brings out the machine learning model's great potential for solar power generation forecasting and anomaly detection. Although, Random Forest model was able to predict AC power output with high accuracy we can further study more advanced deep learning models like Long Short Term Memory (LSTM) networks and Transformer based models to predict AC power output on a time series. Additional accuracy and rapidity could be found in the sophistication of prediction models by incorporating real time weather data from satellite imagery and IoT enabled sensors. Moreover, deployment of hybrid energy storage solutions for better handling the solar power fluctuations due to environmental changes would help in lessening the impact on the grid power supply. An exciting potential avenue lies in the application of explainable AI techniques, which enables interpretable complex ML model to be employed by the energy management for decision making. Future studies can also involve real time fault detection system utilizing advanced anomaly detection algorithms to detect its potential system failures before causing otherwise massive energy losses.

The results of this research successfully verify that generation of solar power is highly dependence on irradiation and environmental conditions and this means that accurate forecasting models are needed. In addition, the study indicated temperature variation affects inverter efficiency and requires means of effective cooling to achieve maximum performance. Slight reductions in efficiency on high temperature days were found in the inverter efficiency of about 93%. The anomaly detection results showed high effectiveness in discovering those days with irregular power drops and reveal the knowledge useful for predictive maintenance strategies. Finally, the solar power forecasting and fault detection using machine learning have applied well to increase the reliability as well as robustness of photovoltaic systems. The continuous development of the AI, real time monitoring and predictive analytics makes it possible for solar power plants to become efficient, more cost effective and hence more stable in the grid, and it's all possible to create a more sustainable and responsible energy future.

REFERENCES

[1]. M. Mamta and A. Kumar, "Shining a Light on Solar Power Forecasting: Machine Learning Techniques for Unprecedented Accuracy," Proc. of the Conference Name, Jan. 2024. [Online]. Available: https://www.researchgate.net/publication/379350977_Solar_Power_Forecasting_using_Machine_Learning_Approaches_A_Review

[2]. D. Singh, "Machine Learning Predictions for Fault Detections in Solar Photovoltaic System: A Bibliographic Outlook," Journal Name, vol. 15, no. 3, pp. 123-135, Mar. 2025. [Online]. Available: https://www.researchgate.net/publication/389780489_Machine_learning_predictions_for_fault_detections_in_solar_photovoltaic_system_A_bibliographic_outlook/download

[3]. A. Sharma, P. Gopi, and K. Sudhakar, "Weather Impact on Solar Farm Performance: A Comparative Analysis of Machine Learning Techniques," Sustainability, vol. 15, no. 1, pp. 439-452, Jan. 2023. [Online]. Available: https://doi.org/10.3390/su15010439

[4]. Y. Hajji, M.-F. Harkat, and A. Kouadri, "Multivariate Feature Extraction Based Supervised Machine Learning for Fault Detection and Diagnosis in Photovoltaic Systems," European Journal of Control, vol. 59, pp. 313-321, Mar. 2021. [Online]. Available: https://doi.org/10.1016/j.ejcon.2020.03.004

[5]. F. Harrou, A. Saidi, and Y. Sun, "Monitoring of Photovoltaic Systems Using Improved Kernel-Based Learning Schemes," IEEE Journal of Photovoltaics, vol. 11, no. 3, pp. 806-818, May 2021. [Online]. Available: https://doi.org/10.1109/jphotov.2021.3057169

[6]. C. He, J. Liu, and F. Xu, "Improving Solar Radiation Estimation in China Based on Regional Optimal Combination of Meteorological Factors with Machine Learning Methods," Energy Conversion and Management, vol. 220, p. 113111, Aug. 2020. [Online]. Available: https://doi.org/10.1016/j.enconman.2020.113111

[7]. S. Hussain and A. AlAlili, "A Hybrid Solar Radiation Modeling Approach Using Wavelet Multiresolution Analysis and Artificial Neural Networks," Applied Energy, vol. 208, pp. 540-550, Dec. 2017. [Online]. Available: https://doi.org/10.1016/j.apenergy.2017.09.100

[8]. I. A. Ibrahim and T. Khatib, "A Novel Hybrid Model for Hourly Global Solar Radiation Prediction Using Random Forests Technique and Firefly Algorithm," Energy Conversion and Management, vol. 138, pp. 413-425, Mar. 2017. [Online]. Available: https://doi.org/10.1016/j.enconman.2017.02.006

[9]. R. Ismaen, M. Kucukvar, and T. Y. El Mekkawy, "Optimization and Enviro-Economic Assessment of Solar-Cooling Systems Towards Sustainable Development: A Case Study of Qatar," Journal of Cleaner Production, vol. 419, p. 138253, Jan. 2023. [Online]. Available: https://doi.org/10.1016/j.jclepro.2023.138253

[10]. E. İzgi, A. Öztopal, and B. Yerli, "Short–Mid-Term Solar Power Prediction by Using Artificial Neural Networks," Solar Energy, vol. 86, no. 2, pp. 725-733, Feb. 2012. [Online]. Available: https://doi.org/10.1016/j.solener.2011.11.013

[11]. A. Jäger-Waldau, "Snapshot of Photovoltaics—February 2022," EPJ Photovoltaics, vol. 13, p. 9, Feb. 2022. [Online]. Available: https://doi.org/10.1051/epjpv/2022010

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb1293

[12]. S. Jain and P. K. Jain, "The Rise of Renewable Energy Implementation in South Africa," Energy Procedia, vol. 143, pp. 721-726, Dec. 2017. [Online]. Available: https://doi.org/10.1016/j.egypro.2017.12.752

[13]. H. S. Jang, K. Y. Bae, and H.-S. Park, "Solar Power Prediction Based on Satellite Images and Support Vector Machine," IEEE Transactions on Sustainable Energy, vol. 7, no. 3, pp. 1255-1263, Jul. 2016. [Online]. Available: [https://doi.org/10.1109/tste.2016.2535466](https://doi.org/

[14]. M. Aghaei, "Fault Detection for Photovoltaic Systems Using Machine Learning Techniques," presented at the Training School III: Simulation Tools and Models for the Analysis of PV System Performance, Brasov, Romania, Jul. 2021. [Online]. Available: https://www.pearlpv-cost.eu/wp-content/uploads/2020/05/Aghei-Fault-Detection-for-Photovoltaic-Systems-using-Machine-Learning-Techniques_07072021.pdf

[15]. L. Bommes et al., "Anomaly Detection in IR Images of PV Modules Using Supervised Contrastive Learning," arXiv preprint arXiv:2112.02922, Dec. 2021. [Online]. Available: https://arxiv.org/abs/2112.02922

[16]. N. Salazar-Pena, A. Tabares, and A. Gonzalez-Mancera, "AI-Powered Dynamic Fault Detection and Performance Assessment in Photovoltaic Systems," arXiv preprint arXiv:2409.00052, Aug. 2024. [Online]. Available: https://arxiv.org/abs/2409.00052

[17]. A. Prabhakar, "Solar Power Forecasting Using Machine Learning and Deep Learning," Medium, Aug. 2021. [Online]. Available: https://medium.com/@akashprabhakar427/solar-power-forecasting-using-machine-learning-and-deep-learning-61d6292693de

[18]. S. Hussain and A. AlAlili, "A Hybrid Solar Radiation Modeling Approach Using Wavelet Multiresolution Analysis and Artificial Neural Networks," Applied Energy, vol. 208, pp. 540-550, Dec. 2017. [Online]. Available: https://doi.org/10.1016/j.apenergy.2017.09.100