

Enhancing Solar Power Generation Prediction: A Novel Grading Approach Leveraging Random Forest and Decision Tree Models

Rishabh Kushawah¹, Akash Arya², Komal Gahtori³, Sneha Lata Singh⁴, Divisht Jaiswal⁵, Neha Bisht⁶

^{1,3,4,5,6}Department of Computer Science & Engineering, Dr. A.P.J Abdul Kalam Government Institute of Technology, Tanakpur, Uttarakhand, India

²Department of Computer Engineering, G.B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand, India

ABSTRACT

Solar power generation prediction is crucial for sustainable energy management, as weather conditions significantly influence irradiance and temperature. This paper introduces the mean grading approach, which integrates qualitative Solar Power Generation Grades (SPGG) from ensemble model with quantitative regression prediction. The SPGG scale spans 1–16, reflecting weather impacts on solar output. Experimental results validate the framework's effectiveness: the ensemble model achieved training accuracy of 99.25% and testing accuracy of 97.90%, while the RandomForestRegressor yielded an MSE of 0.292, RMSE of 0.540, MAE of 0.104, and R^2 of 0.974. The approach enhances energy scheduling, grid integration, and renewable adoption, advancing sustainable systems.

Key Words: Solar power generation, mean grading approach, ensemble model, regression model, SPGG scale, energy optimization.

1. INTRODUCTION

As we recognize the indispensable role of electricity in driving economic growth, it's noteworthy that our country ranks as the world's 3rd largest producer and 4th largest consumer of electricity. With a total installed capacity of approximately 415 Gigawatts (GW) in 2022, comprising 15% solar energy, 10% wind energy, 13% hydropower, and 10% natural and other gases alongside traditional carbon-based sources, the energy landscape is diverse. Notably, industries constitute the largest share of electricity consumption (approximately 45%), followed by secondary households (24%) and the agriculture sector (18%). The imperative of integrating a substantial amount of renewable energy into the power grid underscores the significance of accurate weather forecasting for efficient grid management and renewable energy integration. This integration is crucial for minimizing greenhouse gas (GHG) emissions from conventional power plants, such as carbon dioxide (CO₂), methane (CH₄), and nitrous oxide (N₂O), thereby enhancing the sustainability and efficiency of the transmission and distribution system. Solar energy emerges as a primary renewable energy source due to its ubiquitous availability worldwide. However, obtaining precise solar data remains challenging due to its location-dependent nature, presenting a critical global challenge for researchers. In light of these challenges, leveraging advanced machine learning (ML) techniques becomes imperative to effectively predict solar power generation and manage renewable energy resources. In this context, we aim to explore and elucidate advanced ML research approaches to address future challenges and identify robust ML techniques for solar power generation prediction.

1.1 Smart Grid and Renewable Energy

The evolution of traditional electrical grids into sophisticated Smart Grid systems marks a crucial shift towards efficient energy management and optimization. In response to escalating energy demands and environmental concerns, governments worldwide are spearheading initiatives to modernize grid infrastructure. Notably, the Indian government's National Smart Grid Mission (NSGM) reflects a concerted effort to harness advanced technologies for enhancing power generation and distribution networks. Central to this transformation are Smart Grid solutions, characterized by bidirectional communication and an array of cutting-edge features such as automation, energy monitoring, and control mechanisms. These innovations aim to synchronize power generation with distribution in real-time or future scenarios, ensuring optimal utilization of resources.

Among all, Solar energy become more popular and highly demandable sustainable energy because of abundant and free availability. but they require more research as it is very much dependent on weather conditions like solar energy needs direct and full sun light in solar PV and wind energy require a powerful wind speed. Now a days India has fourth largest market of renewable energy in the world. According to a review in 2022 a total of 14.21 GW renewable energy is produced during January to October 2022 and in Dec 2023 the total installed capacity of Renewable energies 180.79 GW in which solar power is alone 73.31 GW and India has set an target at the of 500 GW of non-fossil based energy by 2030. To achieve these targets there are so many solar power projects are initiated by Government of India (GOI) such as "Development of Solar Parks and Ultra Mega Solar Power projects" which aims a target of total 40 GW capacity by March 2024. Although so many projects are there for the development and integration of RE but still there is a big Scope of research in this field as generation of renewable energy is depends directly or indirectly on the weather conditions. in our country as climate change is the main concern or challenge as generation of Solar energy is principally dependent on weather conditions, hence we need to predict the weather conditions along with various atmospheric parameters. Keeping the same in mind we need a deep analysis of various weather parameters for the purpose of accurate forecasting in order to achieve an efficient and powerful integration of solar plants. Machine learning and Deep learning are the most efficient tools that will help us to get our goal through a scientific study of algorithms.

Smart Grid solutions hold the promise of addressing several critical challenges plaguing conventional grid systems. By leveraging automation and advanced monitoring capabilities, these solutions can effectively reduce transmission and distribution losses, manage power during peak demand periods, and enhance overall service quality and reliability. Moreover,

they enable streamlined asset management, facilitate seamless integration of renewable energy sources (RESs), and improve accessibility to electricity. Among the various RESs, solar energy emerges as a particularly viable and increasingly popular option, given its abundance and minimal environmental impact.

Here in this paper, we present regression-based estimation of solar power generation based on weather forecasting. Our proposed framework aims to address the challenges associated with solar power integration in Smart Grid systems by employing regression-based estimation methods and ML algorithms. By leveraging smart meter data and historical energy datasets, our framework endeavors to predict solar power generation levels accurately. Through a comparative analysis of regression models, we seek to identify the most suitable ML algorithm for renewable energy prediction within the Smart Grid context. Ultimately, our research endeavors to enhance Smart Grid performance by leveraging advanced ML techniques for efficient solar energy management and integration.

The objectives of this paper is:

- Develop a comprehensive methodology for calculating Solar Power Generation Grade (SPGG) values based on weather conditions and their impact on solar energy production.
- Utilize machine learning techniques to analyze historical weather data and predict SPGG values accurately, considering both qualitative and quantitative factors.
- To conduct a comparative analysis of the proposed framework to identify the most effective ML algorithm for predicting solar power generation within.
- Investigate the potential applications of the SPGG approach in optimizing solar energy integration within Smart Grid systems, aiming to enhance grid efficiency and reliability.

1.2 Overview of ML Techniques

This section endeavors to undertake a comparative analysis of various machine learning (ML) algorithms, focusing primarily on those pertinent to weather predictions. Initially, machine learning can be categorized into three subcategories: Supervised ML, Unsupervised ML, and Reinforcement ML. Supervised learning operates by leveraging historical data, known as training data, to derive insights and generate predictions based on established patterns. Given our research's emphasis on smart meter data analysis and prediction, this section delves into well-established supervised learning models, aiming to identify the most suitable ML algorithm for renewable energy prediction research. Notably, supervised learning techniques are commonly employed to address Regression and Classification problems. Regression methods, integral to statistical modeling, analyze labeled datasets to ascertain relationships between variables and predict continuous values, uncovering correlations between dependent and independent variables. Various ML models rooted in regression methods include simple linear regression, multiple linear regression, support vector regression (SVR), Decision Tree, Random regression (RF), and k-nearest neighbor (KNN), among others. On the other hand, Classification methods are deployed for prediction and forecasting endeavors, wherein trained datasets are categorized into discrete values, with output variables representing discrete values. Common ML techniques based on

classification methods comprise logistic regression (LR), kernel SVM, support vector regression (SVR), Decision Tree classification, random forest classification (RF), k-nearest neighbor (KNN), and Recurrent neural network (RNN), among others. While both methods offer avenues for prediction in machine learning, their distinction lies in their approach towards various machine learning problems.

The transition towards sustainable and renewable energy sources, particularly solar power generation, has become imperative in addressing global energy challenges and mitigating climate change. Accurate prediction of solar power generation is crucial for effective energy management and optimization, especially as the adoption of solar energy systems continues to increase. Weather conditions play a significant role in determining the efficiency and reliability of solar power generation, with factors such as cloud cover and temperature influencing solar irradiance levels. In response to these challenges, this paper proposes a comprehensive framework for predicting solar power generation by integrating ensemble modeling and regression techniques. This framework aims to provide a holistic approach to forecasting solar power generation by combining qualitative assessments of weather conditions with quantitative estimations of solar energy production potential. Through a combination of data preprocessing, feature engineering, and machine learning model development, the proposed methodology offers a systematic and robust approach to predicting solar power generation under varying environmental conditions. By leveraging historical weather data and advanced modeling techniques, the proposed framework seeks to enhance the accuracy and reliability of solar power generation predictions, thereby facilitating informed decision-making in energy management and contributing to the advancement of sustainable energy systems.

2. METHODOLOGY

The approach uses a manner that concatenates data preprocessing, feature engineering, machine learning model development, and ensemble technique which predict weather conditions and optimize solar power production. A detailed methodology is set in several key steps, each designed to improve the accuracy and efficiency of making a forecast of the amount of solar power production.

2.1 Dataset Description

Main dataset to be used in the present study solar_data_prediction.csv contains a wide scale of meteorological parameters observed in various geographic points, with 98,602 entries containing meteorological conditions and air quality indices. Others include temperature_celsius, condition_text (e.g., Sunny, Partly, and cloudy), wind_mph, humidity, uv_index, air_quality_PM2.5, among others. In addition, this is done by converting sunrise and sunset times to numbers indicating minutes since midnight for the purpose of converting the data for machine learning.

2.2 Data Preprocessing

The first steps of preprocessing involve filtering out data with weather conditions that are considered irrelevant or which do not affect solar power prediction significantly. Afterwards, the timings of sunrise and sunset from standard time formats are converted into integer representations these represent the total

minutes since midnight. This is all in an effort to further simplify the input features of the model.

2.3 Features Selection:

Feature selection focuses on removing the condition_text from the feature set (X) and designated it as the target variable (y) in the classification tasks. In this regard, this process will separate the inputs (weather and environmental variables) from the output (weather condition text) quite clearly, hence critical in the training of the models.

2.4 Weather Condition Classification

Process:

- **Data Acquisition:** The methodology begins with the collection of a comprehensive weather dataset.
- **Dataset Filtering:**
 - Unnecessary columns that do not contribute to the weather condition prediction are dropped to streamline the dataset.
 - Basically, it means to take away those values that are required not from the target column. In this case, the entry of weather condition that is not relevant or noisy, which serves to influence the model wrongly.
- **Preparation of the Filtered Dataset:** After applying the filter, the dataset should be prepared. It is divided into training and testing set. In this sense, it is the division of the data that leads to training a predictive model with one set of the data while testing the performance of the trained model in another set of unseen data.
- **Model Initialization:** Two machine learning models are initialized for classification tasks:
 - The Random Forest Classifier is configured with parameters such as the number of estimators and tree depth.
 - The Decision Tree Classifier is set up with its respective parameters, including the depth of the tree and the minimum samples required for a split.
- **Ensemble Learning:** An ensemble model is built using a Voting Classifier that combines predictions of Random Forest Classifier and Decision Tree Classifier to come to a consensus for the prediction.
- **Model Training and Evaluation:** It trains the ensemble model with the dataset, and finally, it evaluates the ensemble model. Then, the evaluation studies the accuracy of the model against the actual data related to the meteorological conditions, studying its effectivity and readiness for deployment into the real world.

2.5 Solar Power Generation Grade Prediction Model:

- **Utilization of the Filtered Dataset:** The data from previous steps, obtained after preprocessing, will be used for predicting the solar power generation grade.
- **Feature Augmentation:** Add a new column to the dataset for the grading of solar power generation. The grading shall be based on predefined criteria to map weather conditions against a set of solar power output levels.
- **Creation of the New Dataset:** After appending the old dataset with the new column, the whole dataset had been ready for regression analysis.
- **Application of the Random Forest Regressor:** It is an ensemble model derived from decision trees. The application of the regression model is initiated with parameters, which predicts continuous values matching the solar power generation grade.
- **Model Training:** With the trained model and data, a new dataset can be formed, thus providing the learning to predict solar power generation given the weather conditions and other relevant features.
- **Model Evaluation:** After training, a trained Random Forest Regressor model is tested to verify that the prediction is accurate, which has to be consistent and reliable for practical uses in solar energy management. It shows how well the model can make accurate predictions about solar power generation, which is required prior to any valid application in managing solar energy.

2.6 Solar Power Generation Prediction

This section of the methodology explained the process of combining the outputs of weather condition classification with the solar power generation grade prediction to get the estimate for solar power generation.

- **Weather Condition Classification Model:**

First, classification of weather conditions takes place based on historical weather data through a predictive model. One of the outputs from this model shall be the "Predicted Weather Condition," which is required in the process of assessing possible solar irradiance.

- **Solar Power Generation Grade Prediction Model:**

Equally with the two given predictive models, another one makes a prediction for the "Predicted Solar Power Generation Grade Value." This value would express the expected level of solar power generation, depending on both the average features of the environment and the average weather.

- **Calculation of SPGG (Solar Power Generation Grade) Value:**

It computes the value of SPGG corresponding to the "Predicted Weather Condition," which is one of the predefined conditions. This step contextualizes the condition of weather with an expected level of solar power generation.

- **Calculation of Mean SPGG Value:**

Values thus obtained from these two predictive models are then combined to be used to find the "Mean SPGG Value." This means integrating the qualitative aspect of weather condition prediction with the quantitative estimation from the grade in solar power generation prediction and hence forming a consolidated assessment of the expected output of solar energy.

- **Estimation of Solar Power Generation:**

On the basis of these, the Mean SPGG value is derived, which also represents final estimation information about the solar energy that could be made under forecasted weather conditions and, in the process, give valuable forecasts in the planning and optimization of energy production.

2.7 Model Initialization Parameters

2.7.1 RandomForestClassifier Parameters:

- **n_estimators=100:** This parameter explicitly talks about the number of trees in the forest. Large numbers give more strength and accuracy to the model, but will be more computationally complex. We chose 100 trees as a balance between performance and computational efficiency.
- **max_depth=15:** Each of the trees has been applied a restriction: the maximum depth of 15 levels that have been enforced. This constraint helps in avoiding the model from being very expressive and fitting exactly to the training data.
- **min_samples_split=5:** If a condition is created, then a node has to possess at least five samples before it could make a split. This parameter helps in ensuring that the trees have enough sample number to give meaningful splits.
- **min_samples_leaf=2:** A leaf node should have at least two samples. This configuration helps avoid the model's learning of noise in the training data and therefore enhances generalization.
- **max_features='sqrt':** At every split in the tree, the maximum number of features to consider is a square root number of features. This introduces randomness into the model and therefore does well to achieve diversity among the trees in the forest, hence reducing the variance of the model.
- **max_leaf_nodes=None:** This parameter is set to None, without any restraint on the number of leaf nodes present in a tree, and it fully expands the tree where it is required.
- **min_impurity_decrease=0.0:** This parameter indicates the minimum value that impurity needs to reach for it to split nodes. A node will only split if the impurity after the split is less than the value given. Setting the value to 0.0 indicates that any impurity decrease will be considered.

2.7.2 DecisionTreeClassifier Parameters:

- **max_depth=15:** As in the case of Random Forest Classifier, the depth of the tree is limited to 15 levels so as not to overfit and, at the same time, to ensure the tree does not grow over-complex.
- **min_samples_split=5 and min_samples_leaf=2:** This works equivalently to a Random Forest Classifier with a change in these parameters: the minimum number of samples required to split a node.
- **max_leaf_nodes=None:** This allows growing the tree in complexity and size as needed, but with a hard limit on the number of leaf nodes.
- **min_impurity_decrease=0.0:** The partition is considered if it reduces impurity, so finer the granularity in the classification process.

2.7.3 Voting Classifier:

- **voting:** If 'hard', uses predicted class labels for majority rule voting. Else if 'soft', predicts the class label based on the argmax of the sums of the predicted probabilities. For example, **voting="hard"**.

2.8 Model Development

Two primary models are developed:

2.8.1 Ensemble Classifier for Weather Condition Prediction:

It is an ensemble model that combines classifiers (Random Forest Classifier and Decision Tree Classifier) to predict the weather condition text, leveraging strengths to enhance accuracy and robustness. The ensemble model is trained with a Voting Classifier optimized for best accuracy among all available weather conditions.

2.8.2 Random Forest Regressor for Solar Power Predictions:

This model predicts the solar power generation order, a numeric value indicative of the predicted solar energy output, aiming to establish a quantitative relationship between weather conditions and solar power generation potential using Random Forest Regressor.

2.8.3 Integration and Prediction

The incorporation of weather forecasts and optimization of solar power production is carried out in a sequential approach:

- **Weather Condition Classification:** Using the ensemble classifier, weather conditions are predicted based on selected input features from the dataset.
- **Solar Power Generation Prediction:** The Random Forest Regressor, informed by both the predicted weather condition and other environmental variables, estimates the solar power generation order.

2.9 Solar Power Generation Grade Classification

This study introduces a novel classification system for solar power generation based on weather conditions and the predicted solar power generation order. It categorizes solar power generation potential into several grades, from 'Very high' to 'Negligible', based on predefined criteria. The classification is derived from the mean of the predicted condition's integer value and the solar power generation order, enabling a comprehensive assessment of solar energy production capabilities under various weather conditions.

2.9.1 Grading Based on Weather Conditions

The method used an innovative way of grading to quantify the effect of different weather conditions on solar power generation. The system is illustrated in Table 1: Criteria for grading of power generated under different weather.

Each weather condition is numerically graded from 1 in the grading system. Based on these, the awarded grades to this condition simply reflect the expected solar power output from the condition: The lower grades indicate good solar power output, while the higher ones show poor and less favorable conditions.

For example, the 'Sunny' conditions, graded at 1, are expected to provide the highest output of solar power, while the grading 'Fog,' graded at 16, is supposed to have the least output. These graded standardizations will help in directly relating the weather conditions with the potentiality of power generation.

Table -1: Criteria for Solar Power Generation Grading Based on Weather Condition

Weather Condition	Solar Power Generation Grade
Sunny	1
Partly cloudy	2
Clear	3
Light drizzle	4
Mist	5
Light rain	6
Moderate rain	7
Cloudy	8
Overcast	9
Heavy rain	10
Light sleet	11
Heavy snow	12
Moderate or heavy rain with thunder	13
Moderate or heavy rain shower	14
Moderate or heavy snow showers	15
Fog	16

2.9.2 Grading for Solar Power Generation

For this view, Table 2: Criteria for Solar Power Generation has been introduced as a tool to convert the weather condition grades into action-oriented solar power generation forecasts.

This table classifies the derived values of both the weather condition classification model and the solar power generation prediction model into a number of solar power generation descriptors from "Very high" to "Negligible," according to the mean grade. In other words, "mean grade" is an important metric value joining the predicted "predicted solar power generation grade" with the "predicted weather condition grade value" joined. For instance, mean grades 1 and 12 would suggest 'very high' solar power generation potential if ≤ 1 , and 'extremely low' if 12 to 15.

Table -1: Criteria for Solar Power Generation

Mean Grade	Solar Power Generation
Mean Grade ≤ 1	Very high
$1 < \text{Mean Grade} \leq 3$	High
$3 < \text{Mean Grade} \leq 6$	Medium
$6 < \text{Mean Grade} \leq 9$	Low
$9 < \text{Mean Grade} \leq 12$	Very low
$12 < \text{Mean Grade} \leq 15$	Extremely low
$15 < \text{Mean Grade} \leq 16$	Negligible

All these findings will form part of the predictive models while calculating mean Solar Power Generation Grade. From this, we have assigned grades from the predicted weather condition and the predicted grade of solar power generation, respectively. This thereby ensures that both the qualitative and the quantitative qualitative parts of the weather and the solar power data are observed during the final estimation of the solar power generation. This study identifies the development of such grading tables that systematically predict and classify the solar power generation potential at any location under different weather situations. In other words, they structurally handle the interpretation of complex weather data in such a way as to come up with workable and pragmatic forecasts of solar power. Fig. 1 represents the overall flow of the proposed framework.

This methodology, encompassing data preprocessing, predictive modeling, establishes a robust framework for enhancing solar power generation predictions. By accurately forecasting weather conditions and their impact on solar energy output, this study contributes to optimizing solar power production, thereby supporting the integration of renewable energy sources into the global energy mix.

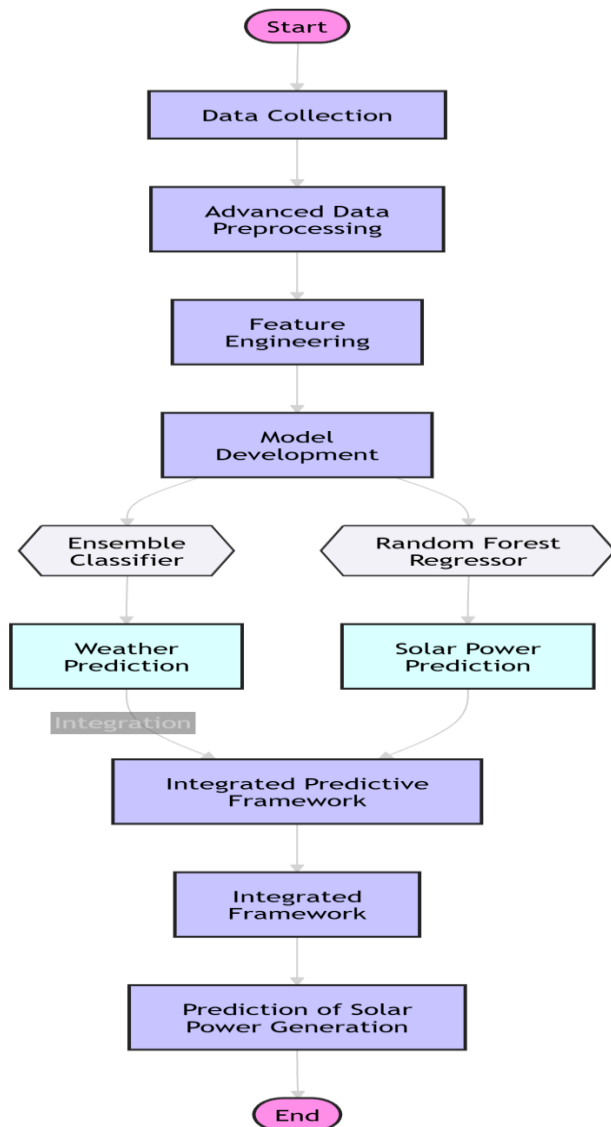


Fig -1: Methodology Flow Chart

3. RESULT AND DISCUSSION

This study aims to develop a robust model for predicting solar power generation by utilizing weather data and advanced machine learning techniques. Our approach involves a two-step process: first, an ensemble classifier predicts weather conditions, and then a separate regressor model estimates the SPGG on a scale from 1 to 16. We then determine the SPGG value corresponding to the forecasted weather conditions using predefined criteria. By averaging both SPGG values, we classify solar power generation into seven categories (very high, high, medium, low, very low, extremely low, negligible) based on predefined criteria. This proposed methodology significantly enhances the accuracy of solar power generation forecasts across a range of weather conditions.

The next section details the results. We will explore how well the classification and regression models performed. We will also examine the predicted solar power generation levels for each weather category and how this can be used for managing solar energy systems. Finally, we will assess the overall

effectiveness of our model and identify areas for future improvement.

3.1 Model Performance Metrics:

3.1.1 Classification Model Evaluation

Metrics:

- **Accuracy:** Accuracy measures the proportion of correctly classified instances out of the total instances.
- **Precision, Recall, and F1-score:** Precision represents the ratio of correctly predicted positive observations to the total predicted positives, while recall measures the ratio of correctly predicted positive observations to all actual positives. F1-score is the harmonic mean of precision and recall, providing a balanced evaluation of the model's performance.
- **Confusion Matrix:** We constructed confusion matrices to visually represent the model's performance by comparing predicted labels against actual labels. This matrix allows us to identify the number of true positives, true negatives, false positives, and false negatives, aiding in understanding the classification model's strengths and weaknesses.

3.1.2 Regression Model Evaluation Metrics:

In the evaluation of our regression model, we employed the following metrics:

- **Mean Squared Error (MSE):** The Mean Squared Error calculates the average of the squared differences between the predicted and actual values. A lower MSE indicates that the model's predictions are closer to the actual values.
- **Root Mean Squared Error (RMSE):** The Root Mean Squared Error is the square root of the MSE, providing an interpretable measure in the same unit as the target variable. It represents the average magnitude of the errors in the predicted values.
- **Mean Absolute Error (MAE):** The Mean Absolute Error calculates the average of the absolute differences between the predicted and actual values. It provides a more straightforward measure of the model's predictive accuracy.
- **R-squared (R²):** The R-squared metric measures the proportion of the variance in the dependent variable (solar power generation grades) that is predictable from the independent variables (weather conditions). It ranges from 0 to 1, with higher values indicating a better fit of the model to the data.

3.2 Weather Classification Result:

The ensemble classification model exhibited strong performance with a training accuracy of 99.25% and a testing accuracy of 97.90%. These high accuracies indicate that the model effectively learned and generalized patterns from the training data to make accurate predictions on unseen testing data. The slightly lower testing accuracy compared to the

training accuracy suggests minimal overfitting, indicating that the model's performance is reliable and consistent across different datasets. These results demonstrate the effectiveness of the classification model in accurately predicting weather conditions based on the input features, which is crucial for subsequent solar power generation forecasts.

The provided table offers insights into the performance of the classification model across different weather conditions, as indicated by precision, recall, and F1-score metrics. Remarkably high precision scores are observed for "Clear" and "Sunny" conditions, implying that the model's predictions for these weather states are highly accurate. The model exhibits notable precision for "Cloudy," "Fog," "Mist," and "Overcast" conditions, suggesting its ability to make precise classifications in diverse weather scenarios. However, it is worth noting that precision scores vary across weather conditions, with lower values observed for "Light drizzle," "Light rain," "Moderate rain," and "Moderate or heavy rain shower," indicating some degree of uncertainty in predicting these conditions accurately. The model demonstrates high recall rates for most weather categories, indicating its capability to correctly identify instances of these conditions. The F1-score, which balances precision and recall, further supports the model's overall effectiveness, with generally high scores across various weather conditions. Despite minor variations in performance across specific weather categories, the classification model exhibits robust performance in accurately predicting weather conditions, reflecting its potential for practical applications in weather forecasting. The ensemble classification model demonstrates an average precision of 0.93, an average recall value of 0.85, and an f1-score of 0.87 across all weather conditions. Table 3 illustrates the evaluation metrics for the classification model.

Table -3: Model Evaluation Parameters of Classification Model

Weather Condition	Precision	Recall	F1-score
Clear	1.00	1.00	1.00
Cloudy	0.99	1.00	0.99
Fog	0.97	1.00	0.98
Heavy rain	1.00	1.00	1.00
Heavy snow	0.83	0.83	0.83
Light drizzle	0.70	0.90	0.79
Light rain	0.71	0.77	0.74
Light sleet	1.00	0.50	0.67
Mist	0.97	0.98	0.98
Moderate or heavy rain shower	0.80	0.94	0.87
Moderate or heavy rain with thunder	1.00	0.48	0.65
Moderate or heavy snow showers	1.00	1.00	1.00
Moderate rain	0.86	0.36	0.51

Overcast	1.00	0.89	0.94
Partly cloudy	0.99	0.99	0.99
Sunny	1.00	0.97	0.98

The confusion matrix has been visualized for both the training and testing datasets in the ensemble classification model. Fig. 2 and Fig. 3 depict the confusion matrix for the testing and training data, respectively. The results obtained from the classification model play a pivotal role in predicting weather conditions, which are integral to forecasting solar power generation. With high accuracy, precision, recall, and F1-score, the model reliably identifies various weather patterns such as "Clear," "Cloudy," "Rainy," or "Overcast." These predictions offer valuable insights into prevailing weather conditions, which directly impact solar irradiance and, consequently, solar power generation. For instance, clear and sunny conditions typically yield optimal solar irradiance levels, resulting in high energy production from solar panels. Conversely, cloudy or overcast conditions can diminish solar irradiance, affecting energy generation capacity. By accurately predicting weather conditions, the classification model contributes to better forecasting of solar power generation, informing stakeholders in the renewable energy sector about expected energy output. This information is crucial for energy grid operators and utility companies to optimize energy management strategies, balance energy supply and demand, and integrate solar energy into the existing energy infrastructure more effectively.

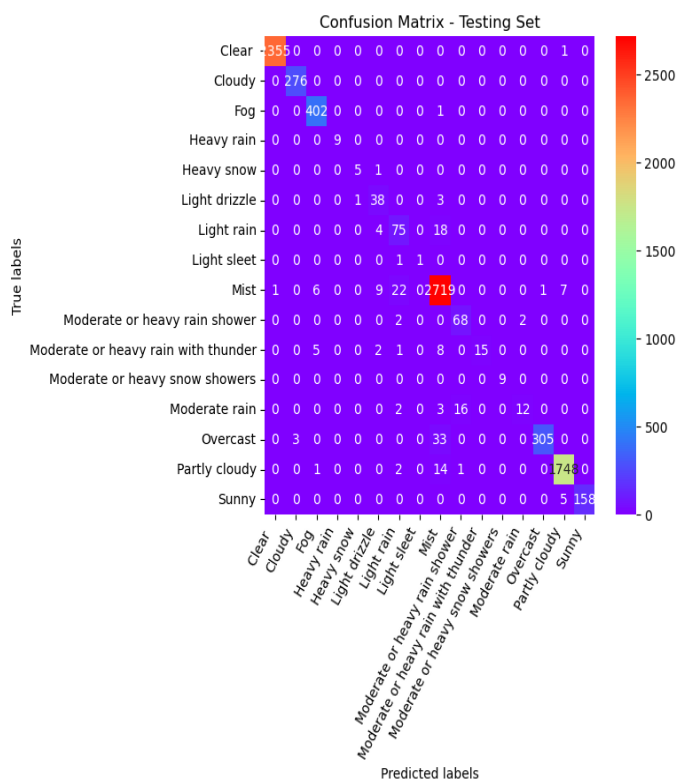


Fig -2: Confusion matrix for ensemble model for test data

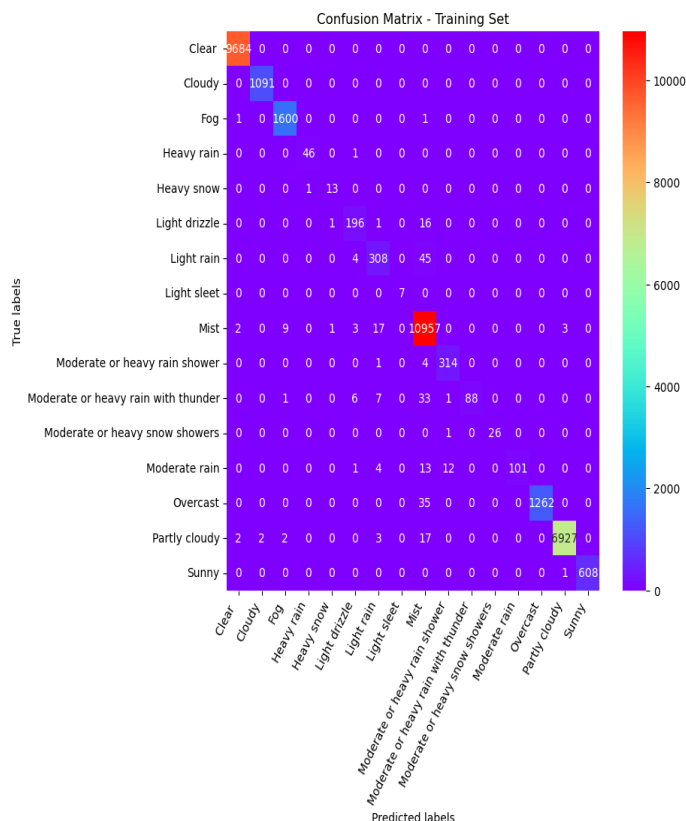


Fig -3: Confusion matrix for ensemble model for train data

3.3 Regression result:

The results obtained from the RandomForestRegressor model provide insights into the accuracy and performance of the solar power generation grade prediction. The Mean Squared Error (MSE) of 0.292 indicates the average squared difference between the predicted solar power generation grades and the actual grades. A lower MSE value suggests that the model's predictions are closer to the actual values, indicating higher accuracy.

The Root Mean Squared Error (RMSE) of 0.540 indicates the square root of the MSE, providing a measure of the average magnitude of the errors in the predicted solar power generation grades. A lower RMSE value implies that the model's predictions are more precise and closer to the actual values.

The Mean Absolute Error (MAE) of 0.104 represents the average absolute difference between the predicted and actual solar power generation grades. Like MSE and RMSE, a lower MAE value indicates better accuracy and closer alignment between predicted and actual values.

The R-squared (R2) value of 0.974 measures the proportion of the variance in the solar power generation grades that is predictable from the input features. A higher R2 value indicates that the model explains a larger portion of the variance in the data, suggesting good model fit.

Overall, the low values of MSE, RMSE, and MAE, coupled with the high R2 value, indicate that the RandomForestRegressor model performs well in predicting SPGG. The model's high accuracy and strong predictive power

make it a valuable tool for forecasting solar energy output, enabling stakeholders to make informed decisions regarding energy production and management strategies. Table 4 displays the different metrics associated with the regressor model.

Table -4: Metrics for Random Forest Regressor Model

Metric	Value
Mean Squared Error	0.2920167482976944
Root Mean Squared Error	0.540385740279751
Mean Absolute Error	0.10403058177039781
R-squared	0.9737857029028918

3.4 Prediction of Solar Power Generation

As previously stated, we have developed a framework for predicting solar power generation. This framework utilizes the mean grade value derived from the weather condition predictions of the ensemble model and the SPGG forecasted by the regressor model. By computing this mean grade value, we can make predictions regarding solar power generation. Below, we illustrate the operational process of this final framework using specific numerical examples. Table 5 illustrates the parameters assigned to the framework to facilitate solar power generation prediction. Upon parameter assignment, the framework predicts the weather condition as "Mist". Following a predefined set of criteria, the framework calculates the SPGG value for the "Mist" condition, yielding a value of 5. Subsequently, the framework incorporates the SPGG value predicted by the regressor model, which is determined to be 3.41. Through a calculation of the mean grade value for both SPGG values, totaling 4.205, the framework arrives at a prediction of solar power generation, categorized as "Medium". Table 6 presents a comprehensive display of all outcomes generated by the proposed framework.

Table -5: Parameter values for implementation of solar power prediction framework

Weather parameter	Parameter value
latitude	23
longitude	78
temperature_celsius	34
temperature_fahrenheit	78
wind_mph	9
wind_kph	28
pressure_mb	1000
pressure_in	37
precip_mm	0
precip_in	0
humidity	67
cloud	59
feels_like_celsius	34

feels_like_fahrenheit	78
visibility_km	7
visibility_miles	3
uv_index	12
air_quality_Carbon_Monoxide	433
air_quality_Ozone	12
air_quality_Nitrogen_dioxide	3
air_quality_Sulphur_dioxide	3
air_quality_PM2.5	2
air_quality_PM10	3
air_quality_us-epa-index	1
air_quality_gb-defra-index	21
hours for sunrise (0-23)	8
minutes for sunrise (0-59)	0
hours for sunset (0-23)	5
minutes for sunset (0-59)	0

Table -6: Prediction of solar power generation by proposed framework

Parameter	Result
Predicted weather condition by ensemble model	Mist
SPGG value for predicted weather condition	5
Predicted SPGG by regressor model	3.41
Mean grade value	4.205
Predicted solar power generation	Medium

3.5 DISCUSSION

The proposed framework for predicting solar power generation is designed to provide accurate forecasts. The ensemble model predicts weather conditions based on the preprocessed data, utilizing a combination of algorithms to enhance prediction accuracy. These predicted weather conditions are mapped to SPGG values using predefined criteria, capturing the potential impact on solar energy production. A regressor model is then employed to estimate the SPGG values for each weather condition. By calculating the mean grade value, which serves as a benchmark for comparison, the framework determines the overall solar power generation potential. Finally, based on the predicted SPGG values and the mean grade value, the framework generates predictions for solar power generation levels.

The proposed framework for predicting solar power generation plays a crucial role in managing smart solar power plants by providing valuable insights and aiding in effective decision-making. With precise forecasts of solar power generation, the framework facilitates better integration of solar energy into the existing grid infrastructure. The framework enables proactive energy management by providing real-time insights into expected solar power generation. Plant operators

can use the model to optimize solar power generation schedules, adjust energy storage capacities, and coordinate grid interactions to maximize energy efficiency and reliability. This can lead to cost savings, reduced reliance on non-renewable energy sources, and increased overall system resilience.

With reliable predictions of solar power generation, the framework promotes the integration of renewable energy sources into the energy mix. This supports sustainability goals, reduces greenhouse gas emissions, and contributes to the transition towards a cleaner and greener energy infrastructure. By providing accurate predictions of solar power generation, the framework helps mitigate risks associated with energy production and distribution. One limitation of the proposed model could be its reliance on the availability and quality of input data. If historical weather data or solar power generation data are incomplete or inaccurate, it may affect the model's predictive accuracy. The model likely makes certain assumptions about the relationship between weather conditions and solar power generation, as well as the stability of these relationships over time. These assumptions may not hold true in all situations or geographic regions, leading to potential inaccuracies in the predictions.

3. CONCLUSION

The proposed framework represents a significant advancement in the field of solar power generation prediction, offering a comprehensive and effective approach to managing smart solar power plants. By integrating ensemble modeling and regression techniques, the framework provides accurate forecasts of solar power generation under varying weather conditions. The experimental results demonstrate the efficacy of the framework in accurately predicting solar power generation levels, with high training and testing accuracies achieved for the ensemble classification model and low error metrics obtained for the RandomForestRegressor model. The applications of the proposed framework extend beyond individual solar power plants to broader energy management strategies, including optimizing production schedules, enhancing grid integration, and promoting renewable energy penetration. By enabling proactive energy management and reducing reliance on non-renewable energy sources, the framework contributes to the advancement of sustainable energy systems and the mitigation of climate change impacts. Future research could focus on improving the accuracy and generalizability of the model by incorporating additional data sources, such as satellite imagery, atmospheric simulations, or real-time sensor data. This could enhance the model's ability to capture localized weather patterns and short-term fluctuations in solar radiation. Another area for future exploration is the development of more sophisticated modeling techniques, such as machine learning algorithms or hybrid models combining physical principles with data-driven approaches. There is potential for expanding the model's scope to incorporate additional factors influencing solar power generation, such as cloud cover, air pollution, or terrain variations. The proposed framework represents a valuable tool for decision-making in solar energy management, offering insights into future solar power generation potential and facilitating the transition towards a cleaner and greener energy infrastructure.

REFERENCES

1. Das, S., Dey, A., Pal, A. and Roy, N., 2015. Applications of Artificial Intelligence in Machine Learning: Review and Prospect. *International Journal of Computer Applications*, 115(9).
2. Máté E. Maros^{1,2*}, Chang Gyu Cho^{1,2}, Andreas G. Junge¹, Benedikt Kämpgen³, Victor Saase¹, Fabian Siegel², Frederik Trinkmann², Thomas Ganslandt², Christoph Groden¹ & Holger Wenz Comparative analysis of machine learning algorithms for computer-assisted reporting based on fully automated cross-lingual RadLex mappings, *Scientific Reports* | (2021)
3. F. K. Dosilovic, M. Brcic, and N. Hlupic, "Explainable artificial intelligence: A survey," 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2018 - Proceedings, pp. 210–215, Jun. 2018.
4. K. Yeturu, "Machine learning algorithms, applications, and practices in data science," in *Handbook of Statistics*, vol. 43, Elsevier, 2020, pp. 81–206
5. Gupta, A. Sharma, and A. Goel, "IJERT-Review of Regression Analysis Models Review of Regression Analysis Models," *IJERT Journal International Journal of Engineering Research and Technology*, Accessed: Jun. 06, 2022.
6. M. A. Iqbal, "Application of Regression Techniques with their Advantages and Disadvantages," vol. 4, pp. 11–17, Sep. 2021.
7. Paul, D. P. Mukherjee, P. Das, A. Gangopadhyay, A. R. Chintia, and S. Kundu, "Improved Random Forest for Classification," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4012–4024, 2018.
8. Neetan Sharma^{1*}, Vinod Puri¹, Shubham Mahajan^{2,3}, Laith Abualigah^{4,5,6,7,8,9}, Raed Abu Zitar¹⁰ & Amir H. Gandomi^{11,12}, Solar power forecasting beneath diverse weather conditions using GD and LM-artificial neural networks, *Scientific Reports* | (2023)
9. R. Meenal¹, Prawin Angel Michael², D. Pamela³, E. Rajasekaran⁴, Weather prediction using random forest machine learning model, *Indonesian Journal of Electrical Engineering and Computer Science* Vol. 22, No. 2, May 2021, pp. 1208~1215
10. Sandeep Pattnaik, "Weather forecasting in India: Recent developments," *Mausam*, vol. 70, no. 3, pp. 453–464, 2019
11. Arya A, Mishra PK (2023) A comprehensive review: advancements in pretrained and deep learning methods in the disease detection of rice plants. *J Artif Intell Capsule Netw* 5(3):246–267
12. R. Meenal, et al., "Sensitivity analysis based artificial neural network approach for global solar radiation prediction in India," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 1, pp. 31–38, 2020, doi: 10.11591/ijeecs.v20.i1.pp31-38.
13. Abyson Mathew, Dr. Juby Mathe, "Weather Forecasting Using the Random Forest Algorithm Analysis" *Proceedings of the National Conference on Emerging Computer Applications (NCECA)-2022* 1 Vol.4, Issue.1
14. N. Singh, S. Chaturvedi and S. Aakhter, "Weather Forecasting Using the Machine-Learning-Algorithm," 2019 International-Conference on the Signal-Processing and Communication in the (ICSC), 2019, 171–174, doi: 10.1109/ICSC45622.2019.8938211.
15. NATIONAL SMART GRID MISSION IMPLEMENTATION FRAMEWORK, National Smart Grid Mission Ministry of Power Government of India
16. Press Information Bureau, Government of India, Ministry of New and Renewable Energy, Report on 20 DEC 2022 5:17PM by PIB Delhi, Year- End Review 2022- Ministry of New and Renewable Energy
17. Indian Renewable Energy Sector: 2022 Year End Review, Energy Portal
18. Neha Sanewal, Vandana Khanna, "Solar Power Prediction in North India Using Different Regression Models, 2023 IEEE World Conference on Applied Intelligence and Computing (AIC), DOI: [10.1109/AIC57670.2023](https://doi.org/10.1109/AIC57670.2023), 29–30 July 2023.
19. T. Rajasundarapandiyaneebanon et al., "Solar Energy Forecasting Using Machine Learning and Deep Learning Techniques" article on Archives of Computational Methods in Engineering, 24 February 2023. Springer, Volume 30, pages 3059–3079, (2023).
20. Adedeji PA, Akinlabi SA, Madushele N, Olatunji OO (2020) Neuro-fuzzy resource forecast in site suitability assessment for wind and solar energy: a mini review. *J Clean Prod* 269:122104. <https://doi.org/10.1016/j.jclepro.2020.122104>
21. Singh, R., Arya, A. & Neha, k. A hybrid lightweight model for real-time wheat disease identification using Unmanned Aerial Vehicles (UAV) imaging. *Multimed Tools Appl* **85**, 221 (2026). <https://doi.org/10.1007/s11042-026-21188-8>
22. C. Vennila,¹Anita Titus,²T. Sri Sudha,³U. Sreenivasulu,⁴N. Pandu Ranga Reddy,³K. Jamal,⁵Dayadi Lakshmaiah,⁶P. Jagadeesh,⁷and Assefa Belay "Forecasting Solar Energy Production Using Machine Learning" Volume 2022 | Article ID 7797488.