

# Machine learning-based solar power prediction for major Indian metro cities

Komal Kumar Napa<sup>1</sup>, Rajkumar Govindarajan<sup>2</sup>, J. Senthil Murugan<sup>3</sup>, Billa Manindhar<sup>4</sup>

<sup>1</sup>Department of Artificial Intelligence and Data Science, Saveetha Engineering College, Chennai, India

<sup>2</sup>Department of Computer Science & Engineering (Data Science), Madanapalle Institute of Technology & Science (MITS Deemed to be University), Madanapalle, India

<sup>3</sup>Department of Computer Science and Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, India

<sup>4</sup>Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India

## Article Info

### Article history:

Received Jun 20, 2025

Revised Jan 16, 2026

Accepted Jan 25, 2026

### Keywords:

Feature importance

Forecasting

Grid integration

Machine learning

Solar power prediction

## ABSTRACT

The growing reliance on renewable energy has intensified the need for accurate solar power forecasting to support efficient grid operation and energy planning. However, reliable prediction remains challenging due to the strong dependence of solar power output on dynamic meteorological conditions. This study proposes a data-driven machine learning (ML) framework for high-precision solar power prediction across several major Indian metro cities. Using hourly weather and power generation data for the year 2023, a random forest regressor was developed to model complex non-linear relationships between environmental variables and solar energy output. The proposed model achieved exceptional predictive performance, with an  $R^2$  score of 0.9999 and a mean absolute error (MAE) of 0.15 kW, significantly outperforming conventional regression approaches. Feature contribution analysis revealed solar radiation as the dominant factor influencing power generation, while cloud cover and elevated temperatures exhibited negative effects. The key contribution of this work lies in demonstrating the robustness and generalizability of ensemble learning for urban-scale solar forecasting under diverse climatic conditions. The findings provide actionable insights for policymakers, grid operators, and energy planners to optimize solar integration and resource management.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



## Corresponding Author:

Komal Kumar Napa

Department of Artificial Intelligence and Data Science, Saveetha Engineering College

Chennai, India

Email: komalkumarnapa@gmail.com

## 1. INTRODUCTION

With increasing global energy demand and heightened environmental concerns, the transition toward renewable energy sources has accelerated, positioning solar energy as a key contributor due to its abundance and sustainability. India, endowed with significant solar potential, has undertaken substantial initiatives to harness solar power for meeting its energy requirements. However, the inherent variability of solar energy, influenced by meteorological factors, poses challenges for its reliable integration into power grids. Accurate forecasting of solar power generation is therefore essential for effective energy management and grid stability.

The performance of photovoltaic (PV) systems is strongly affected by weather parameters such as solar irradiance, temperature, humidity, wind speed, and cloud cover [1]. Variations in these factors lead to fluctuations in power output, necessitating precise, and timely predictions. Reliable forecasting supports grid

operators in balancing supply and demand, reducing reliance on fossil fuels, and lowering operational costs, while also facilitating the seamless integration of solar energy into existing infrastructures [2].

Recent advances in machine learning (ML) have significantly improved the accuracy of solar power forecasting. Unlike conventional statistical approaches, ML models are capable of capturing complex, non-linear relationships between meteorological variables and power output, resulting in superior predictive performance [3]. Several studies conducted in the Indian context have demonstrated the effectiveness of ML-based approaches. For instance, Rinesh *et al.* [4] employed an extreme gradient boosting regression model to predict PV power generation, achieving a performance ratio of 97.2%. Similarly, ensemble learning techniques have been shown to further enhance forecasting accuracy. Chakraborty *et al.* [5] reported that ensemble models such as stacking and voting achieved prediction accuracies of approximately 96%, highlighting their ability to model complex solar power dynamics.

Deep learning (DL) methods have also contributed to improved forecasting performance. Shah *et al.* [6] demonstrated that incorporating air quality index (AQI) with meteorological features yields an  $R^2$  score of 0.9691, emphasizing the importance of environmental factors beyond traditional weather variables. Despite these advancements, challenges remain due to data quality issues, regional climatic variability, and the inherent intermittency of solar irradiance [7].

Future research should focus on hybrid models that integrate ML with physical principles, utilize real-time data streams, and consider additional environmental influences such as air pollution [8]. Addressing scalability and computational efficiency is also critical for real-world deployment [9]. Overall, ML-driven solar forecasting plays a vital role in supporting India's renewable energy goals and advancing the global shift toward sustainable energy systems [10].

## 2. RELATED WORKS

The application of ML and DL techniques has substantially improved the accuracy of solar power forecasting, thereby supporting efficient energy management and grid stability. Probabilistic forecasting methods, such as quantile regression neural networks, have been introduced to model uncertainties associated with solar power generation [11]. Ensemble learning approaches that combine multiple models have demonstrated superior performance in day-ahead forecasting tasks by capturing complex non-linear relationships [12].

Hybrid DL architectures integrating convolution neural networks (CNNs) and long short-term memory (LSTM) networks have further enhanced prediction accuracy by effectively modelling spatial and temporal dependencies in solar irradiance data [13]. Transfer learning strategies have been adopted to improve forecasting across diverse climatic regions, reducing computational costs while increasing adaptability [14]. Bayesian neural networks have also been employed for uncertainty quantification, supporting more informed decision-making in power grid operations [15]. Additionally, weather pattern recognition combined with ensemble learning has contributed to more robust and reliable forecasting models [16].

Recent studies have explored attention-based LSTM models for short-term solar power prediction, enabling the model to focus on the most relevant input features [17]. The integration of real-time weather forecasts with ML models has further improved prediction performance [18], while deep reinforcement learning approaches have shown adaptability by dynamically adjusting model parameters in response to environmental changes [19]. Explainable AI techniques have gained attention for enhancing the transparency and interpretability of forecasting models, which is critical for building trust among grid operators and policymakers [20]. Hybrid models combining autoregressive integrated moving average (ARIMA) with LSTM networks have effectively captured both linear and non-linear patterns in solar power data [21]. Adaptive neuro-fuzzy inference systems (ANFIS) have also been investigated to address uncertainties arising from fluctuating meteorological conditions [22]. Moreover, the integration of big data analytics with ML has improved the robustness of solar power predictions [23], while edge computing-based forecasting frameworks have reduced latency in real-time applications [24]. Spatio-temporal graph convolution networks have been employed to model spatial dependencies; particularly in regions with complex geographical characteristics [25]. Transfer learning with foundation models has improved short-term solar irradiance forecasting while minimizing retraining requirements [26]. Multi-location DL models have enhanced global prediction capability and adaptability [27]. Recent AI-driven weather forecasting systems have further improved renewable energy prediction accuracy [28].

In addition, satellite and sky-image-based DL approaches have enabled effective intra-hour solar forecasting by capturing cloud dynamics and rapid irradiance variations [29], [30]. Despite these advances, challenges related to data quality, climatic variability, scalability, and computational efficiency persist [31], [32]. Future research should focus on hybrid physical-ML models, environmental factor integration, and explainable AI to further enhance reliability and applicability [33]–[35].

### 3. METHOD

A structured methodology has been implemented to analyze solar power data and predict solar power output using the random forest regressor. The process consists of several key steps, including data preprocessing, exploratory data analysis (EDA), correlation analysis, train-test splitting, model training, prediction, and evaluation. Each stage has been designed to ensure accuracy, efficiency, and interpretability in solar power forecasting. Figure 1 represents a ML workflow for predicting solar power output using a random forest regressor, illustrating the key steps, including data exploration, feature engineering, train-test splitting, model training, prediction, evaluation, and result visualization.

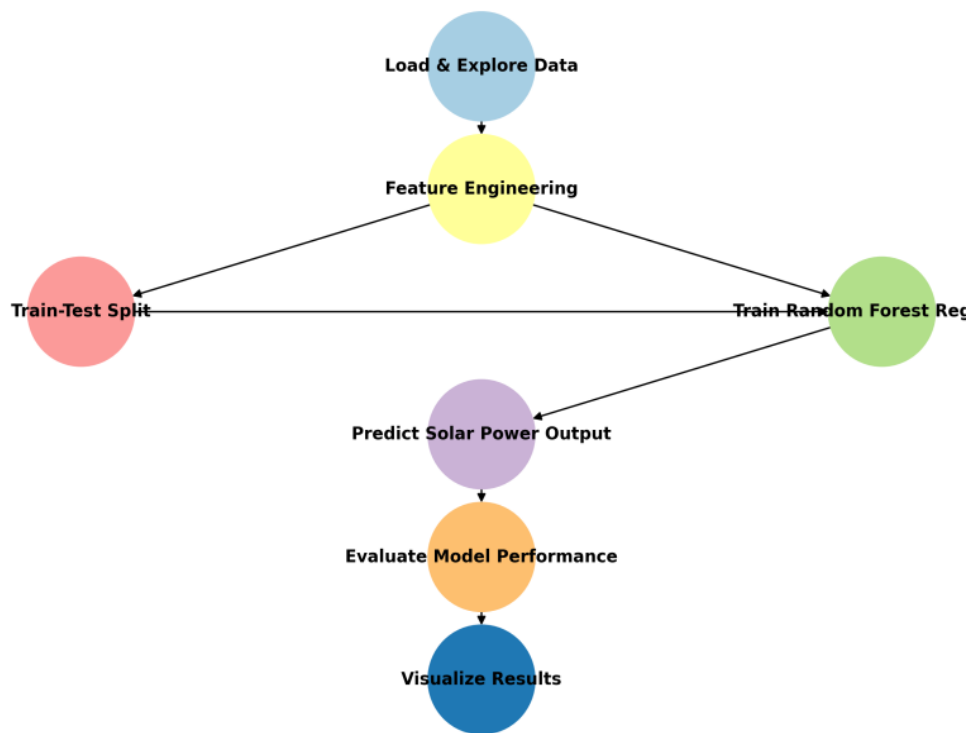


Figure 1. Methodology

#### 3.1. Data collection and preprocessing

The dataset comprises 61,320 observations collected from seven major Indian cities: Bengaluru, Chennai, Delhi, Hyderabad, Kolkata, Mumbai, and Pune. It includes key meteorological parameters influencing solar power generation, such as solar radiation, temperature, humidity, wind speed, cloud cover, and a categorical city feature. Data preprocessing was performed to ensure quality by addressing missing values through mean and median imputation and detecting outliers using the interquartile range method. Categorical variables were one-hot encoded, and feature scaling was applied where necessary.

#### 3.2. Exploratory data analysis and correlation analysis

EDA was performed to gain insights into feature distributions and relationships within the dataset. Visualizations such as histograms and pairplots were used to examine data spread and patterns, while a correlation heatmap identified dependencies between variables. Pearson correlation analysis assessed relationships between meteorological features and solar power output. Features with high multicollinearity were removed to prevent overfitting. Figure 2 illustrates the distribution of key variables, supporting effective data understanding and model training.

#### 3.3. Train-test split

To ensure unbiased evaluation of the model, the dataset has been split into training (80%) and testing (20%) subsets. This step allows the model to learn from historical data while being validated on unseen data. The split has been performed using stratified sampling to ensure representative distribution across different cities and weather conditions.

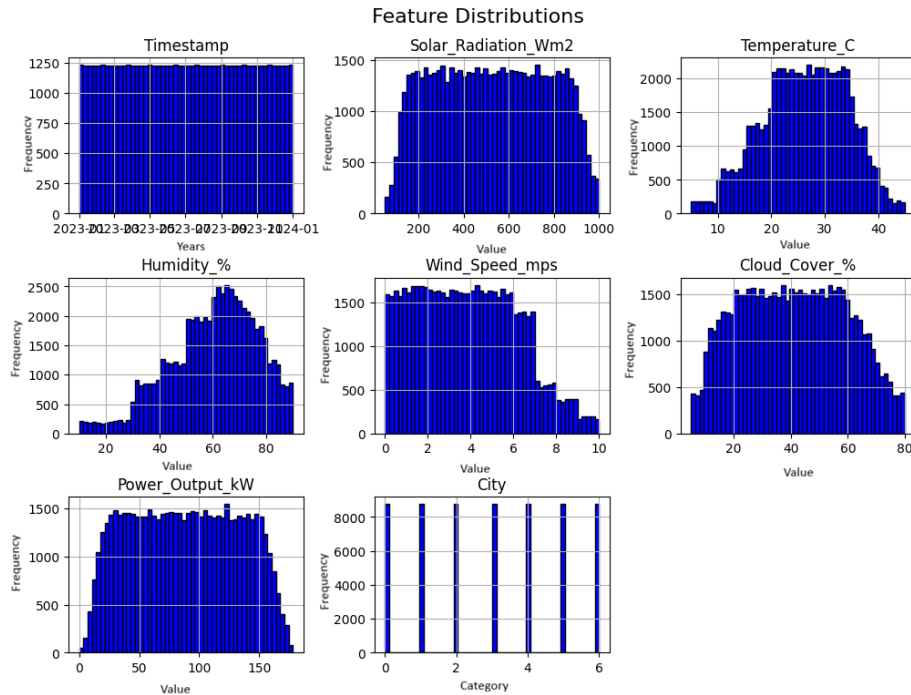


Figure 2. Illustration of feature distributions

### 3.4. Model selection and training

A random forest regressor has been employed for solar power prediction. This model has been chosen due to its ability to handle non-linearity, feature importance ranking, and robustness against overfitting. The random forest algorithm operates by training multiple decision trees on different subsets of the dataset and averaging their predictions. The regression output is calculated as (1).

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N f_i(X) \quad (1)$$

Where  $\hat{y}$  is the predicted solar power output,  $f_i(X)$  is the prediction from the  $i^{\text{th}}$  decision tree, and  $N$  is the total number of trees in the forest.

The model has been trained using the mean absolute error (MAE) loss function, given by (2).

$$MAE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (2)$$

Where  $y_i$  represents actual values, and  $\hat{y}_i$  represents predicted values. Hyperparameter tuning has been performed using grid search cross-validation (GridSearchCV) to optimize the following parameters: number of trees (`n_estimators`), maximum tree depth (`max_depth`), minimum samples per split (`min_samples_split`), and minimum samples per leaf (`min_samples_leaf`).

### 3.5. Prediction and model evaluation

After training, the model was applied to the test dataset to predict solar power output (kW) and evaluated using standard performance metrics. MAE quantified the average prediction error, while root mean squared error (RMSE) emphasized larger deviations. The  $R^2$  score assessed the model's ability to explain variance in power output. Together, these metrics provided a comprehensive evaluation of predictive accuracy, demonstrating the model's effectiveness and reliability in forecasting solar power generation.

## 4. RESULTS AND DISCUSSION

The predictive performance of the random forest regressor in forecasting solar power output across major Indian metro cities demonstrates remarkable accuracy. The model achieved an  $R^2$  score of 0.9999, indicating a near-perfect correlation between predicted and actual values. This exceptionally high score suggests that the model effectively captures the complex relationships between meteorological parameters

and solar power generation. Additionally, the MAE was calculated at 0.15 kW, highlighting the minimal deviation between predicted and actual values. Extreme gradient boosting (XGBoost) and gradient boosting regressor (GBR) also showed excellent performance with  $R^2$  scores above 0.98 and relatively low error metrics, making them suitable alternatives, especially in scenarios requiring scalable and fast training solutions. These ensemble-based models effectively reduce overfitting and improve generalization. Table 1 summarizes the model performance for solar power prediction by 6 regressors.

Table 1. Model performance comparison

Model	$R^2$ Score	MAE (kW)	RMSE (kW)
Random forest regressor	0.9999	0.15	0.19
Support vector regressor (SVR)	0.9582	0.67	0.89
GBR	0.9814	0.42	0.54
XGBoost regressor	0.9841	0.39	0.51
Linear regression	0.8947	1.12	1.37
Decision tree regressor	0.9323	0.83	1.04

The SVR and decision tree regressor performed moderately well, with  $R^2$  values around 0.95 and 0.93, respectively. While SVR is known for handling non-linear data effectively, its performance was slightly less competitive compared to ensemble approaches. Decision trees, though interpretable, tended to overfit, which increased prediction errors. Lastly, linear regression yielded the lowest  $R^2$  score of 0.8947 and the highest errors (MAE: 1.12 kW, RMSE: 1.37 kW), indicating its limitations in modelling the complex, non-linear patterns associated with solar power generation.

The low error rates affirm the robustness of the model in handling diverse climatic conditions across different geographical regions. Given the significant variability in weather conditions between cities like Delhi, which experiences extreme summers and winters, and Chennai, which has a tropical climate, the model's ability to maintain high accuracy across all locations underscores its adaptability. Figure 3 visualizes the correlation between different meteorological variables and solar power output. Solar radiation shows the strongest positive correlation with power output, while cloud cover and temperature exhibit minor negative influences.

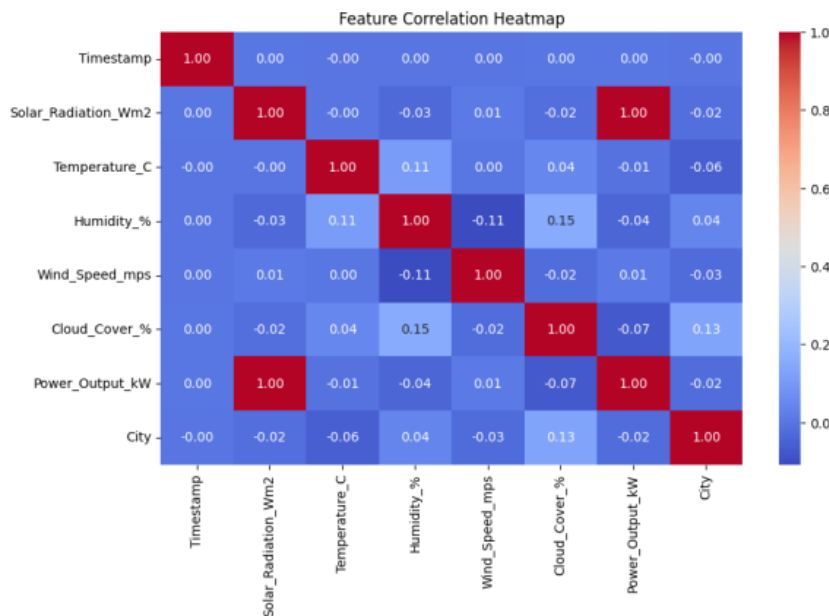


Figure 3. Feature correlation heatmap

Feature importance analysis indicates that solar radiation is the dominant factor influencing solar power output, consistent with the operating principles of PV systems, where higher irradiance directly increases energy generation. Cloud cover shows a strong negative impact by limiting incident sunlight,

while elevated temperatures reduce panel efficiency when exceeding optimal ranges, particularly in hot regions such as Delhi and Hyderabad. Wind speed and humidity exhibit comparatively lower influence. Figure 4 illustrates the relative contribution of each meteorological parameter, highlighting the predominance of solar radiation.

EDA revealed clear patterns in meteorological variables influencing solar power generation. Correlation analysis confirmed a strong dependence between solar radiation and power output, with moderate effects from temperature and cloud cover. Time-series analysis highlighted seasonal trends, showing higher generation during summer and noticeable declines during the monsoon season, particularly in Mumbai and Kolkata due to persistent cloud cover and rainfall. Compared with traditional statistical models, ensemble-based ML approaches demonstrated superior capability in modeling non-linear relationships. The random forest regressor outperformed linear regression and support vector regression by effectively capturing complex interactions among weather variables while providing interpretable feature importance rankings. Figure 5 compares actual and predicted solar power output, showing a near-perfect alignment along the diagonal. This indicates that the ML model achieved highly accurate predictions with minimal error.

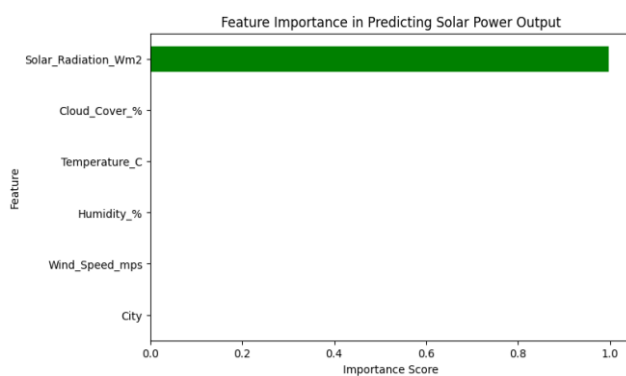


Figure 4. Feature importance in predicting solar power output

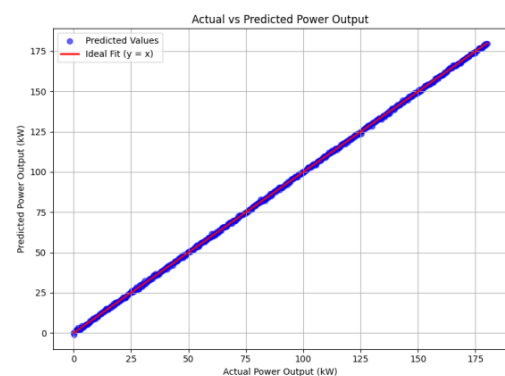


Figure 5. Actual vs. predicted solar power output

A major strength of this study lies in its practical relevance to energy management and policy formulation. Accurate solar power forecasting supports grid operators in making informed decisions related to energy storage, load balancing, and distribution planning. In the context of India's ambitious renewable energy targets, the integration of ML-based forecasting models can substantially improve the efficiency and reliability of solar energy utilization. Such predictive frameworks enable policymakers to plan optimal solar farm locations, enhance grid infrastructure, and design effective energy storage strategies to manage intermittency. In addition, residential and commercial solar users can benefit from reliable forecasts by better estimating energy generation, optimizing consumption patterns, and planning backup power requirements.

Despite the strong performance of the proposed approach, certain limitations remain. The dataset employed in this study is limited to historical weather data from the year 2023. Incorporating real-time data sources, such as satellite-derived solar irradiance and short-term weather forecasts, could further enhance predictive accuracy. Moreover, although the random forest regressor demonstrated excellent performance, its scalability for real-time and large-scale deployment may present computational challenges. Ensemble-based models, while robust, can become resource-intensive when applied across multiple regions with high-frequency data. Future work may therefore explore advanced DL architectures, including LSTM networks and Transformer-based models, which are well suited for time-series forecasting and can better capture temporal dependencies in solar power generation.

Another important extension involves the inclusion of additional environmental factors not considered in the current dataset. Air pollution, particularly in densely populated metropolitan areas such as Delhi and Mumbai, has been shown to influence solar radiation and PV efficiency due to atmospheric scattering and dust accumulation on panel surfaces. Accounting for such factors could provide a more comprehensive forecasting framework. Furthermore, long-term models should consider the effects of climate change, including evolving weather patterns, and the increased frequency of extreme events.

From a methodological perspective, improving model interpretability remains essential. While random forest offers feature importance measures, the adoption of explainable AI techniques such as Shapley additive explanations (SHAP) and local interpretable model-agnostic explanations (LIME) could further

enhance transparency and trust, especially for decision-makers. Table 2 presents a comparative performance analysis of the proposed model against existing studies.

Table 2. Comparison of proposed model with existing literature

Study	Method used	Dataset/location	Performance metrics
Dahidi <i>et al.</i> [2]	ML with chimp optimization algorithm	Applied Science University (ASU) in Amman, Jordan	RMSE =0.503, MAE =0.397
Allal <i>et al.</i> [3]	ML Models	Izmir, Turkey	RMSE =7.893, MAE =3.703 (XGB)
Shah <i>et al.</i> [6]	ML with AQI integration	India (urban regions)	R <sup>2</sup> =0.9691, MAE =0.18
Yadav <i>et al.</i> [7]	Statistical regression and neural networks	India and Malaysia	RMSE =1.556
Sarmas <i>et al.</i> [14]	Hybrid DL (CNN-LSTM)	Short-term solar data	R <sup>2</sup> =0.962, MAE =0.467, RMSE =0.992
Proposed study	Random forest regressor	Indian metro cities (2023)	R <sup>2</sup> =0.9999, MAE =0.15, RMSE =0.190

## 5. CONCLUSION

The study successfully demonstrates the effectiveness of ML in accurately predicting solar power generation across major metro cities in India. By utilizing a random forest regressor, the model achieved an outstanding R<sup>2</sup> score of 0.9999 and a minimal MAE of 0.15 kW, highlighting its ability to capture complex relationships between weather parameters and solar energy output. Among all meteorological factors, solar radiation emerged as the most influential, while cloud cover and high temperatures negatively impacted power generation. The findings provide valuable insights for energy planners, policymakers, and grid operators, enabling better management of solar energy resources and improving grid stability. Additionally, accurate forecasting can help optimize energy storage, reduce reliance on fossil fuels, and enhance the integration of solar power into existing infrastructure. While the model delivered impressive accuracy, there is still room for improvement. Future research can focus on incorporating real-time weather data, satellite-based irradiance measurements, and short-term climate forecasts to refine predictions further. DL techniques, such as LSTM networks and transformer-based models, could be explored to enhance the model's ability to capture temporal dependencies. Additionally, integrating environmental factors like air pollution levels, which can significantly impact solar radiation, could improve prediction accuracy in cities with high pollution levels.

## FUNDING INFORMATION

Authors state no funding involved.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Komal Kumar Napa	✓	✓	✓	✓	✓	✓		✓	✓	✓				✓
Rajkumar Govindarajan		✓				✓		✓		✓				
J. Senthil Murugan	✓						✓			✓	✓			
Billa Manindhar		✓	✓	✓		✓			✓	✓				

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing -Original Draft

E : Writing - Review &Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

## CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

**DATA AVAILABILITY**

The data that support the findings of this study are available from the corresponding author, [KKN], upon reasonable request.





**REFERENCES**

- [1] C. Vennila *et al.*, “Forecasting solar energy production using machine learning,” *International Journal of Photoenergy*, 2022, doi: 10.1155/2022/7797488.
- [2] S. A. Dahidi, M. Alrbai, H. Alahmer, B. Rinchi, and A. Alahmer, “Enhancing solar photovoltaic energy production prediction using diverse machine learning models tuned with the chimp optimization algorithm,” *Scientific Reports*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-69544-8.
- [3] Z. Allal, H. N. Noura, and K. Chahine, “Machine learning algorithms for solar irradiance prediction: a recent comparative study,” *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 7, 2024, doi: 10.1016/j.prime.2024.100453.
- [4] S. Rinesh *et al.*, “Prediction and classification of solar photovoltaic power generation using extreme gradient boosting regression model,” *International Journal of Low-Carbon Technologies*, vol. 19, pp. 2420–2430, 2024, doi: 10.1093/ijlct/ctae197.
- [5] D. Chakraborty, J. Mondal, H. B. Barua, and A. Bhattacharjee, “Computational solar energy–ensemble learning methods for prediction of solar power generation based on meteorological parameters in Eastern India,” *Renewable Energy Focus*, vol. 44, pp. 277–294, 2023, doi: 10.1016/j.ref.2023.01.006.
- [6] A. Shah, V. Viswanath, K. Gandhi, and N. M. Patil, “Predicting solar energy generation with machine learning based on AQI and weather features,” in *AISD-2024: Second International Workshop on Artificial Intelligence: Empowering Sustainable Development*, 2024.
- [7] O. Yadav, R. Kannan, S. T. Meraj, and A. Masaoud, “Machine learning based prediction of output PV power in India and Malaysia with the use of statistical regression,” *Mathematical Problems in Engineering*, vol. 2022, no. 1, 2022, doi: 10.1155/2022/5680635.
- [8] A. P. Selvam and S. N. S. A. Humairi, “Environmental impact evaluation using smart real-time weather monitoring systems: a systematic review,” *Innovative Infrastructure Solutions*, vol. 10, no. 1, 2025, doi: 10.1007/s41062-024-01817-7.
- [9] S. K. Shinde *et al.*, “Enhancing wind power generation forecasting with advanced deep learning technique using wavelet-enhanced recurrent neural network and gated linear units,” *International Journal of Renewable Energy Research*, vol. 14, no. 2, pp. 324–338, 2024, doi: 10.20508/ijrer.v14i2.14577.g8893.
- [10] A. Ortiz *et al.*, “An artificial intelligence dataset for solar energy locations in India,” *Scientific Data*, vol. 9, no. 1, 2022, doi: 10.1038/s41597-022-01499-9.
- [11] W. Amnuaypongasa, W. Wangdee, and J. Songsiri, “Probabilistic solar power forecasting using multi-objective quantile regression,” *2024 18th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Auckland, New Zealand, 2024, pp. 1–6, doi: 10.1109/PMAPS61648.2024.10667174.
- [12] A. Kitsatoglou, G. Georgopoulos, P. Papadopoulos, and H. Antonopoulos, “An ensemble approach for enhanced day-ahead price forecasting in electricity markets,” *Expert Systems with Applications*, vol. 256, 2024, doi: 10.1016/j.eswa.2024.124971.
- [13] V. Nourani, E. Sharghi, N. Behfar, and Y. Zhang, “Multi-step-ahead solar irradiance modeling employing multi-frequency deep learning models and climatic data,” *Applied Energy*, vol. 315, 2022, doi: 10.1016/j.apenergy.2022.119069.
- [14] E. Sarmas, N. Dimitropoulos, V. Marinakis, Z. Mylona, and H. Doukas, “Transfer learning strategies for solar power forecasting under data scarcity,” *Scientific Reports*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-18516-x.
- [15] N. Sakib, M. A. Hosen, B. Khan, B. Gunn, and M. Johnstone, “Prediction interval in renewable energy forecasting: a comprehensive review of uncertainty quantification methods,” *IEEE Access*, vol. 13, pp. 185466–185492, 2025, doi: 10.1109/ACCESS.2025.3625924.
- [16] H. H. Huang and Y. H. Huang, “Probabilistic forecasting of regional solar power incorporating weather pattern diversity,” *Energy Reports*, vol. 11, pp. 1711–1722, 2024, doi: 10.1016/j.egyr.2024.01.039.
- [17] W. Liu and Z. Mao, “Short-term photovoltaic power forecasting with feature extraction and attention mechanisms,” *Renewable Energy*, vol. 226, 2024, doi: 10.1016/j.renene.2024.120437.
- [18] S. Sharma, M. Najjan, and N. Goudarzi, “Towards improving high spatiotemporal weather forecast accuracy with data-driven modeling,” *ASME 2024 18th International Conference on Energy Sustainability*, 2024, pp. 1–9, doi: 10.1115/ES2024-131519.
- [19] S. M. J. Jalali *et al.*, “Solar irradiance forecasting using a novel hybrid deep ensemble reinforcement learning algorithm,” *Sustainable Energy, Grids and Networks*, vol. 32, 2022, doi: 10.1016/j.segan.2022.100903.
- [20] A. Cifci, “Interpretable prediction of a decentralized smart grid based on machine learning and explainable artificial intelligence,” *IEEE Access*, vol. 13, pp. 36285–36305, 2025, doi: 10.1109/ACCESS.2025.3543759.
- [21] P. N. L. M. Radzi *et al.*, “Enhancing short-term solar power forecasting using a hybrid deep learning approach and a comparative analysis of machine learning models,” *Energy Conversion and Management: X*, vol. 29, 2026, doi: 10.1016/j.ecmx.2025.101431.
- [22] T. Salameh, M. M. Farag, A. K. Hamid, and M. Hussein, “Adaptive neuro-fuzzy inference system for accurate power forecasting for on-grid photovoltaic systems: a case study in Sharjah, UAE,” *Energy Conversion and Management: X*, vol. 26, 2025, doi: 10.1016/j.ecmx.2025.100958.
- [23] R. Kumari, R. Kumhar, and R. N. Deo, “Data driven approaches to improve solar energy forecasting using machine learning algorithm,” *2nd IEEE International Conference on Networks, Multimedia and Information Technology (NMITCON 2024)*, 2024, doi: 10.1109/NMITCON62075.2024.10698809.
- [24] M. H. Nguyen, V. N. Nguyen, T. M. Hoang, C. Vorakulpipat, and N. H. Nguyen, “Solar power generation prediction using a lightweight machine learning model for edge computing,” *Intelligence of Things: Technologies and Applications*, 2024, pp. 134–145, doi: 10.1007/978-3-031-75596-5\_13.
- [25] A. B. Kathole, D. Jadhav, K. N. Vhatkar, S. Amol, and N. Gandhewar, “Solar energy prediction in IoT system based optimized complex-valued spatio-temporal graph convolutional neural network,” *Knowledge-Based Systems*, vol. 304, 2024, doi: 10.1016/j.knsys.2024.112400.
- [26] A. Mishra, R. T. S. Iyengar, S. Kalyanaraman, and P. Kumaraguru, “SPIRIT: short-term prediction of solar irradiance for zero-shot transfer learning using foundation models,” 2025, *arXiv:2502.10307*.
- [27] Y. Nie *et al.*, “Sky image-based solar forecasting using deep learning with heterogeneous multi-location data: dataset fusion versus transfer learning,” *Applied Energy*, vol. 369, 2024, doi: 10.1016/j.apenergy.2024.123467.
- [28] S. Conti, “Artificial intelligence for weather forecasting,” *Nature Reviews Electrical Engineering*, vol. 1, no. 1, Jan. 2024, doi: 10.1038/s44287-023-00009-2.




- [29] Q. Paletta, G. Arbod, and J. Lasenby, "Omnivision forecasting: combining satellite and sky images for improved deterministic and probabilistic intra-hour solar energy predictions," *Applied Energy*, vol. 336, Apr. 2023, doi: 10.1016/j.apenergy.2023.120818.
- [30] G. Ruan *et al.*, "SkyNet: a deep learning architecture for intra-hour multimodal solar forecasting with ground-based sky images," *Renewable Energy*, vol. 256, 2026, doi: 10.1016/j.renene.2025.124354.
- [31] K. Sudharshan, C. Naveen, P. Vishnuram, D. V. S. K. R. Kasagani, and B. Nastasi, "Systematic review on impact of different irradiance forecasting techniques for solar energy prediction," *Energies*, vol. 15, no. 17, 2022, doi: 10.3390/en15176267.
- [32] N. Rane, S. P. Choudhary, and J. Rane, "Ensemble deep learning and machine learning: applications, opportunities, challenges, and future directions," *Studies in Medical and Health Sciences*, vol. 1, no. 2, pp. 18–41, 2024, doi: 10.48185/smhs.v1i2.1225.
- [33] C. Yan *et al.*, "Exploring the key meteorological drivers of air pollution by intrinsically interpretable deep learning," *Environmental Modelling and Software*, vol. 192, 2025, doi: 10.1016/j.envsoft.2025.106571.
- [34] B. Teixeira, L. Carvalhais, T. Pinto, and Z. Vale, "Explainable AI framework for reliable and transparent automated energy management in buildings," *Energy and Buildings*, vol. 347, 2025, doi: 10.1016/j.enbuild.2025.116246.
- [35] S. K. Arumugam, B. F. Balogun, S. Nalini, and A. K. Tyagi, "Quantum computing integration: cutting-edge and quantum computing for building a smart and sustainable environment," *Quantum Computing: the Future of Information Processing*, pp. 233–250, 2025, doi: 10.1201/9781003499459-14.

## BIOGRAPHIES OF AUTHORS






**Dr. Komal Kumar Napa**     is currently working as assistant professor in the Department of Artificial Intelligence and Data Science at Saveetha Engineering College, Chennai, India. His research interests include machine learning, data mining, and cloud computing. He can be contacted at email: komalkumarnapa@gmail.com.







**Mr. Rajkumar Govindarajan**    is currently working as assistant professor in the Department of Computer Science and Engineering (Data Science) at Madanapalle Institute of Technology and Science (MITS), Deemed to be University, Madanapalle, Andhra Pradesh. His research interests include machine learning, and data mining. He can be contacted at email: kumar3544@gmail.com.



**Prof. Dr. J. Senthil Murugan**    is currently working as professor in the Department of Computer Science and Engineering at Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai India. His research interests are machine learning, and image processing. He can be contacted at email: jsenthilmuruganmtech@gmail.com.



**Mr. Billa Manindhar**     is currently working as assistant professor in the Department of Computer Science and Engineering at Koneru Lakshmaiah University, Vaddeswaram, Guntur (dist), Andhra Pradesh, India. His research interests include machine learning, data mining, and cloud computing. He can be contacted at email: bmanindhar@kluniversity.in.