İSTANBUL UNIVERSITY CERRAHPAŞA

ORIGINAL ARTICLE
Electrica 2024; 24(2): 272-283

Short-Term Electricity Consumption Forecasting at Residential Level Using Two-Phase Hybrid Machine Learning Model

Alisha Banga, S. C. Sharma

Electronics and Computer Discipline IIT Roorkee Saharanpur Campus, Saharanpur, India

Cite this article as: A. Banga and S. C. Sharma, "Short-term electricity consumption forecasting at residential level using two-phase hybrid machine learning model," *Electrica*, 24(2), 272-283, 2024.

ABSTRACT

The need for electricity has increased as the population, electrical appliances, electric cars, and industrialization have grown. Therefore, an accurate short-term electricity forecasting is required because it is helpful in day-to-day scheduling activities of utility companies like transmission and generation of electric energy. It can make the power grid safe, reduce electricity production costs, and fulfil user needs and economic and environment benefits. A single model may not be able to solve the electricity consumption problem because it contains linear and non-linear data. In this study, a two-phase hybrid machine learning model is developed for electricity utilisation prediction at a residential level. In the first phase, two algorithms, namely extreme gradient-boosting and linear regression are combined to learn the trend, seasonality, randomness, and cyclic components of the data. In the second phase, a voting ensemble model optimized is applied considering the best three models from 11 baseline models and weight parameter of the models is optimized using genetic algorithm. The developed model outperformed all models (baseline and state-of-the-art) considering four performance parameters over the individual household electric power consumption dataset. The proposed model has given Mean Squared Error (MSE) value of 0.025, Room Mean Squared Error (RMSE) value of 0.162, Mean Absolute Error (MAE) value of 0.129, and Mean Absolute Percentage Error (MAPE) value of 15.61 on a daily-level dataset. The proposed model has given a 0.159 MSE value, 0.387 RMSE value, 0.283 MAE value, and 25.07 MAPE value on an hourly-level dataset. Analysis of variance one-way statistical test is applied to show that results are statistically significant.

Index Terms— Electricity load forecasting, ensemble voting regressor, genetic algorithm, GRU, LSTM

I. INTRODUCTION

Electricity is today's world's lifeline. Almost every modern-day activity depends on electricity. Due to the increase in population, electrical appliances, electric vehicles, and industrialization, the electricity demand has increased [1]. As per World Energy Outlook 2021, by 2050, the world electricity demand will increase by almost 80% [2]. Residential electricity consumption accounts for 26.9% of overall electricity consumption [3] and substantially impacts total power usage. The total electricity produced over a period should be equal to electricity consumption or electricity wastage due to technical or non-technical losses [4]. If this balance is not maintained for a while, the power grid can even collapse or face security issues [5]. Electricity demand forecasting can help in deciding how much electricity should be produced in the future. Accurate electricity forecasting can help in providing a stable power supply and have environmental benefits. Ten thousand megawatts of electricity and 1.6 million dollars can be saved if the error is reduced by 1% in the prediction model [6].

Electricity consumption forecasting can be performed at four levels: short-term, real-time, medium-term, and long-term. Generally, real-time load forecasting is over 1 minute; the short term is over 1 hour to 1 week; medium term is over 1 to 10 weeks, and the long term is for 1 to 20 years. Accurate short-term electricity forecasting is helpful in day-to-day scheduling activities of utility companies like transmission and generation of electric energy. It can make the power grid safe, reduce electricity production costs, and fulfill user needs and economic benefits [7]. Medium-term forecasting is useful in deciding how much fuel needs to be purchased, planning maintenance activities, and utility assessment. Long-term forecasting is useful in determining strategic planning, how many new power generation houses need to be constructed, and the changes required in the supply process. Short-term electricity forecasting is difficult due to external factors like holidays, weather, temperature, economy, etc., which play an important

Corresponding author:
Alisha Banga

E-mail:
alishabanga47@gmail.com

Received: June 10, 2023

Revision Requested: November 24, 2023

Last Revision Received: October 30, 2023

Accepted: December 12, 2024

Publication Date: April 17, 2024

DOI: 10.5152/electrica.2024.22107

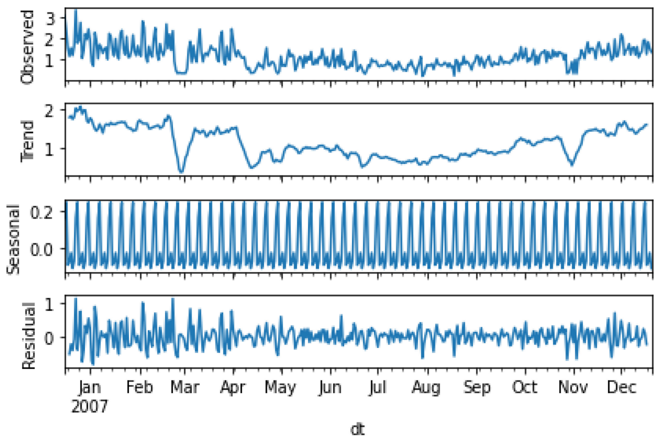

Fig. 1. Different components of electricity consumption parameter (global active power).

role in electricity consumption. Electricity consumption forecasting is a multivariate time series problem because power consumption depends on multiple variables. The data collected through sensors can have missing values, redundancy, uncertainty, etc. [8, 9]. The electricity consumption pattern can be decomposed into four components: seasonal, residual, observed, and trend factor, as shown in Fig. 1. Time series decomposition gives helpful visuals for better understanding the difficulty of analyzing and forecasting energy use in general.

Due to advancements in technology like IoT and smart meters, electricity consumption data can be monitored easily. Data collected through smart meters can be processed using Machine Learning (ML) models to predict electricity demand [10, 11]. In the current scenario, mainly three types of methods, namely, machine learning, statistical, and deep learning, are applied for the prediction of electricity consumption. Statistical methods like AutoRegressive Integrated Moving Average (ARIMA), Seasonal AutoRegressive Integrated Moving Average (SARIMA), exponential smoothing, etc. assume a linear relationship between time series data, which is not the case in electricity consumption data. These models can learn features related to time series but cannot learn nonlinear features, and the generalization ability of unseen data is not good [12]. Electricity consumption prediction based on machine learning models can capture nonlinear features efficiently and generalize well to unseen data. Many researchers in the past have used machine learning and deep learning models for electricity consumption tasks. The XGBoost model performed better than machine learning and deep learning models in electricity consumption prediction tasks [13]. The limitation of the XGBoost model is that it cannot extrapolate trend components of electricity consumption data. The ensemble model performs better than the single regression model [14] because the regression model has to perform a local search to minimize the objective function. An individual model has a chance of being caught in local minima while multiple models have a lower probability of being stuck in local minima because each model has different starting points. Voting is a superior ensemble strategy for the electricity consumption forecasting problem [15].

In this study, we have proposed a two-phase hybrid machine learning model. In the first phase, the linear regression and XGBoost models are combined, linear regression to learn the trend component and the XGBoost model to learn the remaining three components. In the second phase, the genetic algorithm optimizes a voting ensemble regression model, considering the three models out of 11 baseline models, and applies the hybrid model designed in phase one.

The major contributions are as follows:

(i) The dataset is cleaned by replacing missing values with the backward direction interpolation method, and outliers are detected and removed by the interquartile range method. Min-max scaler is used to scale the data from 0 to 1, and the data is resampled at an hourly and daily level.
(ii) A novel two-phase hybrid model is developed to forecast electricity utilization at the short-term level in residential buildings. At the first level, linear regression and gradient-boosting are combined to learn the different components (trend, seasonality, randomness, and noise) of the electricity consumption dataset. In the second level, the three models from the 11 baseline models and the hybrid model designed in phase 1 are combined using the voting ensemble technique, and the weight hyperparameter of the models is tuned with the genetic algorithm.
(iii) The proposed hybrid model outperformed 11 baseline models (eight machine learning and three deep learning) and existing models at daily and hourly levels considering performance parameters: MAE, RMSE, MSE, and MAPE.

The paper is organized into the following sections: relevant research on electricity consumption is presented in section II. The developed approach is given in section III. The dataset description, its cleaning, experimental setup, evaluation measures, comparison with baseline models and state-of-the-art approaches, discussion, and limitations are given in section IV. The conclusion of this study and future scope are given in section V.

II. RELATED WORK

Many techniques have been developed for electrical energy consumption forecasting in the past. These have been divided into mainly two kinds: (i) machine learning-based and (ii) deep learning-based. These two types of techniques are as follows:

A. Machine Learning Based

Banga, Alisha, et al. [16] have applied 15 models (machine learning and classical) to forecast the electricity consumption of household appliances at the hourly and daily levels. They found that the stacking ensemble model have outperformed all models. Fumo, Nelson, et al. [17] have applied simple quadratic regression models and multiple linear regression for energy consumption prediction of a house at the daily and hourly levels. They concluded that the data's time range affects the model's performance. They have found that temperature and solar radiation are important features in evaluating the model's efficacy. Vantuch, Tomas, et al. [18] have applied five machine learning models, XGBoost, ANN, SVR, RF, and FNT, to forecast electricity load at short-term and long-term levels. They extracted features from the dataset using the Pearson correlation coefficient, analog ensemble application, and maximal information coefficient. They have also considered temperatures outside the building as a feature. They have used the grid search technique for hyperparameter tuning. They have concluded from the results that the XGBoost model is the best-performing model among all models applied.

B. Deep Learning Models

Kim, Tae-Young, et al. [19] have proposed a hybrid model involving Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) models to predict electricity consumption at the residential level. CNN is used to extract features affecting energy consumption. The LSTM model is used to model temporal data with erratic trends. They have considered four machine learning (LR, DT, RF, and MLP) and four deep learning algorithms (LSTM, GRU, Bi-LSTM, and Attention LSTM) as baseline models. Ozcan, Le, Tuong, et al. [20] have proposed a hybrid model involving CNN and Bi-LSTM models called EECP-CBL to forecast electricity consumption at the residential level. They have compared their approach with linear regression, LSTM, and CNN-LSTM models. Their model outperformed all other models applied in their study. Kim, Jin-Young, et al. [21] have proposed a model in which energy demand is predicted based on the current situation using the autoencoder model. The model consists of a projector that defines the system state based on the current situation and a predictor, which anticipates the energy demand based on the defined state. They have explained the results using the t-SNE algorithm by visualization of the state. They have compared their model with LR, MLP, RF, DT, LSTM, and stacked LSTM. Bu Seok-Jun et al. [22] have proposed a model involving CNN, LSTM, and a multi-head attention mechanism. Multi-headed attention is used to extract spatiotemporal features. They have compared their model with Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), Support Vector Regression (SVR), ARIMA, MLP, Convolutional Neural Networks- Long Short Term Memory (CNN-LSTM), CNN, LSTM, and approaches reported in the past. They have computed the impact of different attention mechanisms on various neural networks. Sajjad, Muhammad, et al. [23] have proposed a model involving CNN and GRU, called the CNN-GRU model, to forecast electricity consumption at the short-term level. They have compared their approach with LR, SVR, DT, CNN, LSTM, and CNN-LSTM models. They have considered two standard datasets, namely, individual household electric power consumption (IHEPC) and American Electric Power (AEP). Their approach outperformed all models applied in their study and models reported in the literature. Khan, Zulfiqar Ahmad, et al. [24] have proposed a model involving CNN and Long Short Term Memory Autoencoder (LSTM-AE) models called CNN-LSTM-AE to forecast electricity consumption in commercial and residential buildings. The CNN model extracts the features from the dataset, which are passed to the LSTM encoder, producing encoded sequences to the LSTM decoder for energy prediction. Zhang, Junfeng, et al. [25] have proposed a hybrid model involving a transformer and a k-means model to predict power consumption. They applied the k-means algorithm to find a data cluster that contributes more to the predicted value, thus improving the performance of the transformer model. They have compared their approach with LSTM and k-means models. Ullah, Fath U. Min et al. [26] have proposed a hybrid model involving CNN and multilayer Bi-LSTM models called Multilayer Bidirectional Long Short Term Memory (M-BDLSTM) to forecast electricity consumption at the residential level. They have compared their approach with LSTM, BDLSTM, CNN-LSTM, and models reported in the literature and found that their approach performed better among all. Marino, Daniel L et al. [27] have investigated two neural networks, namely, sequence-to-sequence-based LSTM and standard LSTM, to forecast the energy load at the building level. They have concluded from the results that the seq2seq-based LSTM model has performed better among the two. Mocanu, Elena, et al. [28] have explored two stochastic models: Conditional Restricted Boltzmann Machine (CRBM) and Factored Conditional Restricted Boltzmann Machine (FCRBM). They have found that FCRBM model has performed better than SVM, CRBM, ANN, and RNN models. Sinha, Ayush, et al. [29] have proposed a hybrid model, Vector Auto Regressor-Convolutional Neural Networks-Long Short Term Memory (VACL), which combines VAR, CNN, and LSTM models. The VAR model separates the linear patterns from the time series data, the CNN model extracts complex features, and the LSTM model is used for temporal information. They have compared their approach with CNN-LSTM, CNN, MV-KWNN, LSTM, MLP, MV-ANN, VAR, and ARIMAX models. To forecast electricity use at the short-term level, Phyo, Pyae-Pyae, et al. [15] examined 17 machine learning algorithms. They selected the best five performing models out of 17 and applied a voting ensemble regressor to predict electricity consumption. They compared the voting ensemble with the best five ML models and found that the voting ensemble performed better among all models. Januschowski, Tim, et al. [30] have demonstrated why gradient-boosting models have outperformed deep learning models on various forecasting completions like M5, M4, and Kaggle forecasting competitions. The reasons are: a (i) gradient-boosting is more robust than deep learning models and can be used as a black box learner, (ii) tree-based models have built-in functionality to handle real-world data complications like missing values or categorical features, (iii) a wide range of loss functions are available with tree-based models as compared to deep learning models, and (iv) tree-based models are faster to train and more interpretable as compared to deep learning models. Ribeiro, Andrea Maria NC, et al. [13] have explored three machine learning (SVR, gradient-boosting, RF), three deep learning (LSTM, GRU, CNN), and one statistical method (ARIMA) for forecasting electricity consumption of a warehouse at a short-term level and very short-term level. They have concluded from the results that the gradient-boosting algorithm has outperformed all the machine learning, deep learning, and statistical method applied. Elsayed, Shereen, et al. [31] have compared the gradient-boosting regression model with eight deep learning models on the time series problem considering nine datasets. They have found that the gradient-boosting model outperformed all the deep learning models. Abbasi, Raza Abid, et al. [32] have applied a gradient-boosting model to forecast load at a short-term level. They have found that the gradient-boosting model has performed better in terms of computing time and memory resources. Aguilar Madrid et al. [33] have applied five machine learning models (KNN, SVR, RF, MLR, and gradient-boosting) to forecast load at the short-term level. They have concluded that the gradient-boosting algorithm has performed better.

In the past, researchers have applied various machine learning, hybrid models, and deep learning models to forecast electricity consumption. It is observed that the XGBoost and ensemble model have outperformed all the models. The XGBoost model is more robust, takes less training time, and handles real-world data complications easily as compared to deep learning models [30]. XGBoost is not able to extrapolate trends, so the linear regression model, which can extrapolate trends, is combined with the XGBoost model in Phase 1. The linear regression model is used to learn the trend component, and the remaining components (seasonality, cyclic, and randomness) are learned by the XGBoost model. The voting ensemble method has proved its efficacy in short-term electricity consumption forecasting tasks [15]. But the authors have not applied hyperparameter tuning, which can improve the performance of the model further. In the second phase, the genetic algorithm-optimized voting ensemble regression model is applied, considering the three models out of 11 baseline models and the model designed in phase 1.

III. PROPOSED APPROACH

The architecture of the developed model is given in Fig. 2. It has two phases; in the first phase, a hybrid of linear regression and extreme gradient-boosting is developed as shown in Fig. 3; in the second phase, the three models are selected, and the genetic algorithm optimizes a voting ensemble regression model is applied. The two phases of the hybrid model are as follows:

A. Phase 1: Ensemble of Linear Regression with XGBoost Model

Ensemble learning is the process of merging different models to generate a new model. The combined model has performed better than individual models on electricity consumption forecasting, and many more tasks [16, 34]. Time series data contains four components namely, trend, seasonality, cycles, and randomness. One model can be applied to learn all these components simultaneously. Another way is that one model can be applied to learn one component and another model for the remaining components. In this phase, linear regression and XGBoost models are combined to learn the different components of electricity consumption data. Linear regression models are also best suited for time series forecasting because they require fewer assumptions, are easy to interpret, handle data drift efficiently, and take less time to train. The linear regression model is used to learn the relationship between the response variable (global active power) and independent variables, as shown in equation 1.

$$y = \sum_{i=1}^{6} va_i * b_i \tag{1}$$

Where y is the response variable, va_1 to va_6 va_1 *to* va_6 are the independent variables (given in dataset description Section 4.1), b_1 to b_6 are regression coefficients, and a is the intercept. The ordinary least squares method is used to find the weights to minimize the sum of residuals, as given in equation 2.

$$min(S) = \sum_{i=1}^{n} (av - pv)^2 \tag{2}$$

Where S is the sum of squared residuals, *av* is the actual value, and *pv* is the predicted value.

1) XGBoost Model

XGBoost is an improved version of the gradient-boosting decision tree**.** XGBoost uses Taylor expansion to increase the speed of optimization and overcomes the problem of overfitting using the complexity of tree models in regularization terms. The XGBoost model supports parallel processing, which is faster to train and deploy. XGBoost combines multiple models (also called weak learners) sequentially to form a strong learner. As a weak learner, Classification and Regression Tree (CART) model is used. A new model is created on the errors generated by the previous model and reduces the errors. All the models are combined to make the final prediction as given in equation 3. The XGBoost model can learn feature interaction and nonlinearity in the data efficiently as compared to artificial intelligence models [35]. Whereas $F = \{f1, f2, f3, f4 \ldots fn\}$ is the set of base learners, y_i is the predicted value of the ith sample, n is the number of CART models, and f_t(pi) is the predicted value by the tth tree for the ith sample.

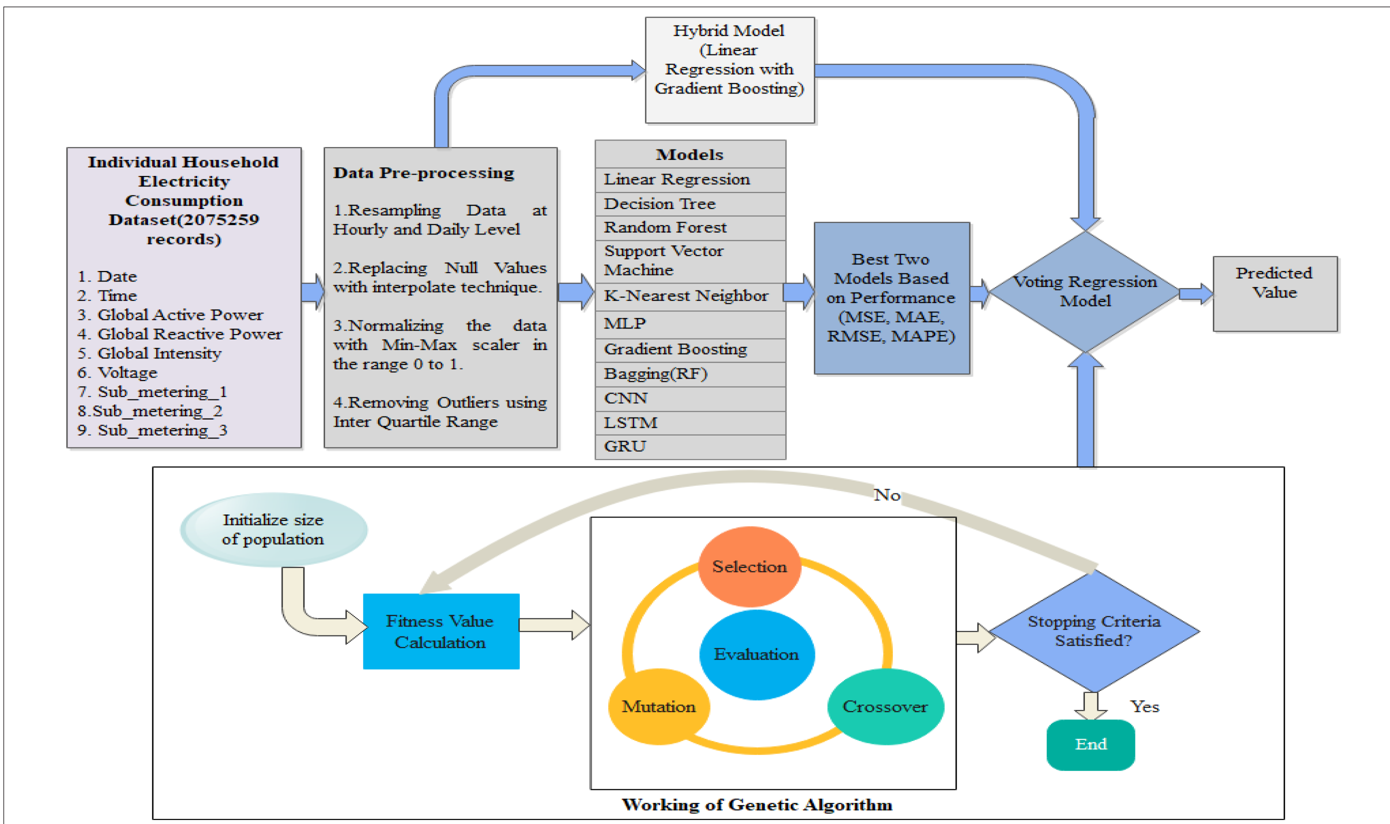

Fig. 2. Voting ensemble model optimized with genetic algorithm (proposed model).

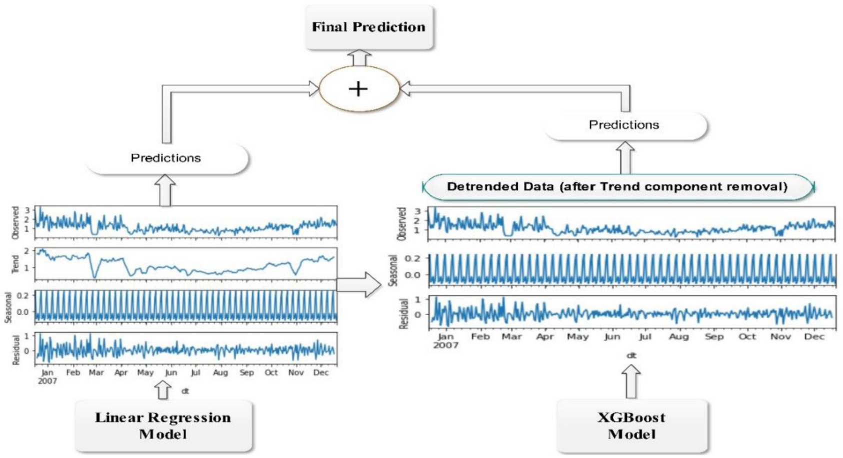

Fig. 3. Phase 1 (hybrid of linear regression and XGBoost model).

$$y_i = \sum_{t=1}^{n} f_t(pi) \quad f_t \in F \tag{3}$$

XGBoost has been used by various researchers in the past for electricity consumption forecasting problems and has produced excellent results.

The electricity consumption dataset consists of four components, namely, trend, seasonality, observed, and random factors. The linear regression model can extrapolate trends but cannot learn interactions, while gradient-boosting can learn interactions and non-linearity but cannot extrapolate trends. Therefore, these two models are combined to overcome their limitations. Firstly, the linear regression model is applied to learn the trend. The target variable is transformed to remove the trend learned by the linear regression model. A gradient-boosting algorithm (XGBoost) is applied to the detrended residuals. The models learn the different components separately, and their predictions are combined. The overall prediction is the sum of predictions generated by the XGBoost model and linear regression, as shown in Fig. 3.

B. Phase 2: Genetic Algorithm Optimized Voting Ensemble Regressor

The weighted average voting regression has shown its efficacy in the electricity consumption problem [15]. A voting regressor is a meta-estimator ensemble that fits numerous base regressor models to the entire dataset. There are two types of voting ensemble: average voting and weighted average voting. In average voting, the weight of every model is the same, and the average of all predictions is the final predicted value. Different weights are assigned to the models in the weighted average method. The final value produced is the weighted average of all values predicted by different models, as given in equation 4.

$$fpv = \sum_{i=1}^{n} pv_i * w_i \tag{4}$$

Where *fpv*, the final is the predicted value, pv_i is the predicted value by the i^{th} model, and W_i is the weight of the i^{th} model. The best two models are selected as base regressors based on performance from the 11 baseline models, and applies the hybrid model designed in phase 1 is given as input to the weighted average voting regression model. The machine learning model's performance depends upon data quality, quantity, and hyperparameter setting. The hyperparameters of a model are tuned using Bayesian optimization, grid search, genetic algorithm, and random search. The genetic algorithm has performed better and takes less time in hyperparameter settings [36, 37]. The weight of each base regressor is a hyperparameter and is tuned using a genetic algorithm [38]. It is a population-based search algorithm. This algorithm begins with the population, which consists of possible solutions called chromosomes. The population contains the possible weights assigned to different models. The elements in the chromosome are called genes.

$$Fitness\_Value = \frac{1}{Root\ Mean\ Square\ Error} \tag{5}$$

The fitness value of the voting ensemble model is evaluated considering weights {(1,1,1), (1,2,1), (1,1,2), (1,2,2), (2,1,1), (2,2,1), (2,2,2), (2,1,2)}. Based on the root mean square value, the fitness value is computed for each model as given in equation 5. In the next step called selection, the models which produced better fitness value are selected as parents, which mate and recombine together to produce offspring for the next generation. The parents are selected with the tournament selection technique in this study. In the tournament selection technique, k individuals from the population are randomly selected, and the best individual out of these is taken as a parent. In the same way, the next parent is selected. The next step is a crossover, in which two parents are selected to produce offspring using the genetic properties of the parents. A multipoint crossover strategy is used in which alternate segments are swapped to produce the offspring. The next step is a mutation, which is a random tweak in the chromosomes to produce new individuals in the population. The random resetting approach is used for mutation in this study. These steps are repeated until 100 iterations are reached or there is no improvement in the population. Table I shows the parameters used in the genetic algorithm as well as the optimum weights obtained for each model.

IV. EXPERIMENTAL RESULTS

This section discusses the dataset considered, data cleaning process, baseline models considered (machine learning and deep learning), experimental setup, evaluation measures, and comparative study with baseline and SOA models at daily and hourly levels.

TABLE I. GENETIC ALGORITHM PARAMETERS AND OPTIMIZED WEIGHT VALUE OOBTAINED

Parameter	Value
Population size	8
Mutation rate	0.25
Tournament size	5
Crossover rate	0.25
Maximum number of generations	100

Optimized weight of models at the hourly level are (1, 2, 2),and at the daily level are (1, 1, 2).

A. Dataset Description and Cleaning

Individual household electric power consumption [39] is considered in this study. The dataset consists of the electricity consumption of a house located in Sceaux near France at the minute level over 47 months (16 December 2006 to 26 November 2010). The dataset description is given in Table II. It is observed at the minute level, which has been resampled at the hourly and daily levels, representing short-term load prediction, as shown in Fig. 4. The dataset considered has various abnormalities like missing values and outliers that may impact the performance of the models. Missing values are filled with the backward direction interpolation method, and outliers are detected using the interquartile range and removed from the dataset. Further, a min-max scalar is applied to the data to scale it from 0 to 1, as given in equation 6. Whereas X_{min} is the minimum value, X_{max} is the maximum value, and X_{scaled} is the scaled value.

TABLE II. METADATA ABOUT THE DATASET USED

Name	Duration	Attributes
IHEPC dataset	December 16, 2006 to November 26, 2010 (around four years). Per-minute observation - total 2 075 259 measurements	Date, time, voltage (volt), sub metering 1, global intensity (Ampere), sub metering 3, global reactive power (kilowatt), global active power (kilowatt), sub metering 2

IHEPC, individual household electric power consumption.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (6)$$

B. Baseline Models

The most frequently used deep learning and machine learning algorithms for electricity consumption prediction [40, 41] are selected as baseline models. The machine learning models considered are as follows: (i) Linear Regression (ii) Decision Tree ('min_weight_fraction_leaf': 0.1, 'min_samples_leaf': 2, 'max_leaf_nodes': 40, 'max_depth': 5), (iii) Support Vector Regressor (kernel = 'rbf', degree = 1) (iv) Random Forest Regressor ('min_samples_split': 10, 'n_estimators': 400, 'min_samples_leaf': 4) (v) MLP(hidden_layer_sizes = (150,100,50), activation = 'relu',max_iter = 300,solver = 'adam')(vi)KNN(n_neighbors = 3)(vii)Bagging(base_estimator = RandomForestRegressor, n_estimators = 10, random_state = 0) and (viii) Gradient Boosting (min_samples_split = 2, n_estimators = 100, learning_rate = 0.1, min_samples_leaf = 1, min_weight_fraction_leaf = 0.0, max_depth = 3). The deep learning models considered are as follows: (i) CNN (2 layers of Conv1D with kernel_size = 2, filter = 64, activation = relu, maxpooling1D, dropout = 0.5, pool_size = 2) (ii) LSTM (number of units = 100, dropout = 0.2, loss = mse, optimizer = Adam) and (iii) GRU (number of units = 100, dropout = 0.2, loss = mse, optimizer = Adam). The grid

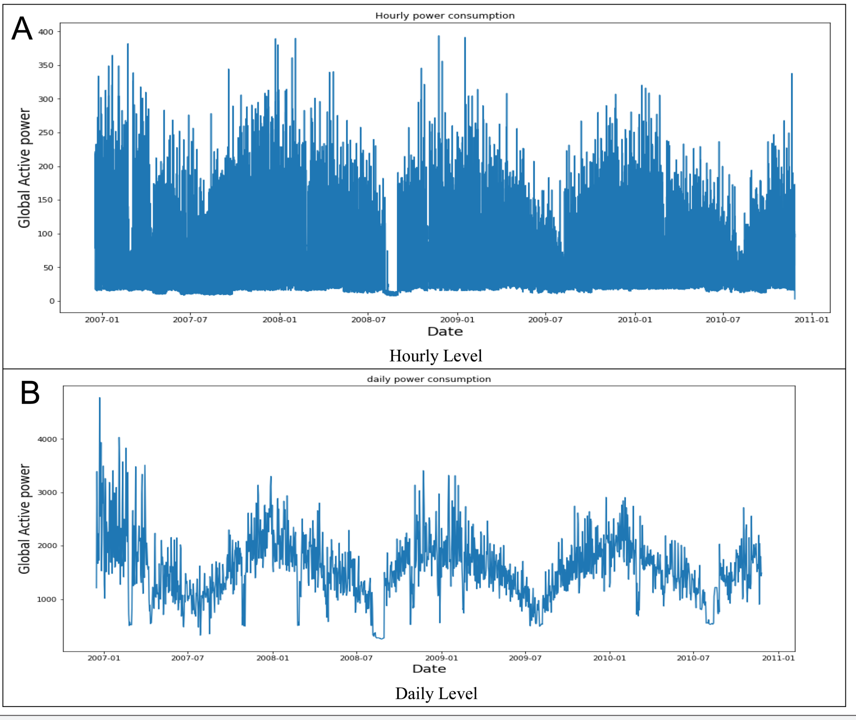

Fig. 4. Electricity consumption data at hourly and daily resolution.

search strategy is used to tune the hyperparameters of these models and optimized parameters are presented.

C. Experimental Setup and Results Analysis

The experiments were performed on Google Colab with the following configuration: central processing unit (CPU) (Intel(R) Xeon(R) CPU @ 2.20GHz) and 13 GB RAM. The models were implemented using the Python language (version 3.6.9), Niapy (version 2.0.0rc18), sklearn (version 1.0.2), Numpy, Seaborn, Keras (V2 2.4), and Tensorflow (version 2.8.0) libraries. The dataset was sampled at a minute level and then resampled at an hourly and daily level. The dataset was recorded over four years, with the first three years of data used for training and for testing remaining one year of data is used. The four metrics, namely, root mean square error, mean squared error, mean absolute error, and mean absolute percentage error, as given in equations 7 to 10, were used to evaluate the models. These four metrics are the most commonly used measures for electricity load forecasting [40].

$$RMSE = \sqrt{\frac{1}{N}\left(\sum_{n=1}^{N}\left(fv_n - av_n\right)^2\right)} \tag{7}$$

$$MSE = \frac{1}{N}\left(\sum_{n=1}^{N}\left(fv_n - av_n\right)^2\right) \tag{8}$$

$$\mathrm{MAE} = \frac{1}{\mathrm{N}}\sum_{\mathrm{n}=1}^{\mathrm{N}}\mathrm{abs}\left(\mathrm{fv_n} - \mathrm{av_n}\right) \tag{9}$$

$$MAPE = \frac{100\%}{N}\sum_{n=1}^{N}abs\left(\frac{fv_n - av_n}{av_n}\right) \tag{10}$$

Where N is the sample count, fv is the predicted value, and av is the actual value.

1) Results Analysis on Hourly Level Dataset

To validate the effectiveness and robustness of the proposed model, experiments considering 11 baseline models were conducted on hourly level data. The results of machine learning, deep learning, and state-of-the-art models are given in Table III. It is observed from the results that our approach (genetic algorithm optimizes a voting ensemble regressor) has outperformed all the baseline models and models reported in the literature with an MSE value of 0.159, RMSE value of 0.387, MAE value of 0.283, and MAPE value of 25.07. In the voting ensemble, the three best-performing models (XGboost, Bagging, and phase 1 model) are taken into consideration as shown in Fig. 5. The XGboost model has given an MSE value of 0.173, RMSE value of 0.416, MAE value of 0.311, and MAPE value of 47.01. The bagging (with the random forest as the base estimator) has given an MSE value of 0.177, RMSE value of 0.421, MAE value of 0.315, and MAPE value of 48.65. The model designed in phase 1 (linear regression + XGboost) has given an MSE value of 0.166, an RMSE value of 0.403, MAE value of 0.285, and MAPE value of 44.71. The weight parameter of these three models is optimized using a genetic algorithm. The decision tree algorithm has performed poorly among all the models with an MSE value of 0.583, RMSE value of 0.764, MAE value of 0.540, and MAPE value of 91.48.

2) Results Analysis on Daily Level Dataset

The results of deep learning, machine learning, existing models, and the developed method on the daily level dataset are given in Table IV. It is observed from the results that our approach has outperformed all the baseline models and models reported in the literature with an MSE value of 0.025, an RMSE value of 0.162, an MAE value of 0.129, and an MAPE value of 15.61. In the voting ensemble model, three best performing models namely, the phase 1 model (combination of XGboost, and linear regression), XGBoost, and linear regression are considered as shown in Fig. 6. The weight parameters of these three models are optimized with a genetic algorithm. XGboost has given an MSE value of 0.032, RMSE value of 0.182, MAE value of 0.146, and MAPE value of 17.48. The linear regression model has given an MSE value of 0.035, RMSE value of 0.186, MAE value of 0.148, and MAPE value of 18.36. The hybrid of linear regression and XGboost has improved the performance with an MSE value of 0.028,

TABLE III. COMPARISON OF DEEP LEARNING, MACHINE LEARNING, EXISTING MODELS, AND THE DEVELOPED APPROACH AT THE HOURLY LEVEL

Models	MSE	RMSE	MAE	MAPE
Machine Learning Models				
Support vector regressor	0.298	0.546	0.454	98.81
Random forest regressor	0.270	0.520	0.413	80.68
Linear regression	0.221	0.470	0.366	49.86
Decision tree	0.583	0.764	0.540	91.48
MLP	0.231	0.481	0.379	74.37
K-NN	0.236	0.486	0.345	49.88
XGBoost	0.173	0.416	0.311	47.01
Bagging (RF)	0.177	0.421	0.315	48.65
Linear regression + XGBoost	0.166	0.403	0.285	44.71
Deep Learning Models				
CNN	0.345	0.587	0.511	91.86
LSTM	0.182	0.427	0.335	48.36
GRU	0.182	0.426	0.333	54.15
Comparison with Existing Models				
Kim, Tae-Young et al. [19]	0.3549	0.5957	0.3317	32.83
Le, Tuong, et al. [20]	0.298	0.546	0.392	50.09
Kim, Jin-Young et al. [21]	0.384	-	0.3953	-
Bu Seok-Jun et al. [22]	0.262	-	-	-
Sajjad, Muhammad et al. [23]	0.22	0.47	0.33	-
Khan, Zulfiqar Ahmad et al. [24]	0.19	0.31	0.47	0.76
Zhang, Junfeng, et al. [25]	-	0.74	-	-
Ullah, Fath U. Min et al. [26]	0.3193	0.565	0.3469	0.291
Marino, Daniel L et al. [27]	-	0.625	-	-
Mocanu, Elena et al. [28]	-	0.663	-	-
Sinha, Ayush, et al. [29]	0.210	0.458	0.317	-
Our Work	**0.159**	**0.387**	**0.283**	**25.07**

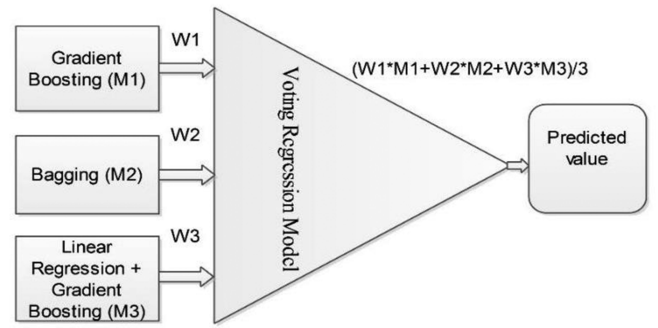

Fig. 5. Hourly level voting ensemble technique.

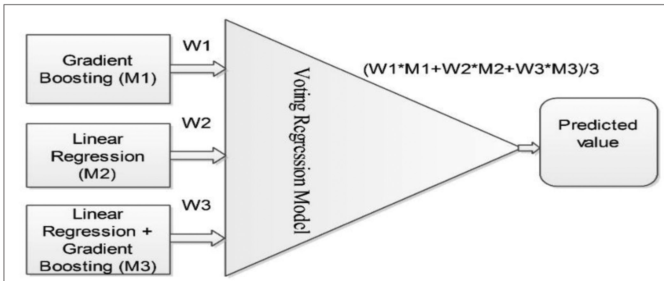

Fig. 6. Daily level voting ensemble technique.

RMSE value of 0.176, MAE value of 0.139, and MAPE value of 16.69. The decision tree algorithm has performed poorly among all with an MSE value of 0.102, RMSE value of 0.319, MAE value of 0.238, and MAPE value of 28.06.

XGBoost model has performed better among the baseline models considered, and the decision tree model performed poorly among all the baseline models on both datasets (hourly and daily). The XGBoost model has given better performance because it computes second-order gradients to find out the gradient direction and uses L1 and L2 regularization approaches to avoid overfitting. The decision tree model has given poor performance because it divides the data into smaller subsets without considering the correlation between features. The decision tree model is sensitive to small changes in data, as this is the case in our dataset in which electricity consumption data keeps changing with the season, holidays, and weather.

TABLE IV. COMPARISON OF DEEP LEARNING, MACHINE LEARNING, EXISTING MODELS, AND THE DEVELOPED APPROACH AT THE DAILY LEVEL

Models	**MSE**	**RMSE**	**MAE**	**MAPE**
Machine Learning Models				
Support vector regressor	0.046	0.215	0.170	21.81
Random forest regressor	0.041	0.202	0.157	18.66
Linear regression	0.035	0.186	0.148	18.36
Decision tree	0.102	0.319	0.238	28.06
Multilayer Perceptron (MLP)	0.042	0.206	0.157	19.88
Bagging (RF)	0.038	0.195	0.151	18.15
Exteme Gradient Boosting	0.032	0.182	0.146	17.48
K-Nearest Neighbors (K-NN)	0.053	0.230	0.181	21.10
Linear Regression +XGBoost	0.028	0.176	0.139	16.69
Deep Learning Models				
Convolutional Neural Networks (CNN)	0.082	0.287	0.228	26.65
Long Short Term Memory (LSTM)	0.043	0.207	0.161	17.92
Gated Recurrent Unit (GRU)	0.040	0.201	0.157	18.02
Comparison with Existing Models				
Kim, Tae-Young et al. [19]	0.1037	0.3221	0.2569	31.83
Le, Tuong, et al. [20]	0.065	0.255	0.191	19.15
Bu Seok-Jun et al. [22]	0.0969	-	-	-
Our Work	**0.025**	**0.162**	**0.129**	**15.61**

The proposed approach has performed better than all the baseline and state-of-the-art models considering datasets at the hourly and daily levels. So, the developed model can be recommended in the intelligent electricity management system to forecast electricity consumption.

Analysis of variance one-way test: This test is used to check whether the results produced are statistically significant or not. The null hypothesis is that models have not performed statistically significantly, and the alternative hypothesis is that models performed statistically significantly. Analysis of variance test gives *P*-value and *F*-statistics as an output. If the *P*-value obtained is less than α (significance level, considered 0.05 in this study), then the null hypothesis is rejected [42]. This test is applied considering four performance parameters, namely MAPE, MSE, RMSE, and MAE, using the statsmodels python library. The *P*-values obtained corresponding to these four performance parameters are 0.0397, 0.0245, 0.0326, and 0.0309, as given in Table V, which is less than 0.05. Hence, the null hypothesis is rejected, and we can conclude that the machine learning, deep learning, existing approaches, and proposed models performed statistically significant.

D. Discussion

The developed hybrid model was compared with existing models on the same dataset at an hourly and daily level as given in Tables III and IV respectively. The proposed method results were compared with references [19-29] at the hourly level on the same dataset. The proposed model has achieved the best values in terms of MSE (0.159), MAE (0.283), and MAPE (25.07) which is around 3%, 5%, and 7% better respectively than existing approaches. The proposed approach results at the daily level were compared with references [19, 20, and 22] and achieved the best values in terms of MSE (0.025), RMSE (0.162), MAE (0.129), and MAPE (15.61) which is 4%, 9%, 6%, and 3% better than existing approaches. The authors [19-29] have applied deep learning models to forecast electricity consumption, but gradient boosting model has outperformed deep learning model in the forecasting problem [30]. It is observed from the results that the XGBoost model, which is tree-based, has outperformed all the baseline deep learning and machine learning models applied to both the datasets. XGBoost is not able to extrapolate trends and learn interactions and non-linearity in the data. Linear regression can extrapolate trends. So, these two models are combined, and the results showed that the ensemble of linear

TABLE V. ANALYSIS OF VARIANCE TEST (ONE-WAY) RESULTS

Source	DF	Sum_sq	Mean_sq	F	PR (>F)
Considering MAPE (Hourly Level Data)					
Models	3.0	4197.948424	1399.316141	3.709069	0.039788
Residual	13.0	4904.494600	377.268815	NaN	NaN
Considering MSE (Daily Level Data)					
Models	3.0	0.005938	0.001979	4.371318	0.024579
Residual	13.0	0.005886	0.000453	NaN	NaN
Considering RMSE (Daily Level Data)					
Models	3.0	0.020132	0.006711	3.975868	0.032639
Residual	13.0	0.021942	0.001688	NaN	NaN
Considering MAE (Daily Level Data)					
Models	3.0	0.011795	0.003932	4.050784	0.030904
Residual	13.0	0.012618	0.000971	NaN	NaN

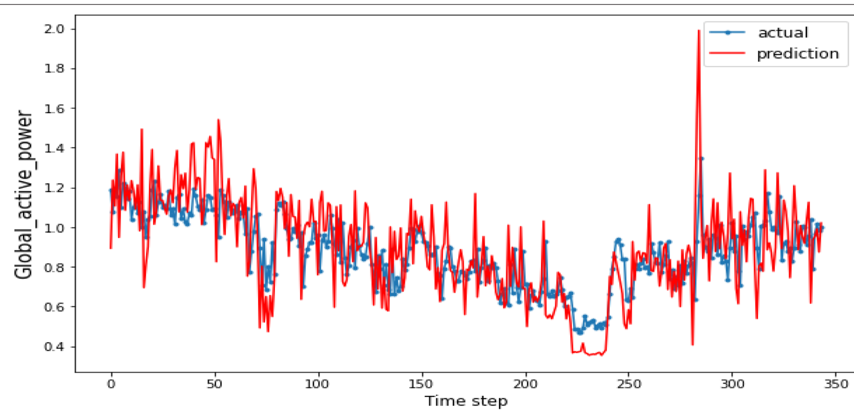

Fig. 7. Actual vs. predicted values at the hourly level on the test dataset.

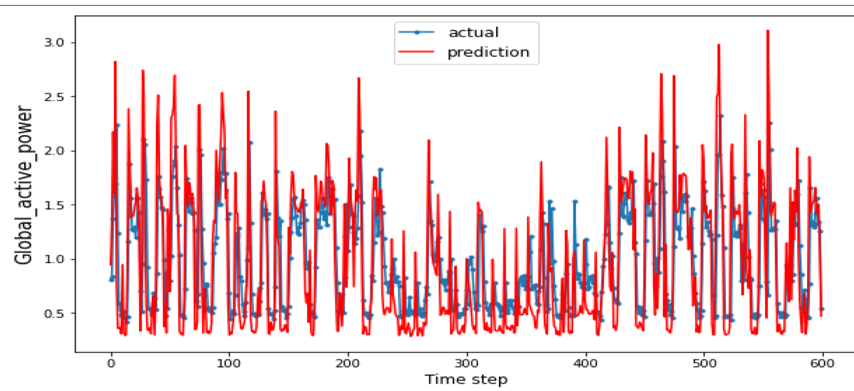

Fig. 8. Actual vs. predicted values at the daily level on the test dataset.

regression and gradient boosting decreases the error of electricity consumption forecasting on both hourly and daily datasets. The major outcome is that the combination of linear regression and XGBoost has performed better than baseline machine learning and deep learning models. Another outcome is that voting ensemble regression further decreases the error of electricity consumption forecasting, and hyperparameter tuning further decreases the forecasting error. The decision tree algorithm has performed poorly among all because it cannot extrapolate the trend component which is present in the dataset considered. Figs. 7 and 8 show the difference between the actual value and the anticipated value by the proposed model at the hourly and daily test datasets. It is observed that the proposed model prediction followed the real value pattern.

F. Limitations

(i) Meteorological information is not considered in this study, which also plays an important role in electricity consumption in residential buildings.
(ii) In this study, electricity consumption forecasting is done only at an hourly and daily level. The electricity consumption prediction at monthly and yearly levels can help utility companies plan how much infrastructure needs to be expanded (new power generation houses are required), strategic planning, and modifications in the supply system.

V. CONCLUSION

In this study, we discussed the significance of predicting electricity consumption and developed an efficient hybrid model to address it. Existing methods were not able to handle linearity and non-linearity in the dataset. Our model is capable of learning linearity and non-linearity in electricity consumption dataset. An individual model has a chance of being caught in local minima while multiple models have a lower probability of being stuck in local minima because each model has different starting points. Therefore, we have also combined three best models from two different categories (machine learning and deep learning) using the voting ensemble technique. The efficacy of the proposed model is evaluated on the IHEPC dataset, publicly available on the University of California Irvine (UCI) machine learning repository. First, the data is pre-processed by removing outliers, filling in missing values, and scaling in the range of zero to one. Furthermore, the two-phase hybrid model is applied to predict electricity consumption at the residential level. The linear regression model is combined with the gradient descent model in the first phase. The voting ensemble model is applied in the second level, considering the best three models out of 11 and the hybrid model designed in phase 1. The weight hyperparameter of each model is tuned using the genetic algorithm. The proposed model is compared with eight machine learning, three deep learning, and state-of-the-art models. The results showed that the hybrid model proposed in phase 1 performed better among all machine and deep learning models applied. The second-level hybrid model further improved the performance, and hyperparameter tuning also improved the performance further. The decision tree algorithm has given the worst performance among all models. In the future, electricity forecasting can be done at the medium-term and long-term levels. A hybrid of gradient boosting and other models from deep learning, such as transformers, can be applied to forecast electricity consumption. The features from different domains like statistical, temporal, and spectral can be combined to predict future electricity consumption. Transformer models (LogTrans, Pyraformer, etc.) can be applied to forecast electricity consumption.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – A.B., S.S.; Design – A.B., S.S.; Supervision – S.S.; Materials – A.B., S.S.; Data Collection and/or Processing – A.B., S.S.; Analysis and/or Interpretation – A.B., S.S.; Literature Review – A.B., S.S.; Writing – A.B.; Critical Review – A.B., S.S.

Declaration of Interests: The authors have no conflict of interest to declare.

Funding: The authors declared that this study has received no financial support.

REFERENCES

1. G. Y. Zhao, Z. Y. Liu, Y. He, H. J. Cao, and Y. B. Guo, "Energy consumption in machining: Classification, prediction, and reduction strategy," *Energy*, vol. 133, pp. 142–157, 2017. [CrossRef]
2. J. Chen, "Energy Information Administration (EIA): Meaning, history," *Investopedia*. Available: https://www.investopedia.com/terms/e/energy-information-admin.asp. [accessed: Oct. 28, 2023].
3. IEA, *Final Consumption – Key World Energy Statistics 2020 – Analysis*. IEA. Available: https://www.iea.org/reports/key-world-energy-statistics-2020/final-consumption. [accessed: Oct. 28, 2023].
4. A. Basit, T. Ahmad, A. Yar Ali, K. Ullah, G. Mufti, and A. D. Hansen, "Flexible modern power system: Real-time power balancing through load and wind power," *Energies*, vol. 12, no. 9, p. 1710, 2019. [CrossRef]
5. A. Estebsari, E. Pons, T. Huang, and E. Bompard, "Techno-economic impacts of automatic undervoltage load shedding under emergency," *Electr. Power Syst. Res.*, vol. 131, pp. 168–177, 2016. [CrossRef]
6. B. F. Hobbs, S. Jitprapaikulsarn, S. Konda, V. Chankong, K. A. Loparo, and D. J. Maratukulam, "Analysis of the value for unit commitment of improved load forecasts," *IEEE Trans. Power Syst.*, vol. 14, no. 4, pp. 1342–1348, 1999. [CrossRef]
7. K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-term load forecasting with deep residual networks," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3943–3952, 2019. [CrossRef]
8. C. Deb, F. Zhang, J. Yang, S. E. Lee, and K. W. Shah, "A review on time series forecasting techniques for building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 74, pp. 902–924, 2017. [CrossRef]
9. K. H. Kim, and S. B. Cho, "Modular Bayesian networks with low-power wearable sensors for recognizing eating activities," *Sensors (Basel)*, vol. 17, no. 12, p. 2877, 2017. [CrossRef]
10. A. Miyasawa, S. Akira, Y. Fujimoto, and Y. Hayashi, "Spatial demand forecasting based on smart meter data for improving local energy self-sufficiency in Smart Cities," *IET Smart Cities*, vol. 3, no. 2, pp. 107–120, 2021. [CrossRef]
11. T. M. Ghazal et al., "Energy demand forecasting using fused machine learning approaches," *Intell. Autom. Soft Comput.*, vol. 31, no. 1, pp. 539–553, 2022. [CrossRef]
12. A. Heydari, M. Majidi Nezhad, E. Pirshayan, D. Astiaso Garcia, F. Keynia, and L. De Santoli, "Short-term electricity price and load forecasting in isolated power grids based on composite neural network and gravitational search optimization algorithm," *Appl. Energy*, vol. 277, p. 115503, 2020. [CrossRef]
13. A. M. N. C. Ribeiro, P. R. X. do Carmo, P. T. Endo, P. Rosati, and T. Lynn, "Short- and very short-term firm-level load forecasting for warehouses: A comparison of machine learning and Deep Learning Models," *Energies*, vol. 15, no. 3, p. 750, 2022. [CrossRef]
14. D. Opitz, and R. Maclin, "Popular ensemble methods: An empirical study," *J. Artif. Intell. Res.*, vol. 11, pp. 169–198, 1999. [CrossRef]
15. P.-P. Phyo, Y.-C. Byun, and N. Park, "Short-term energy forecasting using machine-learning-based ensemble voting regression," *Symmetry*, vol. 14, no. 1, p. 160, 2022. [CrossRef]
16. A. Banga, R. Ahuja, and S. C. Sharma, "Stacking machine learning models to forecast hourly and daily electricity consumption of household using internet of things," *NISC*. Available: https://nopr.niscpr.res.in/handle/123456789/58228. [accessed: Jan. 28, 2022].
1174. N. Fumo, and M. A. R. Biswas, "Regression analysis for prediction of residential energy consumption," *Renew. Sustain. Energy Rev.*, vol. 47, pp. 332–343, 2015. [CrossRef]
18. T. Vantuch, A. G. Vidal, A. P. Ramallo-Gonzalez, A. F. Skarmeta, and S. Misak, "Machine learning based electric load forecasting for short and long-term period," 2018 IEEE 4th World Forum on Internet of Things (WF-IoT). IEEE Publications, 2018. [CrossRef]
19. T.-Y. Kim, and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, pp. 72–81, 2019. [CrossRef]
20. T. Le, M. T. Vo, B. Vo, E. Hwang, S. Rho, and S. W. Baik, "Improving electric energy consumption prediction using CNN and Bi-LSTM," *Appl. Sci.*, vol. 9, no. 20, p. 4237, 2019. [CrossRef]

21. J.-Y. Kim, and S.-B. Cho, "Electric energy consumption prediction by deep learning with state explainable Autoencoder," *Energies*, vol. 12, no. 4, p. 739, 2019. [CrossRef]
22. S.-J. Bu, and S.-B. Cho, "Time Series forecasting with multi-headed attention-based deep learning for residential energy consumption," *Energies*, vol. 13, no. 18, p. 4722, 2020. [CrossRef]
23. M. Sajjad et al., "A novel CNN-GRU-based hybrid approach for short-term residential load forecasting," *IEEE Access*, vol. 8, pp. 143759–143768, 2020. [CrossRef]
24. Z. A. Khan, T. Hussain, A. Ullah, S. Rho, M. Lee, and S. W. Baik, "Towards efficient electricity forecasting in residential and commercial buildings: A novel hybrid CNN with a LSTM-AE based framework," *Sensors (Basel)*, vol. 20, no. 5, p. 1399, 2020. [CrossRef]
25. J. Zhang, H. Zhang, S. Ding, and X. Zhang, "Power consumption predicting and anomaly detection based on transformer and K-means," *Front. Energy Res.*, vol. 9, 2021. [CrossRef]
26. F. U. M. Ullah, A. Ullah, I. U. Haq, S. Rho, and S. W. Baik, "Short-term prediction of residential power energy consumption via CNN and multi-layer bi-directional LSTM Networks," *IEEE Access*, vol. 8, pp. 123369–123380, 2020. [CrossRef]
27. D. L. Marino, K. Amarasinghe, and M. Manic, "Building energy load forecasting using deep neural networks," *IECON*, vol. 2016, 7046–7051, 2016. [CrossRef]
28. E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," *Sustain. Energy Grids Netw.*, vol. 6, pp. 91–99, 2016. [CrossRef]
29. A. Sinha, R. Tayal, A. Vyas, P. Pandey, and O. P. Vyas, "Forecasting electricity load with hybrid scalable model based on stacked non linear residual approach," *Front. Energy Res.*, vol. 9, 2021. [CrossRef]
30. T. Januschowski, Y. Wang, K. Torkkola, T. Erkkilä, H. Hasson, and J. Gasthaus, "Forecasting with trees," *Int. J. Forecasting*, vol. 38, no. 4, pp. 1473–1481, 2022. [CrossRef]
31. S. Elsayed, D. Thyssens, A. Rashed, H. S. Jomaa, and L. Schmidt-Thieme, "Do we really need deep learning models for time series forecasting?" Available: arXiv.org. Available: https://arxiv.org/abs/2101.02118. [accessed: Oct. 28, 2023].
32. R. A. Abbasi, N. Javaid, M. N. J. Ghuman, Z. A. Khan, S. Ur Rehman, and Amanullah, "Short term load forecasting using XGBoost," *Adv. Intell. Syst. Comput.*, pp. 1120–1131, 2019. [CrossRef]
33. E. Aguilar Madrid, and N. Antonio, "Short-term electricity load forecasting with machine learning," *Information*, vol. 12, no. 2, p. 50, 2021. [CrossRef]
34. W. Sun, and B. Trevor, "A Stacking Ensemble Learning Framework for annual river ice breakup dates," *J. Hydrol.*, vol. 561, pp. 636–650, 2018. [CrossRef]
35. A. D. Lainder, and R. D. Wolfinger, "Forecasting with gradient boosted trees: Augmentation, tuning, and cross-validation strategies," *Int. J. Forecasting*, vol. 38, no. 4, pp. 1426–1433, 2022. [CrossRef]
36. L. Xu et al., "Mid-term prediction of electrical energy consumption for crude oil pipelines using a hybrid algorithm of support vector machine and genetic algorithm," *Energy*, vol. 222, p. 119955, 2021. [CrossRef]
37. S. Oh, Y. Kim, and M. F. Mridha, "Optimizing energy consumption prediction models using genetic algorithms," *The J. Contents Comput.*, vol. 3, no. 1, pp. 307–316, 2021. [CrossRef]
38. S. Mirjalili, "Genetic algorithm," *Stud. Comp. Intell.*, pp. 43–55, 2019. [CrossRef]
39. G. Hebrail, and A. Berard, "Individual household electric power consumption," *UCI Machine Learning Repository* [accessed Mar. 28, 2022]. [CrossRef]
40. I. K. Nti, M. Teimeh, O. Nyarko-Boateng, and A. F. Adekoya, "Electricity Load Forecasting: A systematic review," *J. Electr. Syst. Inf. Technol.*, vol. 7, no. 1, 2020. [CrossRef]
41. P. Lara-Benítez, M. Carranza-García, and J. C. Riquelme, "An experimental review on Deep Learning Architectures for time series forecasting," *Int. J. Neural Syst.*, vol. 31, no. 3, p. 2130001, 2021. [CrossRef]
42. J.-S. Chou, and D.-S. Tran, "Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders," *Energy*, vol. 165, pp. 709–726, 2018. [CrossRef]

Alisha Banga received the B.E degree (Electronics and Communication Engineering) and M. Tech degree (Electrical and Electronics Engineering) from Maharishi Dayanand University, Rohtak India, in 2009 and 2011 respectively. She is currently pursuing the PhD degree from Indian Institute of Technology Roorkee, India. Her research interests include Machine Learning and IoT in Smart City applications.

S. C. Sharma received M.Sc. (Electronics), M. Tech. (Electronics & Communication Eng.) and Ph.D. (Electronics & Computer Eng.) from the Indian Institute of Technology Roorkee, India. He is currently working as a professor at IIT Roorkee. He has more than 28 years of research and teaching experience. He has published over three hundred research papers in national and international journals/conferences and supervised 21 Ph.D. students. IIT-Roorkee has awarded him the Khosla research prize for the best research paper. He has worked as a research scientist at FMH, Germany.