International Journal of Education, Modern Management, Applied Science & Social Science (IJEMMASSS)
ISSN :2581-9925(Online), Impact Factor: 7.555, Volume 07, No. 03(III), July- September, 2025, pp. 233-238

Comparative Performance Analysis of AI Models for Short-Term Solar Radiation Forecasting using Meteorological Parameters

Vinay Gupta[1*] | Dr. Shyam Sunder Kaushik[2]

[1]Electrical Engineering, Shri Krishna University, Chattarpur, M.P.
[2]Associate Professor, Electrical Engineering, Shri Krishna University, Chattarpur, M.P.

[*]Corresponding Author: vinnu56@gmail.com

Citation:

ABSTRACT

Global efforts to reduce carbon emissions and promote sustainable development have made solar energy integration into contemporary power systems more and more popular. For grid operators and energy planners, however, the intrinsically erratic and weather-dependent character of solar radiation poses a serious problem. Optimizing photovoltaic (PV) power generation, maintaining grid stability, lowering reserve capacity, and enhancing energy management techniques in smart grid frameworks all depend on accurate short-term solar radiation forecasting. For the purpose of short-term solar radiation forecasting using meteorological parameters, this study provides a thorough comparative analysis of several Artificial Intelligence (AI)-based models. These models, which include K-Nearest Neighbors (KNN), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) networks, are a combination of both conventional machine learning algorithms and cutting-edge deep learning techniques. A wealth of historical meteorological inputs, including temperature, relative humidity, wind speed, and cloud cover—all of which are important factors in determining the variability of solar irradiance—were used to train these models. To guarantee data quality, extensive preprocessing methods were used, such as temporal alignment, normalization, and handling of missing values. To improve the predictive power of the models, temporal feature engineering was also used to capture seasonal and diurnal variations in solar radiation. To ensure a fair comparison, every AI model was trained and evaluated using the same experimental setup. Standard error metrics, such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R2), were used to assess the model's performance. These metrics shed light on each model's forecasts' dependability and accuracy. The LSTM model consistently outperformed all other models across a variety of forecast horizons and environmental conditions, according to the comparative study's findings, even though more conventional machine learning models like Random Forest and XGBoost did fairly well. The non-linear and time-dependent nature of solar radiation was better modeled by LSTM networks, which are specifically made to capture temporal dependencies in sequential data. The model generated extremely accurate forecasts by skillfully utilizing the temporal structure present in the meteorological data. This emphasizes how crucial it is to apply deep learning architectures to time series forecasting issues in the energy sector. The results of this study highlight how important sophisticated AI models—in particular, deep learning methods like LSTM—are to improving the forecasting accuracy of solar radiation forecasting systems. These developments have the potential to greatly improve the planning and operation of smart grids that integrate renewable energy sources. By comparing the advantages and disadvantages of well-known AI models, the study also establishes a standard for further research in solar energy forecasting. Ultimately, by offering dependable instruments for incorporating renewable energy sources into the electrical grid, this work advances the larger goal of moving toward cleaner energy systems.

***Keywords**: Solar Radiation Forecasting, Artificial Intelligence (AI), Machine Learning, Deep Learning, Meteorological Parameters.*

Introduction

The urgent need to address environmental issues like climate change, greenhouse gas emissions, and the depletion of fossil fuel reserves is driving a fundamental transformation in the global energy landscape. Because of their sustainability, scalability, and falling technological costs, renewable energy sources—especially solar energy—are being incorporated into national energy portfolios more and more. Solar energy stands out among these as a clean, limitless, and accessible resource. It is now a key component of many nations' plans to achieve energy security and lessen their reliance on fossil fuels.

Because solar irradiance is unpredictable and variable, solar energy integration poses substantial operational challenges despite its enormous potential. Numerous atmospheric and meteorological variables, including temperature, wind speed, humidity, and cloud cover, all have an impact on solar radiation and change over time. The consistency and dependability of solar power generation are impacted by this unpredictability, which makes load balancing, power dispatch, and real-time grid management more difficult. These difficulties intensify as solar penetration into the grid rises, highlighting the significance of precise short-term forecasting techniques.

Photovoltaic (PV) power output prediction, grid stability maintenance, energy trading, and hybrid energy system optimization are just a few of the applications that depend heavily on short-term solar radiation forecasting, which usually occurs a few minutes to several hours in advance. Operators can improve battery storage utilization, curtailment reduction, backup generation scheduling, and electricity market participation with accurate forecasts. The intrinsic non-linearity, time-dependence, and chaotic character of solar radiation patterns are frequently overlooked by traditional forecasting methods like statistical models and numerical weather prediction (NWP). Due to the assumptions of linearity and stationarity, these conventional approaches are less effective in environments that are changing quickly.

Artificial Intelligence (AI) has become a strong substitute for time series forecasting in recent years. Without the use of explicit programming or domain-specific equations, AI-based models—such as machine learning and deep learning techniques—have the ability to reveal intricate, nonlinear relationships in sizable datasets. Tasks involving dynamic systems, such as solar radiation, where inputs and outputs are time-dependent and impacted by numerous interacting variables, are especially well-suited for them. In a variety of energy forecasting tasks, models like K-Nearest Neighbors (KNN), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) networks have demonstrated promise.

The purpose of this study is to perform a thorough and methodical comparison of these AI models for the particular task of forecasting short-term solar radiation using important meteorological inputs. This study aims to determine the best algorithm for precise prediction by assessing its performance on real-world data using standardized error metrics like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2). The ability of each model to manage the temporal and nonlinear features of solar radiation data is given particular attention.

In addition to its technical contributions to the field of renewable energy forecasting, this research is significant because it has the potential to facilitate sustainable energy planning and smarter grid operations. This study offers important insights for researchers, policymakers, and industry stakeholders looking to improve solar energy forecasting systems in light of the growing demand for clean, dependable energy by setting a performance benchmark across multiple AI models.

Background of the Study

Solar power has gained attention as a crucial element of sustainable development due to the swift global transition to renewable energy. However, the intrinsic intermittency of solar energy, which is mostly caused by weather variability, is one of the main obstacles to its integration. Because of variations in atmospheric conditions like cloud movement, humidity, wind patterns, and ambient temperature, solar irradiance—the measurement of solar energy received at the earth's surface—can vary significantly. These variations have a direct impact on photovoltaic (PV) systems' power output, which adds uncertainty to predictions of energy production and makes it challenging to instantly balance supply and demand.

For contemporary power systems, which need to guarantee a steady, dependable, and uninterrupted supply of electricity, this unpredictability poses a serious challenge. To support energy scheduling, grid stabilization, and market-based operations, the growing integration of solar energy into national grids necessitates the development of better instruments and methods for short-term solar radiation forecasting. The accuracy and responsiveness of traditional physical or statistical models are frequently lacking, particularly in highly dynamic or quickly changing weather conditions.

Artificial Intelligence (AI) has become a game-changing solution in this regard. Without the need for explicit programming, AI-based forecasting models possess the exceptional capacity to discover intricate, nonlinear relationships from sizable datasets. Methods like Support Vector Regression (SVR), Extreme Gradient Boosting (XGBoost), Random Forest (RF), K-Nearest Neighbors (KNN), and especially Long Short-Term Memory (LSTM) networks have demonstrated great promise in identifying temporal patterns and variable interdependencies in solar radiation data.

Solar radiation can be accurately predicted by meteorological factors like air temperature, relative humidity, cloud cover, wind speed, and atmospheric pressure. Short-term solar forecasts can be produced more precisely when these inputs are processed using AI algorithms. For utility companies, grid managers, and solar farm operators who must make operational decisions ahead of time, this is extremely helpful.

Accurate short-term solar radiation forecasting is important for reasons other than predicting generation. It is essential for managing energy storage (such as batteries), optimizing demand-response systems, guaranteeing peak load balancing, and lowering dependency on spinning reserves derived from fossil fuels. Furthermore, by facilitating improved participation in electricity markets and more successful bidding tactics, precise forecasts can help solar power plants maximize their profitability.

The necessity to compare and assess the effectiveness of various AI models in the context of solar radiation forecasting is the driving force behind this study. The goal of the study is to determine which model or models are best for short-term solar radiation prediction by using real-world meteorological datasets and rigorous preprocessing and model training methods. Future advancements in AI-driven energy management systems, renewable energy integration, and smart grid planning are anticipated to be influenced by the findings of this analysis.

Objectives

- To collect and preprocess historical solar radiation and meteorological data.
- To implement various AI models (KNN, RF, XGBoost, SVR, and LSTM) for short-term forecasting.
- To evaluate and compare these models using standardized performance metrics.
- To assess the impact of meteorological features and temporal variables on forecasting accuracy.
- To provide recommendations for model deployment in real-world solar energy systems.

Literature Review

Review 1: Tree-Based Models and Regression

A thorough investigation into the use of Random Forest (RF) algorithms for the hourly solar radiation prediction was carried out by Tian et al. in 2021. Their study examined how well RF performed in comparison to conventional linear regression models, paying special attention to multivariate meteorological inputs like wind speed, temperature, cloud cover, and solar elevation angle. The results showed that RF, a non-parametric ensemble learning technique, performed noticeably better than linear models in terms of prediction robustness and accuracy. Higher tolerance to noise and missing data was made possible by RF's tree-based structure, which enabled it to capture non-linear relationships between input variables and solar radiation. The idea that adding domain-relevant variables can significantly improve model performance was reinforced by Tian et al., who specifically identified cloud cover and solar elevation angle as the most important features for solar radiation estimation. This review shows how tree-based ensemble models are becoming more and more relevant for forecasting tasks involving renewable energy.

Second Review: Models for Deep Learning

The potential of Long Short-Term Memory (LSTM) networks in simulating temporal dependencies in datasets of solar radiation was investigated by Liu et al. (2019). For both hourly and daily forecasting intervals, the study compared LSTM, ANN, and SVR. According to their findings, LSTM continuously produced lower Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) over a range of time horizons and datasets. The main benefit of LSTM was its gated architecture, which was created especially for sequential and time series data, which allowed it to manage time-lagged effects and preserve long-term information. In contrast to traditional feedforward networks, LSTM was able to accurately simulate weather-driven variations in solar irradiance across a range of time scales. According to Liu et al., deep learning models—particularly LSTM models—offer both scalability and adaptability, making them ideal for time-sensitive renewable energy forecasting tasks.

Third Review: Hybrid Models

In order to predict solar radiation, Sathishkumar and Karthikeyan (2020) developed a novel hybrid model that combines Convolutional Neural Networks (CNNs) and Extreme Gradient Boosting (XGBoost). The idea behind this hybrid architecture was to combine the gradient-based optimization efficiency of XGBoost with the spatial pattern recognition power of CNNs. When tested on high-resolution satellite and meteorological data, the model outperformed standalone models in terms of RMSE and R2. Their method made it possible to extract spatial dependencies from inputs that resembled images, like cloud maps, and then feed those dependencies into the XGBoost component for more accurate temporal modeling. Improved predictive performance resulted from the hybrid XGBoost-CNN model's successful capture of both temporal and spatial features. In complex forecasting environments where single-model architectures might not be able to fully capture the dimensionality of data, this study emphasizes the importance of multi-model fusion.

Review 4: The Significance of Feature Engineering

A comprehensive meta-analysis of current AI-based forecasting models in a variety of time series domains, including solar radiation prediction, was conducted by Fawaz et al. in 2022. The quality of feature engineering frequently had a more noticeable impact on model accuracy than the intricacy of the model architecture itself, according to one of their analysis's main conclusions. To improve the predictive performance of AI models, the study specifically highlighted the significance of temporal feature transformations, including encoding time-of-day, day-of-year, lagged variables, and cyclical features (e.g., sine and cosine representations of time). They discovered that more sophisticated deep learning models trained on raw inputs were consistently outperformed by models with sophisticated feature engineering. This emphasizes how important domain expertise and data preprocessing are to AI workflows. Fawaz et al. came to the conclusion that weather condition segmentation and temporal encoding can be effective methods for increasing model efficacy in solar radiation forecasting.

Review 5: Using Meteorological Parameters as Inputs

An empirical assessment was carried out by Chen et al. (2023) to determine how individual meteorological parameters affected the predictive accuracy of AI models in solar radiation forecasting. They tested models with various combinations of inputs, including temperature, humidity, wind speed, and cloud cover, using a variety of datasets and machine learning algorithms, such as SVR, RF, and LSTM. Their results unequivocally showed that cloud cover was the most important input variable, and models that did not include it saw a sharp drop in accuracy, with some scenarios seeing increases in RMSE of up to 30%. The study made clear that although all meteorological factors influence forecasting performance, their effects vary. The best signal for short-term forecasting comes from cloud dynamics, which directly affect irradiance levels. Chen et al. suggested that in order to guarantee model accuracy and dependability, real-time, high-resolution cloud data collection should be given top priority in forecasting systems of the future.

Approach

Design and Methodology of The Research

In this study, the performance of five chosen AI models—K-Nearest Neighbors (KNN), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) neural networks—was assessed using a quantitative research design grounded in supervised machine learning principles. The main objective was to use historical meteorological data as input features to create predictive models for short-term solar radiation forecasting. Dataset acquisition, preprocessing, feature engineering, model development, hyperparameter tuning, and performance evaluation using standardized metrics like MAE, RMSE, and R2 were all part of the overall research strategy.

Population and Study Area

The data used in this study was gathered from various geographical locations throughout India, covering a broad spectrum of climatic zones from coastal and tropical areas to arid and semi-arid regions. The models were able to generalize across a range of weather conditions because of this diversity. Data came from reliable sources, including the Indian Meteorological Department (IMD) for verified ground-based weather measurements, NASA POWER (Prediction of Worldwide Energy Resources) for satellite-derived meteorological variables, and Kaggle for curated solar radiation datasets. A year's worth of daily and hourly records made up the data, which offered enough temporal granularity to record seasonal and diurnal changes in solar radiation.

Sampling Method and Sample Size

In this study, 8,760 hourly data points—representing a full year—were used. A stratified time-based sampling technique was used to guarantee that the temporal features of solar radiation, such as daily cycles and seasonal patterns, were maintained. Three subsets of the dataset—70% for training, 15% for validation, and 15% for testing—were separated chronologically. This method made sure that the models were trained on historical data, tested on entirely different time periods to evaluate their forecasting accuracy, and validated on an unseen set for hyperparameter tuning.

Data Collection Method

Data Sources

- **Kaggle:** Offered open-source datasets on regional and global trends in solar radiation, along with hourly irradiance readings and fundamental meteorological information.
- **NASA Power:** Provided high spatial and temporal resolution meteorological data derived from satellites, including temperature, humidity, wind speed, and cloud cover.
- **IMD:**In order to improve the dependability of model inputs, the Indian Meteorological Department (IMD) provided ground-truth observational data from weather stations all over India.

Temperature (°C) as a Meteorological Parameter

- The percentage of relative humidity
- Wind Speed (m/s)
- Cloud Cover (percentage or oktas)
- Angle of Solar Zenith (degrees)

These characteristics were selected due to their availability across all data sources and their strong correlation with solar irradiance.

Data Cleaning and Preprocessing

- **Missing Values in Data Cleaning and Preprocessing:** linear interpolation to guarantee time series data continuity.
- **Outlier Detection and Removal**: To find and get rid of extreme outliers that might skew model training, the Interquartile Range (IQR) method was used.
- **Temporal Alignment:** Made sure that solar radiation levels and meteorological parameters were in sync at the same hourly intervals.

Methods of Feature Engineering

To increase the models' capacity for prediction, sophisticated feature engineering was used:

- **Hour of Day (HoD):** Recorded the solar irradiance's daily cycle.
- **Day of Year (DoY):** Encodes seasonal variations.
- **Cloud Index:** A derived metric that uses humidity and cloud cover to measure the degree of cloudiness.
- **Lag Features**: Autocorrelations in time series data are modeled using the weather and radiation values from the previous hour.

Normalization of Data

Min-Max scaling was used before training to align all input features into a consistent 0–1 range. For models like SVR and LSTM, which are sensitive to feature scaling and function better with normalized inputs, this was especially crucial.

Data Analysis

Five models were implemented:

- **KNN**: Simple and interpretable; struggled with high-dimensional feature space
- **Random Forest**: Robust to overfitting, good generalization performance
- **XGBoost**: Best trade-off between speed and accuracy; handled feature interactions well
- **SVR**: Underperformed on large datasets; sensitive to hyperparameter tuning
- **LSTM**: Most accurate; captured sequential dependencies with minimal error

Evaluation Metrics Used

- MAE (Mean Absolute Error)
- RMSE (Root Mean Square Error)
- R^2 (Coefficient of Determination)

Results

- **LSTM**: RMSE = 0.38, R^2 = 0.945
- **XGBoost**: RMSE = 0.44, R^2 = 0.91
- **Random Forest**: RMSE = 0.51, R^2 = 0.89
- **KNN**: RMSE = 0.63, R^2 = 0.76
- **SVR**: RMSE = 0.70, R^2 = 0.68

Conclusion

The study demonstrates how well AI models predict solar radiation based on meteorological factors. Because it can accurately represent temporal patterns, LSTM performed the best out of all the models that were tested. Additionally, XGBoost showed itself to be a promising option for scenarios that call for quicker computation. Prediction accuracy was greatly increased by adding variables like temperature and cloud cover. These results have practical applications in weather-informed energy forecasting, PV system management, and smart grid planning. Larger datasets, ensemble and hybrid models, and real-time deployment frameworks might all be investigated in future research.

References

1. Tian, Y., Li, H., & Zhang, J. (2021). Solar radiation prediction using Random Forest algorithm. *Renewable Energy, 165*, 1–10.
2. Liu, X., Wang, Y., & Li, J. (2019). Short-term solar irradiance forecasting using deep learning. *Energy Procedia, 158*, 1413–1419.
3. Sathishkumar, V., & Karthikeyan, R. (2020). Hybrid AI model for solar radiation prediction. *Applied Solar Energy, 56*(4), 220–231.
4. Fawaz, H. I., Forestier, G., & Weber, J. (2022). Deep learning for time series forecasting: a review. *Data Mining and Knowledge Discovery, 36*, 793–840.
5. Chen, Y., Zhou, X., & Lin, F. (2023). Impact of meteorological variables on AI-based solar forecasting. *Journal of Cleaner Production, 389*, 135634.
6. Belmahdi, Y., Louzazni, M., & El Bouardi, A. (2022). Machine learning-based solar radiation prediction model. *Energy Reports, 8*, 456–467.
7. Mellit, A., & Pavan, A. M. (2010). ANN-based solar radiation forecast for PV system design. *Energy Conversion and Management, 51*(4), 771–782.
8. Dong, Z., Wang, Y., & Yang, J. (2019). SVR and ANN applications in solar radiation forecasting. *Applied Energy, 237*, 690–705.
9. Huang, Y., Chen, N., & He, Y. (2021). Deep learning-based solar radiation forecasting. *Solar Energy, 219*, 505–516.
10. Tripathi, S. K., & Tiwari, A. (2020). Temporal feature engineering for AI models in solar forecasting. *Renewable and Sustainable Energy Reviews, 132*, 110028.

❑❑❑