Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, No. 6, June : 2023

HYBRID FUSION MODEL FOR HOURLY SOLAR RADIATION PREDICTION USING MACHINE LEARNING TECHNIQUES

Nishok K. R. , M.Tech [Data Science], Information Technology Department ,Kumaraguru College of Technology, Coimbatore-49.
Rajathi N. , Professor, Information Technology Department,Kumaraguru College of Technology, Coimbatore-49.
Vanitha V. , Professor, Information Technology Department,Kumaraguru College of Technology, Coimbatore-49.

Abstract
Solar energy is a crucial element in the development of solar energy systems. Residential and utility scale solar energy penetration has been growing at an exponential rate. However, the operation of the electrical grid is severely hampered by its stochastic character. To effectively balance, plan, and optimize the power grid, it is crucial to have advanced knowledge of the anticipated amount of solar energy that will be produced. Accurate prediction of hourly solar radiation is critical for effective planning and management. The reliability and efficiency of the modern smart grid relies on accurately predicting the amount of solar energy that will be generated. Despite being able to calculate the path and energy of the sun using physical principles, forecasting the growth and generation of solar energy remains an immensely challenging topic in both the realms of physical simulation and artificial intelligence. In this study, for feature optimization and hourly solar energy forecasting, we suggested a mixture of fusion model. Slime Mould Algorithm (SMA) is used to extract attributes. After that the advantages of the K-Nearest Neighbor (KNN), Random Forest (RF), LightGBM (LGBM), and Deep Neural Network (DNN) are combined the prediction of solar radiation needs to be more accurate. So, the results demonstrate the effectiveness of the model in predicting hourly solar radiation and can be used in energy planning and management

Keywords: KNN, Random Forest, Solar energy prediction, Slime Mould Algorithm, Hybrid fusion model, LGBM

1 Introduction

The rapid growth of renewable energy sources, particularly solar energy, presents both opportunities and challenges for power grid management. Solar energy output prediction and its integration into smart grid power balancing have emerged as critical areas of research and development. Accurate prediction of solar energy generation enables efficient planning and operation of power systems, while effective power balancing ensures the reliable and optimal utilization of available resources [1].
Undoubtedly, one of the most significant initiatives for environmental and social benefit is a revolution in green energy. Among the most likely alternatives to fossil fuels among all types of renewable resources is sunshine [2]. Due to its affordability, abundance, and sustainability, solar energy has become a practical substitute for conventional energy sources. Forecast information is required for the optimal planning of energy output, the stability and balancing of the electrical grid, and the trade of solar energy. In contrast to current quality standards, the approach presented in this study achieves higher forecast accuracy [3]. The main reason behind this is that numerous variables, including as the sun's position, the state of the weather, the properties of photovoltaic panels and others, have an impact on the actual solar production [4]. In the subsequent section, the methodologies, results of the comparative study, were providing valuable insights for researchers, practitioners, and policymakers working in the fields of renewable energy, smart grids, and power system management.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, No. 6, June : 2023

2 Methodology

The proposed approach for solar energy output prediction combines the use of the bioinspired SMA Algorithm, KNN, RF, LGBM and DNN. The workflow of the proposed methodology is represented in figure 1.

Dataset UsedThe Horizontal Photovoltaic Power Output Data from Kaggle website [5]is used in this study. The dataset includes hourly measurements of solar radiation and corresponding meteorological conditions including temperature, humidity, cloud cover, and wind speed.

This file contains the electricity generated by horizontal solar panels set up in 12 Northern Hemisphere locations over a 14-month period. Location, date, sampled time, month, hour, season, humidity, ambient temperature, solar panel power output, wind speed, and cloud ceiling are among the independent variables shown in each column. The table 1 shows the description of the data source.

Table 1 Description of the data source

Dataset Source - Horizontal Photovoltaic Power Output Data	
Features in the Dataset	17
Number of samples	21045

Out of 21,000 records, 80 % is used for training and 20% is used for testing.The work flow of the proposed methodology is given in figure1.

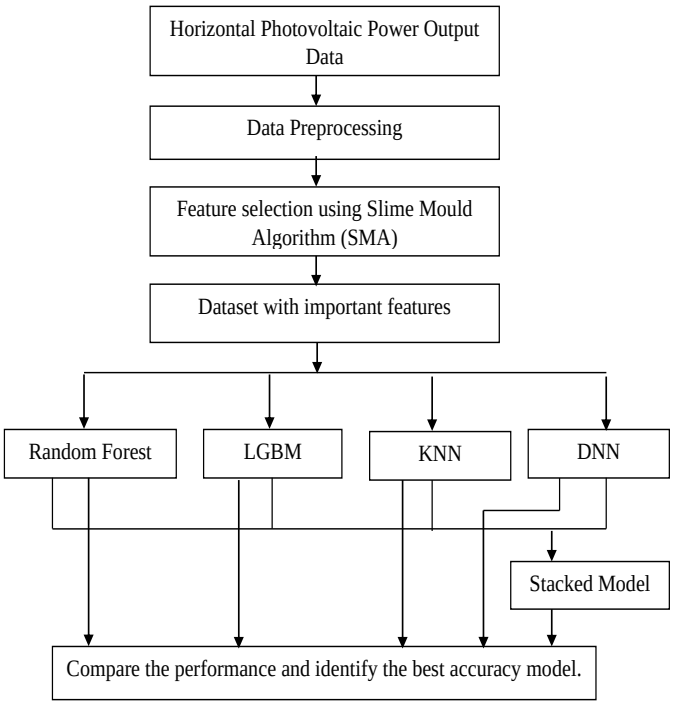

Fig 1 Workflow of the Proposed Work

Data pre-processing

Before passing the data to the machine learning algorithms, the variables that are present in the data are investigated, displayed, and pre-processed. Pre-processing is done on the data to get rid of duplicates, outliers, & missing values. Outliers are often identified and eliminated using statistical

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, No. 6, June : 2023

approaches, whereas values that are absent are usually substituted using interpolation techniques. Duplicates are identified and removed to avoid bias in the analysis [6]. The data is also normalized to ensure that all features have the same scale. This is done to prevent some features from dominating others in the machine learning models. The relationship between the power output and the features that are offered in figure 2.

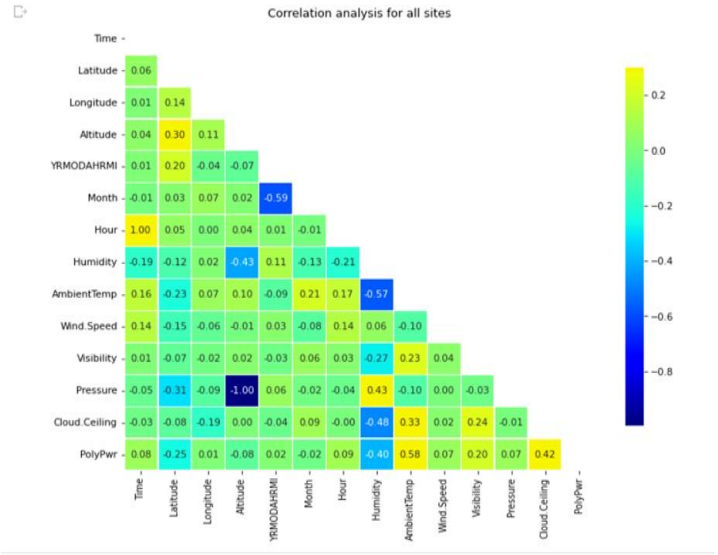

Fig 2. Correlation Analysis between the Power output and the features

Utilising the Slime Mould Algorithm for selecting features (SMA)For feature selection, the study employs the Slime Mould Algorithm (SMA), which builds a web of tubes linking food sources and slime mould cells. The algorithm identifies the most frequently used tubes, indicating the shortest path. However, in this research, the SMA is modified to select the most significant characteristics for predicting solar radiation. The modified algorithm ranks the features based on their relevance to the target variable, enabling the identification of the best features for forecasting solar energy output. This process is achieved through a wrapper feature selection technique utilizing the SMA as a meta-heuristic algorithm [7].

The following are the three stages of the suggested feature selection solution using the SMA approach:

- Initialization Phase: The initialization phase involves generating an initial set of options that represent different combinations of characteristics. This step is crucial for determining the quality and convergence of the best solution. The fitness value is determined by evaluating a random population.
- Update Phase: During the update phase, the fitness function is used to assess new positions, and if a position offers a higher-quality solution than the previous one, it is upgraded. The search possibilities are increased by using the opposite based learning (OBL) method to simultaneously investigate two alternatives. This approach increases algorithm variety, avoids suboptimal solutions, and allows for comparison between new and previous solutions [7].
- Termination stage: In the termination stage, the algorithm continues running until it satisfies the specified ending conditions and reaches the maximum allowed function evaluations. This process ultimately identifies the most optimal set of features by selecting the best solution found through the SMA process.

Model Development

Four machine learning models are used for solar radiation prediction: KNN, Random Forest, LightGBM, and DNN.

K-Nearest Neighbor Model

A popular ML approach for classification and regression applications is called the K-Nearest Neighbour (KNN) model. This model is based on the principle of similarity between instances, which assumes that similar instances are likely to have similar outputs [8]. When it comes to predicting the amount of solar energy produced at a specific time, the KNN model can be employed. This model utilizes historical solar energy output data and other pertinent factors to estimate the quantity of solar energy that can be generated.

Light Gradient Boosting Machine

LightGBM is a gradient boosting framework used for classification, regression, and ranking problems. It is a tree-based algorithm that works by constructing multiple decision trees iteratively and boosting the weak learners to improve the accuracy of the final model [9]. The hyper parameters of the LGBM model were tuned using the Bayesian optimization method to improve its performance.

Random Forest

Random Forest can be used to resolve classification and regression problems. It functions by building a number of decision trees and making accurate projections based on their combined predictions. [10]. In other words, as part of an ensemble learning strategy called Random Forest, predictions are made with greater accuracy since several decision trees are used. In order to lessen overfitting and boost the model's generalisation performance, each decision tree in the forest is trained using a random portion of the training data and a random subset of the features.

Deep Neural Network (DNN)

Popular ML method known as a Deep Neural Network (DNN) finds use in a variety of fieldslike identifying images, understanding speech, and processing natural language.Several layers of neurons make up this particular form of neural network, which allows it to learn complex patterns and features from the input data [11]. DNN works by taking the input data and passing it through multiple layers of neurons, each of which performs a nonlinear transformation on the input. The information generated by one level is passed on to the following level, and this cycle repeats until reaching the last level, which generates the final outcome of the model.

Hybrid Fusion Model

The Hybrid Fusion Model is a ML approach that makes use of several different model’s predictions [12]. The Hybrid Fusion Model combines the predictions of four machine learning models K-Nearest Neighbors (KNN), LightGBM, Random Forest, and DNN. The Hybrid Fusion Model works by training each of the fourML models based on training data, followed by combining their predictions using a weighted average. The weights assigned to each model are calculated using a genetic algorithm, which fine-tunes the weights to lessen the discrepancy between the production that was expected and what was actually produced. The genetic algorithm searches for the optimal combination of weights that produces the best overall performance.

The development of the HFM involved a careful selection of machine learning algorithms, hyper parameter tuning, and model selection to optimize the accuracy of the solar radiation predictions [12]. The use of stacking allowed for the combination of the strengths of multiple algorithms to further improve the accuracy of the predictions. The final HFM was evaluated on a separate test set to ensure its generalizability and suitability for real-world applications.

Model Training

Model training involves using a machine learning algorithm to learn patterns in a dataset. Divide the dataset into a training set and a validation set as part of the procedure. The validation set assesses the model's performance on fresh data while the training set instructs the model. This division is crucial to prevent over fitting [13]. After dividing the data, an appropriate machine learning algorithm is selected, and its hyper parameters are set. Hyper parameters affect the algorithm's behaviour and are determined through trial and error or automated techniques. The algorithm is then trained using the training set, with parameters iteratively updated to reduce the discrepancy between expected and actual results. Regular evaluation on the validation set ensures progress and prevents over fitting. Finally, the trained model is assessed using a separate test set to verify its effectiveness on new data. In this research, the Hybrid Fusion Model utilized machine learning models trained on a specific dataset [14]. Their hyper parameters were optimized through grid search, and the best-performing models were selected to create the ultimate Hybrid Fusion Model.

Model Evaluation

Model evaluation involves analysing the effectiveness of a machine learning model on a given dataset. The purpose of this assessment is to gauge the model's ability to accurately predict outcomes when presented with new data [15]. It also aims to pinpoint any shortcomings in the model and suggest potential enhancements. The following measures can be employed to assess how well a machine learning model performs.

• Mean squared error (MSE): the squared difference between actual values and anticipated values, on average.

3 Results and Discussions

The study revealed that the Hybrid Fusion Model (HFM) outperformed individual ML models in forecasting hourly solar radiation. The HFM was created by combining predictions from KNN, LGBM, Random Forest, and DNN models using a stacking method. Evaluation metrics indicated that the HFM had the lowest RMSE and MAE, additional to having the best R-squared on both validation and test sets. This highlighted its superior accuracy in predicting solar radiation values compared to the individual models. Furthermore, the study showed that feature selection with the Slime Mould Algorithm (SMA) enhanced the effectiveness of machine learning models, resulting in more accurate predictions of solar radiation values.

In conclusion, the study suggests that the HFM, developed through the stacking method and feature selection with SMA, offers an effective approach to predict hourly solar radiation values. R-squared error provides a prediction of the relationship between the movements of a dependent variable and the independent variables. [16].

The table2 and table 3 displays the scaled important 5 features utilized in predicting the solar power output using the LGBM and Random Forest models.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, No. 6, June : 2023

Table 2. Important features used by LGBM

Feature	Importance (%)
Humidity	100
Pressure	98
Ambient Temperature	96
Wind speed	63
Cloud ceiling	38

Table 3. Important features used by Random Forest

Feature	Importance (%)
Ambient Temperature	100
Humidity	57
Cloud ceiling	52
Pressure	28
Sine of month	27

The top 5 features for both LGBM and RF models include ambient temperature, humidity, cloud ceiling, and pressure. The performances of the models used are presented in table 4 and is graphically represented in the figure 3.

Table 4 Performance of each model used

MODEL	R^2	RMSE	MAE
DNN	0.662	4.142	2.709
LGBM	0.670	4.054	2.753
Random Forest	0.671	4.095	2.787
KNN	0.629	4.403	2.971
Stacked Model	0.681	4.024	2.670

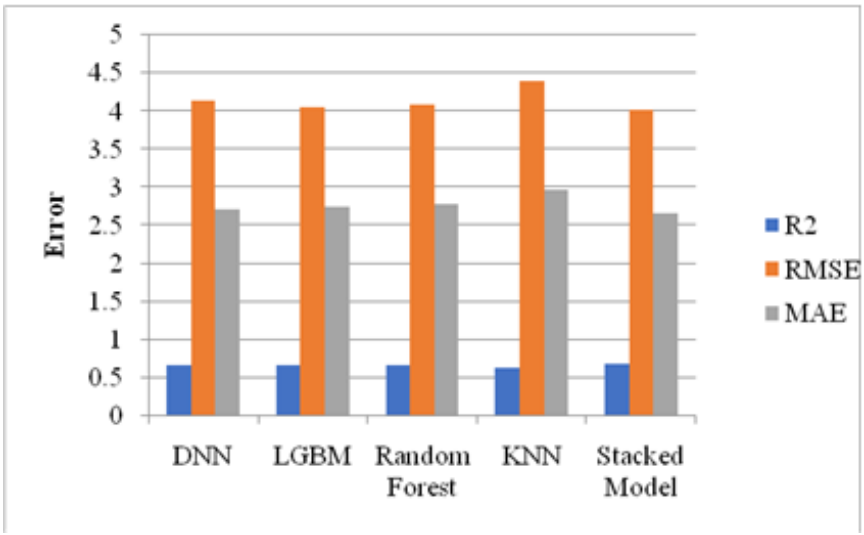

Fig.3 Performance of each model used.

From the performance of each model table, Random Forest and LGBM has more accuracy values which were calculated by coefficient of determination (R-squared). With a 10% improvement above the KNN model, the stacked model performs the best overall. The LGBM model is also the best base model when all metrics are considered.

4 Conclusion

Utilising the HFM to anticipate hourly sun radiation and optimise features, which combines the Slime Mould Algorithm with K-Nearest Neighbors, LightGBM, Random Forest, and DNN models, was developed, and evaluated in this research. The fundamental goal is to create an accurate and robust model for predicting hourly solar radiation values, which is an important factor for designing and operating solar energy systems. The study's findings demonstrated that, in terms of the assessment metrics employed, the hybrid fusion model outperformed the individual machine learning models. In particular, the stacked model outperformed the performance of the individual models, with a R-squared value 0.681 and a MAE value 2.670 W/m2. In conclusion, the hybrid fusion model created in this study offers a strong tool for accurately and robustly forecasting hourly solar radiation value.

References

[1] Aggarwal, S. K., & Saini, L. M. (2014). Solar energy prediction using linear and non-linear regularization models: A study on AMS (American Meteorological Society) 2013–14 Solar Energy Prediction Contest. *Energy*, *78*, 247-256.

[2] JEBLI, I., BELOUADHA, F. Z., & KABBAJ, M. I. (2020, March). The forecasting of solar energy based on Machine Learning. In 2020 International Conference on Electrical and Information Technologies (ICEIT) (pp. 1-8). IEEE.

[3] Rahimi, N., Park, S., Choi, W., Oh, B., Kim, S., Cho, Y. H., ... & Lee, D. (2023). A Comprehensive Review on Ensemble Solar Power Forecasting Algorithms. Journal of Electrical Engineering & Technology, 1-15.

[4] Obiora, C. N., Ali, A., & Hassan, A. N. (2020, October). Predicting hourly solar irradiance using machine learning methods. In 2020 11th International Renewable Energy Congress (IREC) (pp. 1-6). IEEE.

[5] www.kaggle.com/saurabhshahane/northern-hemisphere-horizontal-photovoltaic

[6] Kazem, H. A., Yousif, J. H., & Chaichan, M. T. (2016). Modeling of daily solar energy system prediction using support vector machine for Oman. International Journal of Applied Engineering Research, 11(20), 10166-10172.

[7] Li, S., Chen, H., Wang, M., Heidari, A. A., & Mirjalili, S. (2020). Slime mould algorithm: A new method for stochastic optimization. Future Generation Computer Systems, 111, 300-323.

[8] You, L., & Zhu, M. (2023). Digital Twin simulation for deep learning framework for predicting solar energy market load in Trade-By-Trade data. Solar Energy, 250, 388-397.

[9] Krishnan, N., Kumar, K. R., & Inda, C. S. (2023). How solar radiation forecasting impacts the utilization of solar energy: A critical review. Journal of Cleaner Production, 135860.

[10] Mishra, D. P., Jena, S., Senapati, R., Panigrahi, A., &Salkuti, S. R. (2023). Global solar radiation forecast using an ensemble learning approach. International Journal of Power Electronics and Drive Systems, 14(1), 496.

[11] Zhang, R., Feng, M., Zhang, W., Lu, S., & Wang, F. (2018, November). Forecast of solar energy production-A deep learning approach. In 2018 IEEE International Conference on Big Knowledge (ICBK) (pp. 73-82). IEEE.

[12] Ghimire, S., Deo, R. C., Casillas-Pérez, D., Salcedo-Sanz, S., Sharma, E., & Ali, M. (2022). Deep learning CNN-LSTM-MLP hybrid fusion model for feature optimizations and daily solar radiation prediction. Measurement, 202, 111759.

[13] Shams, M. H., Niaz, H., Hashemi, B., Liu, J. J., Siano, P., &Anvari-Moghaddam, A. (2021). Artificial intelligence-based prediction and analysis of the oversupply of wind and solar energy in power systems. Energy Conversion and Management, 250, 114892.

[14] Kong, X., Du, X., Xu, Z., & Xue, G. (2023). Predicting solar radiation for space heating with thermal storage system based on temporal convolutional network-attention model. Applied Thermal Engineering, 219, 119574.

[15] Deo, R. C., Ahmed, A. M., Casillas-Pérez, D., Pourmousavi, S. A., Segal, G., Yu, Y., & Salcedo-Sanz, S. (2023). Cloud cover bias correction in numerical weather models for solar energy monitoring and forecasting systems with kernel ridge regression. Renewable Energy, 203, 113-130.

[16] Immanual, R., Kannan, K., Chokkalingam, B., Priyadharshini, B., Sathya, J., Sudharsan, S., & Nath, E. R. (2023). Performance Prediction of solar still using Artificial neural network. Materials Today: Proceedings, 72, 430-440.

[17] manceevaluationusing different classifi-ers. Applied Computational Intelligence and SoftComputing,2014.

[18] Yadav, M.,Purwar, R. (2017, January). Hindi handwritten char-acterrecognition using multiple classifiers. In 2017 7th Internation-al Con-ference on Cloud Compu-ting, Data ScienceEngineering-Confluence(pp.149-154).IEEE.

[19] Jabde, M., Patil, C., Mali, S.,Vibhute, A. Comparative Study of Ma-chine Learning and Deep Learning Classifiers on Handwrit-ten NumeralRecognition.