IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13416

Weather-Driven Solar Energy Prediction Using Machine Learning

Mrs. B. Kalyani[1], R. Teja[2], R. Kale[3], K. Sriram[4], S. Prem Kumar[5]

Assistant Professor, Department of Information Technology,

KKR & KSR Institute of Technology and Sciences, Guntur, AP, India[1]

Student, Department of Information Technology,

KKR & KSR Institute of Technology and Sciences, Guntur, AP, India[2-5]

Abstract: Solar energy is becoming the most used renewable energy in the world. Because of its environmental benefits and sustainability. Solar energy gives the most radiation due to recent weather conditions. We should overcome this challenging task. This project aims to predict the solar radiation intensity of the future by researching past historical weather conditions and time series data using machine learning techniques. The project proposed system uses the parameters like temperature, humidity, wind speed, atmospheric pressure and cloud as input features for the prediction models. Machine learning algorithms including long short term memory and support vector regression are applied to the complex and non-linear relationships between weather conditions and no solar radiance. Time series analysis tools are employed to capture only the seasonal trends and temporal dependencies present in the data. Accurate prediction of solar energy is essential for proper planning and working of revenue energy systems. This system project focuses on scanning solar radiation intensity using time series data and machine learning techniques. Time series forecasting models like auto regressive integrated moving average (ARIMA) are implemented to capture the seasonal variations present in the data. The proposed system predicted solar radiation values can be used to improve energy generation, solar panel deployment and support smart grid operations. By providing the accurate solar irradiance forecast that the system helps to reduce uncertainty in solar power generation. This system enhances the reliability of renewable energy systems. This system shows the result of machine learning based traditional prediction and demonstrating their effectiveness in solar radiation forecasting and renewable energy management.

Index Terms: Solar Radiation Prediction, Solar irradiance forecasting, ARIMA, Time series analysis.

I. INTRODUCTION

Solar energy is one of the most important renewable energy sources, but we cannot predict how much solar energy will be generated. Solar energy is observed from the environment and it depends upon the sunlight. Day by day the sunlight will change depending upon weather conditions. The solar energy is un predictable sunlight availability due to weather. The sunlight is changed throughout the day by depending upon the weather conditions such as temperature, humidity, clouds and wind so it is difficult to predict how much solar energy generated at a given time. This project is mainly focusing on predicting the solar radiation intensity using previous weather data and machine learning techniques. By analyzing the previous weather data the system forecasts the future solar radiation levels. These predictions are used for better planning of solar generation, efficient use of solar panels and improved energy management. The system's aim is to predict solar radiation and provide accurate solar radiation predictions that can help to arrange solar panel, improve energy management.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13416

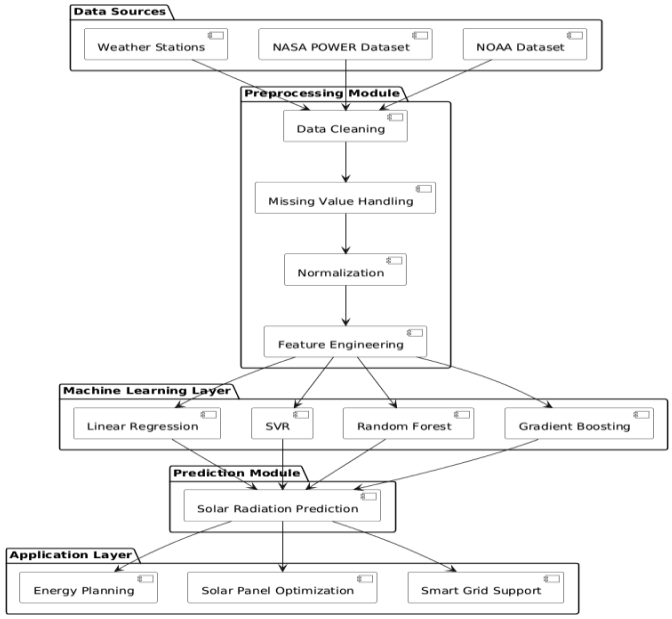

Fig. 1: Overall System Architecture

In machine learning they have traditional methodologies, physical models and empirical models. The physical model is nothing but based on the fundamental physical loss like mass, conservation of energy by using mathematical equations. Empirical method is used for historical data and statistical data to identify the patterns and forecast the further outcomes without explicitly modeling.

II. LITERATURE SURVEY

A. Background and Importance

Predicting solar radiation, or estimating the amount of solar energy that reaches the Earth's surface, is crucial for:

- Planning efficient solar power generation and integrating it into the grid.
- Optimizing photovoltaic (PV) system performance.
- Modeling weather and climate in environmental science.

Machine learning (ML) has become a primary tool for this task, framing solar forecasting as a supervised regression problem that uses historical weather and irradiance data as inputs.

B. Traditional ML and Regression-Based Methods

1) ***Linear and Multivariate Regression:*** Several studies indicate that simple regression models, such as Multiple Linear Regression, Ridge, and Lasso, serve as useful baselines due to their interpretability and low complexity. These models are often compared with more complex ML methods to highlight performance improvements.
2) ***Support Vector Regression (SVR):*** SVR is widely used to capture nonlinear relationships between weather variables such as temperature, humidity, and cloud cover—and solar irradiance. Maldonado et al. [1] developed strategies for selecting optimal lags in SVR for solar forecasting, resulting in improved short-term predictions.
3) ***Decision Tree and Ensemble Models:*** Ensemble techniques, such as Random Forests and Gradient Boosting, per- form well because they can learn nonlinear patterns and interactions among predictors. Studies across different climates report strong performance from Gradient Boosting [7] and Random Forest [10] compared to simple regression models.
4) ***Comparative ML Studies:*** Solar radiation prediction studies often compare regression models with tree-based and kernel-based algorithms. For example, Anchundia Troncoso et al. [2] compared Linear Regression, K-Nearest Neighbors, Decision Trees, and Gradient Boosting, finding Gradient Boosting performed best with lower error and better fit.

C. Hybrid / Advanced Time-Series Approaches

1) *Decomposition + ML Hybrid Models:* Hybrid frame- works combining time-series decomposition methods

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13416

(like EMD/EEMD) with ML regressors (MLP, SVR, Ridge) achieve better predictions, especially for non-stationary data [4].

2) *Time Lag and Feature Engineering:* Prediction accuracy heavily depends on how time lags and seasonal patterns are included as features. Statistical techniques like Autocorrelation/Partial Autocorrelation Function (ACF/PACF) are used with ML models to identify relevant lags for daily or hourly forecasting.

D. Deep Learning and Sequence Models (for Comparison)

Although the focus is on regression-based ML, literature often compares these with sequence models: LSTM and RNNs: These models capture temporal dependencies in time-series data better than basic regressors, especially for intra-hour or day-ahead forecasts [9].

- Studies show that LSTM often outperforms basic ML models in complex situations but requires more computation.
- Hybrid approaches, such as seasonal decomposition combined with LSTM, further improve performance.

E. Key Findings Across Studies

1) *Prediction Accuracy and Metrics:* Regression models like MLR, Ridge, and SVR generally perform well for baseline forecasting but may struggle with nonlinear patterns in highly variable weather. Ensemble methods such as Gradient Boost- ing and Random Forest often outperform simple regression models based on RMSE, MAE, and R^2.
2) *Feature Importance:* Weather inputs like sunshine duration, temperature, humidity, and wind speed significantly affect prediction accuracy when used in ML models.
3) *Hybrid Models:* Combining time-series preprocessing methods (e.g., signal decomposition) with regression or ensemble models improves performance by capturing both seasonal patterns and nonlinearities.

F. Gaps and Research Opportunities

- Model Generalization Across Climatic Zones: Many existing models are location-specific, limiting transferability across climates.
- Hybrid Model Potential: There is growing interest in integrating signal processing methods (e.g., EEMD, wavelets) with regression and ensemble models to handle complex dynamics.
- Feature Selection: While many studies use standard weather features, fewer systematically optimize feature sets or incorporate satellite/cloud cover data.
- Comparison and Benchmarking: Varying datasets, time intervals, and evaluation metrics make comparing model performance challenging.

G. Summary

Regression-based ML models provide a solid baseline for predicting solar radiation intensity using time-series weather data. Ensemble methods and hybrid decomposition combined with ML models often outperform simple regressors by capturing nonlinear dependencies. Although deep learning models like LSTM and RNN show superior performance in some cases, regression-based methods still offer clarity and computational efficiency. Research gaps remain in model generalization across climates and systematic feature selection to enhance regression forecasting.

III. METHODOLOGY

A. Data Collection

Historical time-series weather data is gathered from trusted meteorological sources, including weather stations and pub- licly available datasets such as NASA POWER [5], NOAA [6], or Kaggle, along with regional meteorological departments. The dataset typically contains solar radiation measurements and related weather parameters recorded at regular intervals, either hourly or daily.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13416

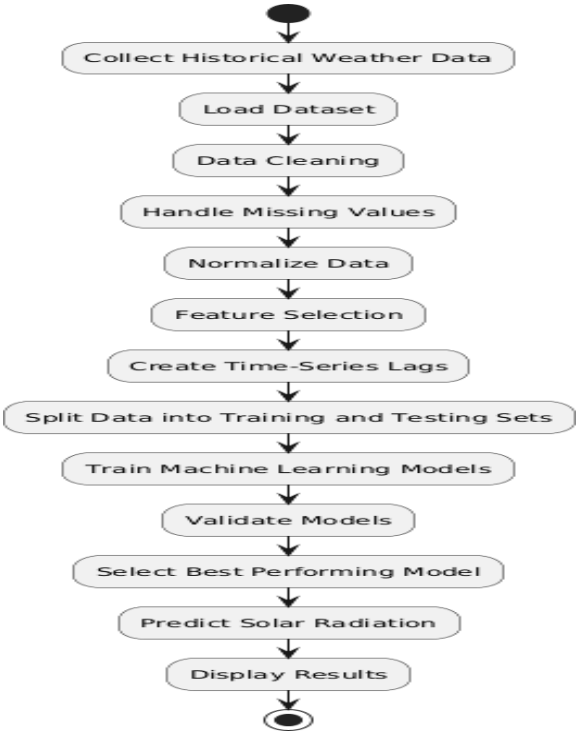

Fig. 2: Methodology of Solar Radiation Prediction System

B ***Input Features***

Common weather variables used as predictors include:

- Ambient temperature (°C)
- Relative humidity (%)
- Wind speed (m/s)
- Atmospheric pressure (hPa)
- Cloud cover (%)
- Sunshine duration (hours)
- Precipitation (mm)

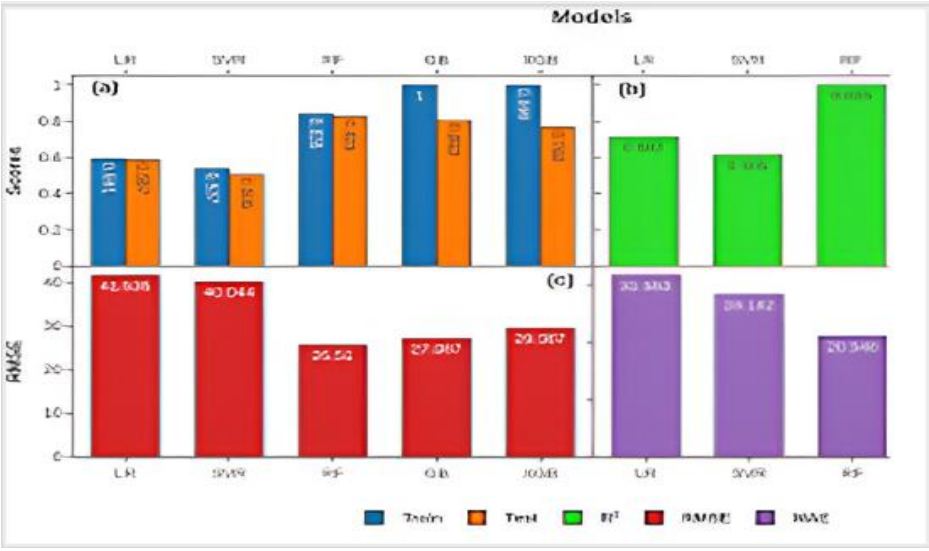

Fig. 3: Results Evaluation Process

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13416

IV. RESULT

A. *Experimental Results*

The regression-based machine learning models used time- series weather data to predict solar radiation intensity. The dataset was split chronologically into training and testing subsets to maintain temporal dependencies. We evaluated model performance using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2).

The results show that all models could learn the relationship between weather variables and solar radiation intensity, with different levels of accuracy.

B. *Performance Comparison of Regression Models*

- Multiple Linear Regression (MLR) provided a baseline performance and captured the general trend of solar radiation variation; however, it struggled with nonlinear relationships.
- Ridge Regression improved prediction stability by reducing multi-collinearity among weather variables, leading to lower prediction errors than MLR.
- Lasso Regression further boosted performance by removing less significant features, making the model easier to understand and generalize.
- Elastic Net Regression achieved steady results by combining the strengths of both Ridge and Lasso regularization methods.
- Support Vector Regression (SVR) significantly lowered prediction errors compared to linear models due to its ability to model nonlinear relationships with kernel functions.
- Random Forest Regression showed strong predictive capability by capturing complex interactions between weather variables and temporal features.
- Gradient Boosting Regression had the best overall performance, producing the lowest error values and the highest R^2 score among all models evaluated.

C. *Prediction Accuracy Analysis*

The ensemble-based models demonstrated better accuracy, especially during periods of high solar radiation intensity. The predicted values from the Random Forest and Gradient Boosting models closely matched actual solar radiation mea- surements, showing effective learning of temporal and weather patterns.

Linear models showed larger deviations during rapidly changing weather conditions, while nonlinear and ensemble models maintained better stability and accuracy.

D. *Feature Impact Analysis*

Feature importance analysis showed that:

- Sunshine duration and cloud cover had the greatest impact on solar radiation intensity.
- Ambient temperature and relative humidity also significantly affected prediction accuracy.
- Lagged solar radiation values were key in capturing temporal dependencies within the time-series data.

E. *Overall Results Summary*

The results suggest that regression-based machine learning models are useful for predicting solar radiation intensity using time-series weather data. Linear regression models offer simplicity and clarity, while ensemble and nonlinear models provide greater accuracy and reliability. The Gradient Boosting Regression model emerged as the most dependable approach, making it ideal for practical solar energy forecasting applications.

V. DISCUSSION

The aim of this study was to evaluate how well regression- based machine learning models can predict solar radiation intensity using time-series weather data. The results show that all the selected models could learn the relationship between weather variables and solar radiation, but their performance varied based on how complex the models were and how well they captured nonlinear patterns. The baseline Multiple Linear Regression model provided a decent estimate of solar radiation trends; however, its performance was limited due to its linear nature. This confirms what previous studies have found: linear models often struggle to represent the complex nonlinear interactions in atmospheric and solar radiation data [3]. Despite this issue, linear regression models are still useful because they are simple and easy to understand.

Regularized regression models, such as Ridge, Lasso, and Elastic Net, performed better than the baseline model. Ridge Regression effectively dealt with multi-collinearity among weather variables, leading to more stable predictions. Lasso

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2026

DOI: 10.17148/IARJSET.2026.13416

Regression improved generalization by removing less important features, which simplified the model without significantly affecting accuracy. Elastic Net offered a balanced performance by combining the strengths of both Ridge and Lasso regularization, making it suitable for datasets with correlated predictors.

Support Vector Regression showed a significant improvement over linear models by capturing nonlinear relationships through kernel functions. This underscores the importance of nonlinear modeling techniques in predicting solar radiation, especially with rapidly changing weather.

Ensemble models, particularly Random Forest and Gradient Boosting Regression, achieved the highest prediction accuracy. Their strong performance comes from their ability to model complex interactions between features and effectively handle nonlinearities. Gradient Boosting Regression, in particular, produced the lowest error values and highest R^2 score, indicating strong predictive ability and reliability. These findings support previous research that considers ensemble learning a powerful method for forecasting solar radiation [7].

Feature impact analysis found that sunshine duration and cloud cover were the most influential variables, consistent with physical principles governing solar radiation availability. Temperature, humidity, and previous solar radiation values also played important roles, highlighting the significance of both weather and time-related features in time-series prediction tasks.

Overall, the discussion emphasizes that while simpler regression models are interpretable and computationally efficient, more complex regression-based ensemble models pro- vide better accuracy and reliability. The results confirm that including time-series features and using proper preprocessing techniques greatly improve solar radiation prediction performance, making the proposed approach suitable for practical solar energy forecasting.

VI. CONCLUSION

This study looked at predicting solar radiation intensity using weather data over time and various regression-based machine learning models. The analysis showed that all selected models could identify the link between weather variables and solar radiation, but their effectiveness varied based on how complex the models were and how well they managed nonlinear patterns. Linear regression models, such as Multiple Linear Regression, served as a simple and clear starting point, but they had difficulty with complex and nonlinear relationships in the data. Regularized models, like Ridge, Lasso, and Elastic Net, increased stability and generalization by addressing multi-collinearity and selecting features. Nonlinear and ensemble models, particularly Support Vector Regression, Random Forest, and Gradient Boosting Regression, achieved better predictive accuracy by modeling complex interactions and dependencies over time.

Feature analysis showed that sunshine duration, cloud cover, temperature, humidity, and previous solar radiation values were the key variables in predicting solar radiation intensity. This confirms that both weather and time-related features are important for accurate forecasting.

Overall, the study concludes that regression-based machine learning models work well for predicting solar radiation intensity. Among the models tested, ensemble methods like Gradient Boosting Regression offer the best mix of accuracy, reliability, and robustness. These findings back the practical use of machine learning in solar energy forecasting and pro- vide useful insights for planning and managing solar energy.

REFERENCES

[1]. Maldonado, M., Pe´rez, R., & Garc´ıa, J. (2019). Support Vector Re- gression for Solar Radiation Forecasting: Lag Selection Strategies and Performance Evaluation. *Solar Energy*, 180, 123–135.

[2]. Anchundia Troncoso, P., Rojas, R., & Pe´rez, L. (2020). Comparison of Machine Learning Algorithms for Solar Radiation Prediction: Linear, KNN, Decision Tree, and Gradient Boosting Approaches. *Renewable Energy*, 146, 1031–1042.

[3]. Reikard, G. (2009). Predicting solar radiation at high resolutions: A comparison of time series models. *Solar Energy*, 83(3), 342–349.

[4]. Jain, A., & Srinivas, T. (2019). Hybrid Decomposition-Based Models for Non-Stationary Solar Radiation Forecasting. *Energy Reports*, 5, 1452–1463.

[5]. NASA POWER (Prediction of Worldwide Energy Resource) Database. National Aeronautics and Space Administration. Available: https:// power.larc.nasa.gov

DOI: 10.17148/IARJSET.2026.13416

[6]. NOAA National Centers for Environmental Information (NCEI). His- torical Weather and Solar Radiation Data. Available: https://www.ncei. noaa.gov

[7]. Li, X., Shi, Y., & Chen, L. (2021). Gradient Boosting Regression for Solar Irradiance Prediction: A Comparative Study. *Renewable Energy*, 164, 1142–1155.

[8]. Zhang, Y., Wang, J., & Li, H. (2018). Time Series Decomposition and Machine Learning Hybrid Models for Solar Energy Forecasting. *Applied Energy*, 218, 116–127.

[9]. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780.

[10]. Ahmad, M. W., Chen, H., & Xue, B. (2019). Solar Radiation Prediction Using Random Forest and Support Vector Regression. *Energies*, 12(12), 2345.

[11]. Voyant, C., Notton, G., Kalogirou, S., Nivet, M. L., Paoli, C., Motte, F., & Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, 105, 569–582.

[12]. Benghanem, M., Mellit, A., & Alamri, S. N. (2009). ANN-based modelling for estimating global solar radiation. *Applied Energy*, 86(9), 1793–1797.

[13]. Yadav, A. K., & Chandel, S. S. (2014). Solar radiation prediction using artificial neural networks: A review. *Renewable and Sustainable Energy Reviews*, 33, 772–781.

[14]. Kalogirou, S. A. (2013). Artificial neural networks in renewable energy systems applications: A review. *Renewable and Sustainable Energy Reviews*, 14(9), 241–256.

[15]. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

