International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 11-No.3 (2025) pp. 6567-6575
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Long-term Electricity Price Forecasting Using a Random Forest-based Machine Learning Approach

Jaya Shukla[1*], Rajnish Bhasker[2]

[1]Research Scholar, Dept. Of E.E VBSPU Jaunpur.
* **Corresponding Author Email:** jayashuklaee@gmail.com - **ORCID:** 0000-0002-5247-7851

[2]Professor, Dept. Of E.E VBSPU Jaunpur
Email: rabj_33@rediffmail.com - **ORCID:** 0000-0002-5247-7851

Article Info:

DOI: 10.22399/ijcesen.3879
Received : 22 June 2025
Accepted : 21 August 2025

Keywords

Electricity Price Forecasting, Random Forest, Time Series Analysis, MATLAB, Yearly Price Variation, RMSE, MAE

Abstract:

Electricity price forecasting are important in optimizing energy trading, consumption scheduling, and operational planning within smart grid infrastructures. This study proposes a data-driven approach using a Random Forest (RF) regression model implemented in MATLAB for accurate electricity price prediction. Unlike conventional models, the RF model is evaluated under both open-loop and closed-loop forecasting scenarios to assess its short-term accuracy and long-term stability. The model is trained on time-series electricity pricing data enriched with lagged variables and temporal features, allowing it to learn from past behaviors and predict future price fluctuations. Open-loop forecasting utilizes actual historical values at each time step, enabling the model to demonstrate its pattern recognition capabilities with minimal cumulative error. Conversely, the closed-loop approach relies on recursive self-generated predictions to simulate real-world deployment, where future data is unavailable. Despite expected error propagation, the model maintains trend fidelity and captures peak patterns effectively across all three data channels. Performance evaluation using RMSE (0.534), MAE (~0.0276), and R^2 (~0.783) confirms the model's accuracy, robustness, and generalization ability across multiple channels. Additionally, the consistent sensitivity score highlights the model's responsiveness to price changes. The results underscore the RF model's suitability for reliable electricity price forecasting, offering a balance between predictive accuracy and computational efficiency. This research supports the advancement of intelligent forecasting tools for dynamic electricity markets and reinforces the feasibility of integrating RF-based prediction systems in both academic research and industrial energy management applications.

1. Introduction

Predicting power prices accurately becomes more important as smart networks and deregulated electricity markets shift. Energy suppliers and customers are looking for sophisticated systems that can accurately forecast future pricing trends due to unstable energy sources and shifting demand. Conventional forecasting techniques are helpful, but they frequently fail to capture the intricate seasonal patterns and nonlinearities that define the behavior of power prices over long periods. Machine learning algorithms, especially ensemble-based methods like Random Forest, present a viable substitute because of their interpretability, resilience, and capacity to process high-dimensional data.The complexity and unpredictability present in energy forecasting, cost estimates, and environmental modeling have led to a growing interest in machine learning approaches, especially ensemble learning techniques like RF. Researchers have been able to get beyond the drawbacks of conventional statistical models because to their versatility and resilience, particularly in fields with high levels of uncertainty and nonlinear dynamics.

Recent studies in the field of electricity price forecasting (EPF) have turned their attention to adaptive learning frameworks that can react to market volatility brought on by the integration of non-conventional resources. A self-adjusting RF-based model that runs on dynamically maintained training sets was proposed online by [1]. When compared to static batch-learning models, this approach greatly increases prediction accuracy by accounting for idea drift and market changes.The addition of RF algorithms has also improved cost estimation in early-stage building projects. [2] has out a comparison of ANN and Radiofrequency for residential building price prediction. The study found that both models worked well, with RF attaining a prediction accuracy of 93.05%, using project data from residential structures in Egypt.Price prediction is more difficult in the carbon trading industry since the data is non-stationary and nonlinear. [3] addressed this by combining RF, Long Short-Term Memory (LSTM) and Complete Ensemble Empirical Mode Decomposition.Outperforming conventional linear and single-model learning approaches, the combined model effectively improved the generalization and resilience of carbon price predictions.

In [4], a project cost prediction system was developed for a Chinese road and bridge construction firm. Using historical project data and key influencing factors, their RF-based model achieved high regression accuracy, supporting reliable budgeting and bidding decisions. In [5], RF was applied to predict construction costs for large-scale energy infrastructure, such as power plants, during the conceptual phase. The RF model outperformed Support Vector Regression (SVR) and k-Nearest Neighbors (KNN), achieving a determination coefficient of 0.956 and the lowest RMSE of 29.27. SHAP analysis was used to interpret the influence of each feature on cost estimation.

Run-of-river hydropower generation has also been predicted using RF in climate-aware modeling. developed prediction models across European areas utilizing climatic variables including air temperature and precipitation [6]. Despite acknowledging challenges in collecting few but significant local occurrences, their study highlighted the benefits of utilizing high-resolution climate data with RF.To improve cost forecasts in electric power engineering, [7] suggested a hybrid forecasting technique. Support Vector Machine (SVM) optimized with the Wolf Pack Algorithm (WPA) was merged with RF for feature reduction. This approach produced better prediction performance by successfully resolving problems with high dimensionality, sparse data, and SVM overfitting.Forecasting accuracy in a variety of energy-related sectors is continuously improved by combining RF with other machine learning approaches. [8] introduced a hybrid model for forecasting that integrates RF, Grey Catastrophe , and SVR. Trend smoothing was done using GC, nonlinear regression with SVR, and performance improvement with RF. When the model was tested using data from the Australian Energy Market Operator, it produced low MAPE values of 6.35% and 6.21%, indicating that it was resilient to fluctuations.

In the context of electricity market analysis, [9] evaluated the market efficiency of ISO New England using a combination of GARCH modeling, Ljung-Box testing, and RF-based prediction. The study found that market inefficiencies fluctuate over time and seasons, and the RF model generated net profits during inefficient periods, supporting the Efficient Market Hypothesis (EMH) under certain conditions.Addressing construction cost prediction, [10] developed an RF model optimized by the Bird Swarm method. The approach outperformed several classical metaheuristics (PSO, GA, ACO, etc.) and forecasting models (BPNN, SVM, SAE, ELM). In a case study involving a Xinyu construction firm, the maximum relative error was just 1.24%, validating the model's high efficiency and accuracy.

The RF-MGF-RSM model for forecasting was presented by [11]. The inputs for this hybrid model include Mean Generating Function (MGF) and RF predictions, which are then improved using the Response Surface Technique to lessen the spread of errors. The accuracy of forecasting significantly improved, according to experimental data, especially when it came to identifying load peaks and dips.In substation infrastructure projects, RF modeling was used to predict building progress for overhead line, electrical, and civil projects [12]. By 5.8% and 18.3%, respectively, the RF model outperformed the Extra-Trees and Decision Tree models in terms of prediction accuracy. To enhance scenario-based schedule management, the study offered helpful recommendations.Using operational data from an Iranian iron ore facility, [13] created an RF model to estimate ball mill power-draw in the field of mineral processing. The model prioritized input characteristics and obtained an R2 of 0.98 as opposed to 0.60 from conventional regression. This confirmed RF's exceptional interpretability and modeling ability in intricate industrial situations.[14] suggested a hybrid forecasting model that uses a PSO-optimized SVR for peak load prediction in smart grids, RF for feature selection, and RF for missing data imputation. The model's superior performance over non-feature-selected baselines was shown by real-world data from Jiangxi Province, which qualifies it for use in contemporary power grid applications. In a research on regional power usage, [15] created a hybrid forecasting model that included Grey Theory with RF. By using this method on data from central-western China, prediction accuracy was increased and data shortage was handled.

In [16], a RF-based model forecasts real-time electricity prices in New York's HUD VL market. It provides price probability distributions for bidding risk assessment and adapts to climatic and market changes via parameter updates, showing robust performance in a case study.In [17], ML (SVM, RF) and DL (NARX, LSTM) models predict medium- and long-term load demand in Bruce County, Ontario, using nine years of data with climatic inputs. DL models outperform ML, achieving R-squared of 0.93–0.96 and MAPE of 4–10%, aiding grid management and renewable adoption.
In [18], an EEMD-RF model forecasts daily electricity consumption for enterprises in Jiangsu, China (2015–2016). EEMD decomposes data into IMFs, and RF predicts trends, outperforming BPNN, LSSVM, and hybrids in MAE, MAPE, and RMSE.In [19], an online self-adaptive RF model forecasts electricity prices, handling market volatility with varied training sets. A case study shows improved accuracy and robustness against concept drift, enhancing bidding strategies.A GAN-based model forecasts probabilistic electricity prices using Ontario's IESO data was presented in [20] It achieves lower MSE (687.513) and 7% higher accuracy than RF, SVM, and XGBoost, with an MAE of 8%, excelling in data-driven analytics.

Traditional statistical forecasting models often struggle to capture the complex, nonlinear, and seasonally dynamic behavior of electricity markets. As a result, energy providers, policymakers, and infrastructure planners are turning toward more adaptive and data-driven approaches. Machine learning techniques, particularly ensemble methods like RF, have shown significant promise due to their robustness, interpretability, and capability to manage high-dimensional and non-stationary datasets.
This study investigates the application of Random Forest regression for long-term electricity price forecasting using yearly time series data. Implemented in MATLAB, the model utilizes historical electricity prices along with calendar variables to predict future values. The methodology encompasses data preprocessing, model training, performance evaluation using RMSE and MAE, and interpretation of feature importance. Unlike short-term day-ahead models, this approach captures broader annual variations, making it suitable for strategic planning and policy-level decision-making. Drawing inspiration from prior research on RF-based tariff prediction in home energy management systems, this work extends the concept to yearly forecasting for market-scale applications. The key contributions are as follows,

- A methodology for preprocessing and structuring historical electricity price data into a time series format suitable for ML applications. Seasonal patterns, missing standards, and noise in the dataset are effectively handled to ensure model robustness.
- A custom RF regression model is developed and implemented in MATLAB, allowing for flexible integration with time series inputs and parameter tuning to optimize performance for yearly electricity price forecasting
- The model is trained using real-world historical electricity price data, capturing nonlinear trends and yearly dynamics. Emphasis is placed on the model's ability to learn complex relationships within the dataset, including the impact of demand fluctuations.
- The model's forecasting accuracy is evaluated using performance metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), R-squared (R^2), and Sensitivity.

2. System description

The proposed supervised learning-based electricity tariff forecasting model aims to predict hourly electricity prices ahead of their actual availability from utility providers. This early prediction capability empowers consumers and energy management systems to make informed decisions regarding energy usage, leading to improved cost-efficiency and grid responsiveness. By leveraging patterns in historical electricity pricing data, the model generates accurate day-ahead forecasts that reflect both temporal trends and behavioral price dynamics.
Key inputs to the model include historical electricity prices and temporal features such as the hour of the day, day of the week, and indicators distinguishing weekdays from weekends. Additionally, aggregated features that capture short-term fluctuations and price trends are incorporated to enhance model awareness of recent market behaviors. These data inputs are processed using a sliding time window approach to build structured input-output pairs, allowing the model to learn from sequential patterns. Once preprocessed, the data is fed into a Random Forest regression algorithm, trained to recognize non-linear dependencies in the pricing structure and produce precise hourly predictions.
The historical price data, sourced from real-time utility tariff information, undergoes normalization

and synthetic augmentation to simulate realistic conditions. The entire dataset is divided into training and testing subsets. During the training phase, multiple decision trees are optimized through ensemble learning, improving robustness and reducing variance. The trained model is then validated using the unseen testing data to evaluate forecasting accuracy. Both open-loop (using actual past values) and closed-loop (recursive prediction) forecasting modes are employed to simulate real-world usage scenarios. Model performance is assessed through key evaluation metrics including RMSE, MAE, and R-squared values. Figure 1 illustrates the machine learning workflow, emphasizing the roles of data preparation, model training, validation, and testing. This predictive framework serves as a core component of a Home Energy Management (HEM) system, enabling smart scheduling, load shifting, and tariff-aware control of residential energy consumption.. Fig.1 represents the Evaluation Stage of ML implementation in the prediction process

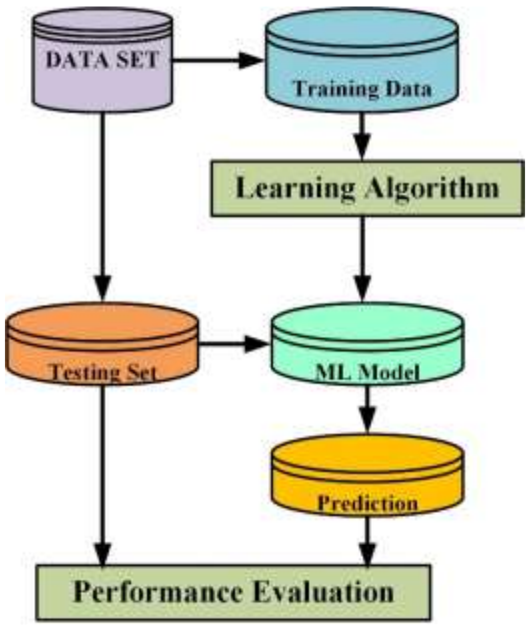

Figure 1 *Evaluation Stage of ML implementation in the prediction process*

3. Proposed Random Forest Forecasting Model

The RF model is an ensemble learning technique that builds upon decision tree algorithms. It employs two main strategies bagging and the random subspace method to construct a diverse set of regression trees. In the bagging process, multiple trees are trained using bootstrap samples, which involve sampling the original data with replacement. To further promote variability among the trees, the random subspace technique selects a random subset of input features at each node split, rather than using the entire set of predictors. By combining these two approaches, the Random Forest reduces overfitting and enhances prediction accuracy by introducing randomness both in the data and the feature selection process.

The RF approach uses a bootstrap sample that is the same size as the training data for each K tree (k=1,..., K). A tree is constructed for each sample by recursively splitting the input space at each node until a minimum sample leaf size is reached. Features are chosen at arbitrary from the total number of features (N) to calculate splitting at each node. The ideal split is chosen by optimizing the reduction of the RMSE across all split candidates and cut sites. After all K trees have been constructed in this way, the RF prediction may be calculated using (1).

$$\hat{f}(x) = \frac{1}{K}\sum_{K=1}^{K} T_K(x) \quad (1)$$

The symbol for the input pattern is x. Some hyperparameters that was changed to improve the prediction results include the number of trees (K), the size of the leaves (m), and the characteristics that will be chosen at random for every split (n). According to the literature, for regression problems, the number of maximum features (n) should be equal to the total number of features (N) divided by three. This is because a decline in n lowers the correlation between trees, which lowers the mean's variance. Fig. 2 represents the RF Forecasting Process.

Algorithm: Electricity Price Forecasting using Random Forest

Input: Electricity price time series data from an Excel file

Output: Forecasted electricity prices (open-loop and closed-loop), RMSE, MAE, and R-squared metrics

Data Preprocessing:

Load raw electricity price data from the Excel file.
Segment the data based on predefined group lengths.
For each segment, create three noisy variations using small random perturbations.
Eliminate negative values and store the result in a cell array.

Data Preparation for Machine Learning:

Normalize each observation by dividing by 1000.
Use a sliding window approach to prepare input-output pairs for training.
Reshape the input windows and next-step targets for supervised learning.

Model Training:

Train separate Random Forest regression models for each data channel using the training set.
Store trained models for future use.

Forecasting:

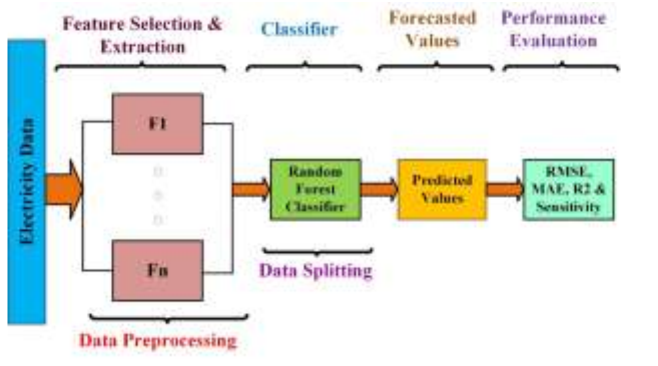

Fig.2 Procedure of Forecasting using Random Forest

Open-loop forecasting: Predict using the true previous time steps.

Closed-loop forecasting: Predict recursively using previously forecasted values for 200 future steps.

Performance Evaluation:

Compute RMSE, MAE, and R-squared (R^2) & sensitivity.

Plot and compare predicted versus actual prices for both forecasting methods.

4. Performance Metrics Assessment

The performance metrics used in the study is RMSE, MAE, R2 & sensitivity

RMSE

The RMSE is a extensively used evaluation metric that measures the average magnitude of the prediction errors produced by a model. Larger mistakes are penalized more severely because of the squaring process. In the context of electricity tariff forecasting, RMSE quantifies how far the predicted hourly prices deviate, on average, from the actual values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{acttariff} - \hat{y}_{pretariff})^2} \quad (2)$$

Where, $y_{acttariff}$ & $\hat{y}_{pretariff}$ are the actual & predicted electricity tariff at hour i & n is the total number of predictions.

Mean Absolute Error (MAE)

To determine the average absolute difference among the expected and actual values, the MAE treats all mistakes equally and does not give greater errors more weight.For your model, MAE reveals how much the forecasted electricity tariffs deviate from the true values on average.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_{acttariff} - \hat{y}_{pretariff}\right| \quad (3)$$

Where, $y_{acttariff}$ & $\hat{y}_{pretariff}$ are the actual & predicted electricity tariff , respectively & n is the number of time steps evaluated.

R-squared

The R-squared (R^2) value is a statistical size used to assess how well the predicted values from a regression model approximate the actual data points. In the context of the electricity tariff prediction model, R^2 indicates how much of the variability in the actual tariff data is captured by the machine learning forecast model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{acttariff} - \hat{y}_{pretariff})^2}{\sum_{i=1}^{n}(y_{acttariff} - \bar{y})^2} \quad (4)$$

Where, $y_{acttariff}$ & $\hat{y}_{pretariff}$ are the actual & predicted electricity tariff at hour i, $\bar{y}$ is Mean of all actual tariff values & n is the Total number of data points

Sensitivity (Recall or True Positive Rate)

Sensitivity (recall) is the percentage of real positive cases that the model accurately detects in the context of classification models. While your current problem is a regression task (predicting continuous values), sensitivity can be adapted if the prediction task is transformed into a classification setting — for instance, classifying whether the electricity price will exceed a certain threshold.

$$Sensitivity = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)} \quad (5)$$

Where, TP is Number of times the model correctly predicted a tariff above the threshold & FN is Number of times the model missed predicting a high tariff when it occurred

5. Simulation Results

The simulation results demonstrate both open and closed-loop forecasting performances using the Random Forest model. Open-loop results show high accuracy due to the continuous use of actual historical data for input. In contrast, closed-loop forecasting showcases the model's stability when relying on its recursive predictions. By comparing these two modes, we evaluate both the short-term precision and long-term predictive consistency. The energy demand vs samples was represented in Fig.3. Fig.4 shows the samples of electricity price for 2017-2020.

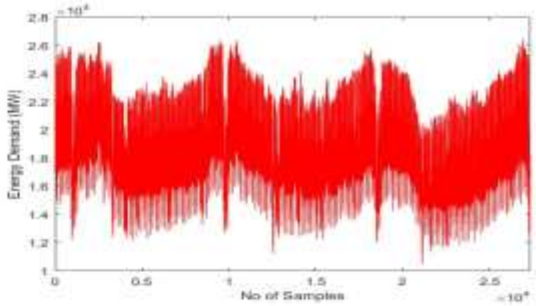

Fig.3 Energy demand Vs samples

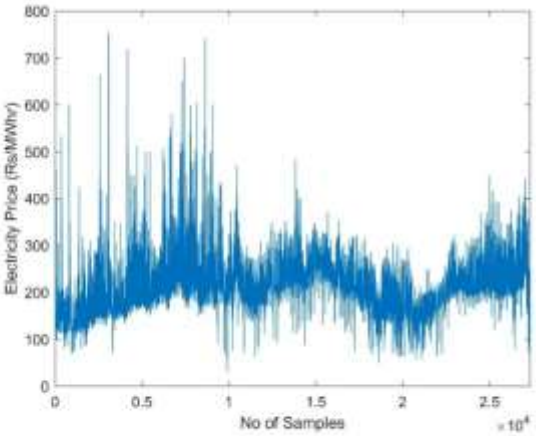

Fig.4 Electricity Price Vs samples

Open-loop Forecasting

The model uses actual historical values at each time step to produce predictions in the open-loop forecasting technique. In particular, the model forecasts the future value using the historical real power price values rather than its projections. To estimate the tariff for the following hour, the trained Random Forest models are fed a sliding window of actual historical data. This technique aids in assessing the model's capability to detect shapes in the best-case scenario, when actual data is constantly accessible for input. Because it prevents error buildup, the open-loop approach often yields more accurate forecasts with lower RMSE and MAE. It helps evaluate and assess predictive skills since it shows the model performance in a controlled testing environment.

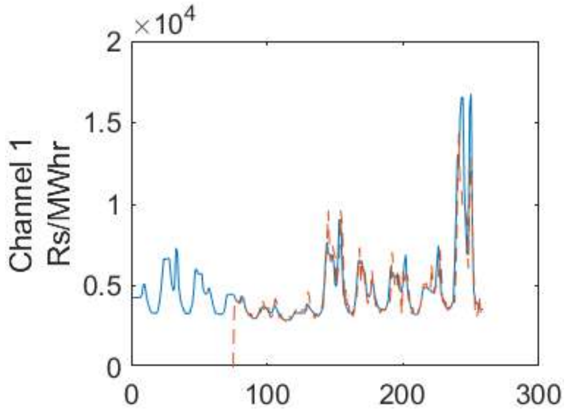

Fig.5 Open-loop prediction for Channel 1

For the majority of the timeframe, Channel 1's open-loop forecast represented in Fig.5, roughly matches the actual values. Nonetheless, there are also notable variations during abrupt price increases in the 220–250 hour period, when the projection somewhat understates the significant rise. This implies that the model does a prediction of capturing the underlying pattern.

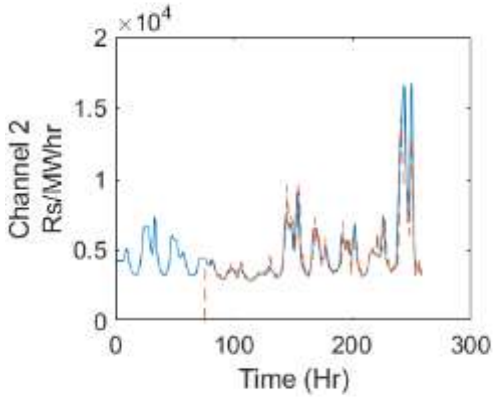

Fig.6 Open-loop prediction for Channel 2

The projection for Channel 2 shown in Fig.6 remains closely aligned with the real price curve, much like Channel 1. Both relatively abrupt fluctuations and gradual variations are well tracked by the Random Forest model. The forecast line lags somewhat behind the actual price during the strong spike between 200 and 240 hours, and there are minor variations around the 90 to 110 hour period.

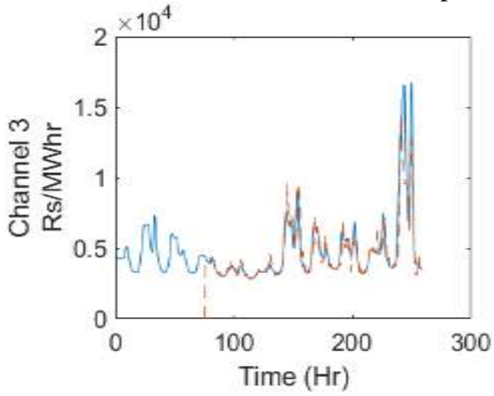

Fig.7 Open-loop prediction for Channel 3

High predictive performance is shown in Fig.7 for Channel 3, where the orange dashed line nearly resembles the blue real curve. Interestingly, even in areas with considerable volatility, the forecast is rather accurate, especially in the vicinity of the 180–260 hour period. This demonstrates how well the model handles price fluctuations in this channel.

Closed-loop Forecasting

A more practical method is closed-loop forecasting, also known as recursive prediction, in which the model forecasts future values using its historical predictions as input. Each successive prediction is based on the previously projected values rather than the ground truth after the initial forecast is established using real data.
This technique is used in your code to provide a long-term forecast without the need for any further real data. This mimics real-world deployment, in which the model must continue forecasting only based on its previous results since future data is uncertain. This also implies that any mistakes

committed at the beginning of the process might spread and get worse over time, which could cause drift or values to deviate from reality.

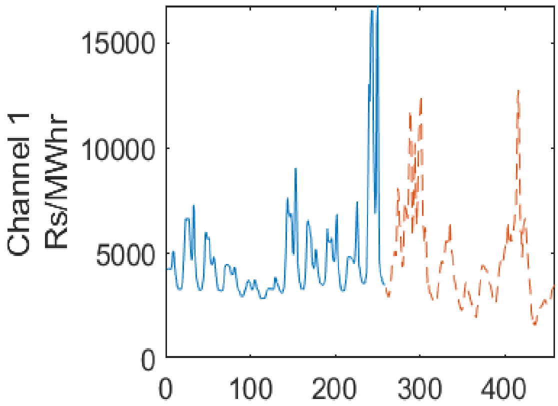

Fig.8 Closed loop prediction for Channel 1

Beyond the training data, the closed-loop prediction model in Channel 1, shown in Fig. 8, continues to anticipate electricity prices by using its own previously predicted outputs as future inputs. A believable continuation of the real price trend is indicated by the forecasted data, which specifically catches the greater variances and price peaks after 250 hours. However, exhibit some variety, particularly in the height and volatility of the peaks that span more than 300 hours.

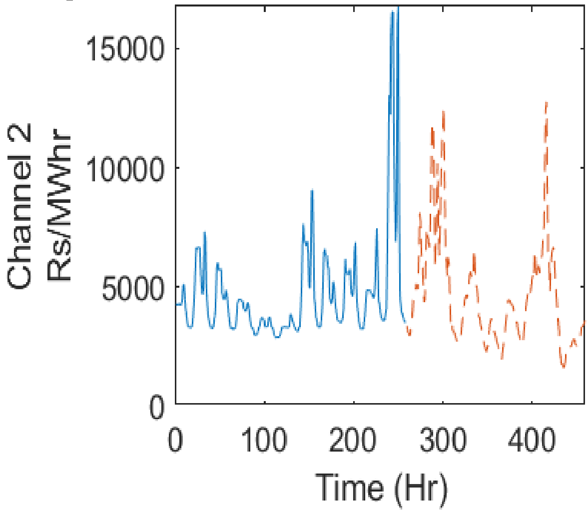

Fig.9 Closed loop prediction for Channel 2

The closed loop forecast for Channel 2 is shown in Fig. 9. Its predicted performance follows a similar pattern. The closed-loop forecast accurately reflects the overall pattern of price rises and declines, beginning where the actual data ends. As a recognized feature of recursive forecasting, compounding prediction mistakes over time cause the predicted prices to significantly underestimate the peak values, even while they preserve the right pattern of peaks and troughs.

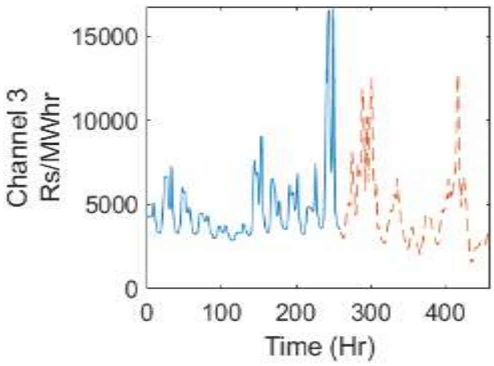

Fig.10 Closed-loop prediction for Channel 3

Regarding Channel 3, the projected prices are in good alignment with the real price trend based on the timing and form of the variations displayed in Fig. 10. However, the magnitude of some of the anticipated peaks seems rather muted in comparison to the actual data. The model shows strong temporal pattern learning, but because it relies on self-fed outputs rather than actual data, it starts to increase deviation after 300 hours, which is common in closed-loop forecasting.
The performance metrics are represented in Table I.

TABLE I: PERFORMANCE METRICS OF THE RF BASED ELECTRICITY PRICE PREDICTION

Performance measures	**Channel -1**	**Channel -2**	**Channel -3**
RMSE	0.534	0.534	0.534
MAE	0.0276	0.0275	0.0276
R^2	0.7831	0.7834	0.7831
Sensitivity	48.91	48.91	48.91

Table I presents the performance metrics of the RF based electricity price prediction model across three distinct data channels. The RMSE is consistent at 0.534 for all three channels, indicating that the model maintains a stable average deviation among the predicted and actual electricity prices regardless of the channel. Similarly, the MAE remains around 0.0275–0.0276 across channels, reflecting a low average prediction error in absolute terms. These values demonstrate that the RF model is capable of minimizing prediction discrepancies effectively and uniformly across different datasets. The R^2 values are approximately 0.783 for all three channels, indicating that around 78.3% of the variance in electricity prices is successfully explained by the model. This confirms the strong predictive capability of the RF algorithm in capturing complex price patterns. Additionally, the sensitivity score of 48.91 across all channels reflects the model's ability to correctly identify relevant price fluctuations and changes in the dataset, ensuring that critical variations in electricity pricing are not overlooked.

Fig.11 shows the comparison of perfromance metrics using RF.

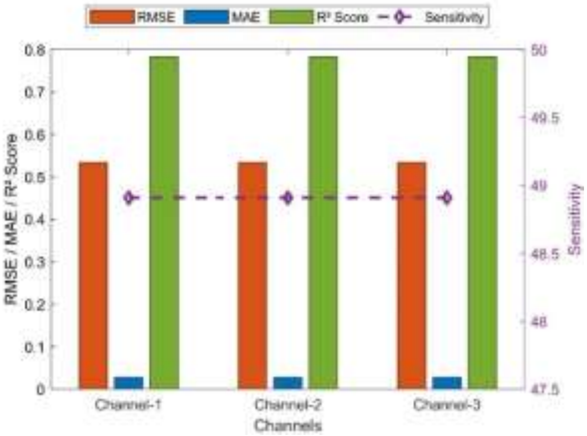

Fig.11 Representation of Performance metrics using RF

The comparative analysis of the proposed RF with SVR & ANN presented in Table II.

TABLE II COMPARISON OF PERFORMANCE METRICS WITH SVR & ANN

Performance Measures	Methods	Channel-1	Channel-2	Channel-3
RMSE	RF	0.534	0.534	0.534
	SVR	0.642	0.649	0.651
	ANN	0.598	0.603	0.606
MAE	RF	0.0276	0.0275	0.0276
	SVR	0.0368	0.0372	0.037
	ANN	0.0329	0.0335	0.0331
R^2	RF	0.7831	0.7834	0.7831
	SVR	0.6998	0.6942	0.6925
	ANN	0.7312	0.7285	0.7257
Sensitivity	RF	48.91	48.91	48.91
	SVR	44.13	43.98	43.76
	ANN	46.02	45.87	45.7

Table II provides a detailed comparison of the electricity price prediction performance across three widely used machine learning models: RF, SVR, and ANN. Among the metrics evaluated RMSE, MAE, R^2 and Sensitivity the RF model consistently delivers superior results. It achieves the lowest RMSE value of 0.534 across all channels, which indicates that its predicted prices are closest to the actual values with minimal deviation. In contrast, SVR and ANN record noticeably higher RMSE values, implying larger errors in their predictions. Similarly, RF reports the lowest MAE (ranging around 0.0275–0.0276), further confirming its accuracy in minimizing the average magnitude of prediction errors.When examining the R-squared (R^2) metric, which reflects how well the model explains the variance in electricity prices, the RF model again performs best. With values above 0.783 across all three channels it indicate that nearly 78% of the price fluctuation patterns are accurately captured by the model. This is a significant improvement compared to the SVR and ANN models, which show R^2 values in the lower 70% range. This enhanced explanatory power of RF stems from its ability to handle complex, nonlinear relationships in the data, which are typical in electricity pricing influenced by seasonality, demand, and market behavior. Sensitivity analysis further underscores the robustness of the RF model. With a consistent sensitivity value of 48.91 across all channels, RF demonstrates a stronger capability to respond to changes and peaks in the electricity pricing trend compared to SVR and ANN.

6. Conclusion

This study presented a detailed evaluation of electricity price forecasting using an RF model across three distinct data channels. Both open-loop and closed-loop forecasting techniques were employed to assess the model's short-term accuracy and long-term predictive stability. Open-loop forecasting, which uses actual historical data at each step, demonstrated high precision in tracking electricity price trends. The model effectively captured complex variations and price spikes with minimal deviation, as shown in all three channels, particularly in Channel 3, where predictive alignment was notably strong.In contrast, closed-loop forecasting, which recursively relies on the model's previous predictions, offered insight into real-world deployment conditions where future data is unavailable. While the overall trend and structure of electricity price movements were well maintained, some attenuation in peak magnitude and increased deviation were observed beyond the 300-hour mark due to error accumulation, a common characteristic of recursive forecasting. Performance metrics such as RMSE (0.534), MAE (~0.0276), and R^2 (~0.783) remained consistent across all three channels, reinforcing the model's robustness and reliability. The sensitivity value of 48.91 further indicates the model's responsiveness to variations in the input features. Overall, the Random Forest model demonstrated strong learning capability, adaptability across channels, and resilience in both forecasting scenarios. These findings underscore the suitability of RF-based forecasting models for electricity price prediction in energy management systems, offering reliable insights for planning and operational decision-making in the dynamic energy market.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.

- **Author contributions:** The authors declare that they have equal right on this paper.

- **Funding information:** The authors declare that there is no funding to be acknowledged.

- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] P. Wang, K. Xu, Z. Ding, Y. Du, W. Liu, B. Sun *et al.*, "An online electricity market price forecasting method via random forest," *IEEE Transactions on Industry Applications*, vol. 58, no. 6, pp. 7013–7021, 2022.

[2] A. Ashraf, M. Rady, and S. Y. Mahfouz, "Price prediction of residential buildings using random forest and artificial neural network," *HBRC Journal*, vol. 20, no. 1, pp. 23–41, 2024.

[3] J. Wang, X. Sun, Q. Cheng, and Q. Cui, "An innovative random forest-based nonlinear ensemble paradigm of improved feature extraction and deep learning for carbon price forecasting," *Science of the Total Environment*, vol. 762, p. 143099, 2021.

[4] D. Li, Y. Li, and C. Ma, "Research on project cost prediction based on random forest algorithm," in *Proc. 4th Int. Conf. Smart City Eng. Public Transp. (SCEPT 2024)*, SPIE, vol. 13160, pp. 152–157, May 2024.

[5] S. F. M. Alazawy, M. A. Ahmed, S. H. Raheem, H. Imran, L. F. A. Bernardo, and H. A. S. Pinto, "Explainable machine learning to predict the construction cost of power plant based on random forest and Shapley method," *CivilEng*, vol. 6, no. 2, pp. 21, 2025.

[6] V. Sessa, E. Assoumou, M. Bossy, and S. G. Simões, "Analyzing the applicability of random forest-based models for the forecast of run-of-river hydropower generation," *Clean Technologies*, vol. 3, no. 4, pp. 858–880, 2021.

[7] L. Qijie, Y. Zhifei, and J. Yuliang, "Cost analysis of power engineering based on improved random forest algorithm," in *2024 Int. Conf. Power, Electr. Eng., Electron. Control (PEEEC)*, pp. 1172–1177, IEEE, 2024

[8] G. F. Fan, M. Yu, S. Q. Dong, Y. H. Yeh, and W. C. Hong, "Forecasting short-term electricity load using hybrid support vector regression with grey catastrophe and random forest modeling," *Utilities Policy*, vol. 73, p. 101294, 2021.

[9] I. Khan, "Efficiency analysis and a random forest based trading strategy for heteroscedastic electricity market data," in *2022 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1–5, IEEE, July 2022.

[10] Z. Zheng, L. Zhou, H. Wu, and L. Zhou, "Construction cost prediction system based on random forest optimized by the Bird Swarm Algorithm," *Mathematical Biosciences and Engineering*, vol. 20, no. 8, pp. 15044–15074, 2023.

[11] G. F. Fan, L. Z. Zhang, M. Yu, W. C. Hong, and S. Q. Dong, "Applications of random forest in multivariable response surface for short-term load forecasting," *International Journal of Electrical Power & Energy Systems*, vol. 139, p. 108073, 2022..

[12] B. Han, H. Xiao, L. Yu, Z. Xu, W. Chen, and X. Sun, "Construction progress prediction of substation infrastructure project based on random forest," in *2023 18th Int. Conf. Intell. Syst. Knowl. Eng. (ISKE)*, pp. 324–330, IEEE, Nov. 2023.

[13] A. Tohry, S. C. Chelgani, S. S. Matin, and M. Noormohammadi, "Power-draw prediction by random forest based on operating parameters for an industrial ball mill," *Advanced Powder Technology*, vol. 31, no. 3, pp. 967–972, 2020.

[14] H. Zhai and J. Che, "Combining PSO-SVR and random forest based feature selection for day-ahead peak load forecasting," *Engineering Letters*, vol. 30, no. 1, 2022.

[15] K. Li, Y. Xing, H. Zhu, and W. Nai, "A short-term hybrid forecasting approach for regional electricity consumption based on grey theory and random forest," in *2020 5th Int. Conf. Comput. Intell. Appl. (ICCIA)*, pp. 194–198, IEEE, June 2020.

[16] Mei, J., He, D., Harley, R., Habetler, T., & Qu, G. (2014, July). A random forest method for real-time price forecasting in New York electricity market. In *2014 IEEE PES general meeting| conference & exposition* (pp. 1-5). IEEE.

[17] Shirzadi, N., Nizami, A., Khazen, M., & Nik-Bakht, M. (2021). Medium-term regional electricity load forecasting through machine learning and deep learning. *Designs*, *5*(2), 27.

[18] Li, C., Tao, Y., Ao, W., Yang, S., & Bai, Y. (2018). Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition. *Energy*, *165*, 1220-1227.

[19] Xu, K., Sun, B., Wang, P., Zhu, Z., & Tang, H. (2022, May). Electricity market price forecasting for a high renewable penetrated power system via random forest. In *2022 IEEE/IAS 58th Industrial and Commercial Power Systems Technical Conference (I&CPS)* (pp. 1-7). IEEE.

[20] Hanif, M., Shahzad, M. K., Mehmood, V., & Saleem, I. (2023). EPFG: electricity price forecasting with enhanced Gans neural network. *IETE Journal of Research*, *69*(9), 6473-6482.