Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 02 , February - 2026

A Review Paper on Solar Power Prediction Using AQI And Weather Forecasting

Prem Suryawanshi
IT Department, Shri Sant Gajanan Maharaj College of Engineering, Shegaon, India

Asst. Prof. F.I. Khandwani,
Research Supervisor
IT Department, Shri Sant Gajanan Maharaj College of Engineering, Shegaon, India

Apurva Kulkarni
IT Department, Shri Sant Gajanan Maharaj College of Engineering, Shegaon, India

Garima Agrawal
IT Department, Shri Sant Gajanan Maharaj College of Engineering, Shegaon, India

Nidhi Agrawal
IT Department, Shri Sant Gajanan Maharaj College of Engineering, Shegaon, India

***Abstract*–Solar power generation is non-continuous and very much dependent on atmospheric conditions such as weather variability and pollution levels. Forecasting solar power accurately is important for effective energy planning, grid stability, and large-scale integration of PV systems. Traditional forecasting approaches (sensor-based) are effective, yet costly, depending on the location. This review paper presents a comparative analysis of software-based solar power prediction using weather forecast and AQI data, which replaces the traditional sensor-based techniques. The study focuses on existing methodologies that use meteorological variables, air pollutant concentrations (PM2.5, PM10, NO_2, SO_2, CO_2, O_3) and solar irradiance indicators for solar power forecasting. This paper reviews machine learning approaches, including regression models, ensemble learning, and stacked architectures, aligning with prediction accuracy, scalability, and practical deployment. This paper also emphasizes the impact of integrating AQI parameters into solar prediction models and demonstrates how effectively the combined weather-AQI features enhance the accuracy of solar power predictions. This study suggests employing a stacked ensemble model based on tree-based learners to align with the trends identified in the literature. The review concludes by identifying key research gaps and future directions, highlighting the potential of AQI-aware, data-driven models for cost-effective, scalable, and reliable solar power forecasting in smart energy systems.**

Key words–solar power prediction; air quality index (AQI); weather forecasting; machine learning; ensemble models; photovoltaic (PV) systems; renewable energy

I. INTRODUCTION

The limited quantity of fossil fuel resources forces us to adapt to renewable energy sources. This makes solar energy one of the most used and deployed energy sources. However, atmospheric factors lead to non-continuous solar power generation. This non-continuous nature of solar power presents significant challenges.

Weather conditions and atmospheric pollution affect the generation of solar power significantly as air pollutants scatter and absorb solar radiation. This leads to reduced PV efficiency and power output. Thus, including AQI parameters improves the accuracy of solar power prediction models.

Earlier, solar forecasting methods used physical sensors, pyranometers, and satellite-based measurements. Although, these methods gave accurate predictions, they were expensive, location-dependent, and needed regular maintenance. Thus, restricting their scalability. To overcome these limitations, software-based prediction systems evolved, which use historical datasets for predicting solar power. These systems are cost-effective and scalable.

The performance of solar power prediction systems has been improved significantly by the advancements in machine learning models. Environmental variables and solar energy output show a non-linear relationship. Regression techniques and ensemble models are capable of capturing this non-linear relationship. Particularly, ensemble and stacked models show better generalization than individual models. In spite of these advancements, many studies only focus on weather factors. This leaves a scope for integrated analysis of weather, AQI, and solar irradiance.

This review paper compares multiple solar power prediction models, including both weather parameters and AQI indicators. The paper studies existing work, and compares various machine learning models. This paper aims to highlight the importance of AQI in predicting solar power.

II. LITERATURE SURVEY

Precise prediction of solar power has been an active area of research. The studies focus on weather-based forecasting, air pollution impact analysis, and machine learning models.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 02 , February - 2026

This section provides literature relevant to solar power production.

The studies mainly focused on the weather conditions. These studies used statistical and machine learning models. Liu and Sun [7] implemented a Random Forest classification technique. This showed improvement in prediction accuracy. Zazoum [10] compared multiple machine learning models for solar power prediction. He highlighted the use of ensemble techniques. Lee et al. [11] further extended this work and implemented deep learning models. It showed better performance than conventional methods.

Sweets et al. [8] studied the long-term losses in solar energy generation. He reasoned about the influence of air pollution. Zhou et al. [9] examined the influence of air pollutants on PV system efficiency. He used the CMAQ model. These studies proved that air pollutants play an important role in reducing solar power generation.

Ghosh et al. [3] studied the impact of cleaner air on India's solar energy production. He concluded that less air pollution could lead to an increase in solar energy output. Galimova et al. [4] studied this perspective further by analysing air pollution mitigation. He highlighted the significance of air pollution in energy planning. These studies showed the necessity of integrating AQI indicators in solar power prediction.

Several studies exclusively combined weather and pollution features for solar power prediction. Chuluunsaikhan and Tserenpurev [1] developed an ML model using weather and air pollution features to estimate solar irradiance and showed improved accuracy. Jia et al. [5] studied multiple machine learning models for predicting solar radiation under varying weather and pollution conditions, and concluded that hybrid feature sets show improved accuracy. Jebli et al. [6] supported this by using Pearson correlation analysis to identify important weather and pollution features that impact the accuracy of solar energy prediction.

Recent enhancements in deep learning and hybrid architectures have further improved forecasting performance. Zhou et al. [2] proposed a DL model that enhanced a solar energy forecasting framework integrating IoT systems. Yeom et al. [15] implemented a deep convolutional LSTM network combined with satellite integrity for short-term solar power prediction. Reproducible research in PV energy prediction is facilitated by the UNISOLAR open dataset [16].

On the other hand, methodological studies, like, [13], [14], and [17], explored advanced statistical and zero-inflated techniques. These studies provided valuable insights into handling skewed and sparse datasets. Chiteka et al. [12] addressed PV soiling mitigation by optimizing cleaning frequency. It reidentified the importance of environmental factors in solar energy.

Shah et al. [18] serve as the base paper for the present study. Their work proposed a solution including an ML model combining AQI and weather features for solar power prediction. This study demonstrated that including AQI enhances the accuracy of prediction.

Overall, the literature shows a clear shift from sensor-dependent predictions to software-based prediction systems. Prior studies confirm the relevance of air pollution in solar power prediction. Some work explores stacked ensemble models that jointly leverage multiple tree-based learners for solar power prediction. This gap motivated the present work.

III. COMPARATIVE ANALYSIS

To predict solar power accurately, models are required to be capable of identifying and capturing nonlinear relationships between weather parameters, AQI features and solar irradiance. A comparative analysis was done to evaluate the effectiveness of different models. Individual machine learning models, ensemble blending techniques, and stacked ensemble learning models were used for the comprehensive study. The performance of these models was assessed using Root Mean Square Error (RMSE), normalized RMSE (nRMSE), and the coefficient of determination (R2 score).

The relationships between solar power generation, weather conditions and AQI parameters are complex, non-linear, and non-stationary. Thus, we used tree-based machine learning models. Also, their ability to provide feature importance helps in interpretability. This supported effective energy planning and environmental analysis.

A. Individual Model Comparison

Three individual tree-based models were used for this study – Random Forest Regressor, XGBoost, and LightGBM. They all were evaluated using the same training and testing sets to ensure fair comparison.

The predictive performance of RandomForest Regressor was shown to be moderate. It demonstrated RMSE values ranging from 72.94 to 89.87 and R2 scores between 0.26 and 0.45, thus showing 45% predicting accuracy. This is demonstrated in the following fig. 1.

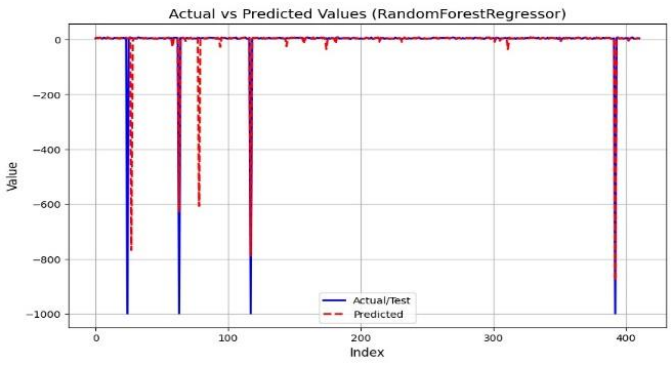

Fig. 1. Random Forest Regressor

Random Forest Regressor model suggested limited capability in capturing the complex relationships between weather parameters and AQI features.

The XGBoost Regressor demonstrated inconsistent performance with RMSE values ranging from 59.79 to 105.02

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 02 , February - 2026

and R2 scores from -0.01 to 0.67. This is shown in the following fig. 2.

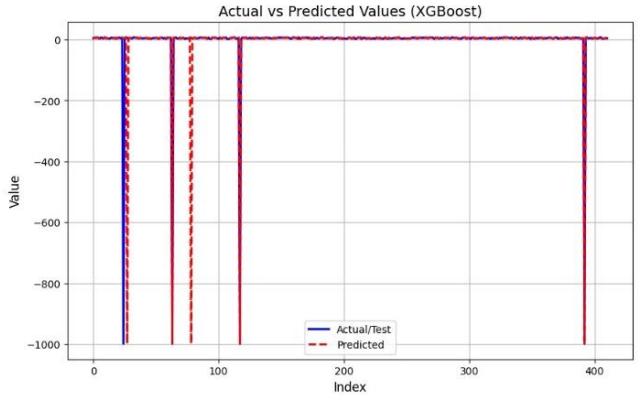

Fig. 2. XGBoost

In contrast, the LightGBM Regressor outperformed the Random Forest and XGBoost models. It achieved RMSE values as low as 50.41, and R2 scores up to 0.77 (77% predicting accuracy). This is demonstrated in the following fig. 3.

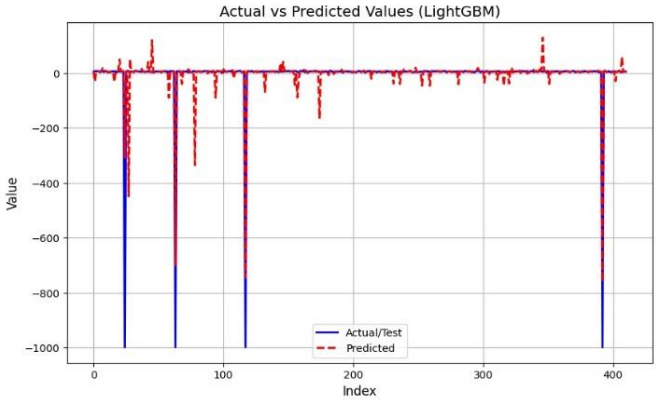

Fig. 3. LightGBM

The following table 3.1 shows the RMSE values and R2 scores for the discussed individual models.

TABLE I. RMSE AND R2 SCORE FOR INDIVIDUAL MODEL

Model	RMSE	R2 Score
RandomForest	72.94	0.45
XGBoost	85.75	0.24
LightGBM	52.75	0.71

B. Blended Ensemble Models

To increase the complementary strengths of individual models, a blended ensemble model was created. This model integrated XGBoost, LightGBM and RandomForest models. The new blended model reduced prediction error than individual learners. This model achieved RMSE values between 57.99 and 64.97 and R2 scores between 0.61 and 0.69. This represented an improvement over the individual models. However, this model is dependent on fixed chosen weights, thus, limiting its ability to adopt different data distributions. The following fig. 4 shows the blended model performance in predicting solar power generation.

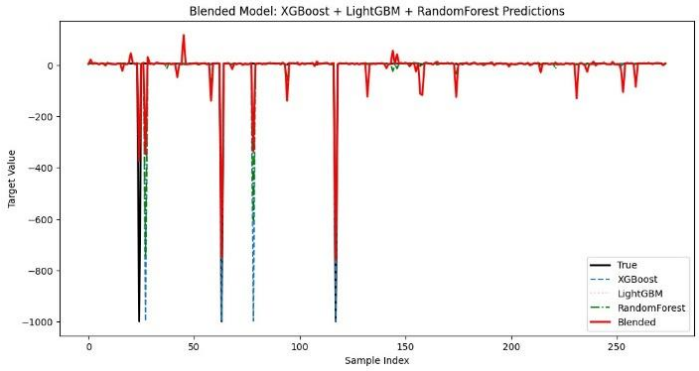

Fig.4. Blended Model

C. CV Optimized Blending

To overcome the limitation of the blended ensemble model, a cross-validation (CV) optimized blending model was implemented. This strategy improved generalization and reduced sensitivity to data splits. But the performance does not improve over that of the blended ensemble model, as shown in the following fig. 5. Thus, this strategy was dropped.

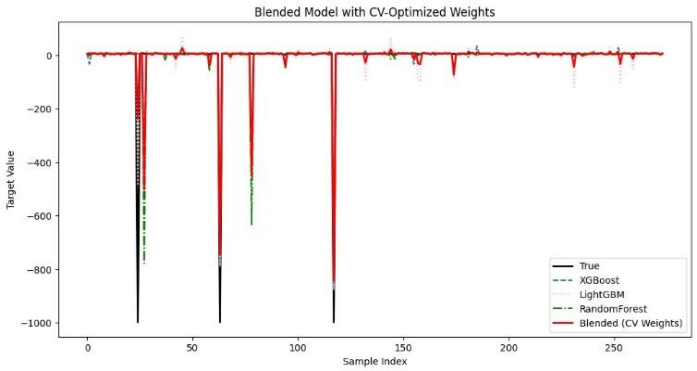

Fig. 5. Blended Model with CV

D. Stacked Ensemble Model

The final approach involved a stacked ensemble learning model. This strategy used the predictions from the base models (XGBoost, LightGBM, and RandomForest) as input features. This resulted in a superior accuracy over blended ensemble models. The stacked ensemble model showed the lowest RMSE of 38.63 and the highest R2 score of 0.86, outperforming all the other models and strategies. It also demonstrated improved robustness and handled outliers efficiently. This is shown in the following fig. 6.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 02 , February - 2026

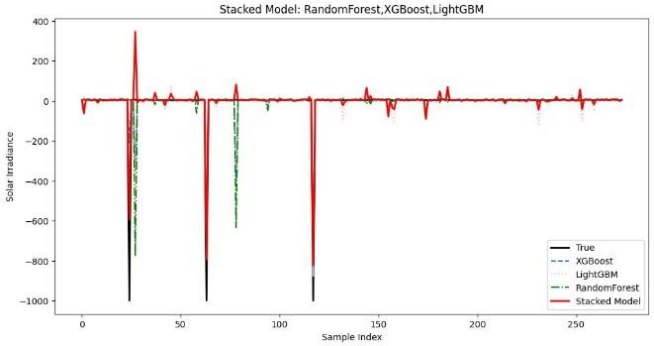

Fig. 6. Stacked Ensemble Model

E. Overall Insights

As the model complexity increases, the prediction accuracy improves. This is shown and proven in this comparative study. The blending ensemble model reduced variance by aggregating predictions. The stacking ensemble model showed the highest prediction accuracy. The following fig. 7 compares these learning models.

These findings align with the proposed methodology in the base paper of this study. It confirmed that integrating weather and AQI features improves prediction accuracy. A stacked ensemble framework provides a reliable, scalable, and cost-effective solution.

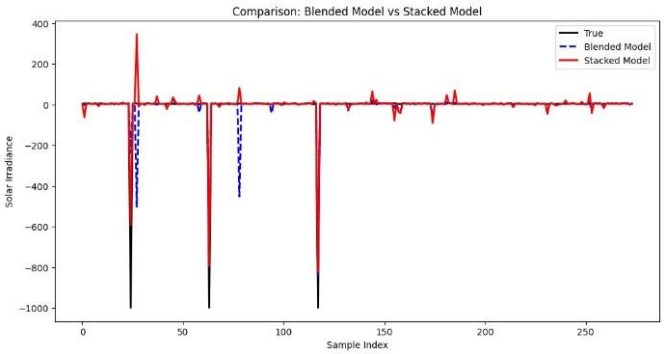

Fig. 7. Comparison: Blended vs Stacked Model

IV. FUTURE SCOPE

The improvement in the accuracy of the stacked ensemble model can be further studied. The dataset can be extended using longer timespans or data extension techniques. A larger and more diverse dataset allows the model to learn richer patterns. Thus, improving the prediction accuracy.

Different ensembling techniques can be explored to enhance the model accuracy. Extensive hyperparameter optimization can improve accuracy significantly. More improvements can be done by using additional environmental features like cloud imagery and satellite-derived solar irradiance.

These improvements would provide a more accurate, reliable, and deployment-ready model. It will be suitable for effective energy planning and smart grid stability. This will allow large-scale renewable energy resource integration.

V. CONCLUSION

This review paper compared different machine learning models that can be used for predicting solar power generation precisely. It also helps to analyze their effectiveness in prediction. This study demonstrated that tree-based models are well-suited for capturing the complex and nonlinear relationships between the weather parameters, air quality indicators and solar irradiance. The ensemble models outperformed the individual models, thus hinting to drop using the individual models for predicting solar energy. The results showed and confirmed that the stacked ensemble model provides the highest R2 score and the lowest RMSE, hence, the best-suited model for solar power prediction. The inclusion of AQI indicators in predicting solar energy proved a significant step in improving prediction accuracy. Thus, this paper concludes that the use of an AQI-aware stacked ensemble learning model makes the prediction more reliable and a scalable strategy for accurate solar power prediction. It also showed the future enhancements that are possible for real-world deployment.

ACKNOWLEDGEMENT

We express our sincere gratitude to the Management of Shri Sant Gajanan Maharaj College of Engineering, Shegaon, for their constant encouragement and commitment to fostering academic and research excellence.

We are deeply thankful to the Principal, Dr. S. B. Somani, for providing a supportive academic environment and the necessary institutional facilities to undertake and complete this research successfully.

We extend our profound appreciation to our research supervisor, Prof. F. I. Khandwani, Department of Information Technology, for invaluable guidance, and insightful suggestions.

We also acknowledge the support of the Head of the Department, Dr. S. D. Padiya, faculty members, and the laboratory staff for their technical assistance and cooperation during experimentation, data collection, and analysis. We are also grateful to the researchers and authors whose work formed the foundation of our literature survey.

Finally, we express our heartfelt thanks to everyone who directly or indirectly contributed to the successful completion of this work.

REFERENCES

[1] Chuluunsaikhan, Tserenpurev. (2021). Predicting the Power Output of Solar Panels based on Weather and Air Pollution Features using Machine Learning. Journal of Korea Multimedia Society. 24. 222. 10.9717/kmms.2021.24.2.222.

[2] Zhou, H., Liu, Q., Yan, K., Du, Y. (2021). Deep Learning Enhanced Solar Energy Forecasting with AI-Driven IoT. Wireless Communications and Mobile Computing, 2021, Article ID 9249387. doi:10.1155/2021/9249387.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 02 , February - 2026

[3] Ghosh, S., Dey, S., Ganguly, D., Roy, S. B., & Bali, K. (2022). Cleaner air would enhance India's annual solar energy production by 6–28 TWh. Environmental Research Letters, 17(5), 054007. doi:10.1088/1748-9326/ac5d9a.

[4] Galimova, T., Ram, M., & Breyer, C. (2022). Mitigation of air pollution and corresponding impacts during a global energy transition towards 100% renewable energy system by 2050. Energy Reports, 8, 14124-14143. ISSN 2352-4847. doi:10.1016/j.egyr.2022.10.343.

[5] Jia, Dongyu & Yang, Liwei & Lv, Tao & Liu, Weiping & Gao, Xiaoqing & Zhou, Jiaxin. (2022). Evaluation of machine learning models for predicting daily global and diffuse solar radiation under different weather/pollution conditions. Renewable Energy. 187. 10.1016/j.renene.2022.02.002.

[6] Jebli, I., Belouadha, F.-Z., Kabbaj, M. I., & Tilioua, A. (2021). Prediction of solar energy guided by Pearson correlation using machine learning. Energy, 224, 120109. doi:10.1016/j.energy.2021.120109

[7] Liu, D., & Sun, K. (2019). Random forest solar power forecast based on classification optimization. Energy, 115940. doi:10.1016/j.energy.2019.115940

[8] Sweerts, B., Pfenninger, S., Yang, S., Folini, D., van der Zwaan, B., & Wild, M. (2019). Estimation of losses in solar energy production from air pollution in China since 1960 using surface radiation data. Nature Energy. doi:10.1038/s41560-019-0412-4

[9] Zhou, L., Schwede, D. B., Appel, K. W., Mangiante, M. J., Wong, D. C., Napelenok, S. L., Whung, P.-Y., & Zhang, B. (2019). The impact of air pollutant deposition on solar energy system efficiency: an approach to estimate PV soiling effects with the Community Multiscale Air Quality (CMAQ) model. Science of the Total Environment, 651(Pt 1), 456–465. doi:10.1016/j.scitotenv.2018.09.194

[10] Zazoum, B. (2022). Solar photovoltaic power prediction using different machine learning methods. Energy Reports, 8(Supplement 1), 19-25. ISSN 2352-4847. doi:10.1016/j.egyr.2021.11.183.

[11] Lee, C.-H., Yang, H.-C., & Ye, G.-B. (2021). Predicting the Performance of Solar Power Generation Using Deep Learning Methods. Applied Sciences, 11(15), 6887. doi:10.3390/app11156887

[12] Chiteka, K., Arora, R., Sridhara, S. N., & Enweremadu, C. C. (2020). A novel approach to Solar PV cleaning frequency optimization for soiling mitigation. Scientific African, 8, e00459. ISSN 2468-2276. doi:10.1016/j.sciaf.2020.e00459.

[13] Kim, D.-W., Deo, R. C., Park, S.-J., Lee, J.-S., & Lee, W.-S. (2019). Weekly heat wave death prediction model using zero-inflated regression approach. Theoretical and Applied Climatology, 137, 823-838. doi: 10.1007/s00704-018-2632-4

[14] Thomas, S. J. (2010). Model-based clustering for multivariate time series of counts. Rice University. ProQuest Dissertations Publishing. (Publication No. 3421317).

[15] Yeom, J. M., Deo, R. C., Adamowski, J. F., Park, S., & Lee, C. S. (2020). Spatial mapping of short-term solar radiation prediction incorporating geostationary satellite images coupled with deep convolutional LSTM networks for South Korea. Environmental Research Letters, 15(9), 094025. doi: 10.1088/1748-9326/ab9467

[16] S. Wimalaratne, D. Haputhanthri, S. Kahawala, G. Gamage, D. Alahakoon and A. Jennings, "UNISOLAR: An Open Dataset of Photovoltaic Solar Energy Generation in a Large Multi-Campus University Setting," 2022 15th International Conference on Human System Interaction (HSI), 2022, pp. 1-5, doi: 10.1109/HSI55341.2022.986947

[17] Feng, C.X. A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. J Stat Distrib App 8, 8 (2021).

[18] Shah, A., Viswanath, V., Gandhi, K., & Patil, N. M. (2024). Predicting Solar Energy Generation with Machine Learning based on AQI and Weather Features. Synapse, Computer Engineering, D.J. Sanghvi College of Engineering, Mumbai, India. arXiv:2408.12476 [cs.LG].