Journal of
Alternate Energy
Sources & Technologies

ISSN: 2230-7982 (Online)
ISSN: 2321-5186 (Print)
Volume 16, Issue 3, 2025
September–December
DOI (Journal): 10.37591/JoAEST

https://journals.stmjournals.com/joaest

Research JoAEST

AI/ML-Based Approach to Solar Irradiance Prediction and Energy Suitability

Kavya Srivastava[1], Krati Garg[1], Palak Agarwal[1,*], Pallavi Gupta[1], Ravi Kumar[1]

Abstract

In this paper, due to challenges in precisely predicting solar irradiance, which is essential for solar power system optimization, we employed six diverse machine learning (ML) techniques: Linear Regression, Decision Tree, Random Forest, Gradient Boosting methods (including XGBoost), and Neural Networks—to analyze and predict outcomes using a dataset containing meteorological and temporal features. Key variables include wind speed, humidity, and temperature, which significantly influence the model's predictive capability. Each method is evaluated for its accuracy, interpretability, and computational efficiency. Linear Regression provides a baseline, Decision Trees offer interpretability, Random Forests enhance robustness, Gradient Boosting improves precision through sequential learning, and Neural Networks capture complex nonlinear patterns. This comparative approach enables selection of the most effective model for the given prediction task. Ensemble models, especially Random Forest and Gradient Boosting, are excellent in capturing intricate data correlations, according to performance measures like RMSE, MAE, and MAPE. This study also highlights the role of feature selection in enhancing the effectiveness of the model. The results in this study show the revolutionary potential of ML in improving solar energy systems and highlight the importance of temporal aspects in improving forecast accuracy. This research focuses on enhancing the reliability of solar irradiance predictions, which in turn support the wider use of solar energy and advancing the global shift toward renewable energy solutions.

Keywords: Solar irradiance prediction, solar energy machine learning, renewable energy, energy forecasting, ensemble learning, random forest, gradient boosting

INTRODUCTION

Because of its growing cost-effectiveness and potential to lower greenhouse gas emissions, solar energy has become one of the most important contributions to the renewable energy mix. Developing techniques for forecasting solar irradiance is a crucial research area for ensuring reliable operations of power systems that uses renewable energy sources [1]. To reach the net-zero energy notion, solar energy is being using in various sectors. As a result, solar irradiance forecasting has become increasingly important for both solar PV systems and passive solar architecture buildings [2]. To evaluate capacity of a region for solar energy utilization, it is necessary to have the data about solar radiation (SR) information of that region [3]. However, constant and dependable power generation is greatly hindered by the erratic and unpredictable nature of solar radiation, which is the primary source of solar energy production. This unpredictability emphasizes the significance of accurate forecasting and optimization methods to optimize the use of solar energy. Solar irradiation

***Author for Correspondence**
Palak Agarwal
E-mail: palakagarwal7094@gmail.com

[1–5]Student, Department of Computer Science Engineering (Data-Science), ABES Engineering College, Ghaziabad, Uttar Pradesh, India

Received Date: July 05, 2025
Accepted Date: August 12, 2025
Published Date: September 19, 2025

Citation: Kavya Srivastava, Krati Garg, Palak Agarwal, Pallavi Gupta, Ravi Kumar. AI/ML-Based Approach to Solar Irradiance Prediction and Energy Suitability. Journal of Alternate Energy Sources & Technologies. 2025; 16(3): 36–48p.

and local meteorological factors, which includes temperature, humidity, and atmospheric pressure, having a significant effect on the energy output of the photovoltaic panels. Solar PV power output's unpredictability and volatility has increased the stability, and quality of system, among other issues [4]. Researchers have developed various kinds of methods for navigating these obstacles, such as Maximum Power Point Tracking (MPPT) algorithms, which optimize PV systems and its output of energy in a range of irradiance situations [5].

Conventional techniques, like Empirical and Numerical Weather Prediction (NWP) models, use meteorological equipment and historical weather data to predict irradiance where accuracy is low [6]. These primary shortcomings of these methods are that they struggle to handle unpredictable data that changes over time and has complex patterns and also fail to generate meaningful input features for prediction when using regression-based models [7]. However, machine learning models—such as Ensemble Learning, XG Boost, Artificial Neural Networks (ANNs) and Decision Tree have shown significant potential in surpassing these constraints. Because of their effectiveness and resilience when dealing with non-linear data, ensemble learning methods like Random Forest (RF), Gradient Boosting Machines (GBMs), and XGBoost have become prominent. Neural networks, especially deep learning architectures, have been utilized along with conventional machine learning approaches to further improve prediction accuracy by capturing intricate temporal and spatial variations in solar irradiance statistics [8]. For machine learning models to function as efficiently as possible, hyperparameter adjustment is essential. To find the optimal configuration, methods like RandomizedSearchCV methodically examine different hyperparameter combinations. Compared to previous combined solar forecasting methods, Deep Learning performs better [9].

Solar forecasting primarily relies on clear-sky persistence method [10]. However, some recent research supports using an optimal blend of climatology and persistence as reference [11]. Employing the most precise naïve method effectively prevents the inflation of prediction accuracy, although makes it harder to compare accuracy metrics. Although performance metrics are regarded as the most effective metrics for evaluating forecaster proficiency, they cannot consistently compare various methods, as there is no assurance that the same method will yield identical skill ratings across different datasets [12]. Therefore, it can be nearly impossible to ascertain the superior accuracy of two distinct methods tested across multiple places, intervals and time based resolutions without replicating and validating on the identical dataset [13].

Statistical approach encompasses all data-driven techniques, from the most recent machine learning models to traditional statistical methods [13]. The models used are straightforward to use in many real-world applications since they are trained using historical data and don't require any knowledge of the PV plant's design specifications. Consequently, models of machine learning are frequently employed for PV power forecasting, according to Antonanzas et al. [14]. In fact, new modern deep NN algorithms can achieve good accuracy even on cloudy days [15].

This study's comparative assessment of machine learning models highlights both their unique advantages and hybridization possibilities. The creation of more accurate forecasting systems may be possible if the pattern identification powers of neural networks are combined with the prediction accuracy of tree-based ensemble models. These hybrid strategies have the potential to completely transform the solar energy industry when accompanied by real-time data assimilation. We have also utilized a strong framework for predicting solar irradiance via the integration of cutting-edge machine learning models, feature selection techniques, and hyperparameter optimization. It offers a thorough evaluation of the ability of different algorithms by comparing them and utilizing historical meteorological data. It is anticipated that outcomes would support further initiatives to improve solar energy systems' dependability and efficiency, encouraging their wide acceptance.

LITERATURE REVIEW

Solar radiation estimation plays a critical role in environmental and agricultural applications. Traditional models, such as the Angstrom-Prescott model, rely on sunshine duration to estimate solar radiation. However, these models often struggle to represent complex nonlinear characteristics in meteorological data. To improve accuracy, advanced machine learning techniques such as Artificial Neural Networks (ANNs) and Support Vector Regression (SVR) have been explored. Studies by Blaga et al. [1] have demonstrated that SVR with polynomial and radial basis function (RBF) kernels provides more precise estimations by better handling nonlinear dependencies.

Machine learning and deep learning models are being used widely for solar forecasting because they can effectively handle large amount of data and capture temporal dependencies. Researchers [3] used data from Errachidia, Morocco, to demonstrate that ANN and Random Forest (RF) models significantly outperform traditional models like Linear Regression (LR) and SVR in short-term solar energy prediction. Their findings were validated using data from Pirapora, Brazil, confirming model reliability across different climate conditions. Similarly, Wang et al. [7] proposed a Deep Recurrent Neural Network (DRNN) approach for solar irradiance prediction, showing superior accuracy compared to conventional methods like Support Vector Machines (SVM) as well as Feedforward Neural Networks (FNN).

To further enhance the stability and accuracy of solar forecasting, researchers have proposed hybrid models. In a study by Ahmad et al. [6], a hybrid deep learning model incorporating Ensemble Empirical Mode Decomposition (EEMD) for data decomposition and an optimized Long Short-Term Memory (LSTM) network fine-tuned via Bayesian optimization, was introduced. Support Vector Regression (SVR) was used to minimize fluctuation errors, improving prediction stability and accuracy by 15%. Additionally, Alzahrani et al. [8] introduced an adaptive forecasting approach based on Portfolio Theory (PT), integrating deep learning models to offset prediction errors. Results from Spain and Brazil demonstrated that DL and the PrevPT model achieved lower Mean Absolute Percentage Error (MAPE) than conventional forecasting techniques.

Artificial intelligence (AI) plays an integral role in optimizing solar energy management and grid operations. The study by Wang et al. [5] explored AI applications in solar and hydrogen-based energy production, balancing supply with demand along with needs of forecasting maintenances. Findings indicated that AI-driven models excel in handling big data, optimizing decision-making processes, and improving cybersecurity in power systems. AI has also been shown to outperform traditional models in smart grid management, resource allocation, and energy efficiency optimization.

According to study by Sharma and Saini [13], a total of 24 machine learning techniques were evaluated for next-day power prediction using NWP data collected from 16 photovoltaic sites in Hungary. Kernel Ridge Regression and Multilayer Perceptron (MLP) were found to be effective, by achieving a forecasting accuracy improvement of 44.6%. The emphasis was placed on choice of predictor and hyperparameter tuning. Another large-scale evaluation by Mellit et al. [15] assessed 68 machine learning algorithms across various climatic regions across the U.S. Tree based models performed consistently well in long-term forecasting, though they lacked daily forecast stability. The study emphasized that no universal model exists, but certain models perform better under specific sky and climate conditions.

Objective

The objective of our study is to develop an accurate as well as efficient model to predict solar irradiance using meteorological parameters. The specific objectives of this study are:

1. Creating forecasting model that uses important climatic parameters like radiation, pressure, humidity, wind direction, and speed to determine solar irradiance.
2. To evaluate and compare different machine learning approaches to find most reliable and precise techniques for solar irradiance prediction.
3. To analyse the influence of various environmental parameters on solar irradiance and assess their significance in predicting solar energy potential.

4. To design and implement a user-friendly interface where users can input meteorological variables and obtain solar irradiation predictions.
5. To assess the suitability of different geographic locations for solar energy generation based on the predicted solar irradiance values.

Dataset Summary

Hourly observations performed for the purpose of predicting sun irradiance make up the dataset used in this research. Among its characteristics are UNIX time, Date, Time, Temperature (°C), Pressure (hPa), Humidity (%), Wind Direction (°), Wind Speed (m/s), Time Sunset (time of sunset in HH: MM: SS), and Radiation (solar irradiance in W/m²). The dataset is composed of 3268 occurrences that were recorded at different times and places. This dataset serves as a comprehensive resource for forecasting solar irradiance using techniques of machine learning (Figure 1).

Methodology

Solar irradiance forecasting methods are generally divided into three main types:

- Approaches based on numerical weather prediction (NWP),
- Techniques using image-data,
- And statistical or machine learning-based models [16].

The procedures used to estimate solar irradiance using machine learning models (3), which include Decision Trees, Random Forest, Neural Networks, XGBoost, and Gradient Boosting (GB), are provided in the techniques used in this study (Figure 2). This method includes thorough model evaluation, hyperparameter adjustment, intense training, and comprehensive data preprocessing. Every element of the process was created to ensure clarity of models, dependability and correctness. The process of data collection and it's preprocessing, model training, hyperparameter optimization, evaluation and result visualization were the various stages of this process. Thus, to improve the model performance, the procedure also assessed ensemble modes including stacking [17].

UNIXTime	Date	Time	Radiation	Temperature	Pressure	Humidity	Wind Direction (Degrees)	Speed	Time SunRise	Time SunSet	Month	Day	Hour	Minute
1475229326	9/29/2016	23:55:26	1.21	48	30.46	59	177.39	5.62	06:13:00	18:13:00	9	29	23	55
1475229023	9/29/2016	23:50:23	1.21	48	30.46	58	176.78	3.37	06:13:00	18:13:00	9	29	23	50
1475228726	9/29/2016	23:45:26	1.23	48	30.46	57	158.75	3.37	06:13:00	18:13:00	9	29	23	45
1475228421	9/29/2016	23:40:21	1.21	48	30.46	60	137.71	3.37	06:13:00	18:13:00	9	29	23	40
1475228124	9/29/2016	23:35:24	1.17	48	30.46	62	104.95	5.62	06:13:00	18:13:00	9	29	23	35

Figure 1. Overview of dataset.

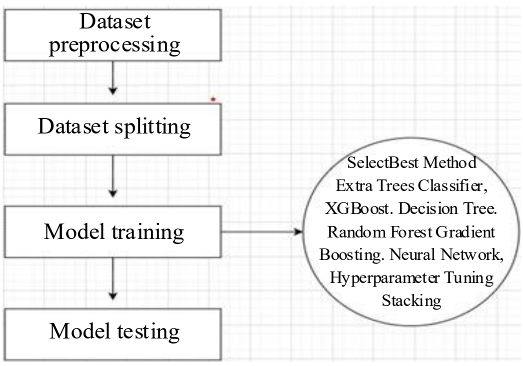


Figure 2. Methodology.

Essential parameters like temperature, daylight duration, relative humidity and solar radiation were among the data collected for this study. Information on inactive solar radiation that was constantly recorded as zero was not incorporated in the statistical evaluation because it did not add to the prediction job and potentially have influenced the model's learning process [18]. The procedure is explained in depth in the components that precede.

Dataset Preprocessing

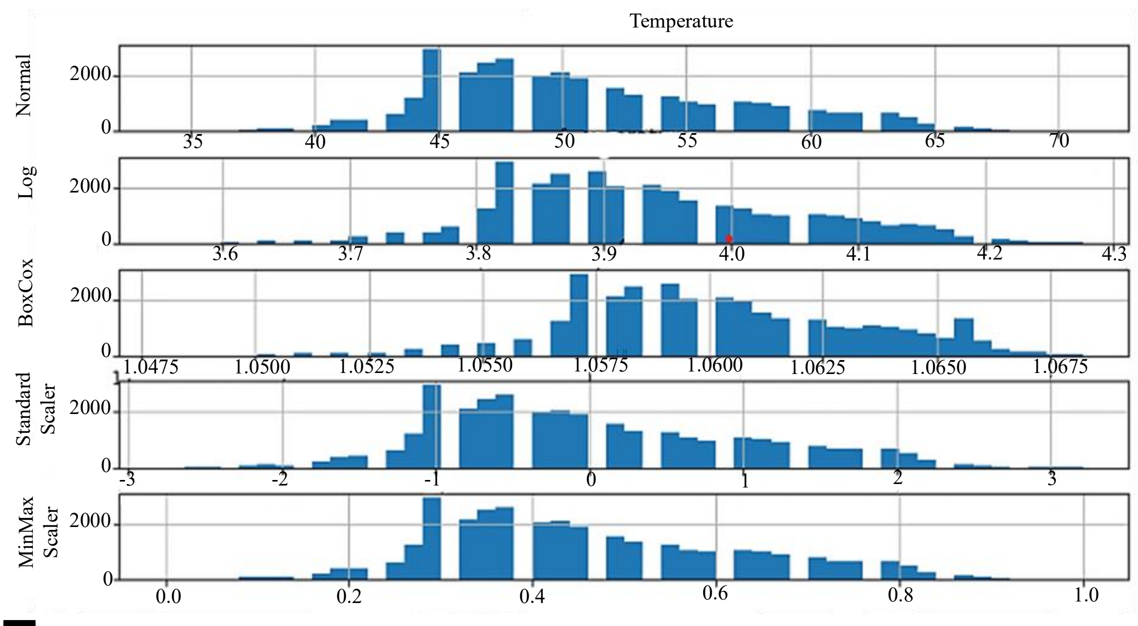
The initial stage was to eliminate items with missing or incorrect values from the dataset, which featured hourly records of the main variables. Statistical estimation approaches, such as incorporating the mean or median values for continuous variables, were used to manage missing values. To enhance data distribution and improve model performance, various transformations were applied to selected features, including Temperature, Pressure, Humidity, Wind Speed, and Wind Direction (Figure 3a,3b,3c). Log transformation was used to reduce skewness by compressing large values and expanding smaller ones, while Box-Cox transformation helped stabilize variance and make the data more normally distributed [19]. Standard scaling was implemented to convert the data to zero mean and unit variance, ensuring compatibility with algorithms sensitive to scale. Additionally, Min-Max scaling was applied to normalize values between 0 and 1, preserving the original distribution while ensuring all features contributed equally to the model. By visualizing these transformations, the most appropriate method for each feature was selected, ultimately improving model accuracy and convergence.

Dataset Splitting

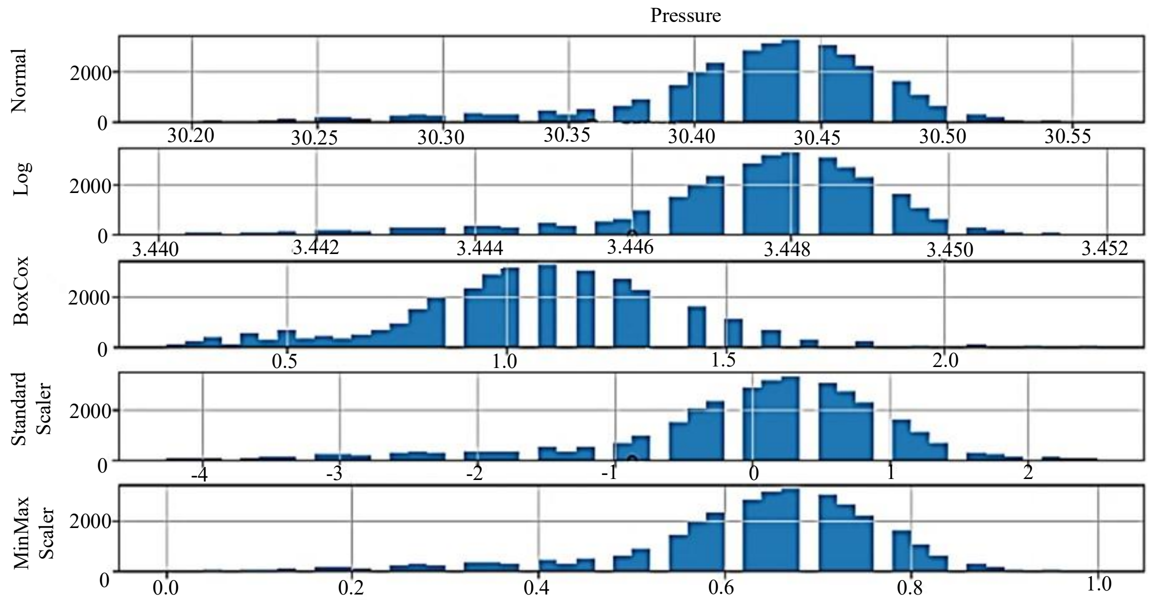
The dataset is divided into two parts: one for testing (20% of data) and other for training (80% of data), with the goal of evaluating and building the model. By ensuring that the models were evaluated on unknown information, this division made it feasible to accurately evaluate how well the models' predicted outcomes. Despite being randomized, the splitting procedure made sure that all of the dataset's temporal patterns were preserved.

Model Training

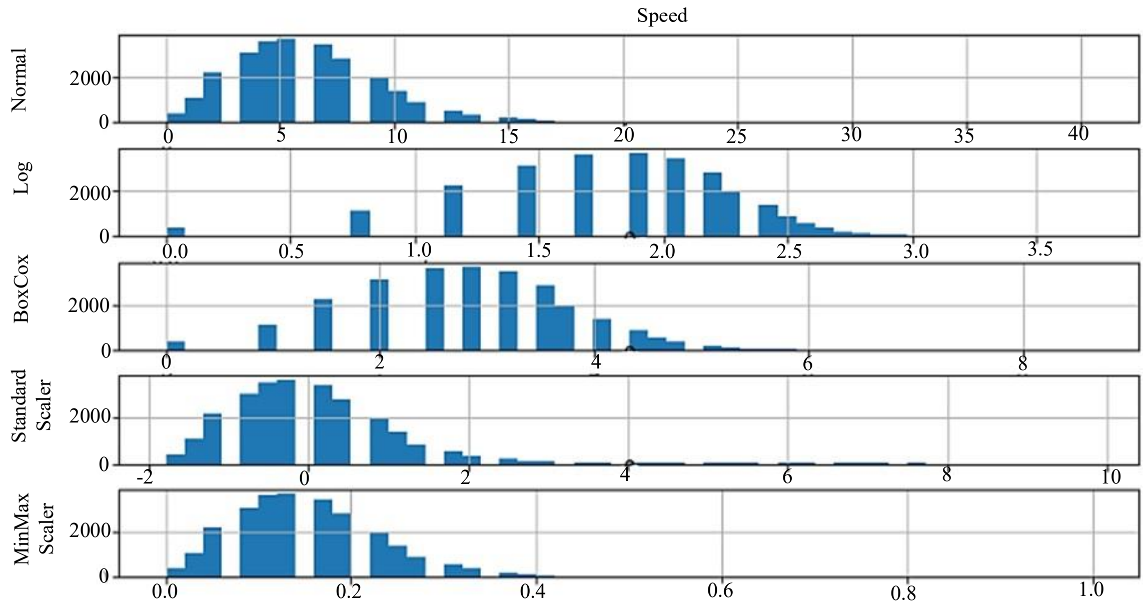
The models used in this research were selected based on their demonstrated effectiveness in managing structured data and their capacity to identify nonlinear correlations. XGBoost, Decision Tree, Neural Network, Random Forest, SelectKBest method, Extra Tree Classifier and Gradient Boosting represent some of the algorithms [20]. The preprocessed dataset was used to train each model, and its performance was optimized.



(a)



(b)



(c)

Figure 3. (a) Various transformations applied to temperature; (b) Various transformations applied to pressure; (c) Various transformations applied to wind speed.

SelectKBest Method

A widely and popular approach for selecting features in machine learning is SelectKBest method uses statistical tests to assess the significance of features in a dataset (Figure 4). It selects by identifying top K features with the highest scores after evaluating every feature according to a grading function. By eliminating redundant or unnecessary properties, this approach guarantees that the most pertinent ones for forecasting the target variable are kept. F-statistics for regression issues, mutual information for identifying non-linear dependencies, and chi-square tests for categorical data are typical scoring functions used with SelectKBest. This approach lowers the dataset's complexity, boosts computing effectiveness, and lessens the chance of overfitting by concentrating on the most important attributes. We find that setHour (hour) has the maximum importance.

Extra Trees Classifier

It is a type of ensemble-based algorithm that improves the performance of standard decision tree models (Figure 5). Unlike Random Forest, which determines the optimal split points based on criteria like information gain or impurity, Extra Trees generates splits by randomly selecting feature thresholds. This added randomness helps reduce variance, making Extra Trees particularly robust against overfitting, even when working with noisy datasets. This approach generates reliable and accurate classification results by averaging predictions from several independent trees. It is frequently used for feature selection and prediction tasks in a variety of specific fields due to its computational efficiency as well as ability to handle high-dimensional datasets.

Using Extra Trees Classifier, we understand that wind direction came out to be most important followed by minute, speed, second, humidity, hour, temperature and so on. The least important feature turned out to be RiseHour which is definitely not that much important while predicting solar irradiance. Figure 4 is the illustration for the same.

XGBoost

A robust ensemble learning technique built on the principles of gradient boosting is called XGBoost, short for Extreme Gradient Boosting. Gradient descent optimization is employed to minimize mistakes at each stage of the progressive construction of decision trees. Through hyperparameter tuning, XGBoost was set up for this study with parameters including learning rate, max dimensions, and n estimators. It was an appropriate fit for this problem because of its capacity to address missing variables and avoid overfitting through regularization (L1 and L2 penalties).

Decision Tree

The Decision Tree Regressor (DTR) had been employed as one of the predictive models for solar irradiance estimation. Being a non-parametric approach, it divides the data repeatedly based on inputs to reduce prediction variance. The model development involved training on a designated training set (X_train, y_train) and validating outcomes using test set (X_test, y_test). To test the performance, key evaluation metrics were computed, including Root Mean Squared Error (RMSE), R-squared (R^2), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE**)**. These indicators offer insights for accuracy of model, where lower RMSE and MAE values indicate better predictive performance, a higher R^2 value reflects a better alignment between the actual and predicted values.

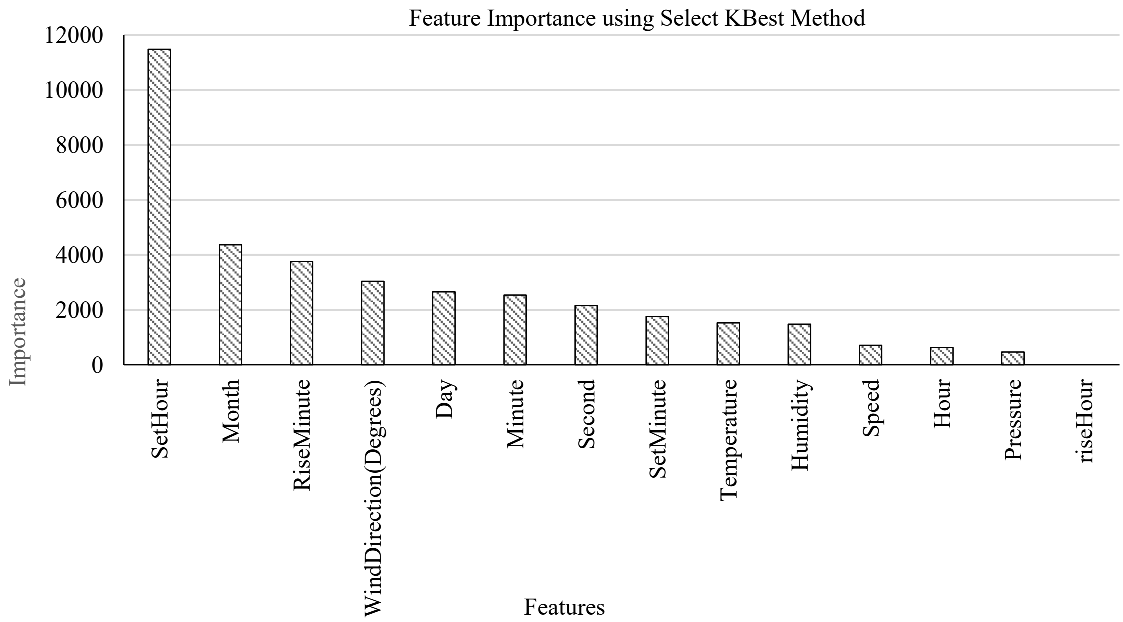


Figure 4. Feature importance using SelectKBest method.

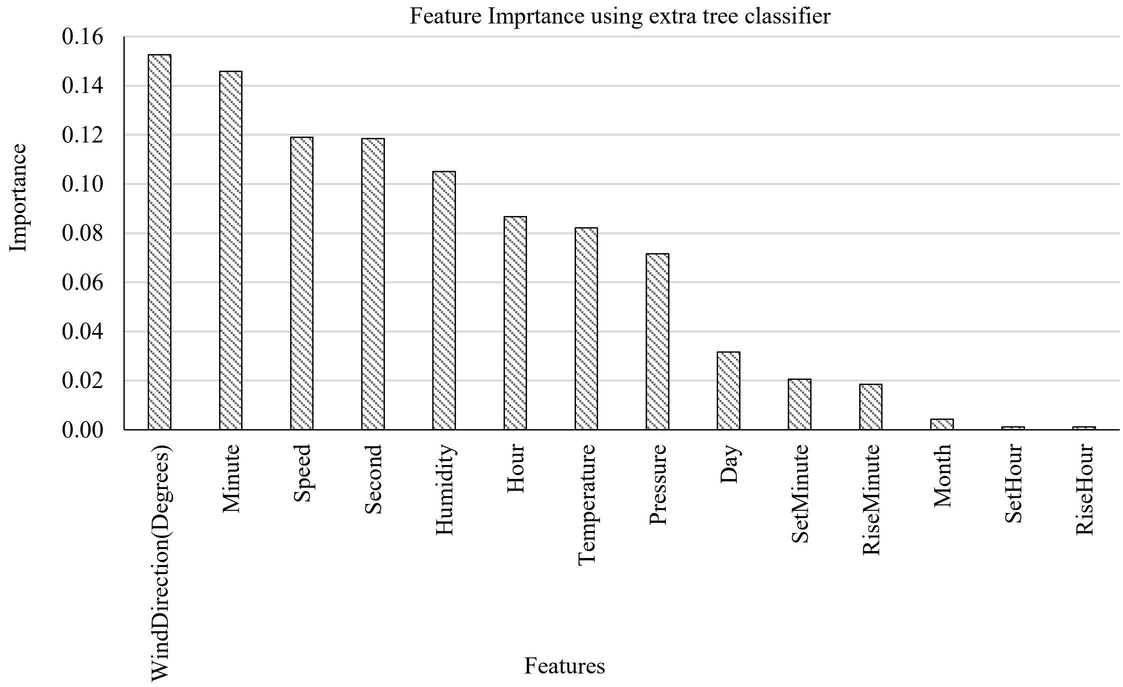


Figure 5. Feature importance using Extra Tree Classifier method.

Likewise, a lower MAPE suggests reduced percentage error. The Decision Tree Regressor's performance was further analyzed in comparison with other models to determine its suitability for solar irradiance prediction.

Random Forest

A subset of dataset is chosen randomly to train each individual tree and results are combined to create the final output (Figure 6). To improve performance, hyperparameters like maximum depth, maximum features, and number of trees (n_estimators) are optimized. This model proved effective in this study due to its strong resistance to overfitting and ability to work effectively with data of high dimensions. Here also, the performance is further analyzed in comparison with other models to determine its suitability for solar irradiance prediction.

Gradient Boosting

It is also an ensemble learning model that optimizes remaining errors, which builds trees in an orderly manner. Although the implementation details are different, it optimizes via gradient descent, just like XGBoost. To regulate the step size and data subset used at each iteration, Gradient Boosting was set up in this study leveraging a learning rate and subsample parameter. Again, the performance was analyzed in comparison with other models to determine the suitability for solar irradiance prediction.

Neural Network

To extract intricate, nonlinear associations from the data, Artificial Neural Networks (ANNs) were utilized. An output layer that predicted sun irradiance, many hidden layers with ReLU activation, and an input layer that represented the meteorological variables made up the ANN architecture. Backpropagation was used to train the network. Again, the performance was analyzed in comparison with other models to determine the suitability for solar irradiance prediction.

Hyperparameter Tuning

RandomizedSearchCV was used to do hyperparameter updating in order to determine the optimal parameter values for every model. Parameters like min_samples_split, max_depth, and n_estimators were adjusted for Random Forest.

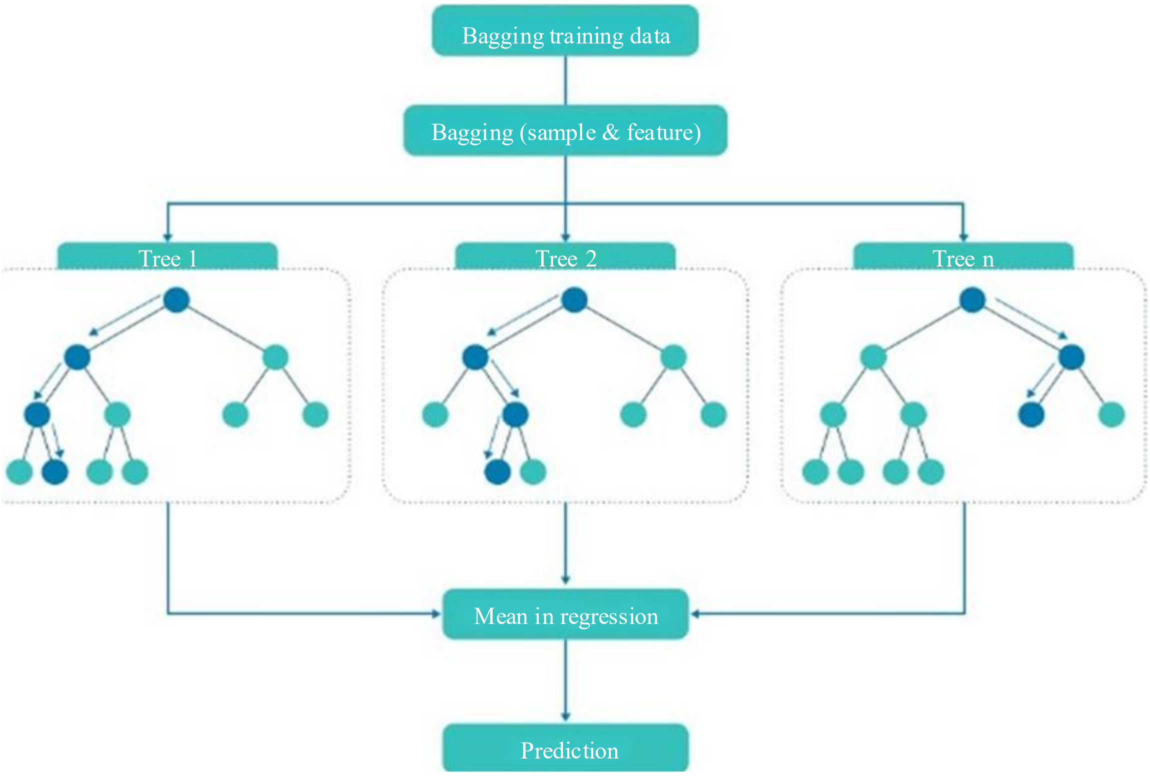


Figure 6. Random Forest regression tree.

Performance and efficiency of computation were optimized by the optimization of XGBoost's learning rate, max_depth, and colsample_bytree. To minimize validation error, adjustments were done to various hyperparameters including learning speed, depth of hidden architecture, batch size and number of processing units in each layer.

Stacking

To combine the predictions of multiple approaches and take advantage of their beneficial strengths, stacking was used. To generate final predictions, the outputs of the initial models were used to train a meta-model, usually a linear regressor, but since the accuracy of random forest came out to be best, we used random forest regressor. Through the reduction of distinct model biases, this method improved accuracy and durability. Additionally researched were hybrid models that combined neural networks along with traditional machine learning techniques. These models captured patterns that were complex in data by combining the flexibility of deep learning with the ability to forecast of machine learning.

Performance Evaluation Parameters

Statistical measures such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-square (R^2) have been implemented to assess the models. It is important to mention that various methods are applied to various forecasting horizons and types of statistical evaluations (SE) [9].

1. *MSE:* **-** It calculates average of squared difference between actual and predicted values.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(yi - \hat{y}i)^2$$

2. *RMSE:* A more accessible metric is offered by RMSE, which corresponds to the square root of MSE.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(yi - \hat{y}i)^2}$$

3. *MAPE:* **-** Prediction error is determined by MAPE as a percentage of observed values.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100$$

4. *R-Squared:* The ability of the model to comprehend the variation in the target variable is indicated by R-squared (R^2).

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

RESULT

In this work and study, we looked how several machine learning models accurately anticipate solar irradiance using a range of weather and time-related features. The evaluated models include Linear Regression, Decision Tree Regressor, Random Forest Regressor, Gradient Boosting Regressor, XGBoost Regressor and Neural Network model incorporated as a scikit-learn compatible estimator for seamless integration into ensemble methods (Figure 7). Each model was rigidly trained and tested, with the accuracy performance evaluated using metrics such as Root Mean Square Error, Mean Absolute Percentage Error and Mean Absolute Error.

The outcomes demonstrate how well various regression models estimate sun irradiance. Given its underlying presumptions regarding linear relationships, it was not surprising that Linear Regression had the lowest accuracy among the tested models. On the other hand, with accuracy ratings of 82.76% and 72%, respectively, Random Forest Regressor and XGBoost Regressor showed excellent prediction ability. After being included into a framework compatible with scikit-learn, the neural network model demonstrated how well deep learning works in finding complex relationships in dataset with an accuracy of 73.98%. Figure 7 demonstrates comparison between different algorithms using bar graph.

Additionally, Figure 8 illustrates that we get the best accuracy with Random Forest of 82.75%. (Output from code).

Hyperparameter tuning using cross-validation and grid search was applied to Random Forest and XGBoost to enhance prediction accuracy. This approach optimized key settings such as the number of estimators, learning rate and tree depth.

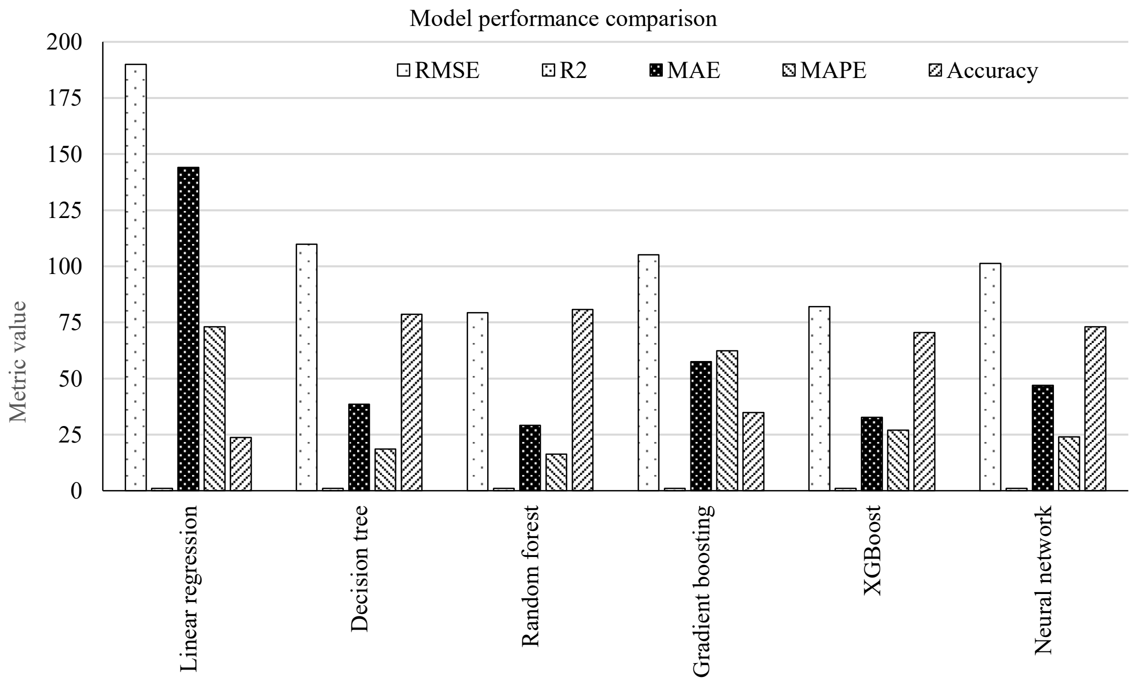


Figure 7. Comparison between different algorithms.

Journal of Alternate Energy Sources & Technologies
Volume 16, Issue 3
ISSN: 2230-7982 (Online), ISSN: 2321-5186 (Print)

Model	RMSE	R²	MAE	MAPE	Accuracy
Linear Regression	193.622836	0.622637	146.566929	75.000000	25.000000
Decision Tree	111.942432	0.873865	39.689686	19.240690	80.759310
Random Forest	81.157963	0.933701	30.250894	17.240112	82.759888
Gradient Boosting	107.508684	0.883659	59.055841	64.000000	36.000000
XGBoost	83.596359	0.929657	33.913749	28.000000	72.000000
Neural Network	103.603109	0.891958	48.317587	25.015898	74.984102

Figure 8. Performance table of different algorithms.

Additionally, model ensembling was implemented using the Stacking Regressor framework for leveraging the strengths of multiple models (Figure 9). Given its superior performance in initial testing, Random Forest Regressor was chosen as the meta-model. This ensemble approach significantly improved the final accuracy to 83.87%, with reduced Mean Absolute Percentage Error (MAPE) of 16.07%. By integrating various models, this method effectively mitigated the limitations of individual models, resulting in a more balanced and robust predictive framework for solar irradiance estimation.

Additionally, the Stacking Regressor architecture was used to build model ensembling, which capitalizes on strengths of various models. The Random Forest Regressor was chosen as the meta-model due to its better performance in the preliminary tests. With a lower Mean Absolute Percentage Error (MAPE) of 16.07%, this ensemble technique greatly increased the ultimate accuracy to 83.87%. This approach successfully reduced the drawbacks of individual models by combining many models, producing a more robust and balanced prediction framework for estimating solar irradiance. Moreover Figure 10, demonstrates the comparative studies between different algorithms vs final stacking model.

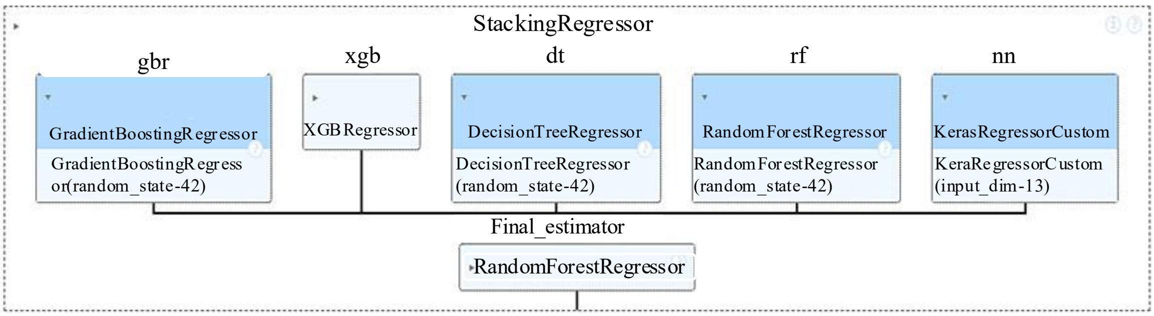


Figure 9. Training results of different algorithms with stacking regressor.

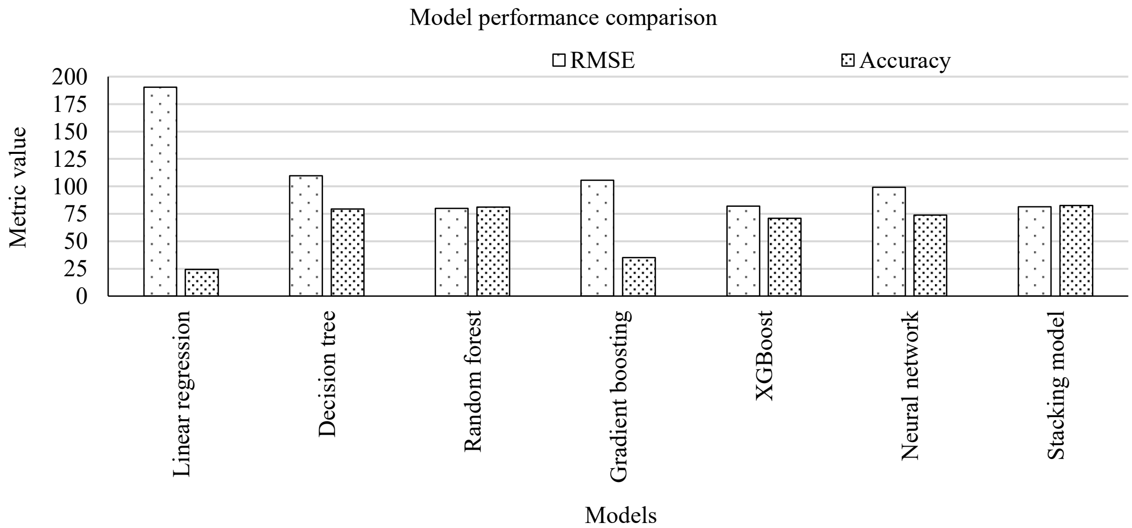


Figure 10. Comparison between different algorithms vs final stacking model.

CONCLUSION

This research aimed to develop an accurate and reliable predictive framework for solar irradiance estimation by comparing multiple machine learning models. Various regression techniques, like *Linear Regression, Decision Tree Regressor, Random Forest Regressor, XGBoost Regressor, and Neural Networks,* were evaluated using key performance metrics such as *RMSE, R^2, MAE and MAPE*. The results indicated that *Random Forest Regressor* outperformed other models by an accuracy of *82.76%,* while *Neural Network* effectively captured non-linear relationships, achieving *73.98% accuracy.*

To further optimize predictions, *hyperparameter tuning using cross-validation and grid search* was performed on selected models, leading to enhanced accuracy. Additionally, *model ensembling using a Stacking Regressor* was implemented, integrating the strengths of multiple models. The ensemble approach, with *Random Forest as the meta-model,* significantly improved the final accuracy to *83.87%,* with a low MAPE of *16.07%,* demonstrating its effectiveness in reducing individual model biases. Multi-Step forecasting of photovoltaic power is still a challenge.

Overall, our study highlights the potential of ensemble learning in solar irradiance prediction, offering a more balanced and robust approach compared to standalone models. Future research could explore *deep learning* and *spatiotemporal analysis* to further enhance predictive accuracy along with adaptability to different climatic conditions. Additionally, several studies have investigated the utilization of solar energy to generate steam used in thermal enhanced oil recovery (thermal EOR) to extract heavy crude oil.

The result of this study provides key takeaways for optimizing solar energy utilization, supporting the advancement of sustainable energy solutions.

REFRENCES

1. Blaga R, Sabadus A, Stefu N, Verscea M, Dughir C, Paulescu M, et al. A current perspective on the accuracy of incoming solar energy forecasting. Prog Energy Combust Sci. 2019;70:119–44.
2. Chakkaravarthy AN, Subathra MSP, Pradeep PJ, Kumar NM. Solar irradiance forecasting and energy optimization for achieving nearly net zero energy building. J Renew Sustain Energy. 2018;10(3).
3. Piri J, Shamshirband S, Chau KW, Tong CW, Malvoni M, Hossain U, et al. Prediction of the solar radiation on the Earth using support vector regression technique. Infrared Phys Technol. 2015;73:179–85.
4. Jebli I, Belouadha F-Z, Kabbaj MI, Tilioua A. Prediction of solar energy guided by Pearson correlation using machine learning. Energy. 2021;224:120109.
5. Wang F, Xuan Z, Zhen Z, Yu Y, Li K, Liu K, et al. A minutely solar irradiance forecasting method based on real-time sky image-irradiance mapping model. Energy Convers Manag. 2020;220:113070.
6. Ahmad T, Zhang D, Cai H, Hui H, Zhang D, Song C, et al. Artificial intelligence in sustainable energy industry: Status quo, challenges and opportunities. J Clean Prod. 2021;289:125449.
7. Wang L, Meng M, Jia X, Zhang Z, Li H, Zhang H, et al. Accurate solar PV power prediction interval method based on frequency-domain decomposition and LSTM model. Energy. 2023;262:125444.
8. Alzahrani A, Shamsi P, Dagli C, Ferdowsi M. Solar irradiance forecasting using deep neural networks. Procedia Comput Sci. 2017;114:304–13.
9. Lima MAFB, Pereira CR, Costa MGF, Fernandes R, Bezerra B, Portela B, et al. Improving solar forecasting using deep learning and portfolio theory integration. Energy. 2020;195:116925.
10. Marquez R, Coimbra CF. Proposed metric for evaluation of solar forecasting models. Sol Energy. 2013;102:313–8.
11. Yang D. Standard of reference in operational day-ahead deterministic solar forecasting. J Renew Sustain Energy. 2019;11:036102.

12. Mayer MJ, Gróf G. Extensive comparison of physical models for photovoltaic power forecasting. Appl Energy. 2021;283:116239.
13. Sharma R, Saini LM. Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. Renew Sustain Energy Rev. 2022;161:112280.
14. Antonanzas J, Osorio N, Escobar R, Urraca R, Martinez-de-Pison FJ, Antonanzas-Torres F. Review of photovoltaic power forecasting. Sol Energy. 2016;136:78–111.
15. Mellit A, Massi Pavan AM, Lughi V, Ayadi O. Deep learning neural networks for short-term photovoltaic power forecasting. Renew Energy. 2021;172:276–88.
16. Yagli GM, Yang D, Srinivasan D. Automatic hourly solar forecasting using machine learning models. Renew Sustain Energy Rev. 2019;105:487–98.
17. Wang J, Ouyang A, Ruan B. Potential solar energy use in the global petroleum sector. Energy. 2017;118:884–92.
18. Roig K, Ayub A, Al-Ammar R, Radwan K. Sensitivity analysis for solar-generated steam for enhanced oil recovery. Soc Pet Eng. 2018.
19. Sandler J, Freeman G, Curtis K, Kumar A. Solar-generated steam for oil recovery: Reservoir simulation, economic analysis, and life cycle assessment. Energy Procedia. 2012;29:176–85.
20. O'Donnell JH, Mueller-Lessmann M, McGregor C. Solar-generated steam for oil recovery: Process integration options, net energy fraction, and carbon market impacts. Energy Procedia. 2015;69:1390–9.

