Journal of Building Engineering 45 (2022) 103406

Contents lists available at ScienceDirect

Journal of Building Engineering

journal homepage: www.elsevier.com/locate/jobe

Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques

Razak Olu-Ajayi [a], Hafiz Alaka [a,*], Ismail Sulaimon [a], Funlade Sunmola [b], Saheed Ajayi [c]

[a] Big Data Technologies and Innovation Laboratory, University of Hertfordshire, Hatfield, AL10 9AB, UK
[b] School of Engineering and Technology, University of Hertfordshire, Hatfield, AL10 9AB, UK
[c] School of Built Environment, Engineering and Computing, Leeds Beckett University, Leeds, LS28AG, UK

ARTICLE INFO

Keywords:
Building energy consumption
Energy prediction
Machine learning
Energy efficiency

ABSTRACT

The high proportion of energy consumed in buildings has engendered the manifestation of many environmental problems which deploy adverse impacts on the existence of mankind. The prediction of building energy use is essentially proclaimed to be a method for energy conservation and improved decision-making towards decreasing energy usage. Also, the construction of energy efficient buildings will aid the reduction of total energy consumed in newly constructed buildings. Machine Learning (ML) method is recognised as the best suited approach for producing desired outcomes in prediction task. Hence, in several studies, ML has been applied in the field of energy consumption of operational building. However, there are not many studies investigating the suitability of ML methods for forecasting the potential building energy consumption at the early design phase to reduce the construction of more energy inefficient buildings. To address this gap, this paper presents the utilization of several machine learning techniques namely Artificial Neural Network (ANN), Gradient Boosting (GB), Deep Neural Network (DNN), Random Forest (RF), Stacking, K Nearest Neighbour (KNN), Support Vector Machine (SVM), Decision tree (DT) and Linear Regression (LR) for predicting annual building energy consumption using a large dataset of residential buildings. This study also examines the effect of the building clusters on the model performance. The novelty of this paper is to develop a model that enables designers input key features of a building design and forecast the annual average energy consumption at the early stages of development. This result reveals DNN as the most efficient predictive model for energy use at the early design phase and this presents a motivation for building designers to utilize it before construction to make informed decision, manage and optimize design.

1. Introduction

Energy efficient and sustainable buildings have become imperative towards saving the environment, as the inefficiency of buildings are the major contributors to world energy consumption and greenhouse gas emission [1]. The high proportion of energy consumed by buildings leads to major environmental problems causing climate change, air pollution, thermal pollution, among others, which deploys a severe impact on the existence of mankind [2]. In the past decades, the demand for energy in buildings has considerably amplified due to the population increase and prompt urbanization [3].

The study to better understand building energy efficiency has enthralled the attention of various researchers, which has engendered new developments through machine learning [4–6]. Building energy consumption forecasting is essential for energy conservation and better decision-making to decrease energy usage. Though, energy prediction remains a complex undertaking in view of a number of features affecting energy consumption such as weather conditions, building physical properties and occupants energy-use behaviour [7]. The American Society of Heating, Refrigeration and Air-Conditioning Engineers (ASHRAE) classified building energy prediction models into two key categories: forward models and data-driven models [8].

Forward models also known as physics-based modelling approach often require a large number of detailed inputs about the building and its environment such as HVAC (Heating, Ventilation and Air Conditioning) system, insulation thickness, thermal properties, internal occupancy loads, solar information, among others [9]. The simulation tools based

* Corresponding author.
E-mail addresses: r.olu-ajayi@herts.ac.uk (R. Olu-Ajayi), h.alaka@herts.ac.uk (H. Alaka), i.sulaimon@herts.ac.uk (I. Sulaimon), f.sunmola@herts.ac.uk (F. Sunmola), S.Ajayi@leedsbeckett.ac.uk (S. Ajayi).

https://doi.org/10.1016/j.jobe.2021.103406
Received 22 July 2021; Received in revised form 29 September 2021; Accepted 1 October 2021
Available online 9 October 2021
2352-7102/

Nomenclature

$F(\bullet)$	decision function
$\varphi(\bullet)$	non-linear function
PE	predicted energy value
AE	actual energy value
AE_m	actual energy values average
$\beta_0 + \beta_1$	regression coefficients estimate
n,i	number
R	attributes
X	nearest neighbour
x	input variable
y	output
ε	threshold
$\in$	radius

on this approach include mainly DOE-2, EnergyPlus and TRNSYS. The parameters required by these models are far too many and usually inaccessible. Therefore, these tools are considered inefficient due to the insufficient amount of required information and its time consumption [1].

In response, data-driven models are based solely on mathematical models and measurements. This model utilizes machine learning algorithms for building energy estimation and it has been proposed in several studies because it does not require a sizeable number detailed inputs about the building [1,3,10–12]. This method is trained on large detailed hourly or sub-hourly readings dataset retrieved from the building management systems and smart meters [13]. The accuracy of these models for predicting building energy consumption is dependent on these three factors: model selected, quantity and quality of the data [9].

In relation to the reduction of inefficient buildings, it is very important to address the origin of the problem by forecasting the energy consumption of buildings before construction [14]. At the early design phase, the customary model for predicting energy consumption is the forward model based on building energy modelling tools rather than adopting the most contemporary and best techniques for estimating energy consumption [15]. Machine Learning (ML) method is recognised as the best suited method for producing desired outcomes in prediction task [16,17].

Researchers suggest that the availability of a building energy system with accurate forecasting, is projected to save between 10 and 30% of total energy consumptions in buildings [3,18]. Thus, the continuous effort to enhance building energy prediction is essential for more efficient buildings. The advancement of data-driven models have produced satisfactory energy estimation results [19]. Although, without the detection of an algorithm that can accurately predict building energy consumption, this will increase greenhouse gas emission, construction of more inefficient buildings, energy demand and decrease in financial savings [20].

In the past decade, a considerable number of machine learning algorithms have been proposed to forecasting of energy consumption in buildings [1,3,10,11,13]. The machine learning algorithm utilized for building energy estimation mainly include Artificial Neural Networks (ANN), Support Vector Machine (SVM) and Decision Tree (DT). These algorithms have been trained and analysed using data from less than 1000 buildings [1,21–24,82]. It is an established fact that the larger the data, the more accurate the result, hence the number of building is not large enough to conclude, as the model has not been fully optimised which in some cases lead to inaccurate model prediction [25–29].

In the field of energy prediction, Artificial Neural Networks (ANN) has gained more recognition due to the exceptional results it produces. It is known to be dominant with big datasets which enable the neural network sufficient data to train model [30]. Runge and Zmeureanu in 2019 reviewed the utilization of ANN for forecasting hourly building energy and concluded that ANN algorithm produces good results in single and multi-step ahead forecasting [9]. Neto et al. conducted a comparative analysis of the prediction performance of ANN and an energy simulation tool called EnergyPlus using a single dataset of two blocks of buildings with six floors. The results revealed that data driven methods (ANN) are better suited for predicting building energy load [31].

Bagnasco et al. explored the use of multi-layer perceptron ANN for predicting electrical consumption in a single hospital facility based on meteorological data and time/day variation. After implementation, ANN forecast performed better during winter [24]. The application of Support Vector Machine (SVM) for predicting the energy consumption of buildings was first proposed by Dong et al. [21]; and it was applied to forecast monthly electricity consumption using four buildings based on meteorological data, Dong et al. revealed that SVM performed better than other related research using neural networks with a coefficient of Determination R^2 higher than 0.99 [21]. Li et al. applied SVM for forecasting hourly cooling load using a single office building. It was concluded that SVM produces a good result for hourly load prediction with a Root Mean Square Error (RMSE) of 1.17% [22,23]. Furthermore, the study by Dong et al. compared ANN and SVM for predicting hourly energy consumption of office buildings using a dataset containing 507 buildings. The meteorological data (dew point, atmospheric pressure outdoor temperature wind speed etc.) and building data (floor area and building type) were used as input features. Dong et al. stipulated that ANN produced better results than SVM with an RMSE of 5.71 and 7.35 respectively [32].

Generally, Decision Tree (DT) does not produce better results than neural networks for non-linear data. However, it's popularity can be ascribed to its ease of use and ability to produce predictive models with interpretable structures [33]. Yu et al. revealed that decision tree method could classify building energy demand levels satisfactorily [34]. In Hong Kong, Tso and Yau conducted a comparative analysis between decision tree, neural networks, and regression method in predicting weekly electricity consumption. It was determined that decision tree and neural networks performs slightly better than the regression method with a root of average squared error (RASE) of 39.36 [33].

Despite the need for accurate predictions to better understand building energy efficiency, none of these studies have explored using more than 1000 buildings to train the model for better prediction performance, and based on the performance metric utilized, none have achieved excellent prediction. It is a proven fact that accuracy highly depends on the algorithm used, quality and quantity of data [9]. In considering the best algorithm, it is deduced that accuracy result of algorithms applied on different dataset are not directly comparable because different data and different situation will produce unique results [35]. Hence, it cannot be concluded that an algorithm is better than the other unless it has been analysed and compared on the same dataset. There are very few studies that have conducted a comparative analysis of the major algorithms on the same data to identify best performing model in building energy forecasting. Hence, this study would utilize the same data and situation to evaluate several algorithms to determine the best method for estimating building energy consumption.

There are several applications of ML algorithms in the field energy consumption of operational buildings without much focus on forecasting energy consumed at the early design phase. The development of a prediction model with excellent performance would enable designers to check energy consumption level of a building design at the early design phase to deduce if there in need for design optimization. This would alleviate the construction of more energy inefficient buildings that are detrimental to the environment. Consequently, this study aims to develop a model using ML algorithms (Artificial Neural Network (ANN), Gradient Boosting (GB), Deep Neural Network (DNN), Random Forest (RF), Stacking, K Nearest Neighbour (KNN), Support Vector Machine

(SVM), Decision tree (DT) and Linear Regression (LR)) that enables designers input key features of a building design and forecast the annual average energy consumption using a larger dataset of multiple buildings. The key objectives of this research are listed below:

- To conduct a comparative analysis of the performance of several machine learning algorithms for predicting annual energy consumption on a large dataset of residential buildings.
- To compare each model's performance in terms of computational efficiency.
- To investigate the effect of building clusters on the feature selected and model performance.
- To investigate the effect of data size on the model performance.

The remainder of the paper is ordered as follows: Section 2 provides a literature review Section 3 gives the related theories for predicting building energy consumption. Section 4 presents the research methodology, which consist of the description of building, data pre-processing, model development and performance measures. Section 5 presents and discusses the result, while Section 6 provides the conclusions and the future work.

2. Literature review

The investigation of the best energy use prediction remains a complex task, as there is no general agreement on the most suitable algorithm to use for energy prediction [36]. In this research, the selected ML algorithms are a combination of the most utilized algorithms and other algorithm that are yet to receive much attention in the field of energy prediction. Each ML algorithms possess its own advantages and disadvantages, for example, ANN and SVM methods often generate better result than DT. Although, DT are recognised for its simplicity and ease of use [33]; Hai-xiang [37]. It is therefore important to understand the strength and weakness of the ML techniques before selection and implementation. Several studies have applied machine learning techniques for predicting building energy consumption and examples of these are summarized in Table 1.

3. Related theories

3.1. *Artificial Neural Networks*

Artificial Neural Networks are the most broadly utilized for predicting building energy consumption [42,43]. ANN is a non-linear computational model that emulates the functional concepts of the human brain [44]. ANN is an effective approach for solving non-linear problems and is dominant with big datasets which enable the neural network sufficient data to train the model [30]. There are several types ANNs such as Back Propagation Neural Network (BPNN), Feed Forward Neural Network (FFNN), Adaptive Network-based Fuzzy Inference System (ANFIS) etc. Among them, feed-forward is the most frequently utilized [39]. Multi-layer Perceptron (MLP) is a function of deep neural network that utilizes a feed forward propagation process with one hidden layer where latent and abstract features are learned [45].

The basic form of ANN consists of three consecutive layers namely input, hidden, and output layer as illustrated in Fig. 1 below. The input layer is used for train the model, the hidden layer is the bridge between input and output layer which can be modified dependent on the type of ANN while the output layer provides the result [43]. There are several types ANNs such as Back Propagation Neural Network (BPNN), Feed

Table 1
Examples of studies that applied machine learning for predicting building energy consumption.

References	Learning algorithm (type)	Building type	Type of energy consumption predicted	Sample size	Performance
[32]	Artificial Neural Network (ANN)	Non-Residential	Hourly load	507 instances	5.71 (RMSE)
[38]	Deep Neural Network (DNN)	Non-Residential	Short-term cooling load	1 Instance	175.7 (RMSE)
	Random Forest (RF)				168.7 (RMSE)
	Support Vector Regression (SVR)				143.5 (RMSE)
	Gradient Boosting Machines (GBM)				136.8 (RMSE)
	Elastic Net (ELN)				286.5 (RMSE)
	Extreme Gradient Boosting Trees (XGB)				129.0 (RMSE)
	Multiple Linear Regression (MLR)				286.8 (RMSE)
[22,23]	Support Vector Machine (SVM)	Non-Residential	Hourly load	1 Instance	1.17 (RMSE)
	Back Propagation Neural Network (BPNN)				2.22 (RMSE)
	the Radial Basis Function Neural Network (RBFNN)				1.43 (RMSE)
	General Regression Neural Network (GRNN)				1.19 (RMSE)
[1]	Random Forest (RF)	Non-Residential	Hourly load	5 Instances	5.53 (RMSE)
	M5 Model trees				6.09 (RMSE)
	Random tree (RT)				7.99 (RMSE)
[21]	Support Vector Machine (SVM)	Non-Residential	Hourly load	4 instances	0.99 (R^2)
[39]	Random Forest (RF)	Non-Residential	Hourly load	1 Instance	4.97 (RMSE)
	Artificial Neural Network (ANN)				6.10 (RMSE)
[40]	Artificial Neural Network (ANN)	Residential	Hourly load	N/S	1.68 (RMSE)
	Support Vector Machine (SVM)				1.65 (RMSE)
	Decision Tree (DT)				1.84 (RMSE)
	General linear regression (GLR)				1.74 (RMSE)
[41].	Stacking	Non-Residential	Hourly load	2 Instances	13.81 (RMSE)
	Random Forest				26.34 (RMSE)
	Decision Tree				19.20 (RMSE)
	Extreme Gradient Boosting				15.37 (RMSE)
	Support Vector Machine				16.12 (RMSE)
	K- Nearest Neighbour				17.81 (RMSE)
This Research	Deep Neural Network (DNN)	Residential	Annual load	5000 Instances	1.16 (RMSE)
	Artificial Neural Network (ANN)				1.20 (RMSE)
	Gradient Boosting (GB)				1.40 (RMSE)
	Support Vector Machines (SVM)				1.61 (RMSE)
	Random Forest (RF)				1.69 (RMSE)
	K Nearest Neighbors (KNN)				2.40 (RMSE)
	Decision Tree (DT)				2.55 (RMSE)
	Linear Regression (LR)				2.59 (RMSE)
	Stacking				2.60 (RMSE)

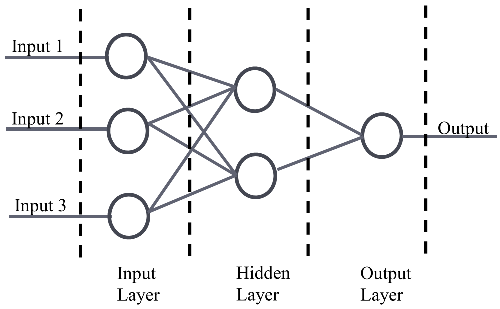

Fig. 1. Feed forward neural network architecture.

Forward Neural Network (FFNN), Adaptive Network-based Fuzzy Inference System (ANFIS) etc. Among them, feed-forward is the most frequently utilized [39]. Fig. 1 displays an illustrative diagram of a feed forward neural network architecture, containing of two hidden layers.

3.2. Support vector machine

SVM is a data mining algorithm, recognised as one of the most robust and accurate methods between all data mining algorithm [46]. SVM is increasingly used in research due to its ability to effectively provide solutions to non-linear problems in various sizes of data (Hai-xiang [47]. The SVM utilized for regression is known as Support Vector Regression (SVR), which has emerged a significant data driven method for forecasting building energy use. The main task in SVR is to create a decision function, $F(x_i)$, by using a historical data also called the training process. For the given input x_i, it is required that the outcome predicted should not differ from the real target y_i greater than the predetermined threshold ε. This function is often expected in form of

$$F(x_i) = (\varepsilon, \varphi(x_i)) + b \tag{1}$$

It is worth highlighting that SVM, or more specifically SVR are superior to other models because its framework is effortlessly generalized for various issues and it can achieve optimum solutions worldwide [48].

3.3. Decision Trees

Decision Tree (DT) is a method of utilizing a tree-like flowchart to partition data into groups. Decision Trees is an adaptable process that could advance with an enlarged amount of training data [49]. In contrast to other data-driven methods, DT is easier to comprehend, and its application does not require complex computation knowledge. However, it often produces major deviation of its predictions from actual results. DT is more suitable for forecasting categorical features than for estimating numerical variables [34].

DT commences execution at the root node where input data are split into various groups dependent on some predictor variables pre-set as splitting criteria. These split data are then dispersed into the branches originated from the root node denoted as sub-nodes. These sub-nodes data will either undergo further or no splits. The internal data node is where further data split is conducted to develop new subgroups. However, the concluding are the leaf nodes that handles the corresponding data groups at the current level as their final outputs. Fig. 2 is an example DT representation utilized for medium annual source energy use per unit floor (kWh/m2/yr) of a non-residential building. In which, the building consumption ratio and gross floor area are selected as predictor variables in the root and internal node respectively, adopted from Wei et al., [48].

3.4. K-Nearest Neighbour (kNN)

K-Nearest Neighbour (k-NN) algorithm is a non-parametric machine learning method that utilizes similarity or distance function d to predicts outcomes based on the k nearest training examples in the feature space [50]. kNN algorithm is one of the common distance functions that works effectively on numerical data [51]. However, KNN is yet to receive much attention in the field of building energy prediction. In the study by Feng et al., KNN produced good results in predicting building energy use with an R^2 of 0.84 [41]. The prediction for KNN as a regressor is performed as follows: From an input of x_i and output y_i is deduced based on the nearest record or most similar (nearest neighbour) $x \in X$. Fig. 3 illustrates the procedure for selecting the nearest neighbour in values of $\in$ .For example, squares labelled 1 and 2 will be chosen on query with radius $\in_2$. contrarywise, no squares will be retrieved on a query with radius $\in_1$. The radius must be increased until one element is selected [52].

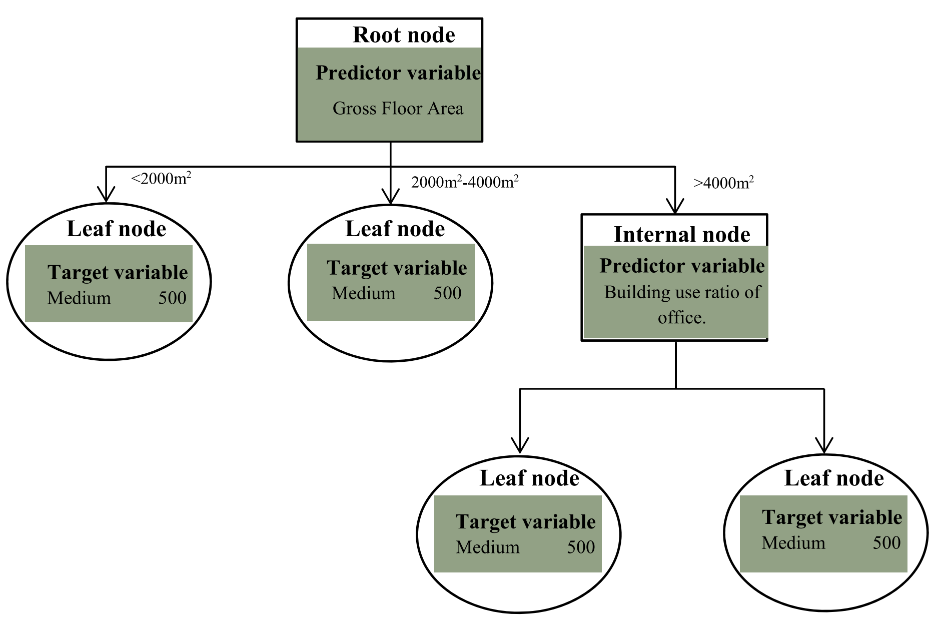

Fig. 2. Illustrative Diagram of a medium annual energy use per unit floor.

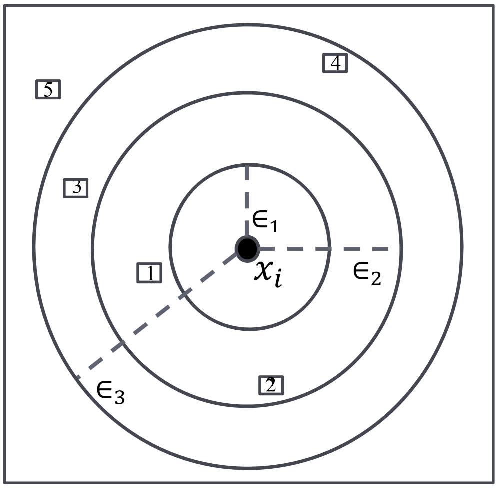

Fig. 3. Example of k-NN Regressor.

3.5. Gradient boosting

Gradient Boosting algorithm is a machine learning method that can be utilized for both regressor and classification problems. This technique algorithm builds model in stages like other boosting systems but generalizes these by enhancing an arbitrary differentiable loss function [53]. The Gradient Boosting method utilizes an ensemble of weak models which collectively form a stronger model. The final model is a function that receives a vector of attributes $x \in R^n$ as input to identify a value $F(X) \in R$. Furthermore, one of the reasons for utilizing GB is based on the preceding reputation of ensemble methods outperforming other machine learning techniques in various situations [54–56]. They are generally recognised as the regressors or classifiers that produce the best out-of-the-box results.

3.6. Linear Regression

Linear Regression method presents the association among variables by fitting a linear equation to the data [57]. Linear regression proffers some advantages such as ease of use, interpretability and so on [58]. The linear fitting is conducted by making all but one predictor variable constant. The relationship between a predictor variable and a response variable does not suggest that the predictor variable causes the response variable, but that there is a significant relationship between the two variables.

$$Y = \beta_0 + \beta_1 X \tag{2a}$$

3.7. Deep learning

In recent years, the deep learning methods are providing powerful techniques to achieve enhanced modelling and better prediction performance [19]. The deep learning method uses deep architectures or multilayer architectures. It is a basic structure of deep neural networks, and the main distinction between Deep Neural Networks (DNN) and shallow neural networks is the number of layers. Generally, shallow neural networks have only two to three layers of neural networks which limits its ability to express intricate functions [59]. Conversely, deep learning has five or more layers of neural networks and presents more efficient algorithms that can further increase the accuracy.

Deep Neural Networks method is considered a superior ML technique due to the addition of multiple hidden layers to the regular ML neural network and it has gained increased attention in various fields such as image recognition [60] and natural language processing [61] among others. However, Deep Neural Networks (DNN) have not received much attention in the field of building energy consumption prediction [62].

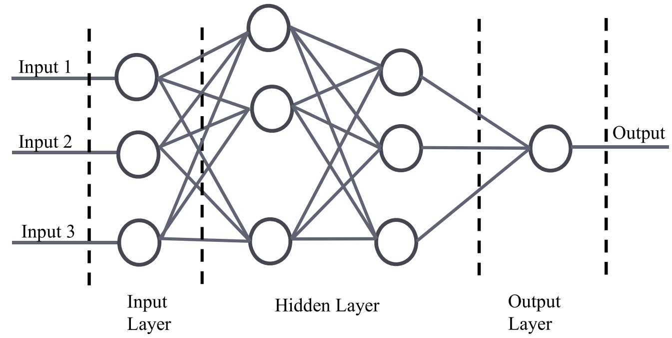

Fig. 4. Deep neural network architecture.

Fig. 4 visualizes the deep neural network architecture.

3.8. Stacking

Stacking method utilizes different machine learning techniques to build its model; then a combiner algorithm is trained to implement the final predictions based on the prediction outcomes of the base algorithms [63]. The combiner algorithm can be referred to as the meta-model and this can be any ensemble technique. The major benefit of the stacking method is the ability to harness the capabilities of a number of good performing algorithms on a classification or regression task and generate predictions that are more accurate than one single model [64]. Fig. 5 shows a general outline of the stacking approach.

3.9. Random forest

Random forest is an ensemble technique that offers several beneficial features. Among these features are [39]; (i) It is built on ensemble learning theory, which enables it to learn both simple and convoluted problems. (ii) It does not demand much hyper-parameter tuning to achieve good performance in comparison to other ML algorithms (e.g, artificial neural network, support vector machine, etc.). (iii) It's default parameters often produce excellent performance. Hence, the Random Forest (RF) method is gaining more attention in the field of building energy consumption [38,39,65,66]. For example [1], conducted the application of Random Forest (RF), model tree and Random Tree (RT) for forecasting hourly energy load. It was concluded that Random Forest (RF) produced the most suitable result.

4. Research methodology

This research explores the utilization of prediction methods for annual energy consumption using a large dataset. The data used in this research was collected in United Kingdom (UK). This estimation approach will adopt nine machine learning algorithms, some of which have been applied in energy use prediction namely Artificial Neural Network (ANN) [39,42,67], Gradient Boosting (GB) [38,41], K Nearest Neighbour (KNN) [41], Deep Neural Network (DNN) [59,62], Random Forest (RF) [39,65], Decision tree (DT) [34,40,41] Stacking, Support Vector Machine (SVM) [32,41,68], and Linear Regression (LR) [40]. These algorithms were selected based on their popularity and generation

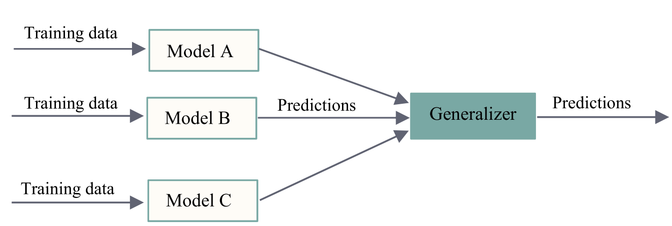

Fig. 5. An example of the stacking approach.

of good performance in the field of energy consumption prediction [32,59,62]. However, this study compares these algorithms on the same dataset as all the selected algorithms have not been directly compared on a single dataset. The development of each model was executed using python programming language on sublime Integrated Development Environment (IDE). Furthermore, this experiment was implemented using python programming language and performed using the following hardware specification (Apple MacBook Air with macOS Big Sur 11.4 and an Apple M1 chip with 16 GB RAM and 8 cores). Firstly, the raw data must undertake certain processes before training and testing of the model such as data cleaning or pre-processing and feature selection to evade possible complexities during the training stage. Furthermore, the energy use prediction framework will consist of four sections as follows: a) Data collection b) Data pre-processing c) Feature Selection d) Model development (training) e) Model Evaluation (Testing) as shown in Fig. 6.

4.1. Data collection

There are two types of data used for model development namely building metadata and meteorological data. The building related dataset was collected from the Ministry of Housing Communities and Local Government (MHCLG) repository. This data contains the metadata and energy consumption data of 5000 different types of residential buildings in the UK as shown in Fig. 7 below.

Building Metadata: The building metadata for 5000 residential buildings located within ten area postcodes in the UK were collected. The data consist of 500 residential buildings from each of the ten different postcode areas (Blackburn, Blackpool, Darlington, Yorkshire, Halton, Hartlepool, Hull, Middlesbrough, Redcar Cleveland and Warrington) for a clear and less ambiguous comparison. The building metadata comprised of only parameters that can be detected and modified during the design stage. This included floor level, roof description, walls description and Number of Habitable Rooms among others as shown in Table 2 below. These were considered as the independent variables while the energy data was considered as the dependent variable. The dataset includes the annual energy consumption of each building for the year 2020. The input variables such as wall and windows type or description are considered important variables as they have effect on the energy consumption of buildings [69–71]. For instance, the appropriate selection of wall or window type can considerably reduce energy consumption [69,71].

Meteorological Data: The meteorological dataset was collected from Meteostat repository, and it contains weather features such as temperature, wind speed and pressure as shown in Table 2 below. According to Ref. [72]; one of the major variables for building energy prediction is meteorological data [72]. This data was collected for ten area postcodes of the residential buildings and the granularity of meteorological data collected was daily average from January 1, 2020 till December 31, 2020. Taking into consideration the applicability of this model beyond UK, the meteorological data was averaged monthly rather than annually in correspondence to the energy consumption data (such that the model prevails in both locations with high and moderate weather conditions).

The building and meteorological variables are enumerated in Table 2 below. The building data is categorized as internal while the meteorological data is categorized as external. The independent variable selected based on related works are included in ID number 1 to 5 in Table 2.

4.2. Data pre-processing

In machine learning, the initial process includes data pre-processing for the preparation of data, although, it is often time consuming and computationally expensive [73]. This process is important to detect the existence of invalid or inconsistent data that can cause error during analysis [44].

Data merging: The utilization of multiple datasets requires data merging. Therefore, the building dataset and meteorological dataset were merged using the common variable (postcode) to match each building data to its respective meteorological data. The datasets were merged using the panda package of the python programming language. The data merged resulted in a total of 285,000 data points.

Data cleaning: The process of data cleaning applied involves the removal of outliers and treatment of missing data. The meteorological data contained few missing values which was resolved by applying the mean value imputation as proposed in 2015 by Newgard and Lewis [74]. This is the replacement of missing values with the mean value of each column. However, the instances in the building dataset with missing values (540 instances) were deleted from the database to avoid ambiguity and complexities during model development (training and testing phase).

Data conversion: The building raw data comprised of some categorical data in variables such as, wall energy efficiency, windows energy efficiency among others (see Table 2). The variables were allocated

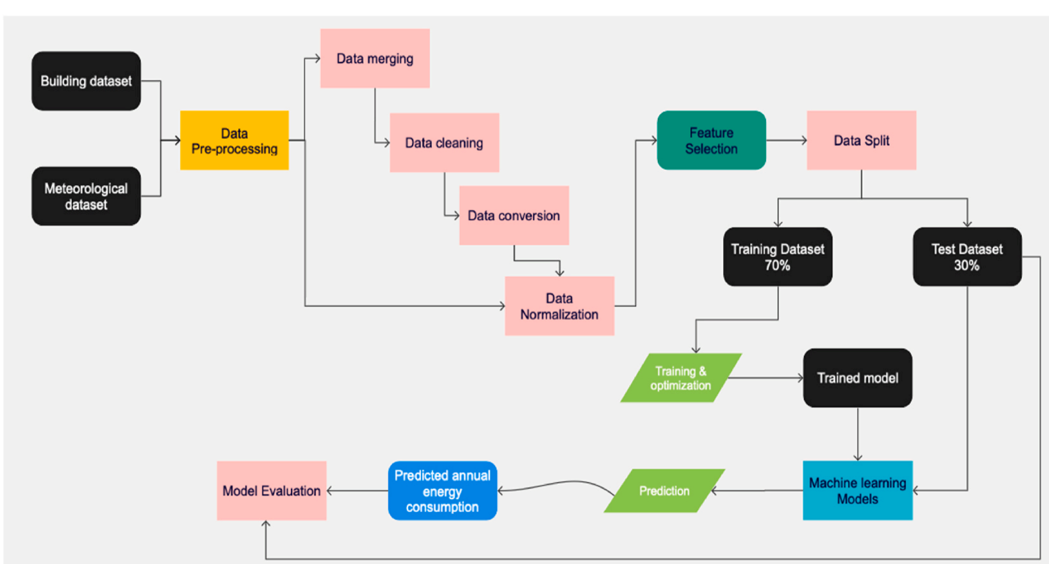

Fig. 6. Flowchart diagram of the prediction framework.

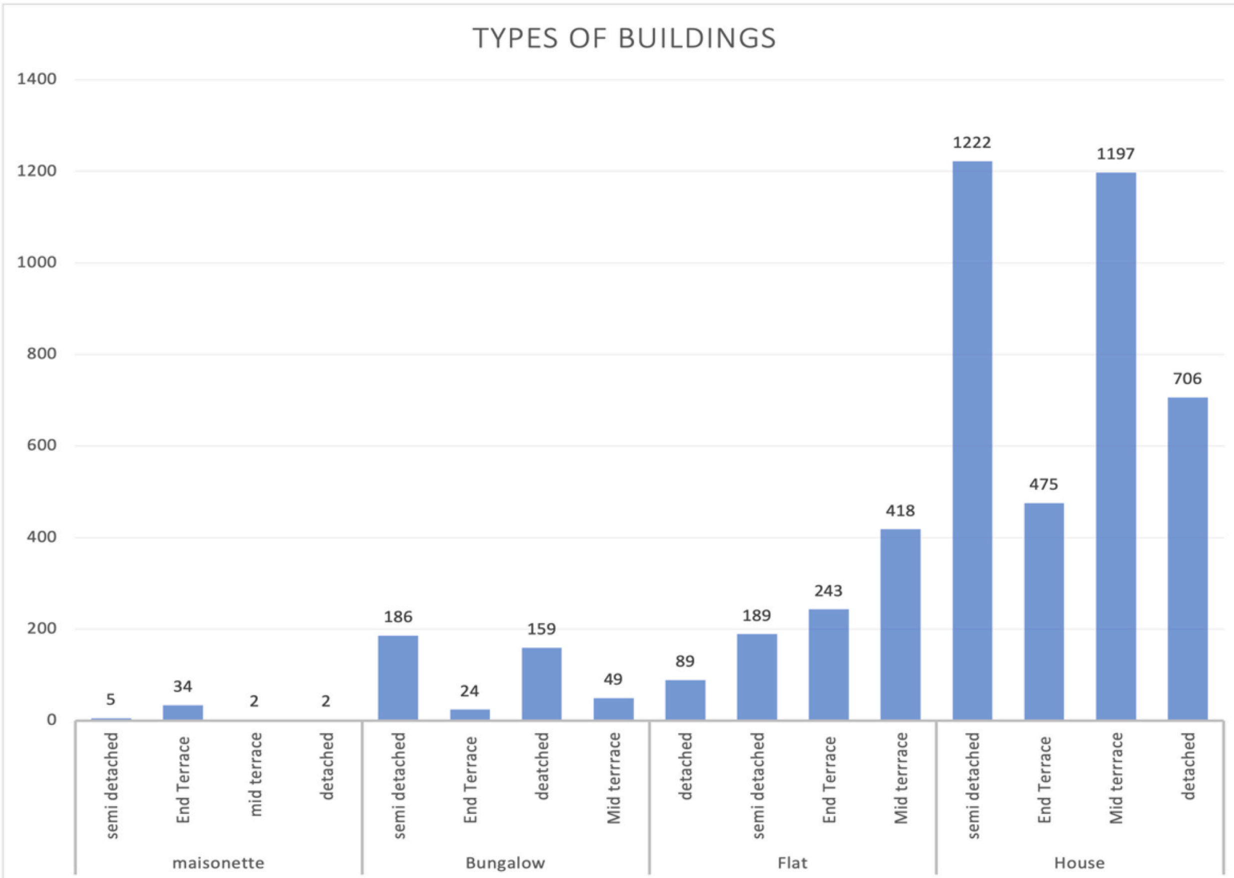

Fig. 7. Graphical representation of the types of buildings utilized.

Table 2
List of Features selected.

ID	Variable	Abbreviation	Internal/External	Type	Label
1	Temperature [^{o}c] (Monthly Average)	Tavg	External	Continuous	Independent (Input variables)
2	Wind speed [km/h] (Monthly Average)	Wspd			
3	Pressure [Hg] (Monthly Average)	Pres			
4	Total Floor Area [m^2]	Total Floor Area	Internal		
5	Property Type	Property Type		Categorical	
6	Glazed Area	Glazed Area			
7	Extension Count	Extension Count		Discrete	
8	Walls Description	Walls Description		Categorical	
9	Floor Description	Floor Description			
10	Floor Level	Floor Level		Discrete	
11	Windows Description	Windows Description		Categorical	
12	Windows Energy Efficiency	Windows Energy Eff			
13	Windows Environmental Efficiency	Windows Env Eff			
14	Walls Energy Efficiency	Walls Energy Eff			
15	Walls Environmental Efficiency	Walls Env Eff			
16	Roof Description	Roof Description			
17	Roof Energy Efficiency	Roof Energy Eff			
18	Roof Environmental Efficiency	Roof Env Eff			
19	Lighting Environmental Efficiency	Lighting Env Eff			
20	Lighting Energy Efficiency	Lighting Energy Eff			
21	Number of Heated Rooms	Number of Heated Rooms		Discrete	
22	Number of Habitable Rooms	Number of Habitable Rooms			
23	Energy Consumption [kWh/m^2]	Energy Consumption		Continuous	Dependent (Output variable)

values [e.g., very good = highest value (5) while very poor = lowest value (0)] to provide a suitable data for the ML algorithm.

Data Normalization: Data normalization is a very common procedure of data pre-processing that eliminates the influence of dimensions as several features often have unrelated dimensions [75,76]. Normalization scales individual samples into a unit norm in order to avoid problems during model development. For instance, if an input variable column contains values ranging from 0 to 5, and another column holds values ranging from 1000 to 10,000. The difference in the scale of the numbers could create problems during model development. Due to the different types of data (e.g., continuous, discrete, and categorical) present in the dataset, it is essential to normalize the data to eliminate the influence of the dimension and avoid difficulties during the model development phase [75]. The building dataset and the meteorological dataset were normalized using sklearn python package normalizer. Sklearn python package is a machine learning library for the python programming language that contains various algorithm's functions including pre-processing techniques such as normalization and standardization among others [77]. Normalization utilizes the formula below to scales down the dataset such that the normalized values fall within the range of 0 and 1.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

where: X is a data point

X_{min} minimum value
X_{max} maximum value
X_{norm} normalized value

4.3. Feature selection

Feature Selection (FS) is related to the accuracy and complexity of the predictive model. FS is an important process in the implementation of ML model development because not all features are impactful. This method is often utilized to identify irrelevant and unimportant features [78]. It is also essential to apply feature selection for optimum model performance [79]. Random Forest (RF) and extra tree classifier were employed for selecting the most suitable input features. Random forest recursively examines the effect of adding or removing a feature on the model and returns the relative dependence value for each variable as displayed in Fig. 8a [32]. Extra tree classifier evaluates random splits over a fraction of features and likewise return relative dependence value and displayed in Fig. 8b [80]. Based on Fig. 8a and b, ten variables of the highest ranking are selected for the development of the model.

The 10 highest ranking variables in both random forest and extra tree are equal. The selected list of variables utilized are as follows: total floor area, extension count, number habitable rooms, number heated rooms, floor description, walls description, roof description, temperature, wind speed and pressure.

4.4. Model development

This research utilized supervised machine learning approach based on regression to forecast the annual energy consumption. After data preparation, the selected variables based on feature importance ranking are fed into the learning algorithms. Thereafter, the data was randomly split into two categories at a ratio of 7:3 namely training group and testing group. The training group containing 70% of total data was utilized to develop a predictive model for each algorithm selected in this research namely Artificial Neural Network (ANN), Deep Neural Network (DNN), Random Forest (RF), Stacking, Gradient Boosting (GB), K Nearest Neighbour (KNN), Support Vector Machine (SVM), Decision tree (DT) and Linear Regression. Furthermore, the testing group containing 30% of total data which was used to test the model.

The Deep Neural Network (DNN) modelled is a feed forward neural network with three hidden layers, each containing 64 neurons. This totals five layer including the input and output layer in the development of this model. Also, the 'RELU' activation and 'Adam' optimiser was utilized. The stacking model combined the base estimator (RidgeCV) with SVM and RF at a random state of 42. Additionally, SVM was developed solely, and the parameters utilized in its development are as follows, 1.0 value of 'C', 0.1 value of Epsilon and radial basis function kernel. Random forest (RF) was developed with 10 estimators while K nearest neighbour (KNN) was developed with five neighbors, uniform weights and leaf size of 30. Furthermore, the gradient boosting model was developed using the least square regression loss function, 0.1 learning rate and 100 estimators.

4.5. Model Evaluation

The performance of each model is evaluated using the following performance measures: R-Squared (R2), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Squared Error (MSE). Among

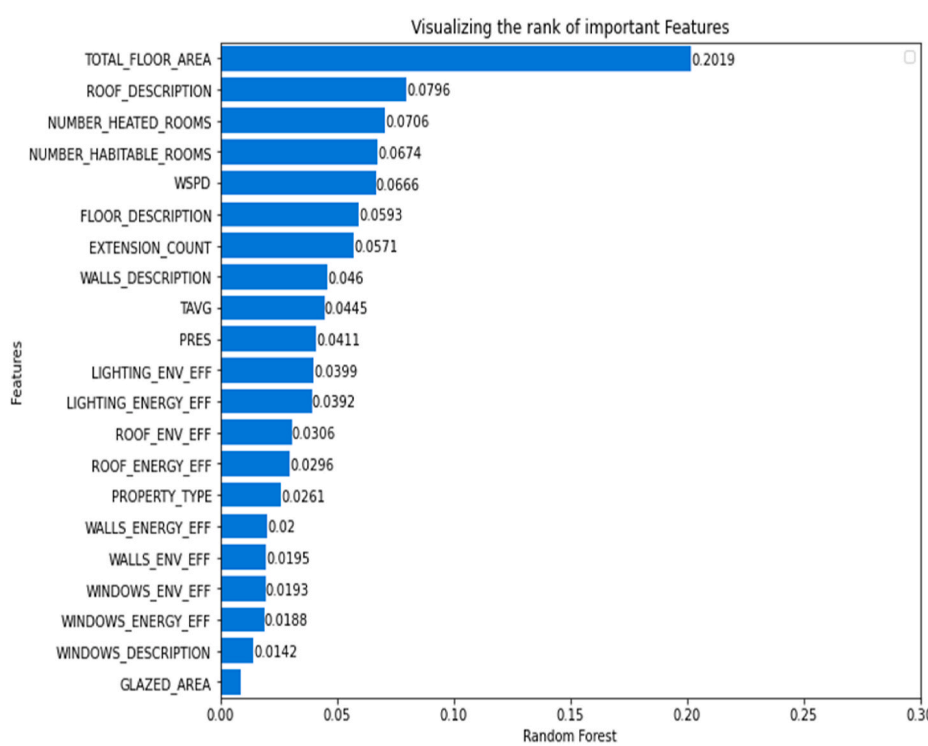

Fig. 8a. Feature importance using Random Forest.

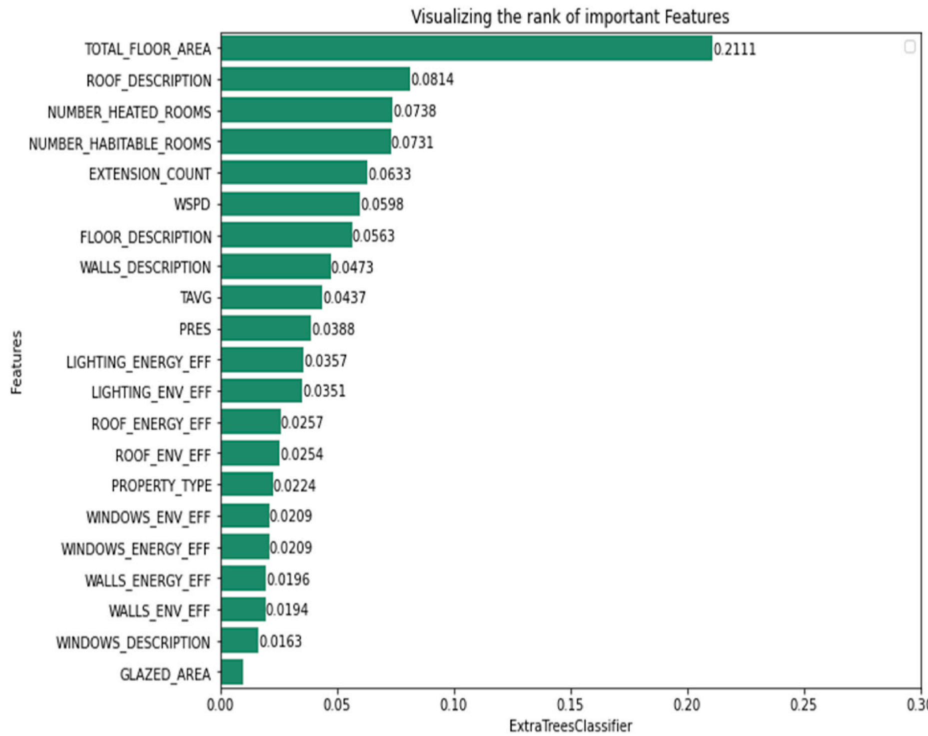

Fig. 8b. Feature Importance using ExtraTree

all the listed evaluation methods, the most often utilized for energy consumption prediction are the MSE and RMSE [1,21,23].

1. Mean Absolute Error (MAE) is a method of calculating the difference between the predicted values and the actual values at each point in a scatter plot. The closer the score is to zero, the better the performance while the higher the score, the worse the performance. It is computed as the average of the absolute errors between predicted and actual effort.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|AE_i - PE_i| \tag{2b}$$

2. Mean Squared Error (MSE) is the measure of squared variation between the estimated values and the actual values. MSE is an assessment of the quality of a predictor. Models with error values closer to zero are consider the better estimation model. It is also known as Mean Squared Deviation (MSD).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(AE_i - PE_i)^2 \tag{3}$$

3. Root Mean Squared Error (RMSE) is also a metric used to calculate the differences between estimated value and the actual perceived value of the model. It is achieved through the square root of the Mean Square Error (MSE).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(AE_i - PE_i)^2} \tag{4}$$

4. R-Squared (R^2) is a statistical measure that determines the proportion of the difference in the target variable that can be justified by the independent variables. It displays the extent to which the data fits the model. R^2 can produce a negative result, however, the best outcome of R^2 is 1.0. It is also known as the Coefficient of determination

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{pred,i} - y_{data,i})^2}{\sum_{i=1}^{n}(y_{data,i} - y_{data})^2} \tag{5}$$

5. Result and discussion

In this study, the analysis of the result has shown major findings between the selected models. The basis of determining the best predictive model stated in section 3.5 stipulates that model holding values closer to zero for MAE, MSE and RMSE are the good predictive model while values closer to one for R^2 produced the best results. In this research, DNN emerged the most efficient model for predicting annual energy consumption. ANN based model, which is recognised for producing good results in large dataset has prevailed and GB based model, which has not received much attention in the field of energy prediction emerged third best predictive model. DNN, ANN, GB, RF and SVM based models are directly comparable likewise KNN, DT, stacking and LR. DNN slightly outperforms ANN and GB with better MAE, MSE, RMSE and R^2. Although, it takes a longer to train the model. The Stacking and LR based model have the lowest performance but consumes a lesser time to train.

Table 3
Performance result for each model.

Model	Training Time	R-squared	MAE	RMSE	MSE
Deep Neural Network	5.2s	**0.95**	**0.92**	**1.16**	**1.34**
Artificial Neural Network	3.7s	0.94	0.95	1.20	1.45
Gradient Boosting	1.8s	0.92	1.10	1.40	1.95
Support Vector Machines	2.0s	0.90	1.22	1.61	2.61
Random Forest	2.5s	0.89	1.32	1.69	2.85
K Nearest Neighbors	1.4s	0.77	1.90	2.40	5.78
Decision Tree	**1.2s**	0.74	1.99	2.55	6.48
Linear Regression	1.4s	0.73	2.02	2.59	6.72
Stacking	1.3s	0.73	2.04	2.60	6.76

Bold represents the best performance.

The most efficient predictive model amongst the compared model in terms of performance measures is indicated in bold in Table 3 below.

The Box plot method was adopted to represent the model's performance using the evaluation measures as shown in Fig. 9a–d. The most efficient predictive model appears to be DNN across all the visualized performance measures implemented. More elaborately, Fig. 9b shows a clear comparison as the R^2 box plot displays DNN interquartile range is the closest to 1.0 while the subsequent box plots shows DNN is the closest to 0.0. There is no significant difference between GB and ANN based model in the R^2 box plot. However, Table 3 clarifies this ambiguity of insignificant variation between ANN and GB with an R^2 of 0.94 and 0.92 respectively. Furthermore, the worst predictive model is stacking as it is least closest to 0.0 for R^2 judging by the outliers of the interquartile range and it is the farthest from 0.0 for MAE and RMSE box plot. Stacking, LR, DT and KNN are considered the most inaccurate predictive models based on the significant variation to the good predictive models DNN, ANN, SVM and GB across all performance measures. Therefore, Fig. 9a–d and Table 3 displays the best predictive model for annual energy consumption as DNN and following this is the ANN, GB and SVM model which also produced good results.

To investigate the effect of the data size on the performance results, the nine models were developed and evaluated using 1000 instances and compared with the performance result as shown in Table 4 below. Afer data processing, the instances were reduced to 908 instances due to the deletion of instances with missing values (92 instances). Nevertheless, the generated result displays Deep Neural Network (DNN) as the most efficient model in annual energy consumption for both instances and decision tree (DT) is the most suitable in term of computational efficiency. However, there are variation in both instances as the models developed using larger data produced better performance result than the model developed with smaller data. Hence, training the model with over 50,000 instances could produce even better accuracy. This further substantiates the statement that the larger the data, the more accurate the result [25–29]. Additionally, in most cases, models developed with larger data consumed a longer time for training such as DNN, ANN and GB among others as shown in Table 4 below, with the most efficient model in terms of performance measures indicated in bold.

Furthermore, a sensitivity analysis was conducted to determine how sensitive the result is to specific types of buildings. The data for houses appears to be the most dominant as shown in Fig. 9. This house dataset contains a total of 3601 instances, which was used to test the effect of houses on the feature selected and the performance results.

After feature importance completion, it was noted that ten highest ranking variables for the house dataset are equal to the variables selected using all types of building dataset as shown in Fig. 8a and b.

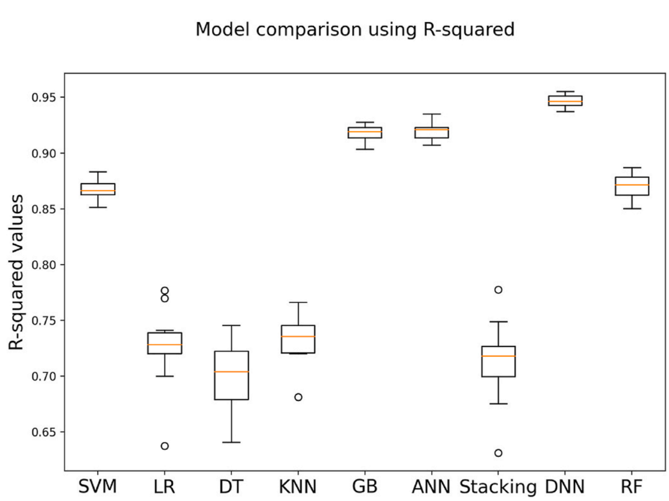

Fig. 9a. R-Squared comparison.

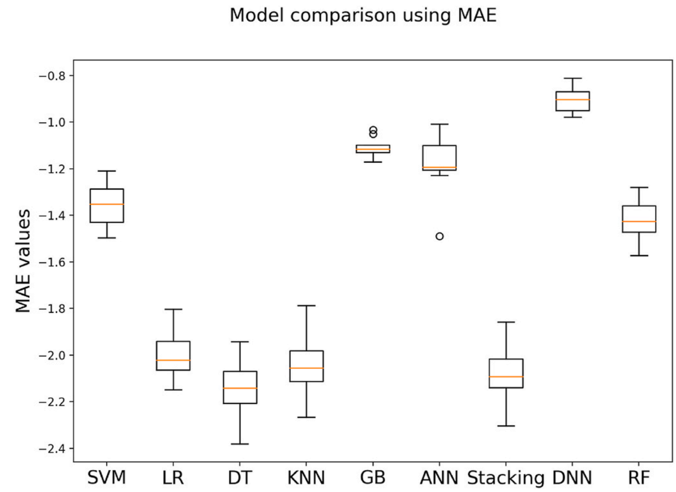

Fig. 9b. MAE comparison.

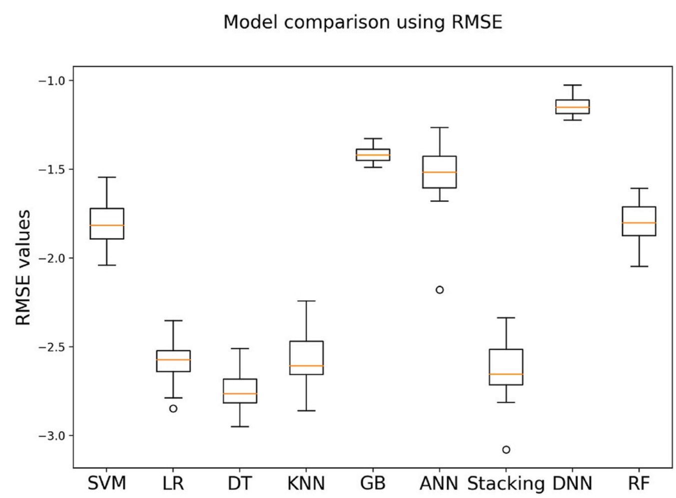

Fig. 9c. RMSE comparison.

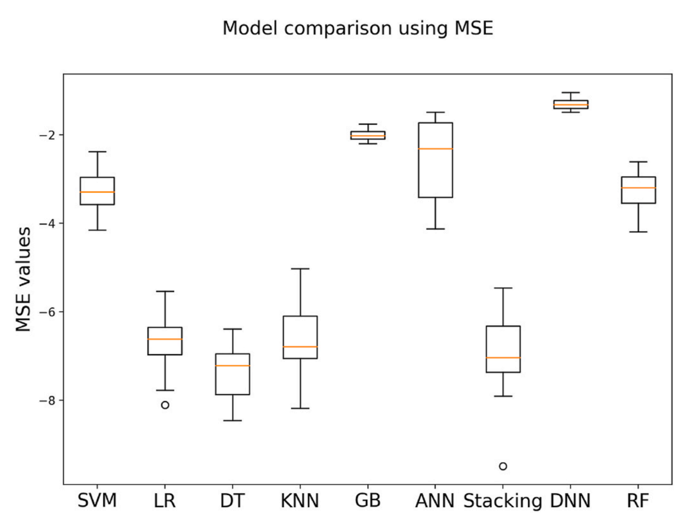

Fig. 9d. MSE comparison.

This indicates that the type of building does not influence the feature selected. Table 5 shows the model performance for houses dataset which does not show a significant difference in comparison with the result

Table 4
Result comparison of models developed using two data sizes.

Model	1000 Instances			5000 Instances		
	Training Time	R-squared	RMSE	Training Time	R-squared	RMSE
Deep Neural Network	3.4s	**0.93**	**1.34**	5.2s	**0.95**	**1.16**
Artificial Neural Network	2.3s	0.80	2.27	3.7s	0.94	1.20
Gradient Boosting	1.6s	0.91	1.51	1.8s	0.92	1.40
Support Vector Machines	1.8s	0.80	2.30	2.0s	0.90	1.61
Random Forest	1.9s	0.83	2.09	2.5s	0.89	1.69
K Nearest Neighbors	1.5s	0.68	2.92	1.4s	0.77	2.40
Decision Tree	**1.4s**	0.61	3.18	**1.2s**	0.74	2.55
Linear Regression	1.7s	0.73	2.63	1.4s	0.73	2.59
Stacking	4.7s	0.65	3.00	1.3s	0.73	2.60

Bold represents the best performance.

Table 5
Performance result for each model based on only houses data.

Model	Training Time	R-squared	MAE	RMSE	MSE
Deep Neural Network	4.2s	**0.95**	**0.95**	**1.19**	**1.41**
Artificial Neural Network	3.5s	0.86	1.56	1.96	3.85
Gradient Boosting	2.2s	0.92	1.15	1.45	2.11
Support Vector Machines	1.8s	0.89	1.33	1.75	3.07
Random Forest	2.6s	0.88	1.43	1.80	3.24
K Nearest Neighbors	1.7s	0.75	2.02	2.57	6.58
Decision Tree	**1.5s**	0.73	2.15	2.71	7.35
Linear Regression	1.7s	0.75	2.03	2.61	6.82
Stacking	1.8s	0.73	2.13	2.71	7.33

Bold represents the best performance.

generated using the dataset containing all different types of building. DNN still outperforms all other model in term of the performance measures while DT outperforms DNN in terms of computational efficiency.

The good prediction result of DNN, ANN, GB, SVM presents a motivation for utilization of this model in the early design phase. However, in comparison with related work, Li et al., and Dong et al., applied SVM for predicting hourly load consumption on less than 5 instances with a result of 1.17 (RMSE) and 0.99 (R^2) respectively [19,21]. This result outperforms the performance of SVM in the study, though this can the subject to the amount of instances used, based on the theoretical rationale proposed by a number of researchers that SVM is recognised for its generation of good result in small dataset [3,23,81]. Furthermore, Dong et al., applied SVM on a larger dataset of 507 instances with a result of 7.35 (RMSE), which performed significantly lower than the performance of SVM in this study using both 1000 and 5000 instances. The ANN produces good results and outperform other studies as shown in Table 1 above using less than 1000 instances. ANN still produces good results in both large and small dataset. Gradient boosting has not received much attention in this field but performs the third-best model among other in terms of performance measures. GB presents promising potential unprecedented results both in performance and computational efficiency.

6. Conclusion

This paper compared the prediction performance of nine machine learning techniques namely DNN, ANN, SVR, DT, GB, LR, Stacking, RF and KNN for predicting annual energy consumption. The good performance result generated in the study have proffered a model that enables building designer predict energy consumption at the early design phase. In general, DNN produced better results than other model for energy use prediction. However, there are other proven efficient predictive models in this study such as ANN, GB and SVM. In terms of computational efficiency, DT produced the best result with a training time of 1.2 s.

Subsequently, a sensitivity analysis was conducted to satisfy the third objective. Using the most dominant building cluster in the dataset, feature selection was implemented, and nine ML models were developed and evaluated. The result shows that the performance of the model is not sensitive to a specific type of building. Hence, building cluster has no significant effect on the feature selected and the performance result. Furthermore, nine ML models were developed using two different sizes of data to satisfy the fourth objective. The result shows that the size of data has significant effect on the performance of the model.

The high performance of these models presents a motivation for building designers to use it at the early design phase to make informed decision, manage and optimize design. Future research should focus on the applying DNN and other deep learning models on an even larger data of greater than 40,000 in anticipation of better performance. Also explore the suitability of other ensemble algorithms for predicting building energy consumption.

Authorship contributions

Conception and design of study: R.A. Olu-Ajayi, H. Alaka. Acquisition of data: R.A. Olu-Ajayi, H. Alaka, I. Sulaimon. Analysis and/or interpretation of data: R.A. Olu-Ajayi, I. Sulaimon. Drafting the manuscript: R.A. Olu-Ajayi, H. Alaka, F. Sunmola. Revising the manuscript critically for important intellectual content: H. Alaka, F. Sunmola, S. Ajayi, I. Sulaimon. Approval of the version of the manuscript to be published (the names of all authors must be listed): R.A. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, S. Ajayi

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A.-D. Pham, et al., Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability, J. Clean. Prod. 260 (2020) 121082, https://doi.org/10.1016/j.jclepro.2020.121082.
[2] B. Dandotiya, Climate-Change-and-Its-Impact-on-Terrestrial-Ecosystems, 2020, https://doi.org/10.4018/978-1-7998-3343-7.ch007.
[3] P. Aversa, et al., Improved thermal transmittance measurement with HFM technique on building envelopes in the mediterranean area, Sel. Sci. Pap. J. Civ. Eng. 11 (2016), https://doi.org/10.1515/sspjce-2016-0017.
[4] Soheil Fathi, et al., Machine learning applications in urban building energy performance forecasting: a systematic review, Renew. Sustain. Energy Rev. 133 (2020) 110287, https://doi.org/10.1016/j.rser.2020.110287.
[5] G. Serale, M. Fiorentini, M. Noussan, 11 - development of algorithms for building energy efficiency, in: F. Pacheco-Torgal, et al. (Eds.), Start-Up Creation, second ed., Woodhead Publishing (Woodhead Publishing Series in Civil and Structural Engineering), 2020, pp. 267–290, https://doi.org/10.1016/B978-0-12-819946-6.00011-4.
[6] L. Li, et al., Impact of natural and social environmental factors on building energy consumption: based on bibliometrics, J. Build. Eng. 37 (2021) 102136, https://doi.org/10.1016/j.jobe.2020.102136.
[7] M. Hamed, S. Nada, Statistical analysis for economics of the energy development in North Zone of Cairo, Int. J. Finance Econ. 5 (2019) 140–160.

[8] American Society of Heating, Refrigeration and Air-Conditioning Engineers, ASHRAE handbook: Fundamentals, in: American Society of Heating, Refrigerating and Air-Conditioning Engineers, ASHRAE, Atlanta, GA, USA, 2009.
[9] J. Runge, R. Zmeureanu, Forecasting energy use in buildings using artificial neural networks: a review, Energies 12 (17) (2019) 3254, https://doi.org/10.3390/en12173254.
[10] C. Robinson, et al., Machine learning approaches for estimating commercial building energy consumption, Appl. Energy 208 (2017) 889–904, https://doi.org/10.1016/j.apenergy.2017.09.060.
[11] K. Li, et al., A hybrid teaching-learning artificial neural network for building electrical energy consumption prediction, Energy Build. 174 (2018) 323–334, https://doi.org/10.1016/j.enbuild.2018.06.017.
[12] Q. Qiao, A. Yunusa-Kaltungo, R. Edwards, Hybrid method for building energy consumption prediction based on limited data, in: 2020 IEEE PES/IAS PowerAfrica. 2020 IEEE PES/IAS PowerAfrica, 2020, pp. 1–5, https://doi.org/10.1109/PowerAfrica49420.2020.9219915.
[13] G. Tardioli, et al., Data driven approaches for prediction of building energy consumption at urban level, Energy Procedia 78 (2015) 3378–3383, https://doi.org/10.1016/j.egypro.2015.11.754.
[14] B.B. Ekici, U.T. Aksoy, Prediction of building energy consumption by using artificial neural networks, Adv. Eng. Software 40 (5) (2009) 356–362, https://doi.org/10.1016/j.advengsoft.2008.05.003.
[15] R. Olu-Ajayi, H. Alaka, Building energy consumption prediction using deep learning, in: Environmental Design and Management Conference (EDMIC), 2021 [Preprint].
[16] Y. Vorobeychik, J.R. Wallrabenstein, Using Machine Learning for Operational Decisions in Adversarial Environments, 2013, p. 9.
[17] V. Ríos Canales, Using a supervised learning model: two-class boosted decision tree algorithm for income prediction, Available at: http://prcrepository.org:8080/xmlui/handle/20.500.12475/557, 2016. (Accessed 31 May 2021).
[18] A. Colmenar-Santos, et al., Solutions to reduce energy consumption in the management of large buildings, Energy Build. 56 (2013) 66–77, https://doi.org/10.1016/j.enbuild.2012.10.004.
[19] C. Li, et al., Building energy consumption prediction: an extreme deep learning approach, Energies 10 (10) (2017) 1525, https://doi.org/10.3390/en10101525.
[20] M.A. McNeil, N. Karali, V. Letschert, Forecasting Indonesia's electricity load through 2030 and peak demand reductions from appliance and lighting efficiency, Energy Sustain. Dev. 49 (2019) 65–77, https://doi.org/10.1016/j.esd.2019.01.001.
[21] B. Dong, C. Cao, S.E. Lee, Applying support vector machines to predict building energy consumption in tropical region, Energy Build. 37 (5) (2005) 545–553, https://doi.org/10.1016/j.enbuild.2004.09.009.
[22] Q. Li, et al., Applying support vector machine to predict hourly cooling load in the building, Appl. Energy 86 (10) (2009) 2249–2256, https://doi.org/10.1016/j.apenergy.2008.11.035.
[23] Q. Li, et al., Predicting hourly cooling load in the building: a comparison of support vector machine and different artificial neural networks, Energy Convers. Manag. 50 (1) (2009) 90–96, https://doi.org/10.1016/j.enconman.2008.08.033.
[24] A. Bagnasco, et al., Electrical consumption forecasting in hospital facilities: an application case, Energy Build. 103 (2015) 261–270, https://doi.org/10.1016/j.enbuild.2015.05.056.
[25] S. Lee, et al., Data Mining-Based Predictive Model to Determine Project Financial Success Using Project Definition Parameters, 2011.
[26] K. Kaur, O.P. Gupta, A machine learning approach to determine maturity stages of tomatoes, Orient. J. Comput. Sci. Technol. 10 (3) (2017) 683–690.
[27] K.R. Dalal, Review on application of machine learning algorithm for data science, in: 2018 3rd International Conference on Inventive Computation Technologies (ICICT), IEEE, 2018, pp. 270–273.
[28] K. Goyal, N. Tiwari, J. Sonekar, An anatomization of data classification based on machine learning techniques, IJRAR 7 (2) (2020) 713–716.
[29] M.A. Kabir, Vehicle speed prediction based on road status using machine learning, Adv. Res. Energy Eng. 2 (1) (2020).
[30] S. Bourhnane, et al., Machine learning for energy consumption prediction and scheduling in smart buildings, SN Appl. Sci. 2 (2) (2020) 297, https://doi.org/10.1007/s42452-020-2024-9.
[31] A.H. Neto, F.A.S. Fiorelli, Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption, Energy Build. 40 (12) (2008) 2169–2176, https://doi.org/10.1016/j.enbuild.2008.06.013.
[32] Z. Dong, et al., Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification, Energy Build. 241 (2021) 110929, https://doi.org/10.1016/j.enbuild.2021.110929.
[33] G.K.F. Tso, K.K.W. Yau, Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks, Energy 32 (9) (2007) 1761–1768, https://doi.org/10.1016/j.energy.2006.11.010.
[34] Z. Yu, et al., A decision tree method for building energy demand modeling, Energy Build. 42 (10) (2010) 1637–1646, https://doi.org/10.1016/j.enbuild.2010.04.006.
[35] J. Demsar, Statistical Comparisons of Classifiers over Multiple Data Sets, 2006, p. 30.
[36] K. Amasyali, N. El-Gohary, Deep Learning for Building Energy Consumption Prediction, 2017.
[37] Hai-xiang Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renew. Sustain. Energy Rev. 16 (6) (2012) 3586–3592, https://doi.org/10.1016/j.rser.2012.02.049.
[38] C. Fan, F. Xiao, Y. Zhao, A short-term building cooling load prediction method using deep learning algorithms, Appl. Energy 195 (2017) 222–233, https://doi.org/10.1016/j.apenergy.2017.03.064.
[39] M.W. Ahmad, M. Mourshed, Y. Rezgui, Trees vs Neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption, Energy Build. 147 (2017) 77–89, https://doi.org/10.1016/j.enbuild.2017.04.038.
[40] J.-S. Chou, D.-K. Bui, Modeling heating and cooling loads by artificial intelligence for energy-efficient building design, Energy Build. 82 (2014) 437–446, https://doi.org/10.1016/j.enbuild.2014.07.036.
[41] R. Wang, S. Lu, W. Feng, A novel improved model for building energy consumption prediction based on model integration, Appl. Energy 262 (2020) 114561, https://doi.org/10.1016/j.apenergy.2020.114561.
[42] A.S. Ahmad, et al., A review on applications of ANN and SVM for building electrical energy consumption forecasting, Renew. Sustain. Energy Rev. 33 (2014) 102–109, https://doi.org/10.1016/j.rser.2014.01.069.
[43] M. Bourdeau, et al., Modeling and forecasting building energy consumption: a review of data-driven techniques, Sustain. Cities Soc. 48 (2019) 101533, https://doi.org/10.1016/j.scs.2019.101533.
[44] K. Amasyali, N.M. El-Gohary, A review of data-driven building energy consumption prediction studies, Renew. Sustain. Energy Rev. 81 (2018) 1192–1205, https://doi.org/10.1016/j.rser.2017.04.095.
[45] J.O. Donoghue, M. Roantree, A framework for selecting deep learning hyper-parameters, in: S. Maneth (Ed.), Data Science, Springer International Publishing (Lecture Notes in Computer Science), Cham, 2015, pp. 120–132, https://doi.org/10.1007/978-3-319-20424-6_12.
[46] X. Wu, et al., Top 10 algorithms in data mining, Knowl. Inf. Syst. 14 (1) (2008) 1–37, https://doi.org/10.1007/s10115-007-0114-2.
[47] Hai-xiang Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renew. Sustain. Energy Rev. 16 (6) (2012) 3586–3592, https://doi.org/10.1016/j.rser.2012.02.049.
[48] Y. Wei, et al., A review of data-driven approaches for prediction and classification of building energy consumption, Renew. Sustain. Energy Rev. 82 (2018) 1027–1047, https://doi.org/10.1016/j.rser.2017.09.108.
[49] P. Domingos, A few useful things to know about machine learning, Commun. ACM 55 (10) (2012) 78–87, https://doi.org/10.1145/2347736.2347755.
[50] José Ortiz-Bejar, et al., k-nearest neighbor regressors optimized by using random search, in: 2018 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC). 2018 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), 2018, pp. 1–5, https://doi.org/10.1109/ROPEC.2018.8661399.
[51] N. Ali, D. Neagu, P. Trundle, Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets, SN Appl. Sci. 1 (12) (2019) 1559, https://doi.org/10.1007/s42452-019-1356-9.
[52] R. Olu-Ajayi, An investigation into the suitability of k-nearest neighbour (k-NN) for software effort estimation, Int. J. Adv. Comput. Sci. Appl. 8 (6) (2017), https://doi.org/10.14569/IJACSA.2017.080628.
[53] V. Flores, B. Keith, Gradient boosted trees predictive models for surface roughness in high-speed milling in the steel and aluminum metalworking industry, Complexity (2019) e1536716, https://doi.org/10.1155/2019/1536716, 2019.
[54] T.G. Dietterich, Ensemble methods in machine learning, in: Multiple Classifier Systems, Springer (Lecture Notes in Computer Science), Berlin, Heidelberg, 2000, pp. 1–15, https://doi.org/10.1007/3-540-45014-9_1.
[55] R.A. Berk, An introduction to ensemble methods for data analysis, Socio. Methods Res. 34 (3) (2006) 263–295, https://doi.org/10.1177/0049124105283119.
[56] C.-X. Zhang, J.-S. Zhang, A novel method for constructing ensemble classifiers, Stat. Comput. 19 (3) (2009) 317–327, https://doi.org/10.1007/s11222-008-9094-7.
[57] N. Fumo, M.A. Rafe Biswas, Regression analysis for prediction of residential energy consumption, Renew. Sustain. Energy Rev. 47 (2015) 332–343, https://doi.org/10.1016/j.rser.2015.03.035.
[58] R. Pino-Mejías, et al., Comparison of linear regression and artificial neural networks models to predict heating and cooling energy demand, energy consumption and CO2 emissions, Energy 118 (2017) 24–36, https://doi.org/10.1016/j.energy.2016.12.022.
[59] L. Lei, et al., A building energy consumption prediction model based on rough set theory and deep learning algorithms, Energy Build. 240 (2021) 110886, https://doi.org/10.1016/j.enbuild.2021.110886.
[60] F. Yu, et al., REIN the RobuTS: robust DNN-based image recognition in autonomous driving systems, IEEE Trans. Comput. Aided Des. Integrated Circ. Syst. 40 (6) (2021) 1258–1271, https://doi.org/10.1109/TCAD.2020.3033498.
[61] S.-R. Leyh-Bannurah, et al., Deep learning for natural language processing in urology: state-of-the-art automated extraction of detailed pathologic prostate cancer data from narratively written electronic health records, JCO Clin. Canc. Inform. (2) (2018) 1–9, https://doi.org/10.1200/CCI.18.00080.
[62] 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, https://doi.org/10.1016/j.rser.2021.110714.
[63] F. Divina, et al., Stacking ensemble learning for short-term electricity consumption forecasting, Energies 11 (4) (2018) 949, https://doi.org/10.3390/en11040949.
[64] X. Wang, X. Lu, A host-based anomaly detection framework using XGBoost and LSTM for IoT devices, Wireless Commun. Mobile Comput. (2020) e8838571, https://doi.org/10.1155/2020/8838571, 2020.
[65] Z. Wang, et al., Random Forest based hourly building energy prediction, Energy Build. 171 (2018) 11–25, https://doi.org/10.1016/j.enbuild.2018.04.008.
[66] Y.-T. Chen, E. Piedad, C.-C. Kuo, Energy consumption load forecasting using a level-based random forest classifier, Symmetry 11 (8) (2019) 956, https://doi.org/10.3390/sym11080956.

[67] X. Li, R. Yao, A machine-learning-based approach to predict residential annual space heating and cooling loads considering occupant behaviour, Energy 212 (2020) 118676, https://doi.org/10.1016/j.energy.2020.118676.
[68] D. Niu, Y. Wang, D.D. Wu, Power load forecasting using support vector machine and ant colony optimization, Expert Syst. Appl. 37 (3) (2010) 2531–2539, https://doi.org/10.1016/j.eswa.2009.08.019.
[69] M.M. Tahmasebi, S. Banihashemi, M.S. Hassanabadi, Assessment of the variation impacts of window on energy consumption and carbon footprint, Proc. Eng. 21 (2011) 820–828, https://doi.org/10.1016/j.proeng.2011.11.2083.
[70] C. Marino, A. Nucara, M. Pietrafesa, Does window-to-wall ratio have a significant effect on the energy consumption of buildings? A parametric analysis in Italian climate conditions, J. Build. Eng. 13 (2017) 169–183, https://doi.org/10.1016/j.jobe.2017.08.001.
[71] M. Marwan, The effect of wall material on energy cost reduction in building, Case Studies in Thermal Engineering 17 (2020) 100573, https://doi.org/10.1016/j.csite.2019.100573.
[72] Y. Ding, X. Liu, A comparative analysis of data-driven methods in building energy benchmarking, Energy Build. 209 (2020) 109711, https://doi.org/10.1016/j.enbuild.2019.109711.
[73] 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, https://doi.org/10.1016/j.dibe.2020.100037.
[74] C.D. Newgard, R.J. Lewis, Missing data: how to best account for what is not known, J. Am. Med. Assoc. 314 (9) (2015) 940–941, https://doi.org/10.1001/jama.2015.10516.
[75] Y. Liu, et al., Energy consumption prediction and diagnosis of public buildings based on support vector machine learning: a case study in China, J. Clean. Prod. 272 (2020) 122542, https://doi.org/10.1016/j.jclepro.2020.122542.
[76] R. Olu-Ajayi, et al., 'Ensemble learning for energy performance prediction of residential buildings, in: Environmental Design and Management Conference (EDMIC), 2021 [Preprint].
[77] J. Hao, T.K. Ho, Machine learning made easy: a review of scikit-learn package in Python programming language, J. Educ. Behav. Stat. 44 (3) (2019) 348–361, https://doi.org/10.3102/1076998619832248.
[78] HaiXiang Zhao, F. Magoulès, Feature selection for predicting building energy consumption based on statistical learning method, J. Algorithm Comput. Technol. 6 (1) (2012) 59–77, https://doi.org/10.1260/1748-3018.6.1.59.
[79] H.A. Alaka, et al., Systematic review of bankruptcy prediction models: towards a framework for tool selection, Expert Syst. Appl. 94 (2018) 164–184, https://doi.org/10.1016/j.eswa.2017.10.040.
[80] T. Joshi, R. Tiwari, Efficient System for Cancer Detection Using Feature Selection Algorithms, 2017.
[81] Qiong Li, Peng Ren, Qinglin Meng, Prediction model of annual energy consumption of residential buildings, in: 2010 International Conference on Advances in Energy Engineering. 2010 International Conference on Advances in Energy Engineering, 2010, pp. 223–226, https://doi.org/10.1109/ICAEE.2010.5557576.
[82] Bing Dong, et al., A hybrid model approach for forecasting future residential electricity consumption, Energy Build. (2016).