Internet of Things 20 (2022) 100636

Contents lists available at ScienceDirect

Internet of Things

journal homepage: www.sciencedirect.com/journal/internet-of-things

Research article

A machine-learning ensemble model for predicting energy consumption in smart homes

Ishaani Priyadarshini [a], Sandipan Sahu [b], Raghvendra Kumar [c], David Taniar [d,*]

[a] School of Information, University of California, Berkeley, USA
[b] Department of Computer Science and Engineering, Bengal Institute of Technology, India
[c] Department of Computer Science and Engineering, GIET University, India
[d] Faculty of Information Technology, Monash University, Australia

ARTICLE INFO

Keywords:
Smart Home
Internet of Things
Decision tree
Random Forest
eXtreme gradient boosting
Ensemble model
Machine learning

ABSTRACT

Smart homes incorporate several devices that automate tasks and make our lives easy. These devices can be useful for many things, like security access, lighting, temperature, etc. Using the Internet of Things (IoT) platform, smart homes essentially let homeowners control appliances and devices remotely. Due to their self-learning skills, smart homes can learn homeowners' schedules and adapt accordingly to make adjustments. Since convenience and cost savings is necessary in such an environment, and there are multiple devices involved, there is a need to analyze power consumption in smart homes. Moreover, increased energy consumption leads to an increase in carbon footprint, elevates the risk of climate, and leads to increased demand in supply. Hence, monitoring energy consumption is crucial. In this paper, we perform an overall analysis of energy consumption in smart homes by deploying machine learning models. We rely on machine learning techniques, like Decision Trees (DT), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and k-Nearest Neighbor (KNN) for predicting the power consumption of multiple datasets. We also propose a DT-RF-XGBoost-based Ensemble Model for analyzing the consumption and comparing it with the baseline algorithms. The evaluation parameters used in the study are Mean Square Error (MSE), R-squared ($R^{2,}$), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE), respectively. The study has been performed on multiple datasets and our study shows that the proposed DT-RF-XG-based Ensemble Model outperforms all the other baseline algorithms for multiple datasets with R^2 around 0.99.

1. Introduction

One of the best applications of Artificial Intelligence (AI) in modern times is in the form of smart homes. Smart homes are residences that rely on internet-connected devices for remote monitoring and device and appliance management [1]. It is also referred to as home automation or domotics and is concerned with providing comfort, security, energy efficiency, and convenience. Smart devices are usually controlled by a smart home application or a networked device. Smart homes are built on the Internet of Things (IoT) platform and incorporate sensors, speakers, smart bulbs, cameras, locks, door openers, etc. [2]. Due to their self-learning skills, they are capable of learning the homeowners' schedules and can adjust accordingly. Additionally, these devices may operate together and share

* Corresponding author.
E-mail address: David.Taniar@monash.edu (D. Taniar).

https://doi.org/10.1016/j.iot.2022.100636

Available online 7 November 2022
2542-6605/© 2022 Elsevier B.V. All rights reserved.

consumer usage data due to automation based on homeowners' preferences [3]. Since these devices are power-driven, they can reduce power consumption and lead to energy-related cost savings [4].

While smart home devices can save a considerable amount of energy, the efficiency can be improvised even further. Many smart speakers and connected cameras consume more power since they add more energy load. Since power consumption relies on many predictable factors like what devices were used before, what devices are being used presently, which product is bought, and how it is used, a hundred percent power saving cannot be guaranteed. The methods to address the issue are being actively researched.

Another major reason that leads to inefficient power consumption is based on poorly constructed buildings. For constructions with a single-pane window with no insulation, deploying a smart thermometer may not be beneficial. In other words, if the building is not designed to save energy, integrating applications and components may be a tedious task. This may lead to greater power consumption. Moreover, as more and more smart homes house intelligent lighting controls, although it may use very little electricity, the fact that it is smart and always connected may lead to more consumption of electricity in general [5]. Hence, there is a need to inspect power consumption for smart homes.

Monitoring power consumption is also necessary for load-balancing power plants. Performing load study is an important aspect of energy monitoring. As there is an increase in energy consumption, it adds to the risk of climate change and increases carbon footprints [6]. The increasing energy costs, in turn, lead to an increase in the demand for energy consumption. Owing to all these factors, there is a need to monitor energy consumption. One of the ways of inspecting power consumption is by taking a look at the prediction data. To reduce power consumption in smart homes, it is necessary to observe prediction trends. In the past several methods have been proposed to monitor power consumption. Some of these methods are in-chip configurations in microprocessing systems, digital power meters, energy auditing tools, delay and power monitoring schemes [7], etc. Moreover, machine learning methods have been applied in abundance to monitor power consumption. Linear Regression [8], Support Vector Machines [9], and Long-short Term Memory [10, 11], etc., are some of the popular machine-learning models that have been relied on in the past for analyzing power consumption. While traditional machine learning models yield satisfactory results for addressing the problem, the models often have limitations. Moreover, overfitting and cost are prevalent in most traditional machine learning algorithms. To address limitations like these, ensemble methods are deployed. Ensemble methods combine many machine learning algorithms to produce one optimal predictive model, thereby enhancing the model's performance and robustness.

The novelty and main contributions of the paper are as follows:

1 We propose an ensemble-based technique for predicting power consumption in smart homes. The ensemble proposed is a combination of Decision Trees (DT), Random Forests (RF), and eXtreme Gradient Boosting (XGBoost).
2 We compare the performance of our proposed ensemble method with several other baseline models such as K Nearest Neighbors (KNN), Decision Trees (DT), Random Forests (RF), and Gradient Boosting (GB). Deploying the ensemble model has two advantages. First, it improves the prediction performance over the other contributing components of the ensemble. Second, it reduces the variance of prediction errors induced by components of the ensemble, thereby addressing any kind of overfitting.
3 The performance of the models is evaluated using multiple statistical parameters such as Mean Square Error (MSE), R-squared Error (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).
4 Extensive analysis has been performed on two different datasets that incorporate readings with a time span of 1 minute of house applications in kiloWatts from smart meters. To the best of our knowledge, this is the first paper highlighting power consumption in smart homes using the DT-RF-GB-based ensemble approach.

The rest of the paper has been organized as follows. Section 2 presents some relevant related works in the area. Section 3 denotes the machine learning algorithms deployed for the ensemble method proposed. In Section 4, we detail the experimental analysis, including the datasets and evaluation parameters. This section also includes the results obtained from the extensive study, along with a comparative analysis. Section 5 discusses conclusions and future works.

2. Related works

In this section, we present a detailed survey of past related works. Since energy management is a global issue, the area has witnessed extensive research over the last few decades. The progress in technology has led to several methods being proposed for energy monitoring as well as management. We highlight some proposed methods in this section and propose an ensemble-based approach in the next section.

Ref. [12] proposed a Multi-output Adaptive neuro-fuzzy inference system (MANFIS) based smart home energy management system for efficient energy management in smart grids. The proposed design aims to reduce electricity costs and reverse power flow. The system is tested with daily data concerning temperature, wind speed, isolation, and controllable and uncontrollable appliances power as input. The output is used for handling energy production and consumption. Results show that electricity cost and power consumption are reduced significantly. Ref. [13] suggested another energy-saving method by integrating big data and machine learning with smart homes. The study deploys the J48 machine learning algorithm, along with WEKA API, for learning user behavior, in terms of energy consumption patterns and thus classifies houses based on energy consumption. RuleML and Apache Mahout have generated recommendations based on user preferences. A case study has been incorporated to manifest reduced energy consumption. Ref. [14] recommended a hybrid robust-stochastic optimization technique for energy management in smart homes. The study conducted for "day ahead" and real-time energy markets incorporate a robust optimization approach for managing the day-ahead market prices. The proposed optimization framework takes into account the real-time energy market, as well as the associated uncertainties, by relying on stochastic programming. The study estimates a profit of day-ahead and real-time markets to be $2.5/day. Ref. [15] proposed a Bio-Inspired Dragonfly Algorithm and Genetic Algorithm for optimizing energy consumption. The study considers two classes of appliances, i.e., Shiftable appliances and Non-shiftable appliances. Simulation results show that the proposed algorithm can minimize electricity costs and a tolerable waiting time. The only drawback of the proposed system is the waiting time. This is because as electricity cost decreases, there is an increase in the waiting time and vice versa. The proposed method also ensures the stability of the grid since the stability of the grid is dependent on the peak-to-average ratio.

Ref. [16] suggested a distributed artificial bee colony for connected appliances for efficient energy consumption in smart homes. In this study, swarm intelligence has been applied to connected devices. Overall decentralized management leads to sharing of information so that individual decisions can be taken, thereby optimizing electricity prices. The proposed approach has been evaluated in a smart home-connected environment, and the simulation manifests load-balancing optimization [17]. recommended a combined Deep learning-IoT-based platform for effective energy management in smart buildings. The YOLO v3 algorithm is used to detect people and count the number of people. Likewise, it is possible to manage the operation of air conditioners in a building. Hence the number of people and status of air conditioners are published on the IoT platform, and decision-making is performed based on energy consumption. Intensive test scenarios support the validation of the study. The proposed model shows appreciable accuracy. [18] presented an artificial bee colony based on non-intrusive appliance monitoring for smart homes. The study, which considers a group of connected consumer electronics loads, has been carried out in two parts. First, data for individual appliances are collected and stored with varying loads. Second, the stored data is used to estimate the individual load current. Simulations in a practical household system have validated search-based optimization.

Ref. [19] proposed multi-objective energy management due to uncertainty in wind power forecasting. The electricity cost is formulated for achieving the best schedule of devices in smart homes. The study employs a multi-objective dragonfly algorithm for optimizing the technical and economic objective functions. Once the optimal Pareto front is deduced, the study relies on an analytical hierarchy process for selecting the best operational schedule for smart homes. The suggested approach is evaluated in a sample smart grid, and numerical results validate that the proposed management method efficiently improves the performance of the smart grid. [20] suggested a fuzzy logic-based approach for optimal household appliance scheduling. The proposed method considers electricity price and load consumption. The daily electricity usage is predicted by a predictive model and Demand Response (DR) scheme. After deploying a Long Short-Term Memory-based (LSTM) optimized predictive model, data is transmitted to a DR fuzzy logic-based controller. The LSTM model outperforms other baseline models and reduces electricity costs significantly. [21] recommended a

Table 1
Summary of the existing research works.

Research	Methodology	Strength	Weakness
Smart Home Energy Management System [12]	MANFIS	Significant reduction in cost of electricity	Curse of dimensionality, computational expense.
Smart Home Energy saving system [13]	Big data and machine learning	Reduces energy consumption	Platform specific, compatibility issues
Energy Management [14]	Hybrid robust-stochastic optimization model	Profitable energy management	Method is brittle, sensitive to change in parameters,Time consuming
Home Energy Management Optimization [15]	Bio-Inspired Dragonfly Algorithm and the Genetic Algorithm	Significant decrease in electricity cost	Increase in wait time, user discomfort, does not work well for all scenarios
Smart Home Energy Management System [16]	Distributed artificial bee colony algorithm	Optimized performance of energy management system	Requires new fitness tests on new parameters, slow
Effective energy management for smart buildings [17]	Deep Learning and IoT based approach (YOLOv3)	Enhanced decision making about energy consumption	Segregating small objects in a group setting is challenging
Enhanced Load Monitoring in Smart Homes [18]	Artificial bee colony algorithm	Efficiant Load Monitoring	Requires new fitness tests on new parameters, slow
Multi object Energy Mangament in Smart Homes [19]	Uncertainty model, dragonfly algorithm	Improved performance of smart grid	Increase in wait time
Otimal household appliance scheduling [20]	Fuzzy Logic and Machine Learning	Reduction in electricity cost, optimal scheduling	Inaccuracy in results due to assumptions
Smart Energy Mangament in Residential Areas [21]	Fuzzy Control System	Reduced energy consumption	Inaccuracy in results due to assumptions
Predicting Energy Consumption [22]	Convolutional Neural Networks (CNN), Bidirectional Long Short Term memory (LSTM), Auto Encoders	Improved performance, computation time and load distribution	Trial and Error experimemnts for selecting optimal hyperparameter values, insufficient data
Predicting Energy Consumption of Oil Pipelines [23]	Fruit fly optimizer, simulated annealing algorithm	High Prediction accuracy and reduced complexity	Algorithm can easily fall into local optimum leading to low convergence
Forecasting building energy consumption [24]	Wavelet Transformation, LSTM	Efficient forecasting for real case electricity consumption	Specific forecasting framework, training latkes longer and more memory
Prediction of heating energy consumption [25]	Pattern Analysis, LSTM	Improvised prediction performance and energy concumption	Training latkes longer and more memory, prone to overfitting

fuzzy control system for energy management in residential buildings. The study is based on environmental data, which is processed by a fuzz control system to recommend minimum energy consumption values. The system relies on a forward chaining Mamdani approach, along with decision tree linearization. As the proposed system generates fuzzy rules, the energy consumption behavior is highlighted. The proposed method manifests improved accuracy as well as faster computation.

Ref. [22] presented a study on using deep learning methods for predicting energy consumption. The study considers commercial and domestic buildings, and the proposed architecture incorporates a hybrid framework constructed using a convolutional neural network (CNN), autoencoder, and bidirectional long short-term memory networks. Experimental analysis suggests improving computation time, and the method achieves satisfactory performance. Ref. [23] proposed a data-driven model for predicting pipeline energy consumption. The study deploys a hybrid support vector that relies on the fruit fly optimizer. The model shows high prediction accuracy and outperforms the other baseline models considered. [24] presented a study on forecasting energy consumption. The study is based on two decomposition algorithms such as empirical mode decomposition and wavelet transformation, long short-term memory networks (LSTM). The analysis has been conducted for twenty buildings in various locations with different functionalities. Results show that LSTM with empirical mode decomposition shows the best performance. [25] suggested LSTM networks for predicting heating energy consumption on operation patterns of buildings. The study highlights three neural networks applied to different operation pattern data. The inputs taken for the three LSTMs were different, and adding additional variables to inputs yielded better results.

Table 1 summarizes the overall strengths and weaknesses of the existing works discussed.

Energy Management is a widely researched problem, and in the past several machine-learning methods have been proposed to encourage efficient energy usage. Since it is a global concern, a much more challenging task is to find methods of energy management that also consider performance efficiency and robustness. While the methods proposed previously shed light on the existence of various methods that can lead to energy management, there is always scope for increased accuracy and endurance, which we present in this study. The study emphasizes ensemble-based techniques (DT-RF-XGBoost) for analyzing energy consumption by combining multiple machine-learning techniques into one predictive model. Finding a good balance between bias and variance is necessary to minimize the total error. An optimal balance between bias and variance ensures no overfitting or underfitting. To understand the behavior of prediction models, there is a need to find an optimal balance between bias and variance, and ensemble models establish the same.

3. Machine learning methods and our proposed ensemble-based method

In this paper, we use machine learning models extensively, including DT, RF, and XGBoost. In addition, we proposed a hybrid method based on a Decision Tree, Random Forest, and eXtreme Gradient Boost, called the DT-RF-XGBoost Ensemble method, which gives much better performance compared to the individual machine learning method.

3.1. Decision Tree (DT)

Decision trees may be defined as supervised machine learning models capable of predicting targets by learning decision rules from features. A decision tree model learns a set of questions based on deducing class labels and is very useful for interpretation. The root node or the first parent of a decision tree undergoes recursive partitioning [26]. Every node in this stage may be split into left and right child nodes, respectively. These nodes can become parents and be split into other nodes [27]. While this can result in a very deep tree due to overfitting, there is a need to prune the nodes. The optimal splitting is decided by the information gain, defined as an objective function that requires optimization using the tree learning algorithm. Decision trees are easy to read, learn and prepare. By creating a comprehensive analysis, decision trees consider all possible outcomes of a decision to conclude.

3.2. Random Forest (RF)

Random Forest regression may be defined as an ensemble learning method based on supervised learning algorithms. An ensemble learning technique incorporates predictions from several machine learning algorithms for making a more accurate prediction concerning a single model. In a random forest model, the trees run parallel, and there is no interaction among them [28]. Several decision trees are constructed while training. The output is the mean of the classes. The algorithm works by picking k data points from a training set, followed by building a decision tree from the points [29]. After choosing the number of trees that must be built, the previous steps are repeated for all the trees. These new trees will predict the output and assign a new data point. These new data points are averaged to get the mean output. A random forest regressor is robust and works with features having non-linear relationships too. However, overfitting is a common problem, hence the number of trees must be chosen correctly.

3.3. eXtreme gradient boosting (XGBoost)

XGBoost is an ensemble learning method where the trees are built sequentially. As the sequencing continues, the errors get reduced. Each tree learns from the previous tree and reconditions the residual error. The most recent tree will have the least residual error in this manner. While the base learners in XGBoost are weak, they contribute vital information for prediction [30]. Hence, the overall boosting technique produces robust learning by combining weak learners. A strong learner also brings down bias and variance. Boosting uses fewer splits for prediction; thus, even small trees are highly interpretable.It is also possible to optimally select parameters through validation techniques, such as k-fold cross-validation. Since many trees may lead to overfitting, it is necessary to choose the stopping criteria. Thus XGBoost may lead to regularization and can effectively handle sparse data.

3.4. Proposed ensemble model: DT-RF-XGBoost ensemble

Our proposed model is an Ensemble Model based on a Decision Tree, Random Forest, and Extreme Gradient Boost (DT-RF-XGBoost Ensemble Model). Traditional machine learning models often run into issues related to performance, efficiency, and overfitting, which can be easily addressed using ensemble methods along with producing extremely accurate predictive models. These learning methods can combine multiple learners. They have supervised learning methods combining weak learners to produce strong ones. Ensemble Learning relies on many machine learning algorithms for building more efficient models for improving the overall prediction accuracy. While weak learners work individually to predict the target outcome, they are not the most optimal models as they are not generalized. They can predict few cases accurately and may not predict target classes and expected cases efficiently. Hence, combining these weak learners leads to formation of a generalized strong model which is optimized well enough to predict the target classes efficiently. Weak learners can be used as building blocks to design complex models as they do not perform well by themselves. When weak learners are combined, the bias variance trade-off is maintained and the ensemble achieves a better performance. The proposed Ensemble Model combines the Decision tree, Random Forest, and XGBoost, which have been discussed in detail previously. Fig. 1 depicts the basic ensemble architecture where multiple models run independently to produce a combined output.

Ensemble-based models ensure the best combination of machine learning algorithms [31]. More than that, ensemble-based models assert that a combination of algorithms will lead to fewer chances of error than a single algorithm. Hence, machine learning models working together will have a better potential for gaining accuracy. This is because diverse classifiers combined together will have a greater potential of gaining higher accuracy compared to non-diverse classifiers. Deploying an ensemble model will lead to improved average prediction performance over the other contributing components of the ensemble. The improved performance of the ensemble is associated with a reduction in the variance of prediction errors induced by ensemble components. Hence it adds to the robustness of the model. In our proposed model, each of the machine learning models (DT-RF-XGBoost) is trained using the same training dataset. All the models are trained and ready to predict. We have passed the same x_test date to each of the trained models (model-1, model-2, model-3), and each model predicts an output value (p1, p2, p3). Next, we combined all the predicted values by generating a mean. Finally, prediction using an ensemble model has been generated.

In the proposed ensemble model, we use a combination of decision trees, random forests and XGBoost algorithms. The decision tree is capable of handling both numeric and categorical data and work well with multi output problems. It does not require extensive preparation of data and the easily explained whitebox model is cost effective. Moreover, it works well eve if assumptions are violated. The limitations of this model lie in its overfitting and unstable nature which is taken care by the random forest component. Random forests work well with large datasets and provide high accuracy. The major limitation associated with random forests is the training time which is conveniently handled by the XGBoost algorithm, which provides a direct route to minimum error. Since it is based on gradient descent, it converges quickly in fewer steps and also leads to improved speed and lower computation cost [39–41]. Thus the

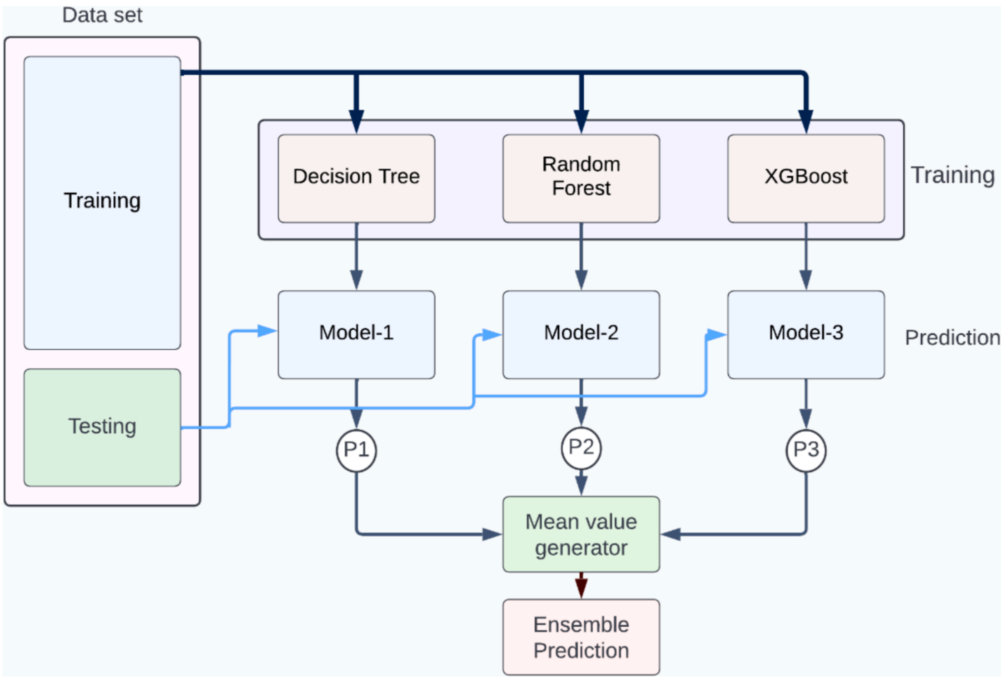

Fig. 1. Proposed ensemble model.

three components of the ensemble model complement provide immensely to the overall ensemble model, in their own ways and handle the limitations of each other, thereby making it an overall robust strong learner [42–44].

4. Experimentation results

In this section, we highlight the datasets followed by the evaluation metrics. Based on the experimental analysis, we also present the experimental results in scatter plots, pair plots, tables, and bar graphs.

4.1. Datasets

In our experimentations, we used two publicly available datasets from 'smart-home-dataset-with-weather-information1' Kaggle, namely HomeC.csv and HomeC1.csv. Both files incorporate readings with a time span of 1 minute of house appliances in kiloWatts (kW) from smart meters and weather conditions of the specific region. The datasets have 32 features each and a total of 503910 data points per dataset. Some of the characteristics are based on time, use [kW], gen [kW], House overall [kW], Home office [kW], fridge [kW]. For the experiment analysis, we are considering two groups, i.e., House overall [kW] and Home office [kW], for both datasets. The dataset has been split into 80% training data and 20% testing data. The datasets have been split into 80% training data and 20% test data taking into account the Pareto principle or the 80-20 rule. It is the most common split when the dataset is large. This split yields statistically meaningful results and good prediction accuracy. Moreover, it contributes to some optimization within the learning. We apply the data to baseline machine learning models DT, RF, XGB, and KNN and evaluate it against the proposed DT-RF-XGB ensemble model, using statistical parameters like *MSE, R^2, RMSE*, and *MAE*, respectively.

4.2. Evaluation metrics

In the experimentations, we used four evaluation metrics: Mean Square Error (MSE), R-squared Error (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MEA).

a *Mean Square Error (MSE)*: Mean Square Error or *MSE* is the mean or average of the square of differences between the real and predicted values:

$$MSE = 1 \Big/ n \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$$

where MSE is the mean squared error, n is a number of data points, Y_i-observed values, and $\widehat{Y}_i$-predicted values.

a *R-squared Error:* *R*-squared error or R^2 is used to determine how close data is to the regression fitted line. It gives the goodness of fit for regression models.

$$R^2 = 1 - \frac{RSS}{TSS}$$

where R^2 is the coefficient of determination, *RSS* is the sum of squares of residuals, and *TSS* is the total sum of squares.

a *Root Mean Square Error (RMSE)*: Root Mean Square Error or *RMSE* is used to determine the error in a model for estimating predictive data.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Predicated_i - Actual_i)^2}{N}}$$

b *Mean Absolute Error (MAE)*: Mean Absolute Error or *MAE* may be defined as the measure of errors for paired observations that express the same phenomenon

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n}$$

where MAE is the mean absolute error, y_i is the prediction, x_i is the true value, and n is the total number of data points.

4.3. Experimental results

In this subsection, we will depict the results for both datasets: Dataset1 (HomeC) and Dataset2 (HomeC1). For Dataset1 (D1) and Dataset2 (D2), we will be specifically analyzing power consumption for fields Home Office [kW] and House overall [kW].

While the datasets look similar and have the same features, the major difference between the two datasets lies in their values. Moreover, the domain values for Home Office and House Overall are different for both datasets. Thus the datasets and the field values have been used independently to perform the experimental analysis. The underlying idea behind using multiple datasets with different field values is to ensure non-independence and observe the patterns and results across multiple datasets for better overall analysis.

The results have been depicted using two data visualization tools, i.e., Pair Plots and Bar graphs. Using pair plots, we can obtain scatter plots for each of the machine learning models for both fields in both datasets. The scatter plot depicts the relationship between two variables, i.e., power consumption concerning time. The comparison between the values has been depicted using bar graphs.

4.3.1. Experimental results from dataset1 (HomeC)

This dataset has two fields: Home Office [kW] and House Overall [kW]. Firstly, we conducted experiments for the Home Office [kW] using individual Machine Learning techniques (e.g., Decision Tree, Random Forest, XGBoost, and KNN) and our proposed Ensemble Model (DT-RF-XGBoost). The results are presented in Fig. 2a-e.

Fig. 2a-d depict the scatterplot relationship between electricity consumption (Home Office) (x-axis) concerning time (y-axis) for Dataset1 (D1) for Decision Tree, Random Forest, XGBoost, and KNN, respectively. All the scatterplots show the values lying in the grid, and the horizontal line at the bottom implies how close the predicted values are to the original values. The line plot is an extension of the same scatter plot graph and shows that most values lie toward the bottom.

Comparing with the competitor methods, namely Decision Tree, Random Forest, XGBoost, and KNN (shown previously in Figs. 2-5), Fig. 6 depicts the Ensemble-based scatterplot relationship between electricity consumption (Home Office) (X-axis) concerning time (Y-axis) for Dataset 1 (D1). Here, **the data is less scattered** and based on the results, we observe that the predicted values here seem the closest to the original values concerning the other models. Again, the line plot is an extension of the same scatter plot graph and shows that maximum values lie towards the bottom. Hence, **the proposed Ensemble Model exhibits the best performance**.

In the second experiment, we conducted experiments for the House Overall [kW] using individual Machine Learning techniques (e.g., Decision Tree, Random Forest, XGBoost, and KNN), and our proposed Ensemble Model (DT-RF-XGBoost). The results are presented in Figs. 7-11.

Figs. 7-10 shows a scatterplot relationship between electricity consumption (House Overall) for (x-axis) concerning time (y-axis) for Dataset1 (D1) using Decision Tree, Random Forest, XGBoost, and KNN, respectively. The scatterplot shows the values lying in the grid, and it appears that the predicted values are close to the original values as they are clustered toward the bottom of the graph. The line plot is an extension of the same scatter plot graph and shows that most values lie toward the bottom. Since the changes occur on a microscopic scale, representing minute changes is almost impossible. Hence, while the graphs seem similar, there are significant differences in the values of performance parameters.

Comparing the four individual methods: Decision Tree, Random Forest, XGBoost, and KNN, as shown in Figs. 7-10, Fig. 11 depicts the Ensemble-based scatterplot. The predicted values, in this case, are **the closest to the original values**, and the line plot is an extension of the same scatter plot graph. Due to changes in the values on a microscopic scale, depicting the subtle differences is challenging; however, the values depict significant differences in the performance metrics of all the models.

4.3.2. Experimental results from Dataset2 (HomeC1)

We conducted the same experiments for Dataset2 (HomeC1); particularly, we ran experiments to evaluate Home Office [kW] and House Overall [kW] using individual Machine Learning techniques, as well as using our proposed Ensemble Model. The results

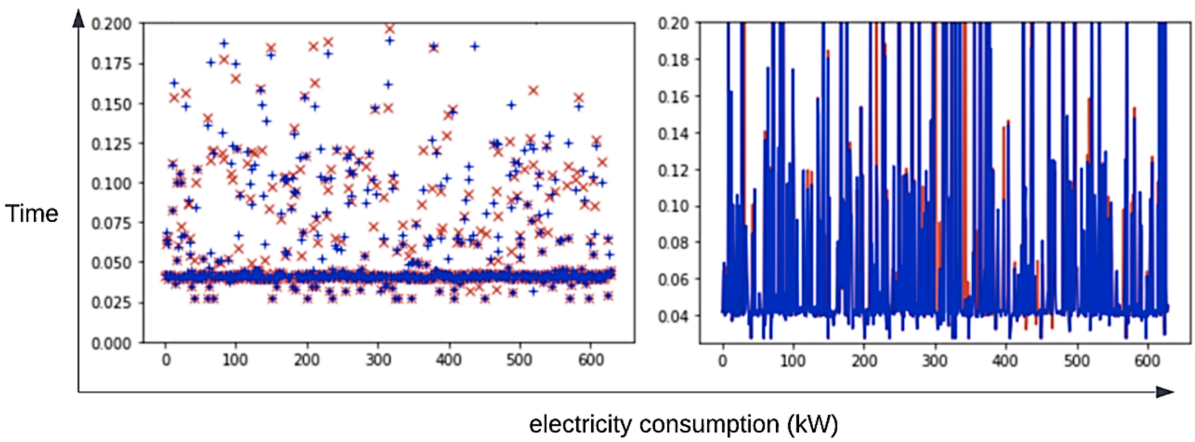

Fig. 2. Decision tree scatter plot and pair plot for home office (D1).

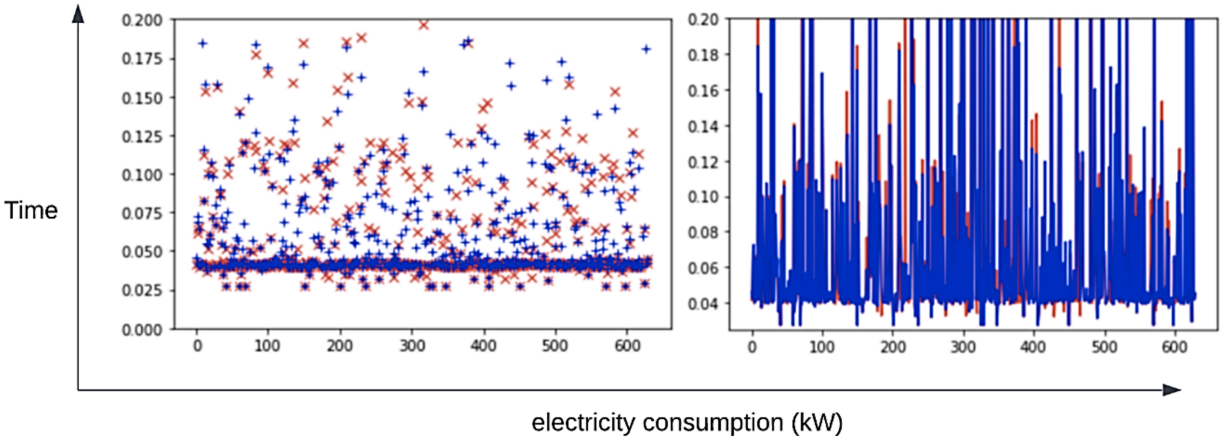

Fig. 3. Random forest scatter plot and pair plot for home office (D1).

presented in Fig. 4a-e are for Home Office [kW].

Figs. 12-15 depicts the scatterplot relationship between electricity consumption (Home Office) for (x-axis) concerning time (y-axis) for Dataset2 (D2), using Decision Tree, Random Forest, XGBoost, and KNN, respectively. The scatterplot shows the values in the grid, and the horizontal line at the bottom implies how close the predicted values are to the original values. For the XGBoost scatterplot (see Fig. 14), while the data points look much more scattered here, the predicted values are seemingly close enough to the original values at the bottom. For the KNN scatter plot (see Fig. 15), fewer are scattered here, and the predicted values are close to the original values at the bottom.

The line plot is an extension of the same scatter plot graph and shows that most values lie toward the bottom.

Fig. 16 depicts the Ensemble-based scatterplot relationship between electricity consumption (Home Office) for (X-axis) concerning time (Y-axis) for Dataset 2 (D2). In this case, the predicted values are the closest to the original values at the bottom. As the changes are made on a microscopic scale, depicting the same on a grid is challenging. However, the performance evaluation values differ significantly from previous models. The line plot is an extension of the same scatter plot graph and shows that most values lie toward the bottom.

For the remaining part of the experiment, we conducted an analysis of the House Overall [kW] using individual Machine Learning techniques (e.g., Decision Tree, Random Forest, XGBoost, and KNN), and our proposed Ensemble Model (DT-RF-XGBoost). The results are presented in Figs. 17-21.

Figs. 17-20 depicts the scatterplot relationship between electricity consumption (House Overall) for (X-axis) concerning time (Y-axis) for Dataset 2 (D2) using Decision Tree, Random Forest, XGBoost, and KNN, respectively. The scatterplot shows the values lying in the grid, and the predicted values are close to the original values. The line plot is an extension of the same scatter plot graph and shows that most values lie toward the bottom. Since the changes occur on a microscopic scale, representing minute changes is almost impossible. Hence, while the graphs seem similar, there are significant differences in the values of performance parameters.

Finally, Fig. 21 depicts the Ensemble-based scatterplot relationship between electricity consumption (House Overall) for (x-axis) concerning time (y-axis) for Dataset2 (D2). In this case, the predicted values are the closest to the original values at the bottom. As the changes are made on a microscopic scale, depicting the same on a grid is challenging. However, the performance evaluation values

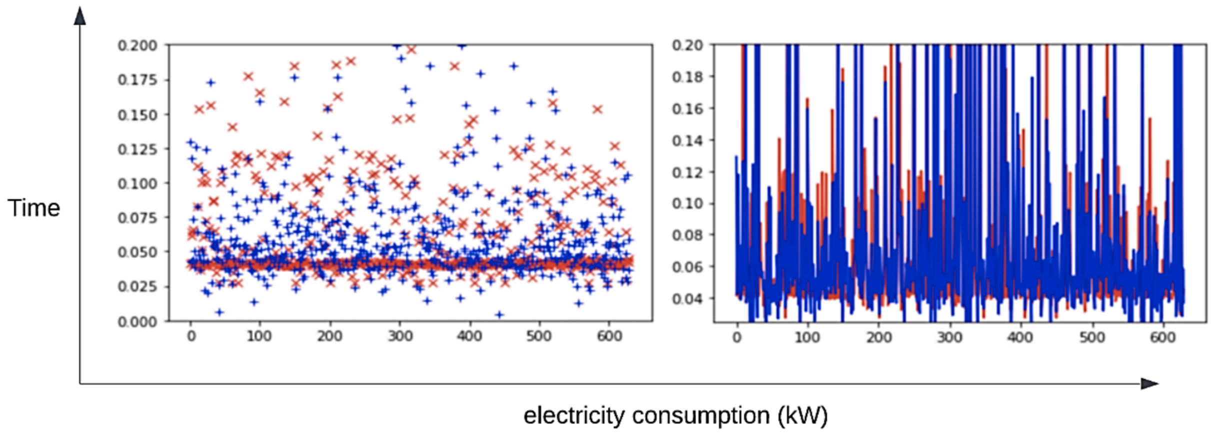

Fig. 4. XGB scatter plot and pair plot for home office (D1).

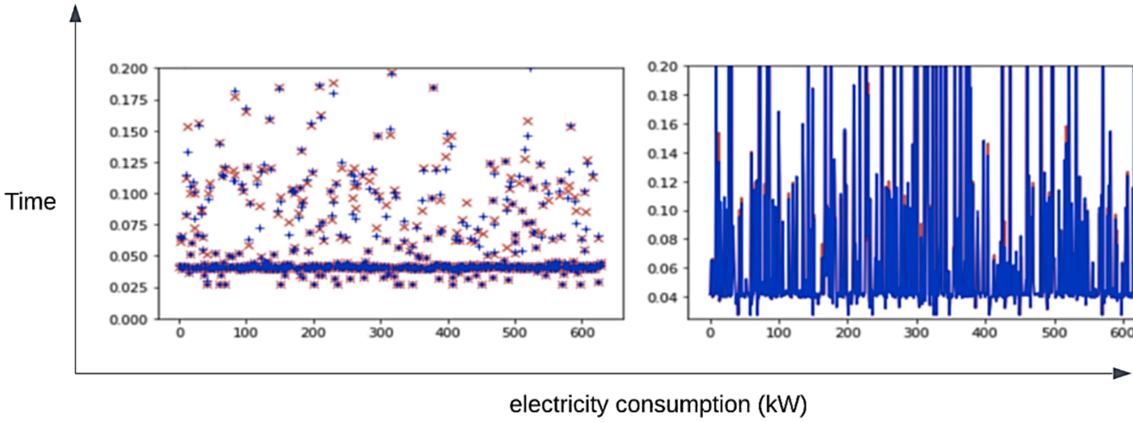

Fig. 5. KNN scatter plot and pair plot for home office (D1).

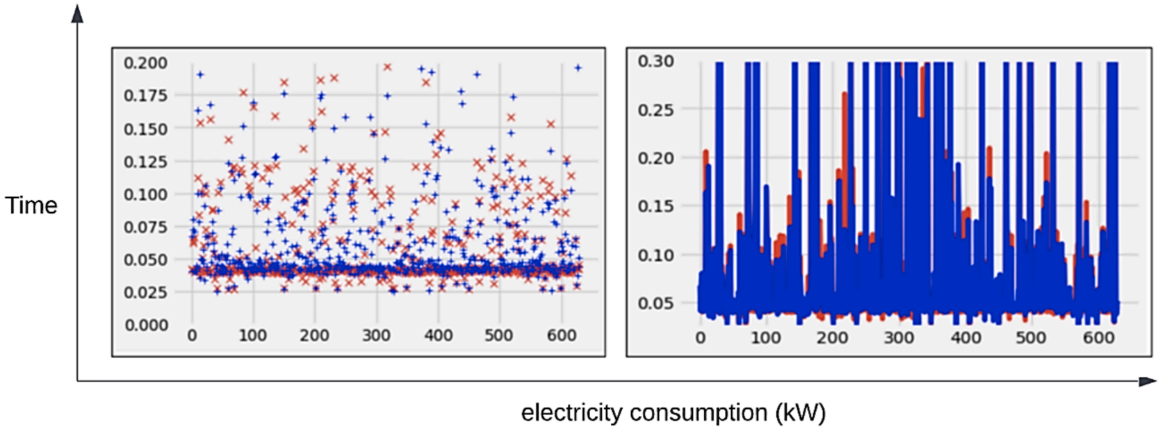

Fig. 6. Proposed ensemble model scatter plot and pair plot for home office (D1).

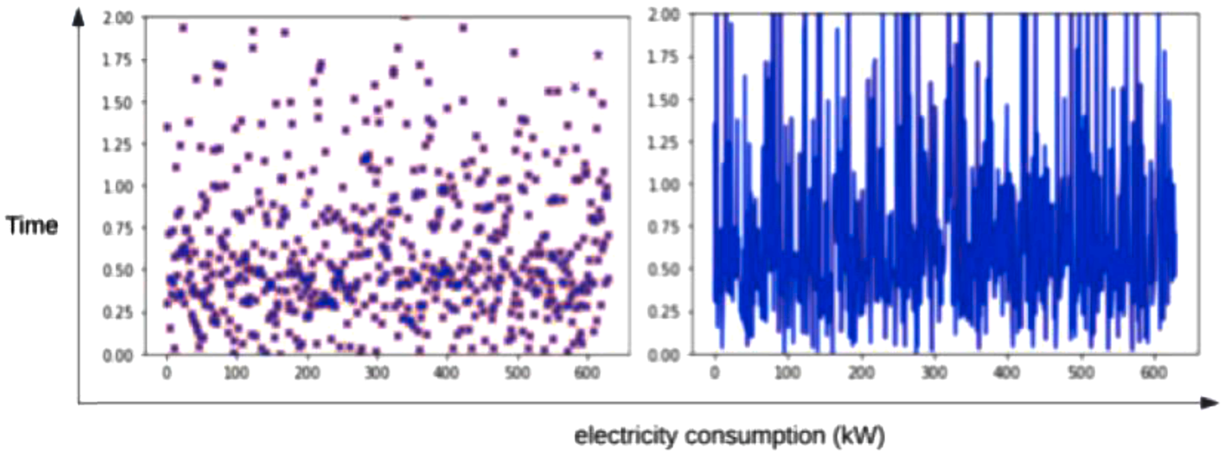

Fig. 7. Decision tree scatter plot and pair plot for house overall (D1).

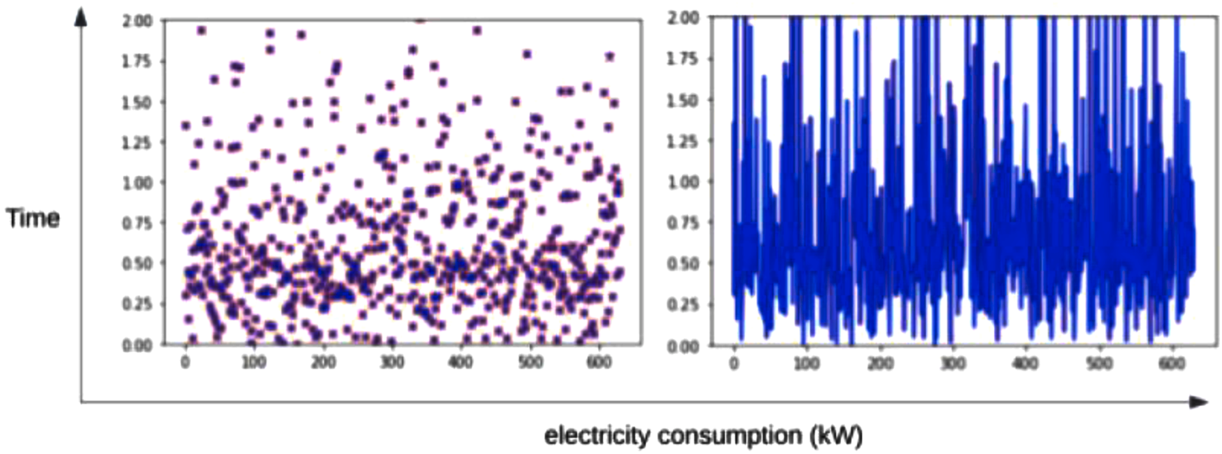

Fig. 8. Random forest scatter plot and pair plot for house overall (D1).

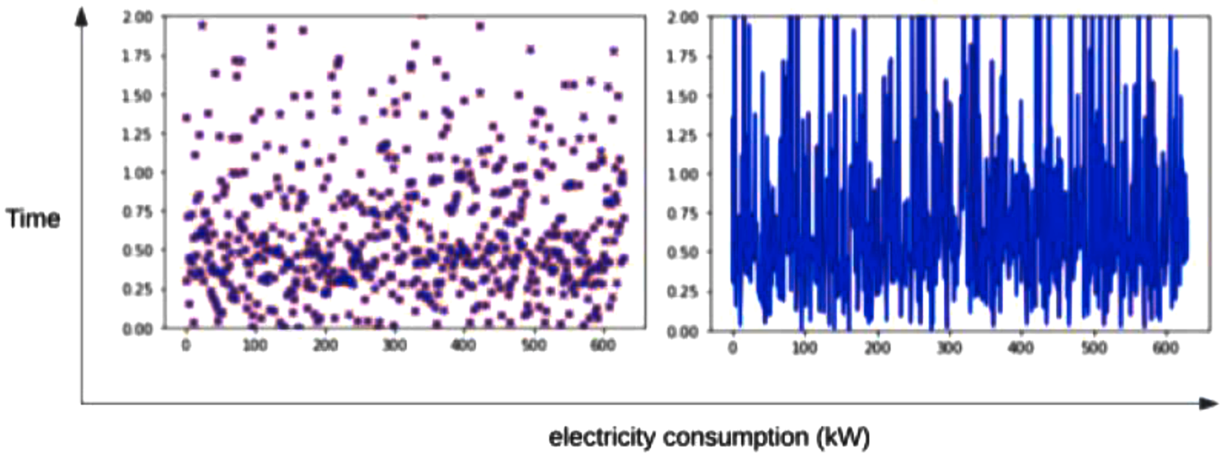

Fig. 9. XG-Boost scatter plot and pair plot for house overall (D1).

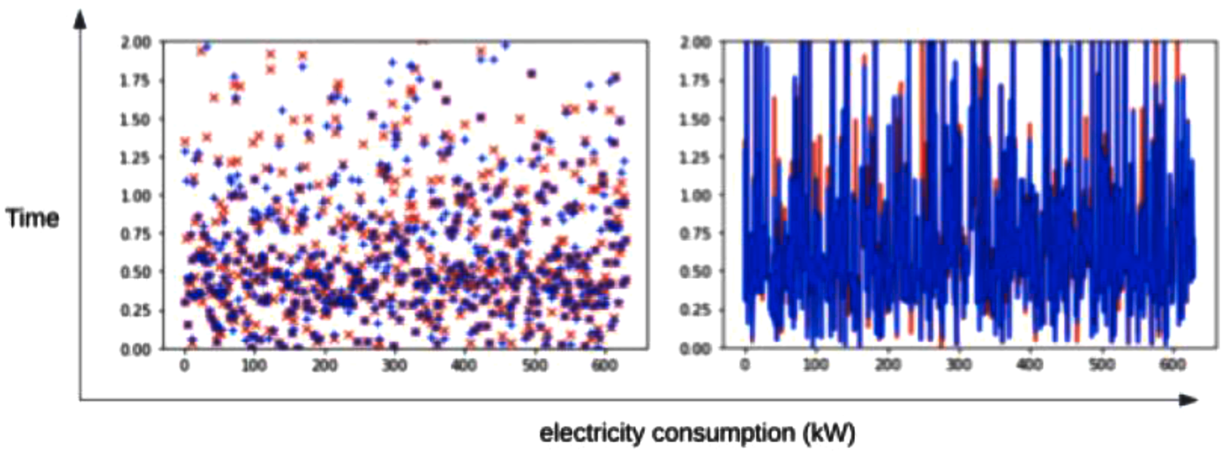

Fig. 10. KNN scatter plot and pair plot for house overall (D1).

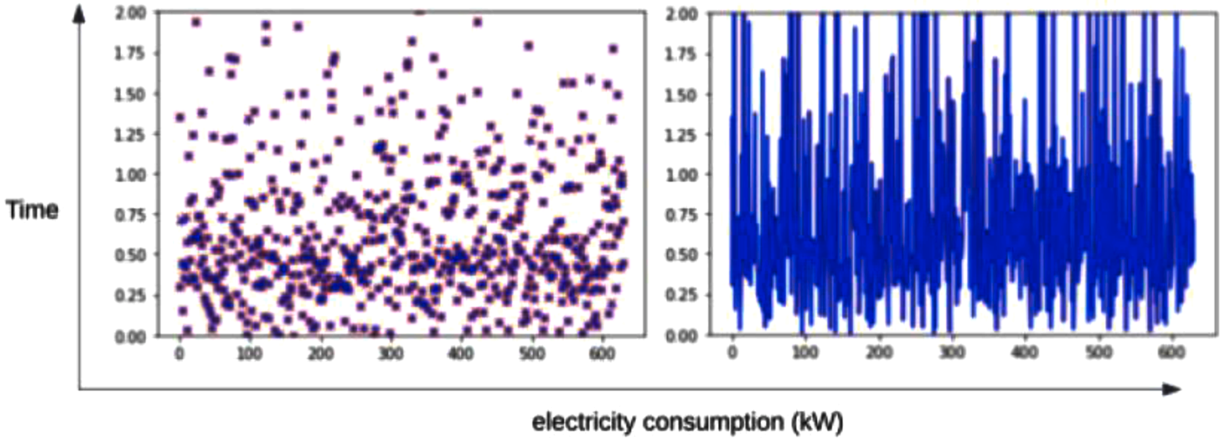

Fig. 11. Proposed ensemble model scatter plot and pair plot for house overall (D1).

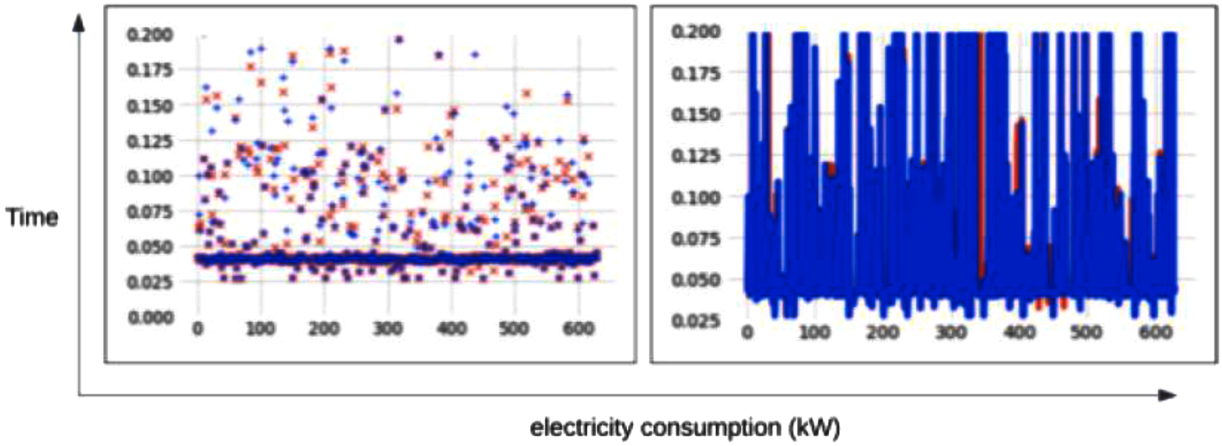

Fig. 12. Decision scatter plot and pair plot for home office (D2).

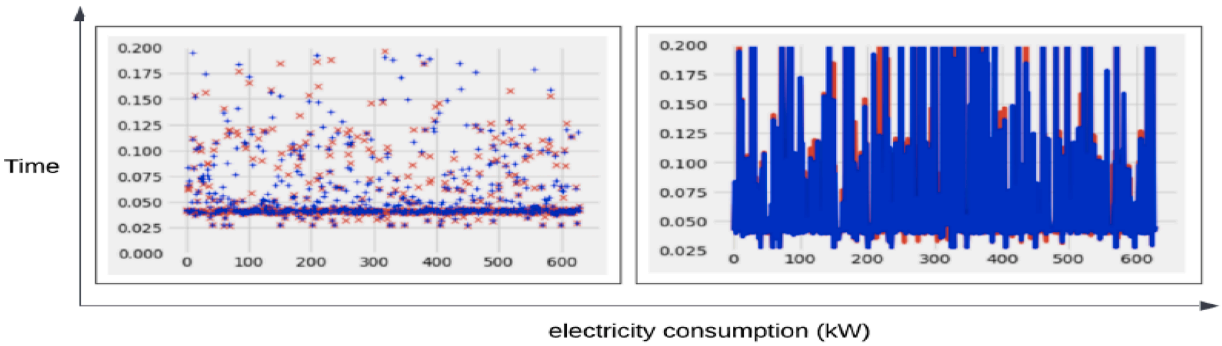

Fig. 13. Random forest scatter plot and pair plot for home office (D2).

differ significantly concerning previous models. The line plot is an extension of the same scatter plot graph and shows that most values lie toward the bottom.

4.3.3. Experimental results from the performance metric

As mentioned in the previous section, the performance evaluation metrics considered for the study are MSE, R^2, RMSE, and MAE. In this section, we will present the performance evaluation values based on the experimental analysis to better compare the models.

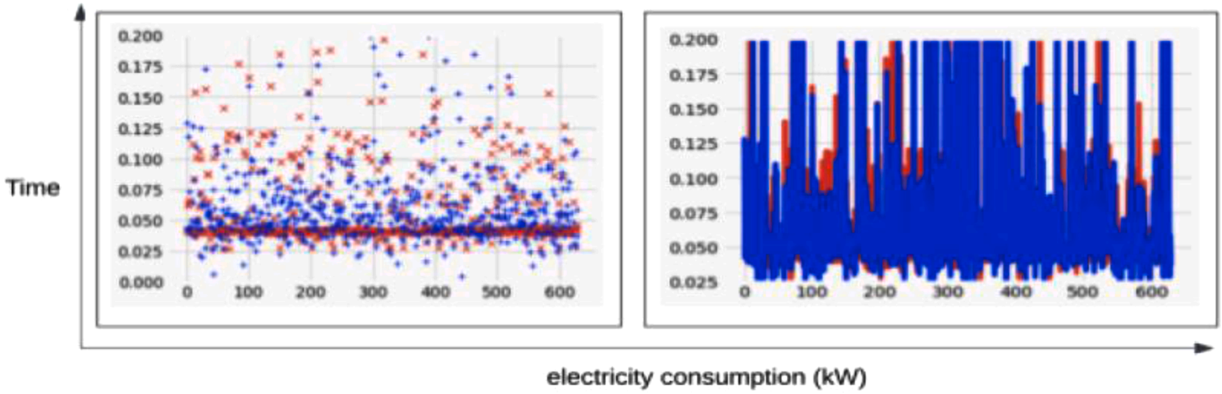

Fig. 14. XGBoost Scatter Plot and Pair Plot for Home Office (D2).

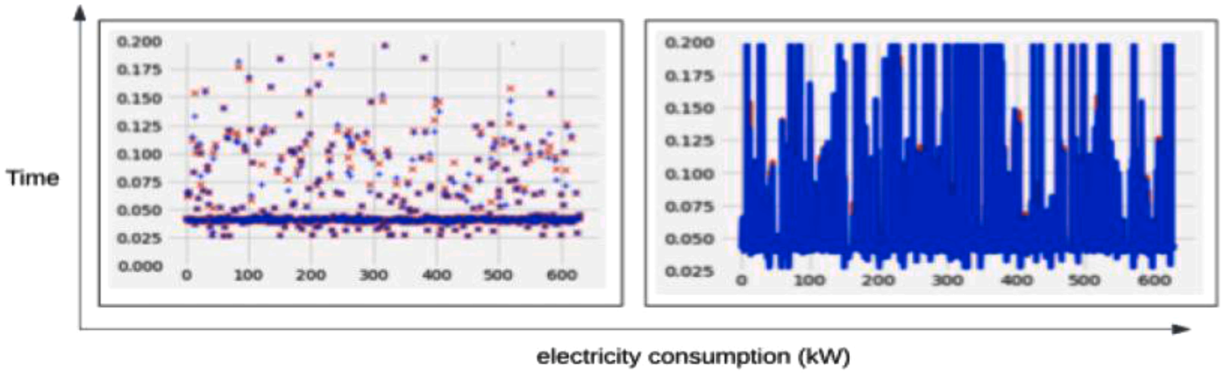

Fig. 15. KNN Scatter Plot and Pair Plot for Home Office (D2).

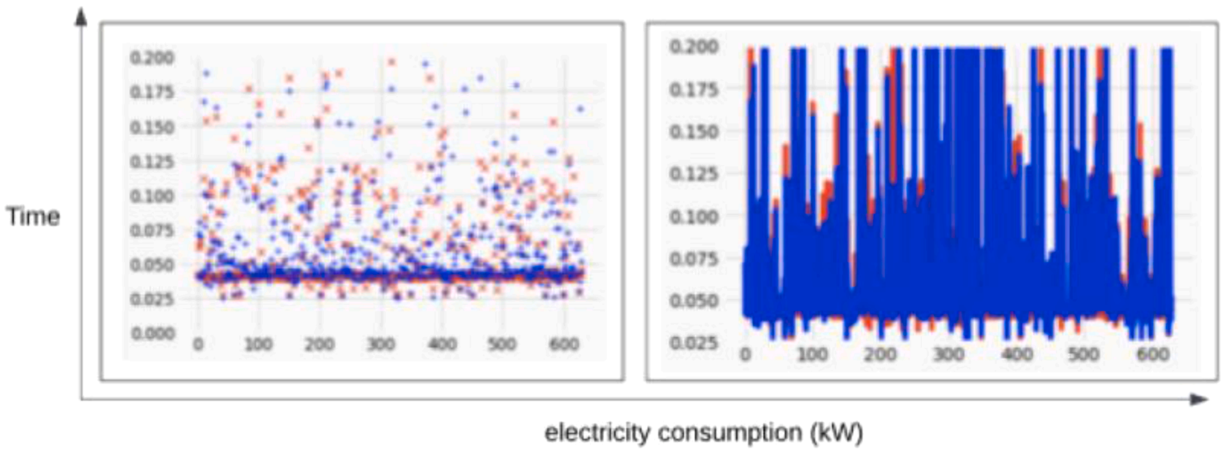

Fig. 16. Proposed ensemble model scatter plot and pair plot for home office (D2).

Table 2 shows the results from Dataset1 (HomeC) Home Office, whereas Table 3 shows the results from Dataset1 (HomeC) House Overall.

From Table 2, we can deduce that the MSE value for the proposed Ensemble Model is up to one order of magnitude better than the Decision Trees and XGBoost. The R^2 value for the proposed Ensemble Model is also the best among all other methods. For the RMSE value, the Ensemble Model performs better than all other methods except KNN. Finally, for the MAE value, the Ensemble Model is better than others, except XGBoost. While RMSE for KNN seems to be the lowest, it seems reasonable for the proposed model. Hence the proposed model outperforms the other baseline models.

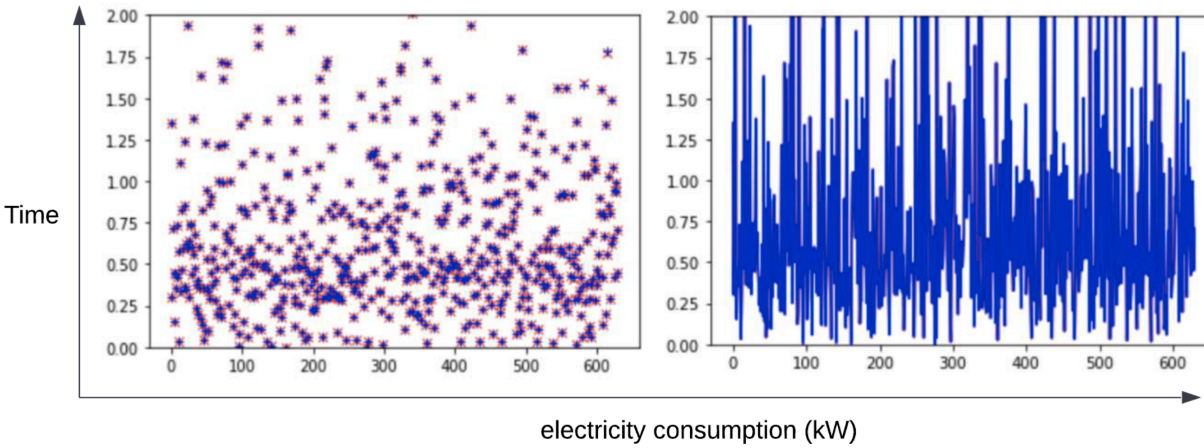

Fig. 17. DT Scatter plot and pair plot for house overall (D2).

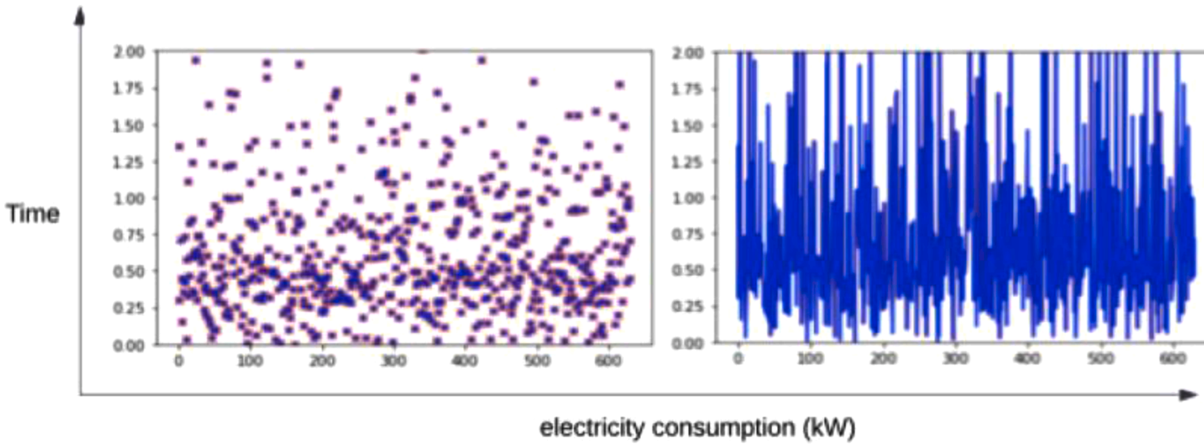

Fig. 18. RF Scatter plot and pair plot for house overall (D2).

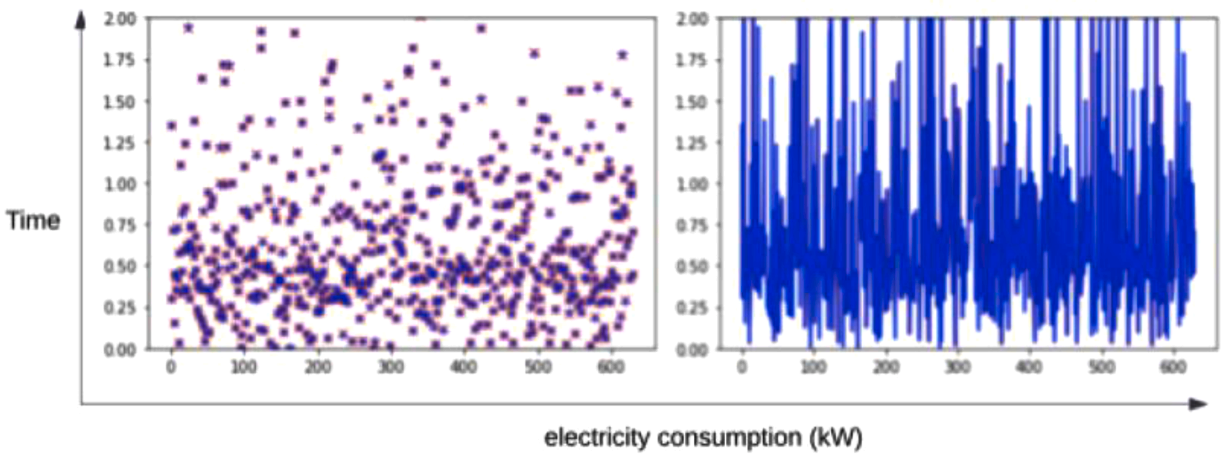

Fig. 19. XGB scatter plot and pair plot for house overall (D2).

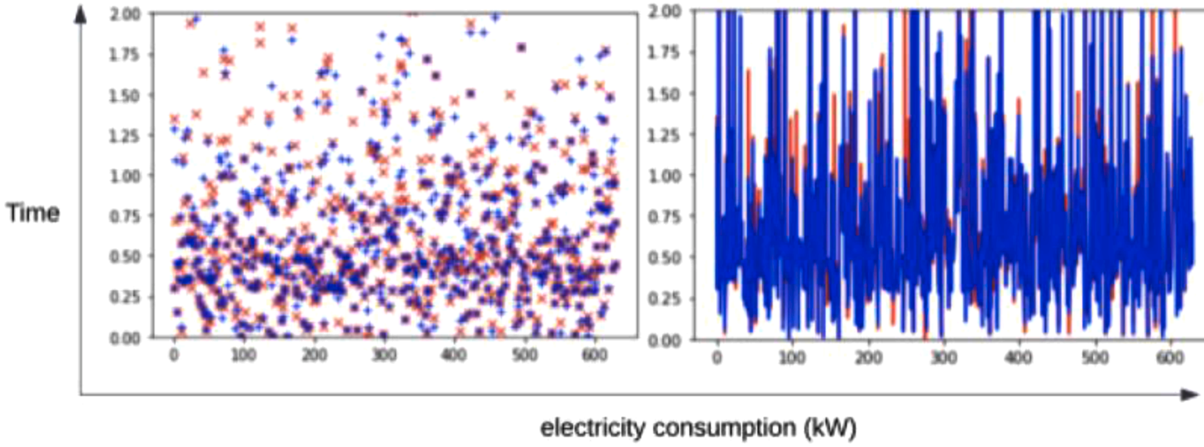

Fig. 20. KNN scatter plot and pair plot for house overall (D2).

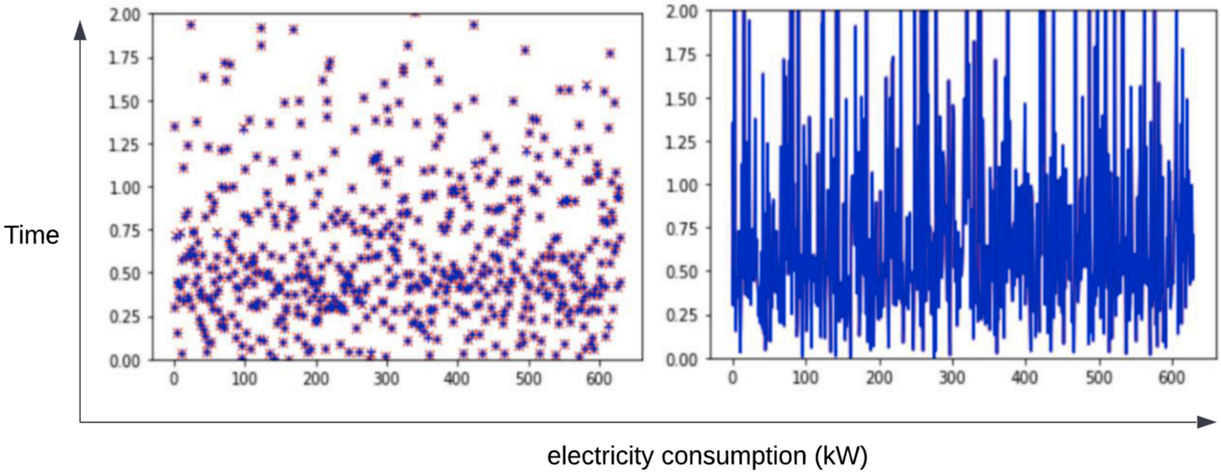

Fig. 21. Proposed ensemble model scatter plot and pair plot for house overall (D2).

Table 2
Performance evaluation for home office (D1).

Machine Learning Algorithm	MSE	R^2	RMSE	MAE
Decision Trees	0.0014	0.8674	0.0383	0.0087
Random Forest	0.0006	0.9393	0.0259	0.0086
XGBoost	0.0012	0.8839	0.0358	0.0202
k-Nearest Neighbor	0.0003	0.9477	0.0116	0.0018
Proposed Ensemble Model	0.0001	0.9846	0.0173	0.0108

Table 3
Performance evaluation for house overall (D1).

Machine Learning Algorithm	MSE	R^2	RMSE	MAE
Decision Trees	0.049800	0.98010	0.00128	0.00026
Random Forest	0.000001	0.98820	0.00082	0.00010
XGBoost	0.000006	0.98390	0.00272	0.00164
k-Nearest Neighbor	0.069850	0.99010	0.26430	0.08870
Proposed Ensemble Model	0.000008	0.99999	0.00072	0.00033

Table 4
Performance Evaluation for Home Office (D2).

Machine Learning Algorithm	MSE	R^2	RMSE	MAE
Decision Trees	0.00147	0.86666	0.03841	0.00878
Random Forest	0.00069	0.93437	0.02630	0.00069
XGBoost	0.00128	0.88398	0.03583	0.02022
k-Nearest Neighbor	0.00113	0.94142	0.01163	0.00187
Proposed Ensemble Model	0.00061	0.98776	0.02486	0.0109

Table 5
Performance evaluation for house overall (D2).

Machine Learning Algorithm	MSE	R^2	RMSE	MAE
Decision Trees	0.01418	0.98998	0.00142	0.00026
Random Forest	0.00004	0.98099	0.00088	0.0001
XGBoost	0.000006	0.96969	0.00272	0.00164
k-Nearest Neighbor	0.06985	0.93556	0.2643	0.0887
Proposed Ensemble Model	0.000002	0.99999	0.00078	0.00033

From Table 3, we can easily see that The R^2 and RMSE values for the proposed Ensemble Model seem to be the best. While Random Forest has the best MSE value and the MAE value, the proposed ensemble model has reasonable MAE and MSE values. Hence the proposed model outperforms the other models in this case.

Tables 4 and 5 focus on Dataset2 (HomeC1). Again, we experimented with the Home Office and House Overall elements of the dataset.

From Table 4, it shows that the MSE and R^2 values for the proposed model seem to be the best in this case. While the RMSE value is the best for KNN, and the MAE value is best for RF, the overall performance of the proposed Ensemble model seems better.

From Table 5, it shows that the MSE, R^2 and RMSE values for the proposed Ensemble model seem to be the best in this case. Although the MAE value is the best for Random Forest, the overall performance of the proposed model seems better.

Based on the experiments above, we can confirm that the proposed Ensemble model outperforms the other baseline models.

The same may be explained using bar graphs (see Figs. 22-23). From Fig. 22, it is evident that the proposed ensemble model achieves the best RMSE value for Dataset1 for both Home Office and House Overall.

Similarly, from Fig. 23, it is evident that the proposed ensemble model achieves the best RMSE value for Dataset2 in terms of Home Office as well as House Overall.

4.4. Comparative analysis

While the experimental analysis on multiple datasets indicates that the overall performance of the proposed Ensemble model outperforms the baseline models for both the fields in both the datasets, we present a comparative analysis of our work with some of the relevant research works done in the past (Table 6).

Based on the overall analysis, we can make certain observations. The significance of this work is multi-fold. We have introduced yet another machine-learning methodology for monitoring. power consumption in smart homes. The proposed method is an ensemble approach incorporating three machine learning models, i.e., DT, RF and XGBoost The approach was adopted to enhance performance issues and robustness, which are often a limitation in traditional machine learning approaches. The performance of the proposed ensemble method was compared to traditional machine learning methods, including the individual components of the ensemble

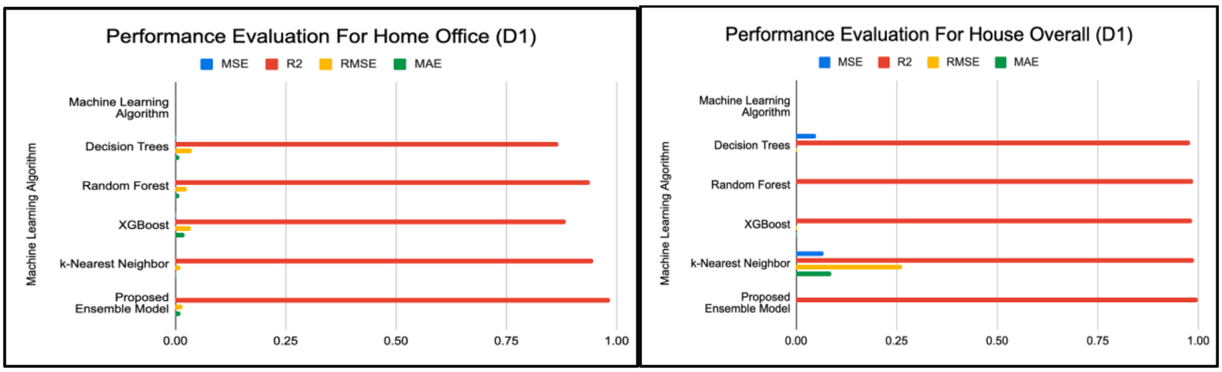

Fig. 22. Performance evaluation for home office and house overall (D1).

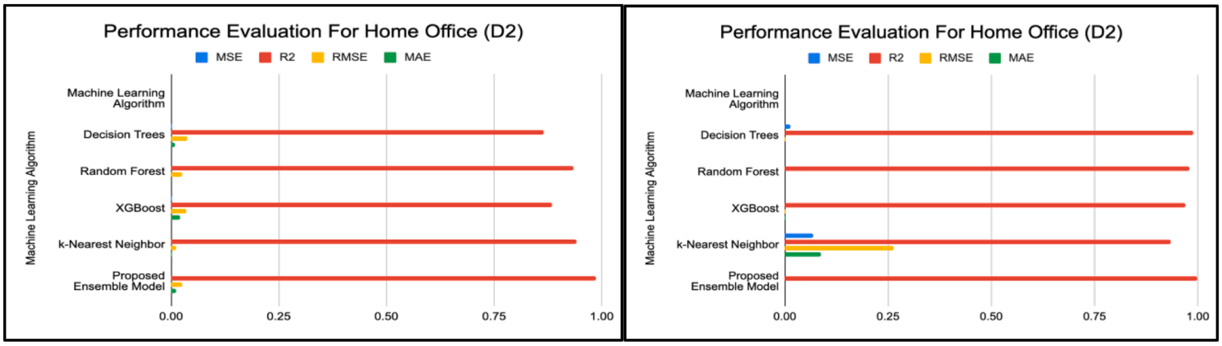

Fig. 23. Performance evaluation for home office (D2).

Table 6
Comparative analysis of our proposed work with related works.

Authors	Proposed Work	Methodologies/ Parameters	Result
Culaba et al. (2020) [32]	Energy Consumption for mix use buildings	Clustering and Forecasting (k-means, support vector machines)	R2~0.0201, MAE~ 0.0163
Khan et al. (2020) [33]	Predicting Energy Consumption for renewable/non-renewable power sources	Multilayer perceptron (MLP), support vector regression (SVR), and CatBoost,	MAE~15.72, MSE~472.96, RMSE~ 21.74
Chou and Truong, (2021) [34]	Energy Consumption Forecasting (MWh)	Time Series ML models	R2~ 0.86, RMSE~ 8.421, MAE~70.919
Shapi et al. (2021) [35]	Energy Consumption prediction for smart building	Support Vector Machine, Artificial Neural Network, and k-Nearest Neighbour	Lowest RMSE for KNN~0.5439, SVM~ 0.5558, ANN~0.5471
Lee et al. (2021) [36]	Prediction of Heating Energy Consumption	Deep Neural Network	R2~0.961
Amasyali and El-Gohary (2021) [37]	Energy consumption prediction in office building	ML algorithms like classification and regression trees (CART), ensemble bagging trees (EBT), artificial neural networks (ANN), and deep neural networks (DNN)	DNN performs the best with 2.97% coefficient of variation
Khan et al (2021) [38]	Energy Consumption Forecasting Model	Ensemble, LSTM and Gated Recurrent Units (GRU)	Mean Absolute Percentage Error (MAPE) values are 4.182 and 4.54
Proposed Model	Predicting energy consumption in smart homes	DT-RF-XGBoost based Ensemble Model	R2~0.9999 across both datasets, likewise, MSE, RMSE and MAE values low across all fields in both datasets

model.

Moreover, the experimental analysis was conducted on multiple datasets to avoid any kind of bias. The analysis yields consistent results across multiple datasets depicting that the proposed method performs better than the individual machine-learning models.

Moreover, the significance has been justified in the comparative analysis as our proposed work shows satisfactory results compared to similar research works in the past.

5. Conclusions and future work

Increased energy consumption has led to an increased carbon footprint and elevated climate change risk. Due to the higher demand for energy across the globe, not only higher costs of energy are incurred, but also there is a constant demand for supply. Hence monitoring energy consumption is necessary to manage energy costs and realize saving opportunities. One of the common ways of monitoring energy consumption is by predicting its usage.

In this study, we have deployed four machine learning algorithms to study the energy consumption in smart homes, i.e., DT, RF, XGBoost and KNN

Moreover, we have proposed a novel Decision Tree- Random Forest- XGBoost-based Ensemble model (DT-RF-XGBoost) for comparing it to the four baseline machine learning algorithms. The study considered two datasets that incorporate readings with a time span of one minute of house appliances in kiloWatts (kW) from smart meters. Each of these datasets has multiple fields. For this study, we have considered two fields from each dataset, which results in an overall four different experimental analyses, hence strengthening our claim. The performance evaluation metrics considered for the study are MSE, R^2, RMSE, and MAE. Our study depicts that the proposed ensemble model outperforms all the baseline algorithms across multiple fields in multiple datasets.

In the future, we would like to monitor energy consumption by relying on other machine learning techniques like Neural Networks and Optimization methods. As machine learning methods continue gaining popularity, it would be interesting to solve such problems using deep learning, time series analysis, and other advanced machine learning algorithms. Energy consumption can be studied on a global scale to analyze the largest sources of energy consumption. Consequently, techniques may be introduced to address the same. Moreover, once the consumption is analyzed, we can look forward to building trustworthy systems for the reliable consumption of energy.

Declaration of Competing Interest

None.

Data availability

Data will be made available on request.

References

[1] H. Kim, H. Choi, H. Kang, J. An, S. Yeom, T. Hong, A systematic review of the smart energy conservation system: From smart homes to sustainable smart cities, Renewable Sustainable Energy Rev. 140 (2021), 110755.
[2] V. Puri, S. Jha, R. Kumar, I. Priyadarshini, M. Abdel-Basset, M. Elhoseny, H.V. Long, A hybrid artificial intelligence and internet of things model for generation of renewable resource of energy, IEEE Access 7 (2019) 111181–111191.
[3] S. Alani, S.N. Mahmood, S.Z. Attaallah, H.S. Mahmood, Z.A. Khudhur, A.A. Dhannoon, IoT based implemented comparison analysis of two well-known network platforms for smart home automation, Int. J. Electr. Comput. Eng. 11 (1) (2021), 2088-8708.
[4] X. Wang, X. Mao, H. Khodaei, A multi-objective home energy management system based on internet of things and optimization algorithms, J. Build. Eng. 33 (2021), 101603.
[5] Y.B. Hamdan, Smart home environment future challenges and issues-a survey, J. Electron. 3 (01) (2021) 239–246.
[6] D. Chakraborty, A. Alam, S. Chaudhuri, H. Başağaoğlu, T. Sulbaran, S. Langar, Scenario-based prediction of climate change impacts on building cooling energy consumption with explainable artificial intelligence, Appl. Energy 291 (2021), 116807.
[7] A. Babuta, B. Gupta, A. Kumar, S. Ganguli, Power and energy measurement devices: a review, comparison, discussion, and the future of research, Measurement 172 (2021), 108961.
[8] E. García-Martín, C.F. Rodrigues, G. Riley, H. Grahn, Estimation of energy consumption in machine learning, J. Parallel Distrib. Comput. 134 (2019) 75–88.
[9] X. Dong, S. Deng, D. Wang, A short-term power load forecasting method based on k-means and SVM, J Ambient Intell Humaniz Comput (2021) 1–15.
[10] S. Verma, S. Singh, A. Majumdar, Multi-label LSTM autoencoder for non-intrusive appliance load monitoring, Electr. Power Syst. Res. 199 (2021), 107414.
[11] S. Kaushik, K. Srinivasan, B. Sharmila, D. Devasena, M. Suresh, H. Panchal, N. Srimali, Continuous monitoring of power consumption in urban buildings based on Internet of Things, Int. J. Ambient Energy (2021) 1–7.
[12] J.S. GK, J. Jasper, MANFIS based SMART home energy management system to support SMART grid, Peer Peer Netw Appl (2020) 1–12.
[13] I. Machorro-Cano, G. Alor-Hernández, M.A. Paredes-Valverde, L. Rodríguez-Mazahua, J.L. Sánchez-Cervantes, J.O. Olmedo-Aguirre, HEMS-IoT: A big data and machine learning-based smart home system for energy saving, Energies 13 (5) (2020) 1097.
[14] A. Akbari-Dibavar, S. Nojavan, B. Mohammadi-Ivatloo, K. Zare, Smart home energy management using hybrid robust-stochastic optimization, Comput. Ind. Eng. 143 (2020), 106425.
[15] I. Hussain, M. Ullah, I. Ullah, A. Bibi, M. Naeem, M. Singh, D. Singh, Optimizing energy consumption in the home energy management system via a bio-inspired dragonfly algorithm and the genetic algorithm, Electronics 9 (3) (2020) 406.
[16] K.H.N. Bui, I.E. Agbehadji, R. Millham, D. Camacho, J.J. Jung, Distributed artificial bee colony approach for connected appliances in smart home energy management system, Process. Expert Syst., Technol. Value Sugar Beet, Prog. Sugar Technol.: Proc. Gen. Assem. C.I.T.S. , 20th 37 (6) (2020) e12521.
[17] M. Elsisi, M.Q. Tran, K. Mahmoud, M. Lehtonen, M.M. Darwish, Deep learning-based industry 4.0 and Internet of Things towards effective energy management for smart buildings, Sensors 21 (4) (2021) 1038.
[18] S. Ghosh, D. Chatterjee, Artificial bee colony optimization based non-intrusive appliances load monitoring technique in a smart home, IEEE Trans. Consum. Electron. 67 (1) (2021) 77–86.
[19] M. Alilou, B. Tousi, H. Shayeghi, Multi-objective energy management of smart homes considering uncertainty in wind power forecasting, Electr. Eng. (2021) 1–17.
[20] S. Atef, N. Ismail, A.B. Eltawil, A new fuzzy logic based approach for optimal household appliance scheduling based on electricity price and load consumption prediction, Adv. Build. Energy Res. (2021) 1–19.
[21] D. Kontogiannis, D. Bargiotas, A. Daskalopulu, Fuzzy control system for smart energy management in residential buildings based on environmental data, Energies 14 (3) (2021) 752.
[22] O. Jogunola, B. Adebisi, K.V. Hoang, Y. Tsado, S.I. Popoola, M. Hammoudeh, R. Nawaz, CBLSTM-AE: A Hybrid deep learning framework for predicting energy consumption, Energies 15 (3) (2022) 810.
[23] H. Lu, Z.D. Xu, M. Azimi, L. Fu, Y. Wang, An effective data-driven model for predicting energy consumption of long-distance oil pipelines, J. Pipeline Syst. Eng. Pract. 13 (2) (2022), 04022005.
[24] S.Y. Chou, A. Dewabharata, F.E. Zulvia, M. Fadil, Forecasting building energy consumption using ensemble empirical mode decomposition, wavelet transformation, and long short-term memory algorithms, Energies 15 (3) (2022) 1035.
[25] J. Jang, J. Han, S.B. Leigh, Prediction of heating energy consumption with operation pattern variables for non-residential buildings using LSTM networks, Energy Build. 255 (2022), 111647.
[26] S. Jha, R. Kumar, M. Abdel-Basset, I. Priyadarshini, R. Sharma, H.V. Long, Deep learning approach for software maintainability metrics prediction, IEEE Access 7 (2019) 61840–61855.
[27] T.A. Tuan, H.V. Long, R. Kumar, I. Priyadarshini, N.T.K. Son, Performance evaluation of Botnet DDoS attack detection using machine learning, Evolut. Intell. (2019) 1–12.
[28] I. Priyadarshini, V. Puri, A convolutional neural network (CNN) based ensemble model for exoplanet detection, Earth Sci. Inf. (2021) 1–13.
[29] N. Pritam, M. Khari, R. Kumar, S. Jha, I. Priyadarshini, M. Abdel-Basset, H.V. Long, Assessment of code smell for predicting class change proneness using machine learning, IEEE Access 7 (2019) 37414–37425.
[30] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, in: *Proceedings of the 22*nd *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.
[31] K.W. Hsu, A theoretical analysis of why hybrid ensembles work, Comput. Intell. Neurosci. 2017 (2017).
[32] A.B. Culaba, A.J.R. Del Rosario, A.T. Ubando, J.S Chang, Machine learning-based energy consumption clustering and forecasting for mixed-use buildings, Int. J. Energy Res. 44 (12) (2020) 9659–9673.
[33] P.W. Khan, Y.C. Byun, S.J. Lee, D.H. Kang, J.Y. Kang, H.S. Park, Machine learning-based approach to predict energy consumption of renewable and non-renewable power sources, Energies 13 (18) (2020) 4870.

[34] J.S. Chou, D.N. Truong, Multistep energy consumption forecasting by metaheuristic optimization of time-series analysis and machine learning, Int. J. Energy Res. 45 (3) (2021) 4581–4612.
[35] M.K.M. Shapi, N.A. Ramli, L.J. Awalin, Energy consumption prediction by using machine learning for smart building: Case study in Malaysia, Dev. Built Environ. 5 (2021), 100037.
[36] S. Lee, S. Cho, S.H. Kim, J. Kim, S. Chae, H. Jeong, T. Kim, Deep neural network approach for prediction of heating energy consumption in old houses, Energies 14 (1) (2021) 122.
[37] K. Amasyali, N. El-Gohary, Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings, Renew. Sustain. Energy Rev. 142 (2021), 110714.
[38] A.N. Khan, N. Iqbal, A. Rizwan, R. Ahmad, D.H. Kim, An ensemble energy consumption forecasting model based on spatial-temporal clustering analysis in residential buildings, Energies 14 (11) (2021) 3020.
[39] A. Lakhan, MA. Mohammed, AN. Rashid, Seifedine Kadry and Karrar Hameed Abdulkareem, Deadline aware and energy-efficient scheduling algorithm for fine-grained tasks in mobile edge computing, Int. J. Web Grid Serv. 18 (2) (2022).
[40] B. Bhola, R. Kumar, BK. Mishra, Internet of things-based low cost water meter with multi functionality, Int. J. Web Grid Serv. 18 (3) (2022) 250–265.
[41] H. Sun, M. Liu, Z. Qing, X. Li, L. Li, Energy consumption optimisation based on mobile edge computing in power grid internet of things nodes, Int. J. Web Grid Serv. 16 (3) (2020) 238–253, https://doi.org/10.1504/IJWGS.2020.109468.
[42] A. Balamane, Scalable Biclustering algorithm considers the presence or absence of properties, *Int. J. Data Warehous. Min.* (IJDWM) 17 (1) (2021) 39–56, https://doi.org/10.4018/IJDWM.2021010103.
[43] H. Li, Z. Liu, P. Zhu, An engineering domain knowledge-based framework for modelling highly incomplete industrial data, *Int. J. Data Warehous. Min.* (IJDWM) 17 (4) (2021) 48–66, https://doi.org/10.4018/IJDWM.2021100103.
[44] T.T. Nguyen, N.L. Giang, D.T. Tran, T.T. Nguyen, H.Q. Nguyen, A.V. Pham, T.D. Vu, A novel filter-wrapper algorithm on intuitionistic fuzzy set for attribute reduction from decision tables, Int. J. Data Warehous. Min. (IJDWM) 17 (4) (2021) 67–100, https://doi.org/10.4018/IJDWM.2021100104.