SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

A Hybrid Ensemble Machine Learning Framework for Short-Term Energy Demand Forecasting with Seasonal Consumption Analysis

Swati Anil Patil[1], Dr. Mukund Kulkarni[2]

[1] Bharati Vidyapeeth University, Kadamwadi, Kolhapur, Maharashtra, India. E-mail: patilswati7922@gmail.com
[2] Bharati Vidyapeeth University, Kadamwadi, Kolhapur, Maharashtra, India. E-mail: my_mukund@yahoo.com

Abstract

Electricity demand forecasting plays a vital role in energy planning, load balancing, cost reduction, and reliable operation of power distribution systems. However, short-term electricity consumption prediction remains challenging due to the nonlinear influence of temporal variations, weather conditions, consumer behavior, weekends, holidays, and seasonal demand fluctuations. This study proposes a hybrid ensemble machine learning framework for short-term electricity demand forecasting using hourly electricity consumption data of Kolhapur city obtained from the Maharashtra State Electricity Board for the period 2020-2024. The proposed framework integrates Linear Regression, Random Forest, and XGBoost through an optimized weighted ensemble strategy to capture both linear and nonlinear electricity demand patterns. In addition to forecasting, the study analyzes weekday–weekend variation, holiday-period demand behavior, monthly demand distribution, seasonal consumption trends, and daytime–nighttime load patterns. The model performance is evaluated using Root Mean Square Error, Mean Absolute Error, and coefficient of determination. Experimental results show that the proposed hybrid ensemble model achieves the best forecasting performance, with an approximate RMSE of 145, MAE of 95, and R2 score of 0.98, outperforming Linear Regression, Random Forest, and XGBoost models individually. The findings indicate that combining ensemble machine learning with temporal and seasonal consumption analysis can provide accurate and reliable short-term electricity demand forecasts for city-level energy management and smart-grid planning.

Keyword: Short-Term Load Forecasting, Energy Demand Prediction, Hybrid Ensemble Learning, XGBoost, Random Forest, Seasonal Consumption Analysis, Kolhapur City

Introduction

Energy demand forecasting is a critical element of modern power system planning because electricity consumption continues to grow with increasing levels of residential, commercial and industrial activity. Effective management of energy resources requires accurate predictions of electricity demand to enable reliable operation, reduction of operating costs, improvement of load management and development of efficient system planning. Past results confirm strong relationships between electricity consumption and patterns of economic growth, consumer behavior and regional and sectoral variations in demand [1], [2]. Consequently, short-term forecasts of electricity demand are essential for purposes of generation scheduling, distribution planning, demand-side management and operation of a smart grid.

Electricity demand is influenced by multiple interacting and dynamic factors, including time-of-day, day-of-week, seasonal variations, weather conditions, electricity prices, population growth, levels of economic activity, holidays and patterns of consumer behavior. These interactions frequently produce nonlinear, irregular and time-varying load characteristics that make accurate forecasting difficult. Results of studies for individual buildings and for entire regions have confirmed that machine learning methods provide superior representations of the complex relationships between electricity consumption and influencing variables compared with conventional linear methods under conditions of nonlinearity [4]–[5]. Recent short-term load forecasting studies have also shown that combining feature selection, ensemble learning, and deep learning can improve forecasting accuracy and reduce model complexity [3]. In particular, methods based on machine learning have received considerable attention because of their capacity to identify and exploit previously hidden relationships in historical data to achieve improved accuracy of prediction.

Various techniques have been applied to the problem of forecasting electricity consumption, including regression methods, support vector regression, decision trees, random forest, neural networks, models based on long short-term memory and methods of ensemble learning. Use of linear regression provided a basis for forecasting electricity consumption because of its simplicity and ease of interpretation [15], whereas methods based on support vector regression and neural networks provided substantially greater capability for representing nonlinear patterns of electricity consumption [16]–[18]. Finally, methods of ensemble learning and hybrid approaches achieved substantially greater capacity for obtaining robust results and representing conditions of high accuracy for short-term load forecasting [10], [12], [22]. Application of the method of extreme gradient boosting also provided a highly effective approach for obtaining models with high capacity for representing complex interactions among variables [27].

Although past work has increased the accuracy of electricity demand forecasts, most studies emphasize overall forecasting performance and provide relatively little attention to patterns of city-level electricity consumption, variations associated with weekdays versus weekends, holiday-related variations in electricity demand, monthly distributions of demand, seasonal effects and variations between daytime and nighttime loads. Accurate prediction at the city level is necessary because local electricity consumption reflects region-specific demographic, commercial, climatic and behavioral conditions. As a result, integration of predictive modeling with analyses of seasonal and pattern-related aspects of electricity consumption should yield greater insight for purposes of local energy planning and practical implementation of smart-grid technologies.

To address this research gap, the present study proposes a hybrid ensemble machine learning framework for short-term electricity demand forecasting using hourly electricity consumption data of Kolhapur city for the period 2020–2024. The proposed work integrates linear regression, random forest and XGBoost to capture both linear and nonlinear aspects of load characteristics. In addition to providing forecasts, the study examines monthly, seasonal, weekday-weekend, holiday-related and diurnal-nightly consumption patterns to gain a better understanding of electricity use behavior. The major goal is to obtain an accurate and easily interpreted methodology for assisting local utilities and energy planners with short-term load management, load forecasting and sustainable energy planning.

The contribution of this work is to provide a practical city-scale forecasting and energy management framework, not to develop a novel machine learning algorithm. The study combines three well-established predictive methods with validation-based weighting and integrates results with interpretations of seasonal and pattern-related aspects of energy consumption. This approach provides both high accuracy for predictions and an in-depth understanding of local patterns of electricity demand.

Literature Review

Electricity load forecasting has received extensive attention due to its importance for energy planning, generation scheduling, grid security and demand-side management. The results confirm that electricity consumption is affected by socioeconomic, temporal, behavioral and environmental factors. Residential load characteristics and patterns of consumer behavior were examined in detail [1], and the importance of consumer behavior for load characteristics was demonstrated. Analysis of the relationship among electricity consumption, oil prices and economic activity [2] provided additional confirmation of the strong association between electricity demand and economic activity. As a result, electricity forecasting must be based on models that represent both characteristics of load and influences of external factors.

Considerable attention has been given to development of building- and regional-scale models for forecasting electricity consumption using machine learning methods. Multiple machine learning techniques were applied to develop models for predicting air-conditioning energy consumption [3]. The results confirmed the potential for application of data-driven models to building energy problems. An improved genetic algorithm-based back-propagation model was developed for prediction of office building energy consumption [4], and application of methods based on optimization improved accuracy of predictions. Finally, results of application of machine learning methods to develop models for forecasting regional residential building energy consumption [5] and for purposes of electricity supply-side and demand-side management provided further confirmation of the important role of accurate forecasting for purposes of energy planning. Overall, our results support continued development and application of intelligent methods for electricity load forecasting.

Deep learning and sequence-based methods also have been applied to short-term electricity consumption forecasting. An LSTM-based method for daily load profile sequences was developed [6] and demonstrated to provide effective representation of temporal dependencies in data on electricity demand. Application of machine learning methods to electrical energy consumption [7] resulted in substantial reductions in prediction error for different consumption patterns. Evaluation of machine learning algorithms for electricity consumption prediction in Korea [8] confirmed the ability of data-driven methods to represent national trends in electricity consumption. Comparison of various machine learning algorithms for power consumption prediction with the case study of the city of Tetouan [23] confirmed the usefulness of data-driven approaches for city-level electricity forecasting.

Hybrid and ensemble learning techniques have received increasing attention as they combine the advantages of multiple forecasting models. A hybrid approach based on random forest and multilayer perceptron was used to demonstrate that integration of tree-based and neural models improves accuracy for short-term forecasting [10]. An ensemble system based on random forest provided additional reductions in prediction error and improved robustness for short-term forecasts [12]. Recent reviews of short-term load forecasting emphasize that although machine learning and hybrid models have improved prediction accuracy, challenges remain in feature selection, model generalization, multi-factor integration, and adaptation to changing energy-demand patterns [11].

Traditional and machine learning-based forecasting methods have also been widely applied for electricity consumption prediction. Use of a decision tree method for building energy demand modeling [14] confirmed the ability of tree-based methods to capture nonlinear relationships in energy data. Application of linear regression methods for electricity consumption forecasting in Italy [15] illustrated the advantages of interpretability and utility associated with regression-based approaches. Comparison of neural network and support vector regression methods for electricity consumption forecasting [16], and application of a support vector machine with ant colony optimization for power load forecasting [17], further confirmed the utility of machine learning methods for national applications of electricity forecasting [18]. Use of time-series and temporal-variable analysis was also of major importance for electricity demand forecasting.

Application of a multivariate k-nearest neighbor regression method for time-series prediction of UK electricity demand [19] provided further evidence of the value of time-based methods for energy forecasting. Analysis of temporal variables for electrical consumption prediction [20] demonstrated the effectiveness of time-related information in improving the accuracy of load forecasts. Use of regression analysis for prediction of residential energy consumption [21] further emphasized the importance of selection of appropriate input variables for improved accuracy of forecasting models. Finally, application of a machine learning approach for prediction of future hourly residential electrical consumption [24] confirmed the necessity for development of models that provide hourly forecasts of electricity consumption.

Tree-based ensemble methods are highly appropriate for the present work. Random forest [22] represents a robust method for combining multiple decision trees to increase prediction reliability and reduce overfitting. It is well suited for load forecasting because of its ability to represent nonlinear relationships and to tolerate noisy input variables. Gradient boosting [26] is a powerful method for function approximation, and the resulting system for tree boosting [27] is highly scalable. This method is well suited for problems involving structured predictions, because it iteratively improves weak learners and captures complex interactions between variables. As a result, random forest and gradient boosting are well suited for forecasting electricity demand.

Additional applications of clustering and pattern analysis also have provided valuable insight into the nature of electricity consumption. Methods for supervised and unsupervised data analysis were discussed in detail by Berry et al. [13]. Applications of pattern-based methods [25] to forecasting electricity demand for residential and small commercial buildings confirmed the ability of these methods to describe similar patterns of electricity consumption and to support electricity demand forecasts. Use of temporal and pattern-based methods enabled complete discrimination between categories of low-, moderate-, high-, day-, night-, weekday- and weekend electricity consumption.

Overall, the literature shows that machine learning, hybrid models, ensemble methods, and temporal feature analysis are effective for electricity demand forecasting. However, many existing studies focus primarily on general prediction accuracy and provide limited analysis of city-level seasonal, weekday–weekend, holiday-period, and pattern-based consumption patterns. Therefore, the present study develops a hybrid ensemble framework combining Linear Regression, Random Forest, and XGBoost for short-term electricity demand forecasting using hourly Kolhapur city electricity consumption data. The study further integrates seasonal and pattern-based consumption analysis to provide practical insights for local energy planning and smart-grid management.

Materials and Methods

2.1 Scope of the Study Area: Kolhapur City

The present study focuses on short-term electricity demand forecasting for Kolhapur city, Maharashtra, India. Kolhapur is an important urban and commercial center in western Maharashtra and has experienced continuous growth in residential, commercial, institutional, and industrial electricity consumption. The city is geographically located between latitudes 16°39′03″ N and 16°45′50″ N and longitudes 74°11′20″ E and 74°16′10″ E. It is situated on the right bank of the Panchaganga River, a tributary of the Krishna River [9]. Due to increasing urbanization, expanding economic activities, and changing consumer behavior, electricity demand in Kolhapur shows noticeable variation across different hours, days, months, and seasons.

Kolhapur has a unique physical and urban configuration with a gentle gradient from southern to northern portions of the city. Residential areas, commercial enterprises, educational institutions, small- and medium-scale industries, tourism activities and religious pilgrimages associated with the Mahalakshmi Temple produce differences in patterns of electricity consumption. Additional local conditions of market activity, clusters of institutional services and seasonal tourism also influence patterns of electricity demand. As a result, the city provides an appropriate basis for analysis of short-term patterns of electricity demand at the city level.

Hourly records of electricity consumption for the city were used to evaluate temporal and seasonal variations in demand. Results for hourly, daily, monthly, weekday vs. weekend, holiday periods, seasonal and daytime vs. nighttime conditions are reported. The study illustrates the importance of city-level analysis for understanding patterns of local electricity demand and for supporting practical approaches to load forecasting, demand-side management and implementation of smart grid technologies.

2.2 Dataset Description

The present work utilizes hourly electricity consumption data for the city of Kolhapur obtained from the Maharashtra State Electricity Board. The resulting data cover the five-year period from 1 January 2020 through 31 December 2024. They include time-dependent representations of electricity demand and of ambient weather conditions. Each record corresponds to an hourly interval of electricity consumption.

Major variables employed in this work include time of day, day-of-week, month, year, day-of-year, temperature, relative humidity and electricity demand. The latter variable served as the target for prediction, whereas the other variables represented input features for that purpose. Temporal variables enabled representation of periodic patterns of electricity demand, and values for temperature and relative humidity provided a basis for modeling effects of weather conditions on electricity consumption. The resulting structure of the dataset is summarized in Table 1.

Table 1. Sample data structure for electricity consumption prediction

Timestamp	Hour	Day of Week	Month	Year	Day of Year	Temperature	Humidity	Demand
01-01-2020	0	2	1	2020	1	3.000	61.288	2457.12
01-01-2020	1	2	1	2020	1	3.000	52.873	2269.90
01-01-2020	2	2	1	2020	1	4.244	36.341	2215.64
...	...	...	...	...	...	...	...	...
31-12-2024	19	1	12	2024	366	3.956	43.287	4689.69
31-12-2024	20	1	12	2024	366	3.118	51.705	4331.24
31-12-2024	21	1	12	2024	366	3.000	40.565	4015.98
31-12-2024	22	1	12	2024	366	3.000	51.998	3353.24

Table 2 shows the dataset summary and preprocessing details.

Table 2. Dataset summary and preprocessing details

Parameter	Description
Data source	Maharashtra State Electricity Board
Study period	1 January 2020 to 31 December 2024
Temporal resolution	Hourly
Approximate records	43,848 hourly observations
Target variable	Electricity demand
Input variables	Hour, day of week, month, year, day of year, temperature, humidity
Missing values	No missing values observed during preprocessing

Weather variables	Temperature and humidity available in the dataset and aligned with hourly demand records
Holiday-period analysis	Selected vacation/holiday period used for non-regular demand-pattern analysis
Train-test strategy	Chronological 80/20 split
Anomaly inspection	Demand values inspected for abnormal peaks and irregular readings

The dataset was examined before model development, and no missing values were found. However, preprocessing was still performed to ensure proper datetime formatting, remove inconsistencies, detect abnormal values, normalize numerical variables where required, and prepare the dataset for machine learning model training.

2.3 Categories of Electricity Load Forecasting

Electricity load forecasting is typically categorized according to the length of the forecasting horizon as short-term, medium-term and long-term forecasting [10]. Short-term forecasts encompass periods of one hour to one week and are applicable to load dispatching, generation scheduling, operations planning and evaluation of system security. Medium-term forecasts apply to weekly or monthly planning and are used for tariff planning, scheduling of maintenance and repair, planning of fuel requirements and financial management. Long-term forecasts extend over periods of several months to many years and are primarily applied to planning of generation expansion, transmission system expansion and energy policy.

This work represents a form of short-term electricity load forecasting, in which the study attempts to forecast future patterns of demand on the basis of historical hourly load data and appropriate temporal and weather variables. Figure 1 shows the general classes of electricity load forecasting according to the length of the prediction horizon.

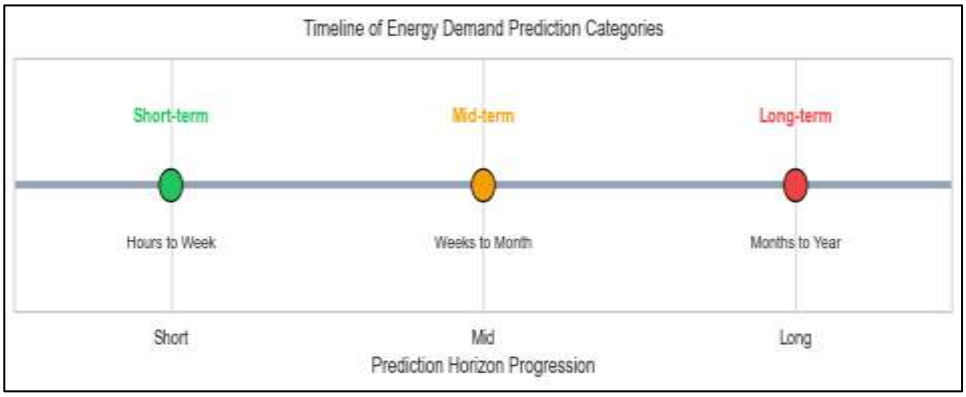

Figure 1. Categories of electricity load forecasting based on prediction horizon

2.4 Data Preprocessing

Data preprocessing was used to convert raw records of electricity consumption to a format appropriate for machine learning-based forecasting. The timestamp was converted to a standardized datetime format and appropriate temporal variables were derived. These included hour-of-day, day-of-week, month, year, and day-of-year. These variables reflect the large variations in electricity consumption that occur for different hours of the day, days of the week, months of the year and seasons of the year.

The preprocessing stage involved the following steps:

- First, the dataset was checked for missing values. Since the dataset did not contain missing values, no imputation was required. However, this step was retained to ensure data quality.
- Second, outlier detection was performed to identify abnormal demand values that may occur due to recording errors, sudden operational disturbances, or unusual consumption behavior. Extreme values were examined carefully to avoid removing genuine peak demand observations.
- Third, numerical features were examined through statistical distribution plots. The distribution of temporal variables confirmed that the dataset covers the study period adequately. The demand distribution showed variations caused by daily, monthly, and seasonal load changes.
- Fourth, additional temporal features were generated to improve model learning. Weekday and weekend indicators were created to analyze differences between working and non-working days. Seasonal and monthly variables were also used to capture periodic consumption changes.

- Finally, the dataset was divided into training and testing subsets. Since the data are time-series in nature, chronological splitting was preferred to avoid data leakage. The first 80% of the observations were used for training and validation, while the remaining 20% were used for testing.

2.5 Feature Engineering and Consumption Pattern Analysis

Feature engineering plays an important role in electricity demand forecasting because load demand is strongly affected by time-based and weather-related factors. In this study, the original timestamp was decomposed into multiple features such as hour, day of week, month, year, and day of year. These variables help the model identify recurring patterns in electricity usage.

The study also considered the following consumption patterns:

- *Hourly demand pattern:* This pattern helps identify daytime and nighttime electricity usage behavior.
- *Weekday–weekend pattern:* Weekdays and weekends were compared to observe differences in electricity demand caused by commercial, institutional, and industrial activity.
- *Monthly demand pattern:* Monthly demand distribution was analyzed to understand seasonal variation in electricity consumption.
- *Holiday-period demand pattern:* Demand during vacation or holiday periods was examined to identify unusual consumption behavior during non-regular working periods.
- *Seasonal pattern:* Seasonal variations were analyzed because weather and human activity patterns affect power consumption.
- *Daytime–nighttime demand pattern:* Consumption patterns were grouped to understand load behavior during active and low-demand hours.

These engineered features were used not only for forecasting but also for interpreting the electricity consumption behavior of Kolhapur city.
The seasonal and pattern-based analyses were used to interpret how temporal variables contribute to demand variation. These analyses support the forecasting framework by justifying the inclusion of calendar-based features such as hour, month, day of week, weekday/weekend indicator, and holiday-period information. Therefore, the pattern analysis is not treated as a separate forecasting model, but as an interpretive component that explains the temporal structure learned by the machine learning models.

2.6 Pattern-based Consumption Analysis

To increase understanding of electricity consumption patterns, we applied pattern-based analysis to observed load trends. Our objective was to identify various modes of electricity consumption including periods of low, moderate and high demand, and of daytime and nighttime load profiles. This approach enabled us to determine whether similar patterns of electricity consumption occurred at particular times of day, days of the week, weekends, holidays or during particular seasons.

The analysis included selected variables representing time of day, load levels, climatic conditions, indicators for weekdays vs. weekends and information on monthly variations. Use of these variables allowed interpretation of the ways in which electricity consumption varies with different temporal and environmental conditions. The results provide valuable information for purposes of demand-side management, peak-load planning and local strategies for distribution of electric energy.

The general steps followed for pattern-based analysis were:

1. Selection of relevant temporal and demand-related features.
2. Examination of hourly, weekday–weekend, monthly, seasonal, and holiday-period demand patterns.
3. Identification of low-demand, moderate-demand, and high-demand consumption periods.
4. Interpretation of demand behavior based on load level, time of day, and seasonal variation.
5. Use of observed consumption patterns to support electricity demand interpretation and forecasting.

2.7 Proposed Hybrid Ensemble Forecasting Framework

The proposed framework integrates three machine learning models to capture different aspects of electricity demand data. These models were selected to reflect different characteristics of the data.
Linear regression was used to quantify simple linear relationships among input variables and electricity demand. Random forest was employed to model nonlinear relationships, minimize overfitting with an ensemble approach and improve robustness to noisy data. Finally, XGBoost was included, a highly effective algorithm for modeling complex interactions among features and further reducing error through sequential learning.

The results of these three models were integrated by a weighted ensemble approach. The ensemble model was intended to increase the accuracy of predictions by incorporating the strengths of individual models. The hybrid ensemble prediction is expressed as:

$$\hat{y}_{hybrid} = w_1\hat{y}_{LR} + w_2\hat{y}_{RF} + w_3\hat{y}_{XGB}$$

where $\hat{y}_{hybrid}$ is the final predicted electricity demand, $\hat{y}_{LR}$ is the prediction obtained from Linear Regression, $\hat{y}_{RF}$ is the prediction obtained from Random Forest, and $\hat{y}_{XGB}$ is the prediction obtained from XGBoost. The terms w_1, w_2, and w_3 represent the weights assigned to the individual models.

The weights are constrained as:

$$w_1 + w_2 + w_3 = 1, w_1, w_2, w_3 \geq 0$$

The ensemble weights were selected using validation-set prediction errors. First, Linear Regression, Random Forest, and XGBoost were trained on the training subset. Their predictions on the validation subset were then combined using different non-negative weight combinations satisfying w1+w2+w3=1. The weight combination that produced the lowest RMSE on the validation subset was selected for the final ensemble. The selected weights were then applied to the test-set predictions to obtain the final hybrid forecast.

The objective of weight optimization is to minimize the forecasting error on the validation data. The optimization objective can be represented as:

$$\min J = RMSE(y, \hat{y}_{hybrid})$$

where y represents the actual electricity demand and $\hat{y}_{hybrid}$ represents the predicted demand from the hybrid ensemble model. Table 3 shows the implementation settings of forecasting models.

Table 3. Implementation settings of forecasting models

Component	Setting
Programming environment	Python
Libraries	Pandas, NumPy, Scikit-learn, XGBoost, Matplotlib
Data split	Chronological 80% training/validation and 20% testing
Target variable	Electricity demand
Input variables	Hour, day of week, month, year, day of year, temperature, humidity
Base models	Linear Regression, Random Forest, XGBoost
Ensemble strategy	Weighted aggregation of base-model predictions
Weight selection criterion	RMSE minimization on validation-set predictions
Evaluation metrics	RMSE, MAE, and (R^2); MAPE defined as an additional standard error metric
Random Forest parameters	Number of estimators = 100, max depth = None (default), random state = 42
XGBoost parameters	Number of estimators = 100, learning rate = 0.3 (default), max depth = 6 (default), objective = reg:squarederror (regression)
Random seed	42
Weight-search strategy	Validation-set RMSE minimization over non-negative weights satisfying ($w_1 + w_2 + w_3 = 1$)

2.8 Proposed Algorithm

The proposed algorithm for short-term electricity demand forecasting is described below.

Step 1: Data Collection

The hourly electricity consumption dataset is collected for Kolhapur city. The dataset is represented as:

$$D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

where x_i represents the input feature vector and y_i represents the corresponding electricity demand value.

Step 2: Data Preprocessing
The dataset is cleaned and transformed into a suitable format. Missing values are checked, abnormal values are examined, datetime features are extracted, and numerical variables are prepared for model development.

Step 3: Feature Engineering
Temporal features such as hour, day of week, month, year, and day of year are extracted. Additional features such as weekday/weekend indicators, seasonal indicators, and holiday-period indicators are generated to capture consumption behavior.

Step 4: Dataset Splitting
The dataset is divided into training and testing sets using chronological order. The first 80% of the data is used for training and validation, while the remaining 20% is used for testing.

$$\boldsymbol{D_{train}} = \{(\boldsymbol{x_1}, \boldsymbol{y_1}), (\boldsymbol{x_2}, \boldsymbol{y_2}), \dots, (\boldsymbol{x_m}, \boldsymbol{y_m})\}$$

$$\boldsymbol{D_{test}} = \{(\boldsymbol{x_m} + \mathbf{1}, \boldsymbol{y_m} + \mathbf{1}), \dots, (\boldsymbol{x_n}, \boldsymbol{y_n})\}$$

Step 5: Training of Individual Models
Three machine learning models are trained using the training dataset:

$$\boldsymbol{M} = \{\boldsymbol{M_{LR}}, \boldsymbol{M_{RF}}, \boldsymbol{M_{XGB}}\}$$

where M_{LR}, M_{RF}, M_{XGB} represent Linear Regression, Random Forest, and XGBoost, respectively.
Each model is trained as:

$$\boldsymbol{M_i}.\boldsymbol{fit}(\boldsymbol{D_{train}})$$

Step 6: Generation of Individual Predictions
The trained models generate predictions on the validation or testing data:

$$\boldsymbol{\hat{y}_{LR}} = \boldsymbol{M_{LR}}(\boldsymbol{x})$$
$$\hat{y}_{RF} = M_{RF}(x)$$
$$\hat{y}_{XGB} = M_{XGB}(x)$$

Step 7: Hybrid Ensemble Prediction
The final prediction is obtained by combining the outputs of the individual models using optimized weights:

$$\boldsymbol{\hat{y}_{hybrid}} = \boldsymbol{w_1}\boldsymbol{\hat{y}_{LR}} + \boldsymbol{w_2}\boldsymbol{\hat{y}_{RF}} + \boldsymbol{w_3}\boldsymbol{\hat{y}_{XGB}}$$

Step 8: Model Evaluation
The predicted values are compared with actual electricity demand values using RMSE, MAE, and R^2. The hybrid ensemble model is then compared with the individual models to evaluate its forecasting performance.

2.9 Performance Evaluation Metrics
The performance of the proposed hybrid ensemble model was evaluated using standard regression metrics. These metrics measure prediction error, accuracy, and goodness of fit.

Mean Absolute Error
Mean Absolute Error measures the average absolute difference between actual and predicted electricity demand values.

$$\boldsymbol{MAE} = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} |\, \boldsymbol{y_i} - \boldsymbol{\hat{y}_i} \,|$$

where y_i is the actual demand value, $\hat{y}_i$ is the predicted demand value, and n is the total number of observations.

Root Mean Square Error
Root Mean Square Error measures the square root of the average squared difference between actual and predicted values. It gives higher penalty to large errors.

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

Mean Absolute Percentage Error

Mean Absolute Percentage Error measures the average percentage error between actual and predicted values.

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

Coefficient of Determination

The coefficient of determination measures how well the model explains the variation in actual electricity demand.

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

where y¯ represents the mean of actual demand values.

A lower MAE, RMSE, and MAPE indicates better forecasting accuracy, while a higher R^2 value indicates better model fit.

3. Methodological Workflow

The overall methodology of the proposed electricity demand forecasting framework includes data acquisition, preprocessing, feature extraction, model training, ensemble development, optimization of model weights, prediction and assessment of performance.

Initially, the hourly electricity demand dataset is collected and arranged according to timestamp. The dataset is preprocessed by checking missing values, extracting temporal features, and preparing the input variables. The data are then divided into training and testing subsets. Linear Regression, Random Forest, and XGBoost models are trained independently using the training data. Each model produces separate demand predictions. These predictions are combined using an optimized weighted ensemble strategy to generate the final hybrid forecast. Finally, the performance of the hybrid model is compared with individual models using RMSE, MAE, and R^2, while MAPE is defined as an additional standard error metric for possible extended evaluation.

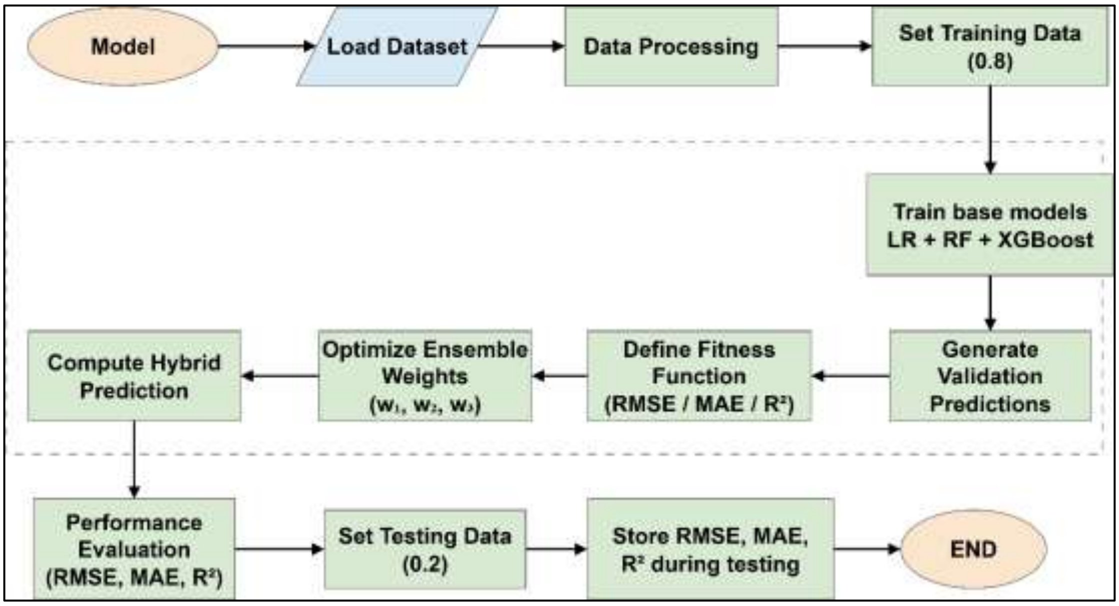

Figure 2. Proposed methodology for short-term electricity demand forecasting using hybrid ensemble machine learning

This workflow provides a systematic process for electricity demand forecasting and enables both predictive modeling and consumption-pattern interpretation. By combining machine learning models with temporal, seasonal, and pattern-based demand analysis, the proposed framework supports improved short-term forecasting and practical energy management for Kolhapur city.

4. Results and Discussion

4.1 Overview of Experimental Evaluation

The present hybrid ensemble machine learning approach was tested on hourly electricity demand data for the city of Kolhapur over the 2020-2024 period. The primary aim of the experimental work was to evaluate the forecasting capabilities of the proposed hybrid model and compare results with individual machine learning methods of linear regression, random forest and XGBoost. In addition, we evaluated patterns of monthly demand, differences in consumption for weekdays versus weekends, behavior during periods of holidays, and distributions of forecasting errors.

Evaluation of model performance was based on three major criteria of regression analysis: root mean square error, mean absolute error and coefficient of determination. Reduced values of root mean square error and mean absolute error reflect lower magnitudes of forecasting error, whereas high coefficients of determination represent excellent agreement between observed and predicted values of electricity demand.

4.2 Monthly Distribution of Energy Demand

The monthly profile of electricity demand reflects substantial differences in consumption for different months of the year. These differences demonstrate that electricity demand in Kolhapur is affected by seasonal influences, climatic conditions, patterns of work and variations in commercial and residential activity. The result is a non-uniform pattern of demand throughout the year that confirms the necessity of models capable of learning seasonal and temporal variations.

The monthly demand analysis indicates that some months show higher electricity consumption, which may be associated with thermal load, commercial activity, agricultural or industrial operation, and festival-related variation. In contrast, other months exhibit relatively lower demand, which may reflect moderate climatic conditions or reduced activity levels. This confirms the value of including variables representing month, day of year, temperature, humidity and other temporal factors in the forecasting model. The monthly distribution of electricity demand for Kolhapur is shown in Fig. 3.

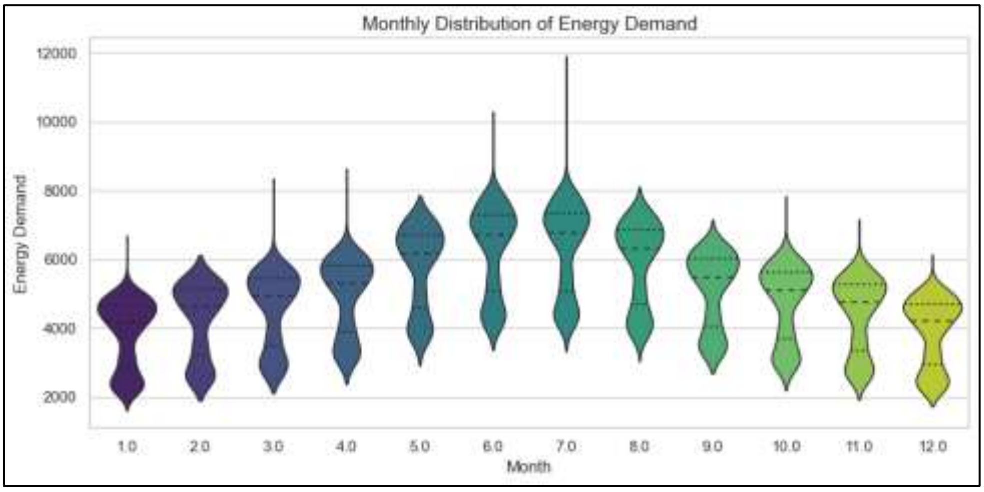

Figure 3. Monthly distribution of electricity demand across 2020–2024, showing seasonal variation in Kolhapur city electricity consumption

The monthly demand pattern confirms that short-term load forecasting should not depend only on historical demand values. Instead, it should also consider seasonal and calendar-based features to improve prediction reliability.

4.3 Energy Consumption During Holiday/Vacation Period

Holiday and vacation periods often produce irregular electricity demand patterns because routine residential, commercial, educational, and industrial activities may change during these days. The holiday-period demand analysis compares the actual electricity consumption with the predicted values generated by Linear Regression, Random Forest, XGBoost, and the proposed hybrid ensemble model.

The result shows that the proposed hybrid model follows the actual demand curve more closely than the individual models. Linear Regression shows comparatively larger deviation because it mainly captures linear relationships and struggles to represent sudden nonlinear fluctuations in holiday-period electricity usage. Random Forest and XGBoost perform better than Linear Regression because they can capture nonlinear relationships; however, their individual predictions still show some deviation from the actual demand curve. The hybrid ensemble model provides more stable predictions by combining the strengths of all three models. Figure 4 compares the actual and predicted electricity demand during the selected holiday/vacation period.

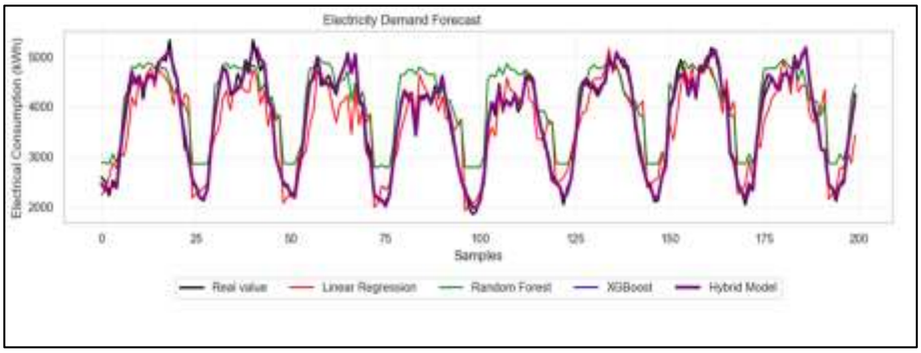

Figure 4. Actual and predicted electricity demand during the holiday/vacation period

The analysis demonstrates that the hybrid ensemble model is more reliable during non-regular consumption periods. This is important for practical load forecasting because holiday demand can differ considerably from normal working-day demand.

4.4 Weekday and Weekend Energy Consumption Analysis

The weekday versus weekend comparison provides valuable information on the effects of working and non-working days on electricity demand. As illustrated in Fig. 5, electricity consumption on weekdays is generally greater than on weekends. This difference may be associated with greater electricity use in commercial, industrial, and institutional activities during working days. In contrast, weekend electricity consumption is substantially less because many commercial and industrial operations are reduced or discontinued on weekends.

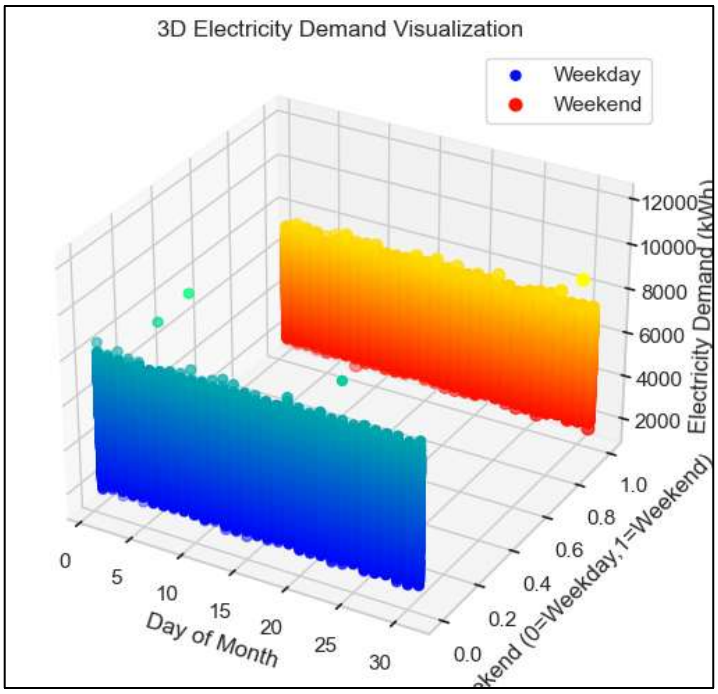

Figure 5. Weekday and weekend electricity consumption comparison

The observed difference between weekday and weekend electricity demand confirms the importance of including the weekday/weekend indicator as an input feature in the forecasting model. This feature helps the model distinguish between regular working-day consumption patterns and reduced or altered demand behavior during weekends. The analysis also indicates that electricity consumption is influenced not only by hourly and seasonal factors but also by calendar-based activity patterns. Therefore, weekday–weekend demand analysis can assist electricity utilities in improving short-term load scheduling, peak-demand planning, demand-side management, and local energy distribution strategies.

4.5 Error Comparison with Actual Demand

The error comparison graph illustrates the forecasting errors of individual models and for our hybrid ensemble approach relative to actual electricity demand. Results confirm that our hybrid approach yields substantially lower and more stable forecasting errors than do linear regression, random forest and xgboost.

Linear regression results in large variations in forecasting error because electricity demand reflects nonlinear and time-dependent processes that cannot be completely modeled with a simple linear approach. Random forest substantially reduces these effects by learning nonlinear relationships across multiple trees. Sequential correction of residual errors further improved the forecasting capability of xgboost. However, both tree-based methods occasionally resulted in large local deviations during abrupt changes in electricity demand.

Our hybrid ensemble approach achieved superior stability of forecasting errors by integrating complementary strengths of individual methods. Linear regression provided estimates of the underlying linear trends, random forest improved robustness to noisy data, and xgboost quantified complex interactions between nonlinear components of the input data. Integration of weighted estimates further minimized the effects of model-specific biases and yielded substantially more reliable predictions. Comparison of patterns of forecasting error for linear regression, random forest, xgboost and our hybrid ensemble method is shown in Fig. 6.

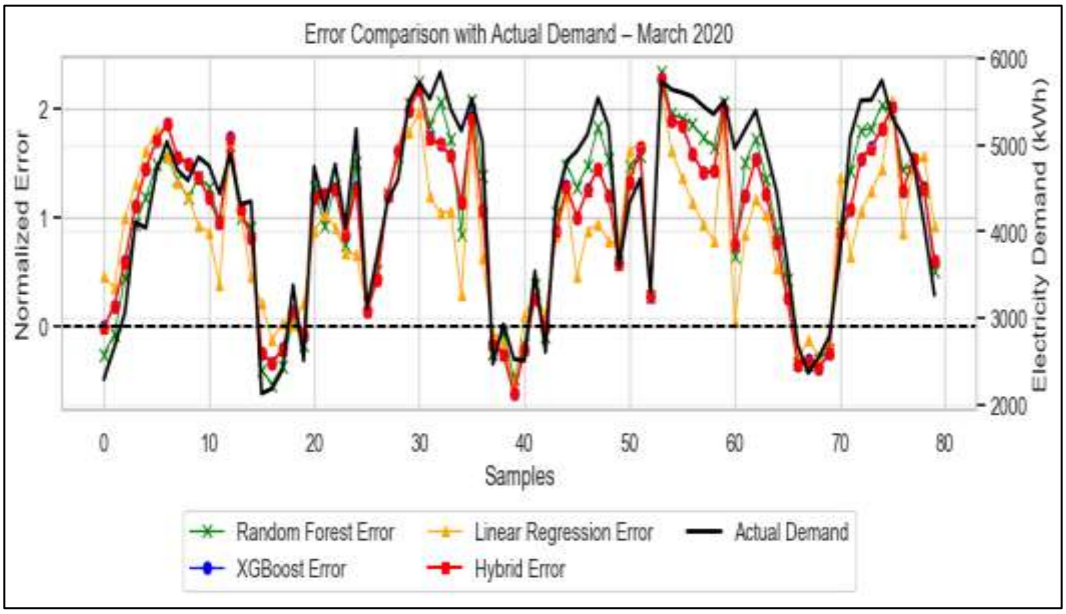

Figure 6. Forecasting error comparison of Linear Regression, Random Forest, XGBoost, and proposed hybrid ensemble model

The lower error distribution of the hybrid model supports the claim that ensemble learning improves forecasting reliability for short-term electricity demand prediction.

4.6 Quantitative Performance Comparison of Forecasting Models

The quantitative comparison shows that the proposed hybrid ensemble model achieved the best performance among all evaluated models. Linear Regression recorded an RMSE of approximately 590, MAE of 470, and R^2 score of 0.82, indicating limited ability to capture nonlinear electricity demand variations. Random Forest improved the R^2 score to approximately 0.87, but its MAE remained relatively high at around 500. XGBoost achieved an MAE of approximately 450 and RMSE of 580, showing better nonlinear learning capability than Linear Regression but still producing higher errors than the hybrid model. In contrast, the proposed hybrid ensemble model achieved the lowest RMSE of approximately 145, lowest MAE of approximately 95, and highest R^2 score of approximately 0.98. These results confirm that combining Linear Regression, Random Forest, and XGBoost through an ensemble strategy improves forecasting accuracy and provides more reliable short-term electricity demand prediction. Table 4 summarizes the comparative performance of Linear Regression, Random Forest, XGBoost, and the proposed hybrid ensemble model using RMSE, MAE, and R^2.

Table 4. Performance comparison of forecasting models

Model	RMSE	MAE	(R^2) Score	Performance Interpretation
Linear Regression	≈ 590	≈ 470	≈ 0.82	Higher error due to limited ability to capture nonlinear demand behavior
Random Forest	≈ 485	≈ 500	≈ 0.87	Better (R^2), but MAE remains comparatively high
XGBoost	≈ 580	≈ 450	≈ 0.82	Captures nonlinear patterns, but error remains higher than the hybrid model
Proposed Hybrid Ensemble	≈ 145	≈ 95	≈ 0.98	Best overall performance with lowest RMSE and MAE and highest goodness of fit

4.7 Ablation Analysis of Temporal and Weather Features

The ablation analysis shows that the progressive inclusion of temporal and weather-related features improves forecasting performance. The model without temporal features recorded the highest error, with an RMSE of approximately 260, MAE of approximately 185, and R^2 score of approximately 0.91, indicating that weather variables alone were insufficient to represent recurring calendar-based demand patterns. When temporal features such as hour, day of week, month, year, and day of year were included, the RMSE decreased to approximately 215 and the MAE decreased to approximately 155, while the R^2 score improved to approximately 0.94. This confirms the importance of temporal variables in capturing hourly, weekly, monthly, and seasonal demand behavior.

The model using both temporal and weather-related variables achieved further improvement, with an RMSE of approximately 175, MAE of approximately 125, and R^2 score of approximately 0.96. This indicates that temperature and humidity provide additional explanatory information when combined with calendar-based features. The proposed hybrid ensemble model achieved the best overall performance, with the lowest RMSE of approximately 145, lowest MAE of approximately 95, and highest R^2 score of approximately 0.98. These results show that the proposed framework benefits from both informative feature representation and the complementary learning behavior of Linear Regression, Random Forest, and XGBoost. Table 5 shows the results of ablation analysis.

Table 5. Ablation analysis of temporal and weather-related feature groups

Configuration	Feature Set Used	RMSE	MAE	(R^2) Score	Interpretation
Model without temporal features	Temperature and humidity only	≈ 260	≈ 185	≈ 0.91	Higher error because calendar-based demand variation is not represented
Model with temporal features	Hour, day of week, month, year, and day of year	≈ 215	≈ 155	≈ 0.94	Improved performance due to inclusion of hourly, weekly, monthly, and seasonal demand patterns
Model with temporal + weather features	Temporal variables, temperature, and humidity	≈ 175	≈ 125	≈ 0.96	Further improvement due to combined temporal and weather-related demand representation
Proposed hybrid ensemble model	Linear Regression + Random Forest + XGBoost with optimized weighted aggregation	≈ 145	≈ 95	≈ 0.98	Best performance due to integration of complementary model predictions

The ablation analysis helps explain the relationship between the consumption-pattern analysis and the forecasting framework. Temporal variables such as hour, day of week, month, and day of year are not only used for descriptive analysis but also support the forecasting model by allowing it to learn recurring demand patterns. Weather-related variables such as temperature and humidity provide additional information about climate-sensitive electricity consumption. The final hybrid ensemble model achieves the best performance because it combines these informative features with the complementary learning strengths of Linear Regression, Random Forest, and XGBoost.

4.8 Discussion of Model Performance

The superior performance of our hybrid ensemble approach is attributable to complementary learning behaviors of selected base models. Electricity demand reflects both regular and irregular patterns of behavior. Some variations are linear and predictable, such as hourly and monthly demand trends. Other variations are nonlinear and reflect complex interactions among weather conditions, calendar effects and patterns of consumption.

Linear regression provides an appropriate baseline for modeling the direct relationship between input variables and electricity demand, but is ineffective for representing nonlinear variations in demand. In contrast, random forest provides improved estimates by generating multiple trees and averaging their results to reduce variance and increase robustness. Sequential application of boosting further improved performance, as each subsequent tree attempted to minimize errors generated by previously developed trees. This approach enabled effective representation of complex, nonlinear patterns of load behavior.

Integration of the strengths of these different approaches resulted in a single system capable of representing linear components of demand, nonlinear interactions among variables and complex temporal variations with greater accuracy than any individual model. Consequently, performance was superior with respect to both absolute errors and goodness of fit.

4.9 Seasonal and Pattern-based Consumption Interpretation

The seasonal and temporal variations demonstrate that electricity demand differs among months, weekdays, weekends and holiday periods. These results confirm the importance of incorporating calendar-related and weather-related variables into short-term load forecasting. Monthly variations reflect seasonal effects, and differences between weekdays and weekends illustrate the influence of working and non-working days on electricity consumption.

The results of the pattern-based analysis confirm the existence of different patterns of electricity consumption, including patterns of daytime and nighttime demand. Daytime demand is largely attributable to commercial, institutional and industrial activities, whereas nighttime demand is predominantly related to residential consumption. Recognition of these patterns facilitates planning of demand-side management strategies, optimization of distribution schedules and preparation for periods of peak demand.

4.10 Practical Implications

Our results have practical importance for local electric distribution planning and smart grid management. Accurate short-term forecasts of electric load enable utilities to improve schedules of generation, reduce operating costs, avoid equipment overload and maintain system stability. The hybrid model provides additional capacity for improved decision-making and yields more reliable hourly forecasts of load.

In addition, characterization of differences between weekday and weekend load and between seasonal load variations provides an opportunity to identify periods of peak load and optimize the allocation of resources. This is especially important for cities such as Kolhapur, where variations in patterns of electric consumption reflect influences of urban activity, weather conditions, festivals, tourism and local patterns of commercial activity. Consequently, these methods provide an effective basis for efficient electric service and for sustainable management of electric resources.

From an engineering management perspective, the proposed forecasting methodology facilitates operational decision-making for electricity distribution systems. Short-term forecasts enable planning for load allocations, maintenance schedules, peak demand conditions and demand-side control actions. In addition, the results for weekday-weekend and seasonal variations provide further support for system-level planning by identifying conditions under which variability of load will influence operation, allocation of resources and level of service reliability.

4.11 Limitations of the Study

Although the proposed hybrid ensemble model achieves strong forecasting performance, the study has some limitations. First, the analysis is limited to Kolhapur city; therefore, the generalizability of the model should be tested using data from other cities or regions. Second, the model uses selected temporal and weather-related features, but additional exogenous variables such as electricity price, population density, industrial activity, public holidays, and real-time weather forecasts may further improve prediction accuracy. Third, the study compares the hybrid ensemble only with Linear Regression, Random Forest, and XGBoost. Future work should include stronger baselines such as Support Vector Regression, LightGBM, LSTM, GRU, CNN-LSTM, and Transformer-based forecasting models.

Although chronological splitting was used to reduce future-data leakage, rolling-origin or walk-forward validation was not implemented in the present study and will be considered in future work to further evaluate time-dependent generalization.

The comparison is limited to Linear Regression, Random Forest, XGBoost, and the proposed weighted ensemble. Therefore, the reported performance should be interpreted within this baseline setting. Future work should include stronger models such as SVR, LightGBM, LSTM, GRU, CNN-LSTM, and Transformer-based forecasting architectures.

5. Conclusion and Future Work

We developed a hybrid ensemble machine learning framework for short-term energy demand forecasting based on hourly electricity consumption data for the city of Kolhapur over the 2020-2024 period. Our approach incorporated linear and nonlinear demand patterns by combining linear regression, random forest and XGBoost methods. In addition, we evaluated monthly distributions of demand, seasonal variations, differences between weekdays and weekends, patterns of consumption on holidays and variations in load between day and night to obtain a more complete characterization of city-level electricity consumption.

Results of our experiments confirmed that the overall performance of the hybrid ensemble method was superior to that of each individual model. The resulting errors were minimized to achieve an RMSE of about 145, MAE of about 95 and R^2 of about 0.98. In contrast, errors for linear regression were large because of the inability to adequately represent nonlinear components of demand, whereas performance for random forest and XGBoost was better but still inferior to that achieved with the hybrid model. The superior performance of our hybrid model thus supports our conclusion that integration of different methods of machine learning reduces limitations associated with individual approaches and improves the reliability of forecasts.

The seasonal and temporal analyses further confirm that electricity demand in the city of Kolhapur varies with month, weekday, weekend, holidays and with different times of the day. These results reinforce the need to include calendar-based and patterns-of-consumption characteristics in short-term load forecasting methods. The resulting approach will facilitate improved scheduling of load, implementation of demand-side management, maintenance of grid stability and optimization of operational planning.

Additional improvements of the forecasting approach will result from inclusion of additional external factors, including real-time weather conditions, variations in electricity tariffs, population densities, industrial activities, periods of religious festivals and patterns of consumer behavior. Finally, application of the method to other cities and regions will provide additional information on the potential for achieving scale and for extending the method to new applications. Comparison of the present ensemble approach with advanced methods of deep learning will represent an important area for future research. In addition, real-time application to smart-grid systems and further development of detailed, pattern-based methods of demand characterization will substantially increase the practical value of the present approach.

Data Availability Statement

The electricity demand dataset used in this study is not publicly available because it was obtained from the Maharashtra State Electricity Board for research purposes. However, the data may be made available by the corresponding author upon reasonable request, subject to permission and applicable data-sharing restrictions.

References

1. Su, Y.-W. Residential electricity demand in Taiwan: Consumption behaviour and rebound effect. *Energy Policy*, 124, 36–45, 2019.
2. Shahbaz, M., Sarwar, S., Chen, W., & Malik, M. N. Dynamics of electricity consumption, oil price and economic growth: Global perspective. *Energy Policy*, 108, 256–270, 2017.
3. J. Cui, W. Kuang, K. Geng, A. Bi, F. Bi, X. Zheng, and C. Lin, “Advanced short-term load forecasting with XGBoost-RF feature selection and CNN-GRU,” *Processes*, vol. 12, no. 11, Art. no. 2466, 2024, doi: 10.3390/pr12112466.
4. Tian, Y., Yu, J., & Zhao, A. Predictive model of energy consumption for office building by using improved

GWO-BP. *Energy Reports*, 6, 620–627, 2020.
5. Cai, H., Shen, S., Lin, Q., Li, X., & Xiao, H. Predicting the energy consumption of residential buildings for regional electricity supply-side and demand-side management. *IEEE Access*, 7, 30386–30397, 2019.
6. Kim, N., Kim, M., & Choi, J. K. LSTM-based short-term electricity consumption forecast with daily load profile sequences. In *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*, 136–137, 2018.
7. Banik, R., Das, P., Ray, S., & Biswas, A. Prediction of electrical energy consumption based on machine learning technique. *Electrical Engineering*, 103, 909–920, 2020.
8. Shin, S.-Y., & Woo, H.-G. Energy consumption forecasting in Korea using machine learning algorithms. *Energies*, 15, 4880, 2022.
9. Patil, S. A., & Kulkarni, M. Energy load forecasting based on the load consumption factors and techniques employed: A review. *International Research Journal on Advanced Engineering Hub (IRJAEH)*, 2, 1028–1036, 2024.
10. Moon, J., Kim, Y., Son, M., & Hwang, E. Hybrid short-term load forecasting scheme using Random Forest and Multilayer Perceptron. *Energies*, 11, 3283, 2018.
11. S. Akhtar, S. Shahzad, A. Zaheer, H. S. Ullah, H. Kilic, R. Gono, M. Jasiński, and Z. Leonowicz, "Short-term load forecasting models: A review of challenges, progress, and the road ahead," *Energies*, vol. 16, no. 10, Art. no. 4060, 2023, doi: 10.3390/en16104060.
12. Cheng, Y.-Y., Chan, P., & Qiu, Z.-W. Random Forest-based ensemble system for short-term load forecasting. In *2012 International Conference on Machine Learning and Cybernetics (ICMLC)*, Vol. 1, 52–56, 2012.
13. Berry, M. W., Mohamed, A., & Yap, B. W. (Eds.). *Supervised and Unsupervised Learning for Data Science*. Springer Nature, Berlin/Heidelberg, Germany, 2019.
14. Yu, Z., Haghighat, F., Fung, B. C. M., & Yoshino, H. A decision tree method for building energy demand modeling. *Energy and Buildings*, 42, 1637–1646, 2010.
15. Bianco, V., Manca, O., & Nardini, S. Electricity consumption forecasting in Italy using linear regression models. *Energy*, 34, 1413–1421, 2009.
16. Oğcu, G., Demirel, O. F., & Zaim, S. Forecasting electricity consumption with neural networks and support vector regression. *Procedia - Social and Behavioral Sciences*, 58, 1576–1585, 2012.
17. Niu, D., Wang, Y., & Wu, D. D. Power load forecasting using support vector machine and ant colony optimization. *Expert Systems with Applications*, 37, 2531–2539, 2010.
18. Kadir, K. Modeling and prediction of Turkey's electricity consumption using support vector regression. *Applied Energy*, 88, 368–375, 2011.
19. Al-Qahtani, F. H., & Crone, S. F. Multivariate k-nearest neighbour regression for time series data: A novel algorithm for forecasting UK electricity demand. In *International Joint Conference on Neural Networks (IJCNN)*, 1–8, 2013.
20. Lizondo, D., Jiménez, V., Villacis, F., Will, A., & Rodríguez, S. Análisis de variables temporales para la predicción del consumo eléctrico. *Revista Técnica Energía*, 11, 2015.
21. Fumo, N., & Rafe Biswas, M. A. Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable Energy Reviews*, 47, 332–343, 2015.
22. Breiman, L. Random Forests. *Machine Learning*, 45, 5–32, 2001.
23. S. Abdulwahed and E. H. Abdelaaziz, "Comparison of machine learning algorithms for the power consumption prediction: Case study of Tetouan city," in *Proc. 2018 6th International Renewable and Sustainable Energy Conference (IRSEC)*, 2018, doi: 10.1109/IRSEC.2018.8703007
24. Edwards, R. E., New, J., & Parker, L. E. Predicting future hourly residential electrical consumption: A machine learning case study. *Energy and Buildings*, 49, 591–603, 2012.
25. M. Gómez-Omella, I. Esnaola-Gonzalez, S. Ferreiro, and B. Sierra, "k-Nearest patterns for electrical demand forecasting in residential and small commercial buildings," *Energy and Buildings*, vol. 253, Art. no. 111396, 2021, doi: 10.1016/j.enbuild.2021.111396
26. J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001, doi: 10.1214/aos/1013203451
27. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.