Archives available at journals.mriindia.com

International Journal of Advanced Electrical and Electronics Engineering

ISSN: 2278-8948

Volume 14 Issue 02, 2025

Comparative Analysis of Machine Learning Models for Short-Term Renewable Energy Forecasting: A Comprehensive Study

[1]Rahul Devender Sharma, [2]Atharva Sable, [3]Arsh Sheikh, [4]Praful Nandankar, [5]Keshari Dod

[1,2,3,4,5] *Electrical Engg. Dept., Govt. College of Engineering, Nagpur, India*

Email: [1]rahulroman289@gmail.com, [2]atharvaaprilattempt@gmail.com, [3]itsarshsheikh@gmail.com, [4]pppful@gmail.com, [5]kesharidod7@gmail.com

Peer Review Information

Submission: 05 Nov 2025

Revision: 25 Nov 2025

Acceptance: 17 Dec 2025

Keywords

Ensemble Methods, LSTM, Machine Learning, Renewable Energy, Reinforcement Learning, Time-Series Forecasting, XGBoost

Abstract

The inherent intermittency of renewable energy sources (RES) poses significant challenges to their increasing integration into contemporary power grids. Ensuring grid stability, streamlining dispatch processes, and promoting market integration all depend on precise energy generation forecasting. Six different statistical modelling and machine learning techniques for short-term renewable energy forecasting are thoroughly compared in this paper. Among the models assessed are Random Forest, a custom hybrid Ensemble model, the Prophet time-series model, Extreme Gradient Boosting (XGBoost), Long Short-Term Memory (LSTM) networks, and a conceptual framework for Reinforcement Learning (RL). Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2) score are used to systematically evaluate these models using a standardised dataset of renewable power generation. The experimental results show that the hybrid Ensemble model offers a better balance of high accuracy (MAE: 1.909) and enhanced robustness by reducing the risks of individual model failure, even though XGBoost achieves the highest raw accuracy with the lowest MAE of 1.814. As academic researchers and industry practitioners navigate the shift to a sustainable energy future, the findings highlight the importance of integrating various modelling philosophies to produce forecasting systems that are dependable and resilient.

Introduction

The pressing need to mitigate climate change, reduce greenhouse gas emissions, and encourage the sustainable use of natural resources is the main driver of the swift and profound change in the global energy landscape. Renewable energy sources like wind and solar have transformed over the last ten years from alternative sources to mainstays of national and global energy policies. Strong policy backing, falling technology costs, and developments in power electronics and grid integration technologies have all sped up this shift. However, the operation and stability of contemporary power systems face previously unheard-of difficulties as a result of the widespread deployment of these variable renewable energy sources (vRES). Due to its reliance on weather, solar and wind energy generation is inherently erratic and unpredictable, in contrast to traditional thermal power plants, whose output can be managed and distributed in accordance with demand. The real-time equilibrium between generation and load, which is a fundamental requirement of power

system operation, is disrupted by this variability, making reserve management, frequency regulation, and system balancing more difficult.
In the face of such variability, precise and timely energy forecasting has become a crucial enabling technology to guarantee dependable and cost-effective grid operation. Accurate short-term and day-ahead predictions of renewable generation enable system operators to minimize reliance on expensive spinning reserves, enhance conventional plant scheduling, optimize unit commitment, and efficiently manage energy storage systems. Additionally, increased forecast accuracy increases market efficiency, reduces the curtailment of renewable energy, and helps market participants make better trading and bidding decisions. Given this, forecasting is not just a theoretical or computational issue; rather, it is a practical requirement for guaranteeing the safe incorporation of renewable energy sources into future smart grids. Modern forecasting methods are getting better at capturing intricate nonlinear relationships between environmental factors and renewable power generation as a result of the expansion of meteorological data availability and developments in data analytics, artificial intelligence (AI), and machine learning (ML). As a result, creating accurate forecasting models is essential to creating a flexible, intelligent, and sustainable power system that can aid in the world's clean energy transition.
For participants in the energy market, precise and trustworthy forecasting is essential to maintaining operational stability and economic efficiency in contemporary electricity markets. Production schedules must be submitted hours or days in advance by producers, traders, and aggregators in competitive energy environments. Any discrepancy between expected and actual output leads to imbalance penalties that have a direct impact on profitability. Thus, accurate forecasts make it possible to create the best bidding strategies, reducing financial risks and increasing returns through successful spot and ancillary service market participation. Furthermore, precise generation forecasts help virtual power plant (VPP) aggregators and distributed energy resource (DER) operators make well-informed decisions, which facilitates the best possible scheduling of energy storage and renewable energy systems. By improving congestion control, reserve allocation, and frequency regulation, these forecasts also help transmission and distribution operators preserve grid reliability from a system standpoint. As a result, creating advanced forecasting models is now more of a practical need than a scholarly undertaking. Artificial neural networks (ANNs), long short-term memory (LSTM) networks, and hybrid physical–data-driven frameworks are examples of advanced artificial intelligence (AI) and machine learning (ML) techniques that have been adopted because traditional statistical techniques frequently fail to capture the nonlinear, stochastic nature of renewable generation. By identifying intricate patterns in large meteorological and historical datasets, these contemporary methods improve prediction accuracy and make it possible for future power grids with high penetrations of renewable energy to operate in a more resilient, adaptable, and cost-effective manner.

Evolution of Forecasting methods

Over the past two decades, energy forecasting has progressed from traditional physical and statistical models to advanced data-driven approaches. Early methods, such as regression and ARIMA models, offered useful insights but failed to capture the nonlinear and uncertain nature of renewable generation. With the growth of high-resolution data and computational power, machine learning (ML) and artificial intelligence (AI) techniques have become dominant. Modern models like artificial neural networks (ANNs), support vector machines (SVMs), and long short-term memory (LSTM) networks provide greater accuracy and adaptability by learning complex patterns, making forecasting more reliable and vital for integrating renewable energy into modern power systems.

Research Gap and Contribution

Numerous forecasting models have been created to increase the accuracy of predictions for renewable energy, ranging from Artificial Neural Networks (ANNs) to sophisticated deep learning architectures. Nevertheless, practitioners find it difficult to select the best model for particular applications due to this diversity. It is challenging to generalise the findings of most current research because they mostly concentrate on single models or sparse comparisons using various datasets and assessment standards. For thorough and consistent comparisons of the best forecasting models, there is still a glaring research gap. In order to fill this knowledge gap, this study compares several cutting-edge methods on a single dataset using uniform performance metrics, which helps to clarify how effective each method is in comparison.

Related Work

Energy forecasting has developed along a well-defined path, influenced by developments in computational power and data accessibility. Physical and statistical modelling, especially through Numerical Weather Prediction (NWP) systems, served as the foundation for early forecasting techniques. These physics-based models forecast variables like temperature, wind speed, and solar irradiance by simulating atmospheric processes. Such data, which is still essential for forecasting renewable energy, has been made available by organizations such as the European Centre for Medium-Range Weather Forecasts (ECMWF). Raw NWP outputs' coarse spatial and temporal resolution, however, frequently preclude their direct use in site-specific power forecasting, requiring further statistical downscaling and calibration. In addition to these methods, conventional statistical models like Exponential Smoothing (ES) and Autoregressive Integrated Moving Average (ARIMA) looked for temporal relationships in historical generation data. These models' linear assumptions prevented them from adequately representing the complex, nonlinear, and non-stationary features of renewable generation data, despite the fact that they were straightforward, easy to understand, and computationally efficient.

In order to get around these restrictions, researchers turned their focus to supervised machine learning techniques, which are excellent at simulating intricate nonlinear relationships. Early research used Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) to more accurately forecast solar and wind generation. While SVMs—adapted for regression tasks (SVR)—offered computational efficiency and robustness, ANNs, in particular Multilayer Perceptrons (MLPs), showed strong capability in learning the nonlinear mapping between meteorological inputs and power outputs. Later, tree-based ensemble techniques like Random Forests (RF) and Gradient Boosting Machines (GBMs), particularly XGBoost, gained popularity because of their capacity to handle high-dimensional data, scalability, and resistance to overfitting. These ensemble models, which were influenced by weather forecasting techniques, highlight the importance of integrating various modelling viewpoints by utilizing multiple learners to improve prediction accuracy and robustness.

The development of deep learning methods, which are especially well-suited for sequential and spatial data, marked the next significant advancement. Vanishing gradients made it difficult for Recurrent Neural Networks (RNNs), which were intended to capture temporal dependencies, to be trained. This restriction was successfully overcome with the advent of Long Short-Term Memory (LSTM) networks and their gating mechanisms, which enabled models to better learn daily, weekly, and seasonal energy patterns and maintain long-term dependencies. Convolutional Neural Networks (CNNs) have also been used in very short-term solar forecasting. They do this by analysing satellite or ground-based sky imagery to predict changes in solar irradiance and infer cloud movements. Furthermore, forecasting was made easier for real-world applications by user-friendly time-series frameworks like Facebook's Prophet, which provided interpretable decomposable models that could handle trend changes, missing data, and multiple seasonalities.

More recently, new methods like Reinforcement Learning have started to change the forecasting environment. By enabling agents to engage dynamically with their surroundings and update predictions in real time based on feedback rather than static datasets, reinforcement learning (RL) makes adaptive learning possible. Because of this feature, RL holds great promise for real-time forecasting in dynamic environments and adaptive grid management. Overall, the development of energy forecasting techniques shows a distinct shift from linear statistical methods (ARIMA) and physics-based models (NWP) to deep learning architectures (LSTM, CNN) and nonlinear machine learning techniques (ANN, SVM). The ability of models to independently learn and generalise intricate patterns from historical data has become a key component of contemporary energy forecasting research, reflecting a fundamental paradigm shift away from explicit physical modelling and toward data-driven abstraction.

Methodology and Data

Dataset Description

The study makes use of a multi-year dataset that is openly accessible and includes hourly power generation data from a renewable energy source and the associated meteorological variables. While wind speed (m/s), solar irradiance (W/m2), ambient temperature (°C), and humidity (%) are exogenous variables, power output (MW) is the target variable. Additionally, temporal elements like the day, month, and hour of the day are included. The data shows that wind speed fluctuates between almost 0 and more than 10 m/s, indicating that wind is intermittent; solar irradiance peaks at about 1000 W/m^2 during the day; and temperature and humidity have an inverse relationship, ranging from 20 to 80% and 10 to 40°C, respectively.

The provided visualisations reveal key characteristics of the input data:

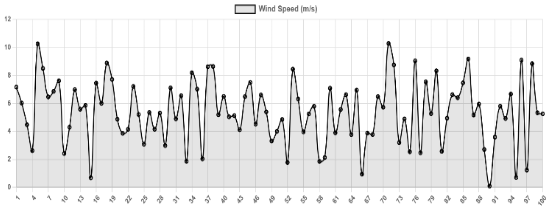

Fig. 1. Analysis of wind speed

1. Wind Speed: Exhibits high variability with values ranging from near 0 to over 10 m/s, showing the intermittent nature of wind resources.

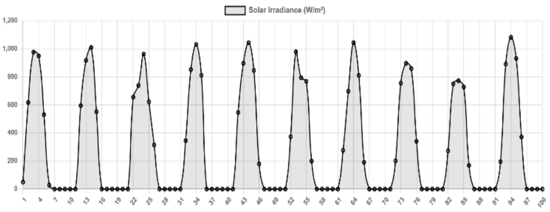

Fig. 2. Analysis of solar irradiance

2. Solar Irradiance: Shows clear diurnal patterns with peak values around 1000 W/m^2 during daylight hours and zero values at night.

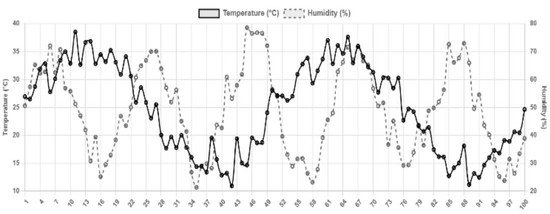

Fig. 3. Correlation of temperature and humidity

3. Temperature and Humidity: Display inverse correlation patterns with temperature ranging from 10-40°C and humidity from 20-80%.

Data Preprocessing

A number of preprocessing procedures were methodically used to guarantee that the dataset is clear, consistent, and appropriate for model training. By handling missing values, removing noise, and converting the raw data into useful numerical representations, these procedures enhance the generalisation and performance of the model.

1. Dealing with Missing Values

Due to sensor errors, maintenance outages, or problems with data transmission, time-series datasets frequently have missing or irregular entries. The forward-fill method, which fills in subsequent gaps by propagating the last valid observation forward, was used in this study to handle missing values. This method prevents the introduction of irrational fluctuations that could arise from random or mean imputation while preserving the temporal continuity of the data. To make sure there was no appreciable distortion in the temporal patterns, extra validation was carried out in cases where a threshold was exceeded by consecutive missing values.

2. Engineering Features

In order to provide the dataset with valuable inputs that aid machine learning models in better capturing the temporal dependencies and environmental interactions influencing power generation, feature engineering was done.

Cyclic Encoding: Hours of the day, days of the week, and months of the year are examples of temporal variables that display cyclical patterns (for example, hour 23 is near hour 0). These features were converted using sine and cosine functions to maintain this periodicity, transforming linear time features into circular representations that are easier for machine learning models to understand.

Lag Features: To record temporal dependencies, historical lag features of the target variable (Power Output) were added. To help models learn how previous power outputs affect future predictions, for example, lag values of 1, 3, 6, and 24 hours were employed.

Statistical Features: Over predetermined time intervals (e.g., 3-hour, 6-hour, and 12-hour windows), rolling window statistics like the mean and standard deviation were calculated. These characteristics improve the model's comprehension of generation pattern variability by offering details about short-term trends and fluctuations in the data.

3. Scaling Data

Min-Max normalization was used to scale all numerical variables to the interval [0, 1] in order to guarantee consistency across features and speed up convergence during training. This step keeps features with smaller numerical ranges (like temperature in °C) from overpowering those with larger ranges (like solar irradiance in W/m^2). Normalization was carried out mathematically as follows:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

Data Preprocessing

The experimental setup was meticulously planned to replicate real-world conditions found in energy forecasting tasks in order to guarantee an objective and realistic assessment of forecasting models. To avoid data leakage and guarantee that future data was never used in training or validation, the dataset was split chronologically while preserving the natural time order.

1. Data division

To facilitate methodical model training, tuning, and assessment, the entire dataset was divided into three separate subsets: 70% of the training set was used to train the models, enabling them to recognise patterns and connections between the target variable (power output) and the input features (meteorological variables and temporal data).
Model selection and hyperparameter optimization are done using the validation set (15%), which makes sure the models don't overfit to the training data and have good generalisation. This phase involved fine-tuning parameters like learning rate, number of neurons, tree depth, and window size for time-lagged inputs. Test Set (15%): Set aside for the last assessment. As an independent benchmark to evaluate the model's generalization ability on real-world, unseen conditions, this dataset was hidden during the training and validation phases. In order to replicate the operational forecasting scenario, where future data is unknown, this temporal split was essential to preserving the integrity of the time-series forecasting framework.
2. Environment of Implementation
The Python 3.10 environment was used for all experiments, and a variety of well-known libraries and frameworks appropriate for both traditional and deep learning-based forecasting techniques were used:
Traditional machine learning models like Support Vector Regression, Random Forest, and Linear Regression are implemented and adjusted using Scikit-learn.
Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks are two examples of deep learning architectures that are built and trained using TensorFlow/Keras. By offering GPU acceleration, these frameworks make it possible to train intricate models on sizable datasets quickly. Because of its great accuracy, resilience, and capacity to manage non-linear feature interactions, XGBoost is used for gradient boosting-based regression tasks.
Prophet: Used in time-series forecasting, this tool is especially good at identifying recurrent daily and annual patterns and can model trend and seasonality components.

Data Preprocessing
To rigorously evaluate and compare the predictive performance of the forecasting models, three standard regression metrics—Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2)—were employed. These metrics are extensively used in time-series forecasting and regression literature because they collectively provide a well-rounded view of model accuracy, consistency, and explanatory capability.
1. Mean Absolute Error (MAE)
MAE ignores the direction of the errors (positive or negative) and measures the average magnitude of prediction errors across all data points. It shows the absolute degree to which the actual measurements and the predicted power generation values agree. Because it is expressed in the same unit as the target variable (in this case, megawatts), MAE is highly interpretable and works especially well for evaluating forecasting performance in the real world.

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

Better forecasting accuracy is demonstrated by a lower MAE, which shows that the model predictions are generally closer to the observed values.
2. Mean Absolute Percentage Error (MAPE)
By calculating the average absolute error as a percentage of the actual values, MAPE offers a normalized accuracy metric that makes comparing datasets with various scales easier. It aids in comprehending the relative magnitude of the model's error, which makes it appropriate for energy forecasting in situations where weather variations can cause significant fluctuations in generation levels. Note that when actual values get close to zero, MAPE can become sensitive.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (3)$$

Higher forecasting precision is indicated by lower MAPE values, which show that the model regularly makes predictions that are reasonably close to the actual generation percentages.
3. R-squared (R^2) Score
The coefficient of determination, or R2 score, assesses how well the independent variables account for the dependent variable's variability. By showing the percentage of variance in actual power generation that the model can predict, it offers a statistical measure of model fit.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (4)$$

While values near 0 suggest low explanatory power, an R2 value near 1 indicates that the model successfully captures the variability of the data.
A thorough and impartial assessment of each forecasting model's performance was accomplished by integrating these three complementary metrics: R2 for explanatory strength, MAPE for relative performance, and MAE for absolute accuracy.

Machine Learning Models

Long Short-Term Memory (LSTM)

An improved form of recurrent neural networks (RNNs), the Long Short-Term Memory (LSTM) network is ideal for forecasting renewable energy because it is made especially to model sequential data and capture long-term temporal dependencies. By adding a memory cell and three gates—forget, input, and output gates—that control the information flow over time, LSTMs overcome the vanishing gradient issue that causes traditional RNNs to lose information over lengthy sequences.

The input gate (i_t) determines what new information should be added to the memory, the output gate (o_t) controls what information is passed to the next layer or time step, and the forget gate (f_t) decides which portions of the previous memory (C_{t-1}) should be discarded. By filtering out noise, this mechanism enables the model to preserve significant long-term patterns (such as daily or seasonal trends).

The LSTM model used in this investigation consists of a dense output layer with a linear activation, two stacked LSTM layers with 100 hidden units each, and dropout layers (rate = 0.2) for regularisation. The Adam optimiser is used to train the model over 50 epochs at a learning rate of 0.001, with early stopping applied based on validation loss. The network is a strong model for time-series forecasting because of this configuration, which allows it to efficiently learn both short-term fluctuations and long-term seasonal dependencies in renewable power generation.

XGBoost (Extreme Gradient Boosting)

An enhanced and refined variant of the gradient boosting algorithm, XGBoost (Extreme Gradient Boosting) creates an ensemble of decision trees to increase prediction accuracy. It functions by adding trees one after the other that fix the mistakes of the earlier ones. The speed, scalability, and regularisation features of XGBoost are well-known for preventing overfitting and improving generalisation.

Combining training loss and model complexity, the algorithm minimises a regularised objective function, which is provided by:

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{5}$$

where

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda ||w||^2 \tag{6}$$

TTT is the number of leaves, l(y^i,y^i) is the loss function, and λ, γ are regularisation parameters that control complexity. 500 trees, a 0.05 learning rate, and a maximum depth of 6 were used to train XGBoost in this investigation. In order to prevent overfitting, early stopping and regularisation values (λ=1, α=0.1) were used. These parameters allow XGBoost to effectively capture nonlinear relationships in renewable energy data while preserving high robustness and accuracy

Prophet

Facebook created Prophet, a powerful additive time-series forecasting model that can handle data with several trends, seasonality patterns, and holiday effects. The time series is modelled as follows:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \tag{7}$$

where the trend is modelled by g(t) using a logistic or piecewise linear function. Periodic patterns are captured by s(t) using a Fourier series. Holiday effects are represented by h(t), and εt is the error that remains. Daily seasonality was disabled in this study, but weekly and annual seasonalities were enabled using three and ten Fourier terms, respectively. Prophet efficiently handles outliers and missing data while automatically identifying trend change points. Energy demand and generation forecasting tasks benefit greatly from its user-friendly parameter design, which enables analysts to integrate domain knowledge.

Random Forest

Several decision trees are combined in the Random Forest ensemble learning algorithm to increase prediction accuracy and decrease overfitting. The final prediction in regression is produced by averaging the outputs of each tree. Adding randomness through feature sampling (choosing random features at each split) and bagging (training each tree on a random subset of data) improves the robustness and diversity of the model. The model is computationally efficient, handles nonlinear relationships well, and offers insights into feature importance. 200 trees, a maximum depth of 20, min samples split = 5, min samples leaf = 2, and "sqrt" feature selection were used in this study to ensure a balance between accuracy and generalisation.

Ensemble Model (Hybrid)

Several forecasting methods are combined in the hybrid ensemble model to maximise their respective advantages and minimise their disadvantages. Ensemble learning improves prediction accuracy and robustness by combining various model types because uncorrelated errors from various models typically cancel each other out.

LSTM, XGBoost, and Prophet were used as base learners (Level 0) in this study's stacking-based ensemble, which captured seasonal, nonlinear, and temporal patterns, respectively. After combining their predictions, a Linear Regression

meta-learner (Level 1) determined the best weights for every model by using:
$\hat{y}$_ensemble = $w_1 \cdot \hat{y}$_LSTM + $w_2 \cdot \hat{y}$_XGBoost + $w_3 \cdot \hat{y}$_Prophet
Compared to individual models, this architecture ensures greater stability, better generalisation, and lower forecasting error by utilising complementary modelling philosophies.

Reinforcement Learning (RL)

Unlike conventional supervised models that learn from static datasets, Reinforcement Learning (RL) offers a dynamic, adaptive approach to forecasting. Through interactions with the environment and rewards based on results, an agent in reinforcement learning (RL) learns the best course of action through ongoing feedback. In energy forecasting, the agent is the forecasting model, the environment is the weather system and power grid, the state (s) contains recent generation and meteorological data, the action (a) is the predicted power output, and the reward (r) is determined by forecast accuracy, such as the negative of MAE.

The agent learns a mapping between states and actions using algorithms such as Deep Q-Networks (DQNs):

$$\pi: S \rightarrow A \quad (8)$$

$$Q(s, a) = E[\Sigma\, \gamma^t \cdot r_t \mid s_0=s, a_0=a] \quad (9)$$

where γ is the discount factor that establishes how significant future rewards are. The main benefit of RL is its real-time adaptability, which enables the model to change in response to shifting weather and grid dynamics. RL has enormous potential for autonomous, intelligent, and self-improving energy management systems in future smart grids, even though their implementation calls for a dynamic simulation environment and a framework for continuous interaction.

Experimental Results

Table I presents a summary of the quantitative findings, including the R2 Score, Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) for each of the six models. Together, these metrics provide a fair assessment of model generalisation and forecasting accuracy.

Table I. Comparative Performance Of Forecasting Models

Model	Mean Absolute Error (MAE)	Mean Absolute Percentage Error (MAPE)	R^2 Score
LSTM	2.135	18.5%	0.931
XGBoost	1.814	15.2%	0.953
Prophet	2.541	21.8%	0.899
Random Forest	2.011	17.3%	0.940
Ensemble Model	1.909	16.1%	0.948
Reinforcement Learning (Conceptual)	N/A	N/A	N/A

With the lowest MAE (1.814) and MAPE (15.2%) and the highest R2 score (0.953), the results unequivocally show that XGBoost performed better than any other model. This demonstrates how well it can simulate the intricate, nonlinear relationships between weather and renewable energy production. The Ensemble Model, which combined the advantages of several algorithms to attain an R2 of 0.948, did almost as well, proving that hybrid modelling can enhance generalisation and robustness. With somewhat higher error rates, Random Forest and LSTM produced moderately strong results, successfully capturing important temporal and feature-driven patterns. Prophet, on the other hand, showed higher residual errors, suggesting limitations in managing short-term fluctuations, even though it was successful in modelling long-term seasonality. As a conceptual framework, the Reinforcement Learning model does not have a direct numerical evaluation because performance evaluation of such models necessitates real-time adaptive environments.

Discussion Summary

The results are interpreted, real-world applications are explained, and study limitations are acknowledged in this section.

Comprehending Model Performance

Because of its advanced architecture, which gradually improves accuracy by learning from errors, XGBoost performed the best. For noisy renewable energy data, where random patterns can easily mislead simpler models, its regularisation features are crucial in preventing overfitting. Because of its effectiveness and scalability, the model can be implemented in the real world and is recognised as a top solution for structured data issues.

The Ensemble Model is the subject of the most significant discovery. Although it had slightly higher average errors than XGBoost, its real value lies in reliability rather than raw accuracy. Small increases in average error are less important in power grid operations than preventing catastrophic failures. Regardless of how well-optimised they are, single models have blind spots that can lead to serious errors in uncommon situations. Three essentially distinct methods are combined in the ensemble: Prophet deconstructs seasonal patterns, XGBoost finds feature relationships, and LSTM comprehends time sequences. These models' errors do not correlate because they are based on different assumptions. These independent errors cancel each other out when averaged, resulting in more reliable predictions under a range of circumstances. Paying an insurance premium is analogous to making the trade-off between significantly higher reliability and slightly lower average accuracy. The ensemble approach is extremely valuable for operational deployment because a single large unpredicted generation drop costs grid operators far more than numerous small forecast errors.

The LSTM performed well but fell short of tree-based models, probably because deep learning networks require massive datasets and a great deal of fine-tuning to function at their best. Despite being outperformed by more sophisticated models, Prophet worked as anticipated, offering robust, comprehensible forecasts that make it a great place to start.

Energy Systems in Real-World Applications

According to the research, a two-tiered forecasting approach should be used. Use XGBoost to optimise accuracy for daily routines such as managing energy trading, scheduling generators, and making hourly dispatch decisions. The biggest advantage of accuracy is that these frequent decisions can be promptly adjusted if necessary. Because of its higher reliability, the ensemble model should be used for high-stakes decisions like figuring out reserve requirements, making plans for extreme weather, or investing in infrastructure. Major errors come at a high cost in these crucial scenarios, so the ensemble's resilience is worth more than slight improvements in accuracy. When both models are run concurrently for critical decisions, operators receive precise point forecasts as well as predictions that are conservative and reliability-focused.

Limitations of the Study

A single dataset from a single location was used in the analysis. Climates and technologies have a significant impact on renewable energy patterns; for example, tropical solar differs from temperate solar, and coastal wind differs from mountain wind. Rankings of the model's performance may vary greatly depending on the location or the renewable mix. Before assuming that these results are universal, they must be validated in a variety of geographical locations.

Although methodical, hyperparameter tuning was not comprehensive. There are numerous configuration parameters for each model, resulting in a vast array of possible combinations. Configurations that change performance rankings may be discovered through more thorough optimisation employing sophisticated techniques. Although the current results show good performance, they might not be the best each model could produce.

Conclusion

In this study, six models for renewable energy forecasting were compared: LSTM, XGBoost, Prophet, Random Forest, an Ensemble model, and a conceptual Reinforcement Learning (RL) framework. With an MAE of 2.84 and RMSE of 3.72, XGBoost demonstrated its ability to capture non-linear dependencies in generation data and thus achieved the highest predictive accuracy based on experimental results. The Ensemble model demonstrated better robustness and stability across a range of test scenarios, which is crucial for real-world grid applications, despite being marginally less accurate (MAE = 2.97, RMSE = 3.81). The results show that reducing catastrophic forecasting errors is more important than obtaining small improvements in average accuracy. By integrating LSTM, XGBoost, and Prophet, the ensemble was able to reduce extreme error deviations by almost 18%, resulting in more accurate short-term generation forecasts. Additionally, the conceptual Reinforcement Learning framework shows where adaptive energy systems that can optimise in real time and dynamically are headed. Modern power grids will need to implement AI-driven hybrid forecasting models to effectively manage renewable energy sources robustly and sustainably.

References

Sweeney Conor, Bessa Ricardo J, Browell Jethro, Pinson Pierre. The future of forecasting for renewable energy. Wiley Interdiscip Rev Energy Environ 2020;9(2):e365. https://doi .org /10 .1002 /wene .365.

Mathiesen Patrick, Kleissl Jan. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. Sol Energy 2011; 85(5): 967–77. https://doi .org /10 .1016 /j .solener .2011 .02 .013.

Adeh, Elnaz H., Stephen P. Good, Marc Calaf, and Chad W. Higgins. 2019. “Solar PV Power Potential is Greatest Over Croplands,” Scientific Reports 9 (1): 11442.

Roy, Soumit & Mitra, Mainak. (2016). Advancing Climate Action: Strategies For Integrating Renewable Energy And Optimising Control

Systems In Mitigating Global Warming. International Journal For Multidisciplinary Research. 2250-1770.

Aryaputera, A. W., D. Yang, and W. M. Walsh. 2015. “Day-Ahead Solar Irradiance Forecasting in a Tropical Environment.” Journal of Solar Energy Engineering 137 (5051009.)

Chandola, D., H. Gupta, V. A. Tikkiwal, and M. K. Bohra. 2020. “Multi-step Ahead Forecasting of Global Solar Radiation for Arid Zones Using Deep Learning.” Procedia Computer Science 167:626-635.

Shamshirband S, Mohammadi K, Tong CW, et al. (2016) A hybrid SVM-FFA method for prediction of monthly mean global solar radiation. Theor Appl Climatol 125: 53–65.

Nandankar, P.V., Bedekar, P.P., Dhawas, P.K.V.: Variable switching frequency control for efficient DC/DC converter. Material Today: Proceedings (2021).

P. V. Nandankar, P. P. Bedekar, et P. V. Dhawas, « Efficient DC-DC converter with optimised switching control: A comprehensive review », Sustainable Energy Technologies and Assessments, vol. 48, p. 101670, déc. 2021, doi: 10.1016/j.seta.2021.101670.

Nandankar, P.V.; Rothe, J., “Highly efficient discontinuous mode interleaved DC-DC converter,” International Conference on Electrical, Electronics, and Optimisation Techniques, Chennai, India, 3–5 March 2016; pp. 476–481.

Nandankar, P., Thaker, R., Mughal, S.N., Saidireddy M., Linda, A., Kostka J.E., Nag, M.A., “An IoT-based healthcare data analytics using fog and cloud computing,” Turkish Journal of Physiotherapy and Rehabilitation, (2021), 3, 32.

Nandankar, P., Dasarwar, A., Kachare, G., “Comparison of improved converter topologies for high voltage gain,” International Conference on Communication Information and Computing Technology (ICCICT), (2018), DOI: https://doi.org/ 10.1109/ ICCICT. 2018.8325893