sustainability

MDPI

Article

Prediction of Cooling Load of Tropical Buildings with Machine Learning

Gebrail Bekdaş [1,*], Yaren Aydın [1], Ümit Isıkdağ [2], Aidin Nobahar Sadeghifam [3], Sanghun Kim [4] and Zong Woo Geem [5,*]

[1] Department of Civil Engineering, Istanbul University-Cerrahpaşa, Istanbul 34320, Turkey; yaren.aydin1@ogr.iuc.edu.tr
[2] Department of Informatics, Mimar Sinan Fine Arts University, Istanbul 34427, Turkey; umit.isikdag@msgsu.edu.tr
[3] Department of Civil Engineering, Curtin University Malaysia, Miri 98009, Malaysia; aidin.sadeghifam@curtin.edu.my
[4] Department of Civil and Environmental Engineering, Temple University, Philadelphia, PA 19122, USA; sanghun.kim@temple.edu
[5] Department of Smart City, Gachon University, Seongnam 13120, Republic of Korea
* Correspondence: bekdas@iuc.edu.tr (G.B.); geem@gachon.ac.kr (Z.W.G.)

Abstract: Cooling load refers to the amount of energy to be removed from a space (or consumed) to bring that space to an acceptable temperature or to maintain the temperature of a space at an acceptable range. The study aimed to develop a series of models and determine the most accurate ones in the prediction of the cooling load of low-rise tropical buildings based on their basic architectural and structural characteristics. In this context, a series of machine learning (regression) algorithms were tested during the research to determine the most accurate/efficient prediction model. In this regard, a data set consisting of ten features indicating the basic characteristics of the building (floor area, aspect ratio, ceiling height, window material, external wall material, roof material, window wall ratio north faced, window wall ratio south faced, horizontal shading, orientation) were used to predict the cooling load of a low-rise tropical building. The dataset was generated utilizing a set of generative and algorithmic design tools. Following the dataset generation, a series of regression models were tested to find the most accurate model to predict the cooling load. The results of the tests with different algorithms revealed that the relationship between the predictor variables and cooling load could be efficiently modeled through Histogram Gradient Boosting and Stacking models.

Keywords: cooling load; building; predictive modelling; energy efficiency

check for updates

Citation: Bekdaş, G.; Aydın, Y.; Isıkdağ, Ü.; Sadeghifam, A.N.; Kim, S.; Geem, Z.W. Prediction of Cooling Load of Tropical Buildings with Machine Learning. *Sustainability* **2023**, *15*, 9061. https://doi.org/10.3390/su15119061

Academic Editor: Graziano Salvalai

Received: 2 May 2023
Revised: 26 May 2023
Accepted: 2 June 2023
Published: 3 June 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The famous Brundtland Commission Report of the UN defines sustainability as meeting the needs of the present without compromising the ability of future generations to meet their own needs [1]. To make life on earth sustainable by consuming less and polluting less is one of the most important responsibilities of humans. The majority of the energy sources used today come from fossils. In fact, in recent years, more environmentally friendly and sustainable methods have been developed to produce green energy or to consume less. New energy sources are sustainable to a certain extent. However, the most important issue is to reduce energy consumption and to ensure that the same work is carried out with less energy and, thus, energy efficiency is achieved. In collective living areas, most of the energy is used for the heating and cooling needs of indoor spaces.

Global warming refers to the impact of human activities on the climate, especially the burning of fossil fuels and large-scale deforestation, which causes the emission of greenhouse gases, such as carbon dioxide, into the atmosphere [2]. As a result of the increase in energy demand with the recent global warming, it has become important to

re-evaluate the old methods and the measures taken in terms of energy efficiency. In this context, it is seen that the implementation of concepts such as green buildings, green industry, and passive houses has increased. Governments have made serious regulations on these issues [3].

The envelope of the buildings has been constantly developed due to the factors such as temperature, humidity, winds and solar radiation of the outside air. These factors are evaluated separately in hot weather and cold weather conditions. Especially the radiation effect of the sun and air temperature are important factors in terms of indoor thermal comfort in both weather conditions. In hot climate regions, the hot period lasts longer than the cold period. For this reason, the hot periods of these regions are important in terms of cooling load of buildings. Climatic changes due to environmental pollution in recent years (greenhouse effect, global warming, ...) bring energy expenditures and cooling to the forefront [4]. The use of climatic data in house design is very important in ensuring energy efficiency in the building. The design of buildings according to different climatic characteristics is effective in the formation of suitable indoor conditions and energy conservation. For example, a courtyard with an external part of the building reduces the cooling load by providing abundant air flow [5].

The calculation of the cooling load in every building has become necessary due to the increase in energy costs and to prevent climate changes caused by fuel energy consumption in buildings. Especially for buildings located in hot and tropical climate regions, it becomes especially important to determine and evaluate cooling loads in the design stage. On the other hand, the gradual increase in outside temperatures due to global warming leads to an increase in cooling load, especially in buildings located in hot-humid climates. This high cooling demand leads to more greenhouse gas emissions and thus supports the global warming process [6]. This situation generates a vicious circle which worsens the situation by causing more and more greenhouse gas emissions every day. Therefore, the cooling load should be calculated carefully. Energy efficiency should estimate and limit the energy consumption for the cooling of buildings.

As indicated in Table 1, the production industry was the sector with the highest energy consumption globally. Buildings were the second in terms of global energy consumption. Only this indicator can point out that serious energy savings can be achieved by enabling energy efficiency in the design and use of buildings. To contribute to the energy savings in the building industry, our study mainly focuses on facilitating the design phase of buildings through the prediction of energy consumption parameters, such as cooling load, with a simple but efficient approach.

Table 1. Total energy consumption by sectors (Trillion Btu) [7].

Sector	2020	2021	2022
Residential	20,555	20,841	21,807
Commercial	16,751	17,451	18,155
Industrial	31,247	32,474	32,912

The amount of energy used in buildings at the urban scale varies regionally but accounts for approximately 40% of total consumption [8,9]. This issue clearly shows that buildings are responsible for a large part of energy consumption that cannot be ignored. Therefore, efficient building design with energy-saving features and auxiliary tools that support these practices can be used to improve the energy efficiency of buildings and may be helpful in alleviating its use.

The main purpose of this study is to determine the cooling load in buildings located in tropical climates. Because when the cooling load of the buildings is known, a design can be made according to the required cooling load, the energy to be consumed by the cooling devices can be reduced, and a sustainable design can be realized. Malaysia is a country located in Southeast Asia. Located between one and four degrees north latitude, Malaysia's climate is equatorial. Tropical forests cover 70% of the country. Because of the

effect of monsoon winds and continuous rains, especially between January and May, the country's humidity increases during this season. The daily temperature is between 21 °C and 32 °C in the lowlands, while it is lower in the higher regions [10]. Malaysia is known for its hot and humid climate. As a solution to these climatic conditions, air conditioners (AC)s and AC systems, in general, are widely used in all regions of the country. The wide use of AC systems results in high energy consumption in buildings. In countries such as Malaysia, buildings should be designed to minimize the energy consumption related to the cooling of indoor spaces. Establishing the functional link between the architectural properties of the building and its energy consumption is crucial to facilitate energy-efficient building designs. In this study, the following aspect of architectural design is utilized for the prediction framework we have implemented and tested.

There are various types of houses available in the market in Malaysia, but the terraced house type typically accounts for 41% of the total residential property stock in 2018, which is available in one and multi-story types [11,12]. It is followed by low-cost houses (including low-cost house, low-cost flat and flat) with a 30 percent portion, which is purposely constructed for the low-income group. Terraced houses are the most common residential type, classified under low- to medium-cost housing [13], which are preferred by developers due to the speedy construction methods [14]. As the design is based on the British terraced house, it did not take into account local climatic conditions and cultures when it was brought to Malaysia [15]. The scope of this study was double-story terrace houses. The double-story terrace houses are the most common type of low-rise buildings and the biggest fraction of both existing supplies and newly planned residential units in Malaysia.

The case study was carried out on an intermediate double-story terrace house to represent the simulation of conventional low-rise residentials in Malaysia. Figure 1 is the representative building used in this study. The building was located in Skudai, Johor Bahru (latitude 1°32′ N, longitude 103°40′ E), with a total floor area of 200 m^2. The building had a rectangular shape floor plan and aspect ratio of (width/length) 1/2 and a ceiling height of 3.8 m. This model is a typical terrace house with an indoor space layout of a living-cum-dining area (DL), kitchen (K), a guest room with one bathroom (WC) on the ground floor as well as one master bedroom (MB), two smaller bedrooms together with a hall area (corridor) along with two bathrooms (WC) on the first floor. The house was facing the South–North direction, with large windows in the front facade (South). The building structure is a reinforced concrete structure with a brick-infilled frame. The roof is pitched, covered with clay tiles and non-insulated walls and roofs. Cement sand renders covered the façade. The walls are built from brick, and the surface is plastered. A single-glazed window frame was made of aluminum, and Table 2 describes the base model materials and thermal properties.



Figure 1. Representative building used in the study [16].

Table 2. Thermal characteristics of the case study base model building's envelope [16].

Variable	Description	U-Value (W/m^2·K)
Roof	Clay roof tiles without insulation	1.72
Ceiling	Suspended joist and plaster ceiling	4.32
Floor	150 mm thick RC concrete floor with ceramic tiles	2.90
Internal Wall	115 mm red clay brick wall both side both side cement mortar plaster	2.75
External Wall	115 mm red clay brick wall plaster with cement mortar on both side	2.65
Window	6 mm thick single glazed with aluminum frame	5.77
Door	Solid timber flush door with frame	2.17

These aspects were the total floor area, aspect ratio, ceiling height, window material, external wall material, roof material, window wall ratio north faced, window wall ratio south faced, horizontal shading, and orientation of the building. The literature defines the Cooling Load as the building's energy consumption, or the amount of energy required to keep the environment at a constant temperature [17]. The number of variables affecting cooling load calculations is very high. This study focuses on the following subset of these variables that appear as the key ones based on findings in our previous research [16] (Figure 2).

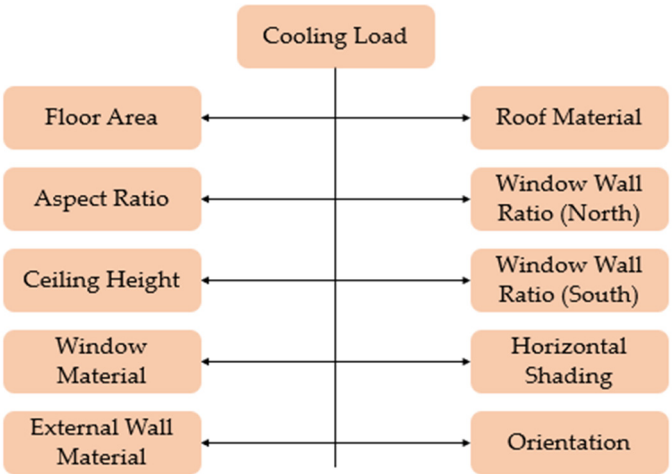

Figure 2. Predictor Variables for Cooling Load.

Floor area: The floor area is the floor area of the region where the cooling load will be calculated [18]. As it increases, the cooling load will increase too.

Aspect ratio: Aspect ratio is the ratio of the width to the height of the building. Thanks to the optimum aspect ratio, the building is shaded in hot weather, and the energy consumption required for cooling is reduced [19].

Ceiling height: As the height of the ceiling increases, the air volume in the room, which directly affects the dynamics that affect the cooling load, also increases. This also affects cooling efficiency [20].

Window material: Appropriate thermal comfort conditions can be achieved with glass selection according to the characteristic features of climate zones [21].

External wall material: One of the major factors that make up the cooling load is the total heat gained from the external walls [22].

Roof material: The roof material used is an important factor that has an impact on the cooling load. For example, when comparing a traditional roof with a green roof, it is known that the green roof application saves energy and has less negative impact on the environment [23].

Window wall ratio: The window-to-wall ratio is the ratio of the window area to the entire façade surface area. In regions with climatic conditions where heating energy demand is high, solar energy gain increases as the window/wall ratio increases [24].

Horizontal shading: With horizontal shading installation, the cooling load is reduced compared to the case without shading [25]. Elements such as balconies, overhangs, etc., are horizontal shading elements [26].

Orientation: Energy efficiency can be achieved with the right building orientation. Among the building orientation types, the cooling load increases in the perimeter zones orientated towards the west façade [27].

Heat loads have long been calculated manually and using the instantaneous calculation method, which assumes that heat gains are converted into instantaneous cooling loads. Although this method is simple and fast, it neglects processes such as heat storage and radiation transfer and, therefore, has little reliability [28]. There are many methods for cooling load calculations. Figure 3 shows the relationship between the American Society of Heating, Refrigerating, and Air-conditioning Engineers (ASHRAE) cooling load calculation methods in terms of complexity and accuracy.

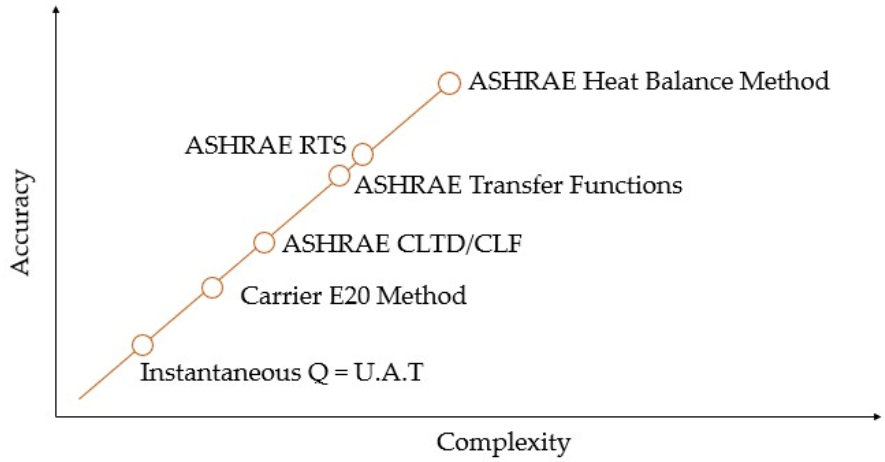

Figure 3. Relationship between complexity and accuracy for cooling load calculations [28].

From Figure 3, it can be seen that ASHRAEs Heat Balance Method has the highest complexity and accuracy. The Heat Balance Method, using the finite differences approach, calculates the inner surface temperatures of each surface in detail, as well as the solar gains, and makes the closest estimation of the heating and cooling load with the inclusion of natural ventilation, shading, HVAC equipment, and thermal mass [29]. The fact that accuracy increases as the complexity increases has led to the search for methods that are less time-consuming and complex. The calculation of the cooling load is more complex than the heating load due to the presence of dynamic responses and thermal mass [30].

In recent years, studies carried out utilizing machine learning methods have achieved very accurate results in the estimation of cooling loads. For example, Xuemei et al. [31] developed the Least Square Support Vector Machine (LS-SVM) for cooling load prediction, and when compared with Back Propagation Neural Network (BPNN), LS-SVM provided a higher accuracy with less error. Similarly, Li et al. [32] utilized a support vector machine (SVM) to predict the hourly building cooling load and achieved effective results. Gao et al. [33] used extreme learning machine (ELM) and random forest (RF) together to predict the cooling load of large commercial buildings. Sha et al. [34] compared the performances of different ML algorithms in predicting cooling load and showed that gradient tree boosting

(GTB) achieves the most accuracy with fewer errors. Ngo [35] applied an ML method for the prediction of cooling loads of buildings, based on data from 243 buildings and observed high accuracy in predictions. Rana et al. [36] proposed a data-driven approach that has shown greater accuracy than gradient tree boosting (GTB). Xuan et al. [37] used the Chaos approach and Wavelet Decomposition (WD) with the Support Vector Regression separately to predict the cooling load, and the results showed that the hybrid forecasting models perform better than the single ones. Zingre et al. [38] applied long short-term memory (LSTM) to estimate cooling load and demonstrated the predictive potential of this method when the data are in the form of a time series.

In this study, we implemented and tested several foundational machine learning models (Linear Regression, Decision Tree, Elastic Net, K Nearest Neighbor, Support Vector Machines) and ensemble machine learning models (Random Forest, Gradient Boosting, Histogram Gradient Boosting, Voting, Stacking) to determine the model with the best performance in the prediction of the cooling load based on architectural aspects of a tropical building. Python [39] programming language v3.6 was used for the cooling load estimation experiments. Anaconda 3 [40] was preferred as the development environment. Numpy [41] and Pandas [42] libraries were used to prepare the data for the training, the sci-kit-learn library was used to develop machine learning models, and the Matplotlib [43] library was utilized for data visualization.

2. Materials and Methods

The dataset utilized in this study is a 10,000-row subset of the dataset generated in [16]. K-Fold Cross Validation (KFCV) method was used as the training/validation strategy. A set of machine learning algorithms consisting of foundational algorithms (linear regression, decision tree regression, elastic network regression, K nearest neighbor regression, and support vector regression) and ensemble learning algorithms (random forest regression, gradient boosting regression, histogram-based gradient boosting regression and voting and stacking) is implemented in the training and validation stage. In the final stage, performance metrics were used to explore the effectiveness of the trained models. At this stage, different evaluation metrics such as coefficient of determination (R^2), mean square error (MSE), root mean square error (RMSE), and mean absolute error (MAE) were checked to discover the most accurate/efficient machine learning model. Finally, the trained models are stored. A schematic representation of the model evaluation process is provided in Figure 4.

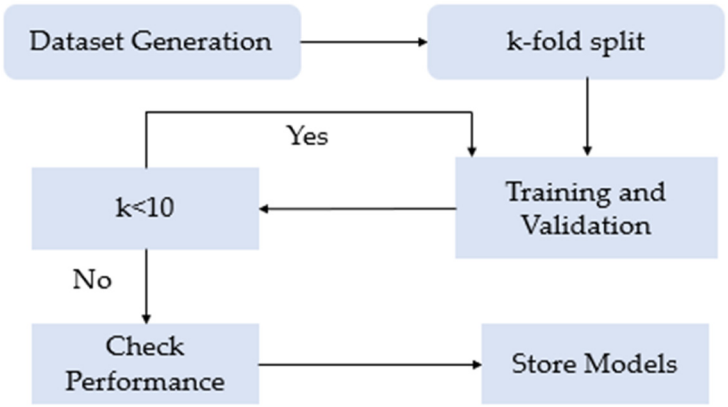

Figure 4. Model evaluation process.

2.1. Dataset Generation and Exploratory Data Analysis

Generating large and reliable datasets is pertinent in developing the prediction model as such datasets lead to better accuracy. In this study, two simulation methods namely the BIM application and Monte Carlo simulation are used in order to generate a large dataset. Simulations are carried out to collect reliable and verified data with an emphasis

on covering the whole range of available values. The four steps in generating a large and reliable dataset involve: (1) preparing a 3D base model for energy consumption (cooling load) simulation using the BIM application; (2) developing the dataset using the BIM application; (3) generating more data using the Monte Carlo method, and (4) synthesizing the data derived from the BIM application and Monte Carlo simulation into one large dataset [16].

The dataset was developed by one of the authors. The dataset includes the variables that have an impact on the cooling load of single-story terrace houses in Malaysia, along with the cooling loads of those buildings. The independent variables of the dataset include Floor Area (FA), Aspect Ratio (AR), Ceiling Height (CH), window (Glazing) material (WI), External Wall (Insulation/no Insulation) material (WA), Roof (Insulation/no Insulation) material (RO), Window Wall Ratio North (rear) faced (WWRN), Window Wall Ratio South (front) faced (WWRS), Horizontal Shading (SH) and Orientation (OR), where the dependent variable is Cooling Load. The variables of the dataset and their range were identified based on the design standards of residential buildings in Malaysia. The ranges of the variables were determined based on the literature review, standards, and semi-structured interviews. Once the variable ranges were determined, a Monte Simulation was run to select random values from the pre-determined value ranges stochastically. The EnergyPlus software (2010 version) was used to perform energy simulations and calculate annual building energy consumption (Cooling Load) for each design scenario that was generated in the Monte Carlo simulation. The dataset in this research is a randomly selected subset of the original dataset of 90,000+ rows and is composed of 10,000 rows. All variables were continuous. Measures of central tendency and dispersion for predictor variables are given in Table 3, and the histogram and boxplot of all variables in the dataset are provided in Figures 5 and 6. An explanation of all the variables was provided in the previous section. As the dataset was developed with a Monte Carlo Simulation based on discrete value ranges determined earlier based on standards and interviews, the values of all variables have discrete distributions. Although the values of independent variables are discrete, the dependent variable, Cooling Load, has a continuous distribution enabled and supported by a large number of samples generated as a result of the data generation with simulation and generative design tools.

Table 3. Measures of central tendency and dispersion for predictor variables.

Variable	Mean	Min	Max	Standard Deviation
Total floor area (FA) (m^2)	203.95	150	250	38.16513
Aspect ratio (AR)	0.544677	0.25	1	0.262597
Ceiling height (CH) (m)	3.00709	2.5	4	0.525065
External wall (WH) (U value *)	1.609684	0.29	2.65	0.81686
Roof (RO) (U value *)	1.819416	0.2	2.74	0.965887
Glazing (WI) (U value *)	2.991862	1.76	5.77	1.404832
WWR ** North Faced (WWRN) (%)	36.038	10	90	26.97448
WWR South Faced (WWRS) (%)	37.159	10	90	27.96496
Horizontal shading overhang (SH) (m)	0.96262	0	4	1.212151
Building orientation (OR) (°)	80.664	0	360	86.36747
Cooling Load (kWh)	108.741	64.174	294.223	42.103

* The U value is a measure of the insulating capacity of the material. (Total heat transfer coefficient $W/(m^2 \cdot K)$.
** WWR: Window wall ratio.

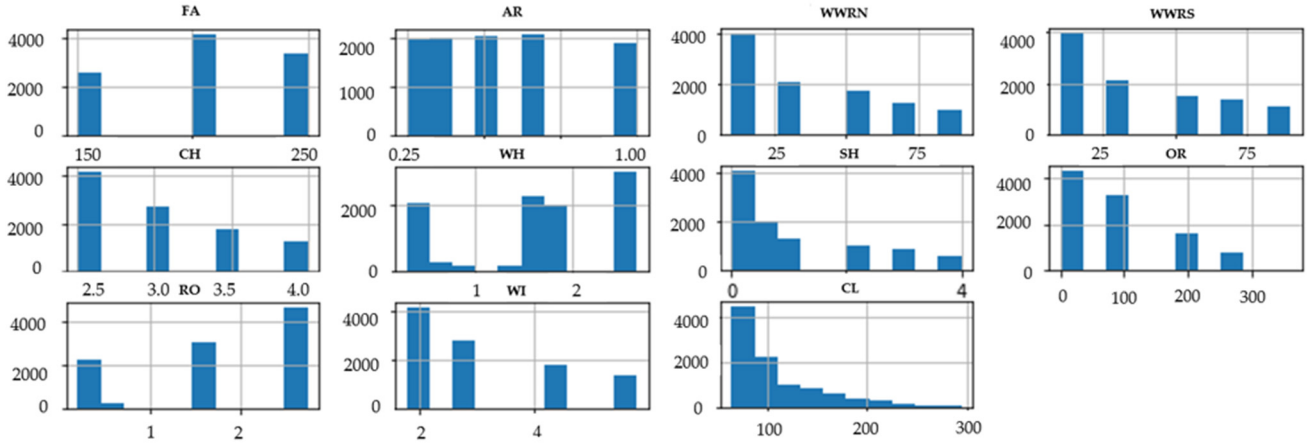

Figure 5. Histogram plot of variables.

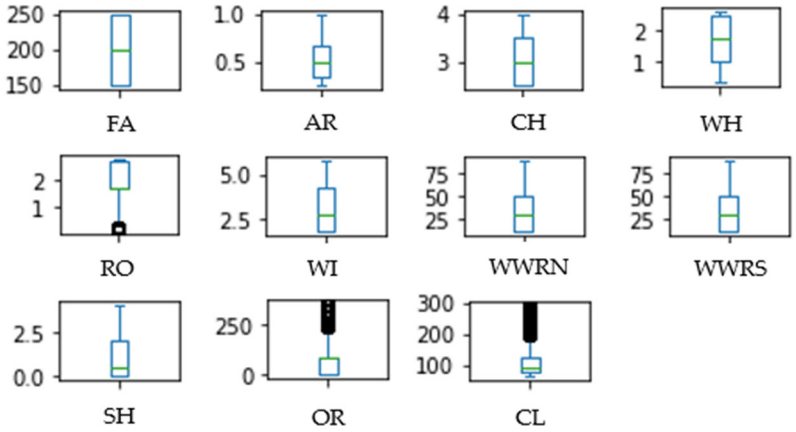

Figure 6. Boxplot of variables.

2.2. Machine Learning Models

Many years ago, researchers began to focus on the prediction of building cooling load in buildings by using BIM tools such as Energyplus and DOE 2. The operational energy prediction can be expedited using Building Information Modeling (BIM). Nevertheless, energy simulation tools have their own challenges such as the size and number of parameters that need to be considered. If the number of parameters and their variables are large, then this method takes a long time and requires more intention, and it will be very challenging. A small mistake due to carelessness would cause the wrong results, thus the wrong judgment of cooling load prediction. Furthermore, the problems with the dataset that is generated through the simulation are erroneous design assumptions and errors within modeling tools and their application. Due to the reasons above, researchers have to put a direct and significant focus on data-based energy consumption prediction for buildings taking into account data properties and artificial intelligence (machine learning) algorithms.

After the generation of the dataset, several machine learning (ML) models were implemented and tested. Machine learning is a sub-branch of artificial intelligence and refers to the artificial learning process through a set of algorithms that can continuously learn structurally and make meaningful inferences from large and complex data. Machine learning methods have gained importance in today's world [44]. Machine learning algorithms are used in many fields today [45–52]. Supervised learning, which is a type of machine learning, refers to the implementation of classification and regression algorithms where a dependent variable is known in advance. Classification algorithms are implemented when the dependent variable is binary or categorical. Regression algorithms are utilized when the type of the dependent variable is continuous. Different validation strategies can be

used in training the models. Two well-known strategies in this regard are train-test split and KFCV. In the train/test split strategy, the data are divided into 2 or 3 subsets (train/test or train/validation/test), trained with the former ones, and tested with the test dataset which the model has not seen before. Another validation strategy is K-Fold Validation where the dataset is trained and tested by dividing into k different parts.

In this study, the dataset is trained with different regression algorithms as the dependent variable (Cooling Load) is in a continuous nature/scale. The algorithms implemented are summarized in the following section. Additionally, the general machine learning process is shown in Figure 7.

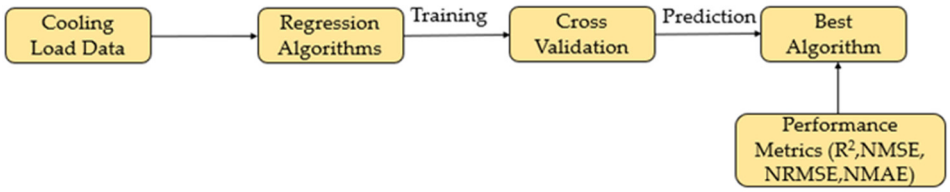

Figure 7. Machine learning process.

2.2.1. Foundational Algorithms

Linear Regression (LR): Linear regression is used to reveal the cause-and-effect relationship between a dependent variable and one or more independent variables [53]. The regression model associates the dependent variable with the independent variable or variables through a linear function [54]. Multiple linear regression (MLR) is used when there is more than one independent variable [55]. Decision Tree Regression (DTR): A decision tree is a nonparametric prediction model that can be used to represent both classifiers and regression models [56]. The method is widely used because the decision rules used in the creation of the mentioned tree structures are understandable. The decision tree performs a simple decision-making process by transforming complex data into a gradual state with a multi-stage and sequential approach to solve the classification and regression problem [57]. Elastic Network Regression (Elastic Net): Regularization in regression is used to avoid the overfitting of the data, especially when there is a large variance between train and test set performances. Well-known regularization methods in ML include LASSO(L1) and Ridge(L2). The Elastic Net method has emerged to tackle some of the shortcomings of the LASSO (L1) regularization method. Elastic Net uses LASSO (L1) regression and Ridge (L2) regression together [58]. K Nearest Neighbor Regression (k-NNR): The k Nearest Neighbor algorithm can be used for both classification and regression. Briefly defined, the data of unknown class is compared with other data in the training set, and a distance measurement is made. According to this calculated distance, the most optimal class is found for the data that has not yet been assigned to a class [59]. Support Vector Regression (SVR): Support vector regression is the use of support vector machines (SVM) for regression. In SVM, the optimum separation hyperplane is found to separate the classes from each other, and the distance between the support vectors of different classes is maximized [60].

2.2.2. Ensemble Algorithms

Ensemble learning is the joint decision-making of more than one learning algorithm. Unlike previous machine learning models, more than one learning model is run on the same data set and a joint decision is made according to certain rules [61]. Random Forest Regression (RF): In the RF method, a random forest is generated with multiple decision trees. In the algorithm, the decision trees are combined to obtain a more accurate and stable forecast [62]. Gradient Boosting Regression (GBR): In the gradient boosting algorithm, the goal is to iteratively predict and improve the error rate. In this model, we tried to fit the loss function found in the previous step to a negative gradient vector. To reduce the error, the incremental reduction results are added to the function [63]. Histogram-based Gradient Boosting Regression (HGBR): Histogram-based gradient boosting is a combination of a gradient boosting machine and a histogram-based algorithm [64]. The

histogram-based gradient boosting algorithm provides faster training of decision trees [65]. Voting: In the voting method, the class predictions made by various classifiers are voted on, and the class that obtains the most votes as a result of the voting becomes the class prediction of the majority [66]. Voting in regression is to combine different machine learning regressor models to return predicted average values. This is useful for balancing the individual weaknesses of the models [67]. In voting classification, more than one classification algorithm can be trained with the same training set, or a single algorithm can be trained with the same data set using different parameter values, and different classification models are created, and the final output value is produced by putting the outputs obtained from the models into the voting mechanism [68].

Stacking: The stacking ensemble learning method is based on the principle of generating a higher-performing prediction from these predictions by accepting the predictions of different types of classifiers as input for the metaclassifier [69]. The stacking ensemble learning method accepts the predictions of different types of classifiers as input for the metaclassifier, producing a higher-performing prediction from these predictions [70]. In stacking in classification, the output from different classifiers is passed as input to a metaclassifier for the final classification task [71].

2.3. K-Fold Cross Validation

The validation strategy for this study was selected as KFCV. KFCV helps determine a more accurate estimate of model prediction when compared with the train/test set validation strategy. This strategy helps in reducing the variance of the performance estimate and allows the use of more data for training. KFCV also avoids overfitting, as it exposes the model to different subsets of the data. In our implementation of KFCV, the number of folds (k) chosen is 10. In the 10-fold cross validation method, the dataset is divided into 10 equal parts. In each iteration, 9/10 of the dataset is used for training and the reserved 1/10 is used for testing. In each iteration, a different part of the data is reserved for testing. The basic structure of 10-fold cross validation is shown in Figure 8.

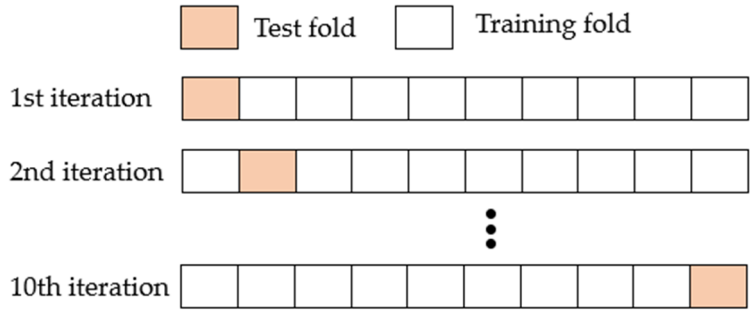

Figure 8. k-fold cross validation strategy [72].

The most important advantage of 10-fold cross validation is that all samples in the dataset are used in the training and testing phases. Thus, the positive or negative effects that some examples may have had during the education process are eliminated. Each sample is guaranteed to be used 9 times for training and 1 time for testing. As a result, 10-fold cross validation tests whether the accuracy value of the success of the model is random (or by chance) and provides a verified success rate.

2.4. Model Performance Metrics

It is important to know the performance of the established models to understand which model performs better than the others. There are several metrics defined in the literature to evaluate the performance of regression algorithms/models. Some of these metrics are given in Table 4. In the formulas presented in Table 4, n represents the number of predictions; y_i represents the actual value of the ith observation, x_i represents the predicted value of the ith observation, and $\bar{x}_i$ represents the average of the predicted values.

Table 4. Performance metrics used in this study.

Performance Metrics	Description
Coefficient of determination	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}\left(y_i - \bar{x_i}\right)^2}$
Mean Squared Error	$MSE = \sum_{i=1}^{n}\frac{(y_i - x_i)^2}{n}$
Root Mean Squared Error	$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(y_i - x_i)^2}{n}}$
Mean Absolute Error	$MAE = \frac{\sum_{i=1}^{n}\lvert y_i - x_i\rvert}{n}$

The following metrics are used to evaluate the models in this study. Coefficient of Determination (R^2): This coefficient is a measure that measures how well a statistical model predicts an outcome. It measures the percentage of variability within the values of the dependent variable that can be explained by the regression model. A high R^2 indicates a good regression model fit. As the number of data increases, the reliability of R^2 also increases.

Mean Squared Error (MSE): MSE is a metric that measures the error rate of the prediction model as the square of the error. A small difference between the actual and predicted value indicates a good prediction. The closer the MSE value is to 0, the better the forecast is [73].

Root Mean Squared Error (RMSE): The RMSE can be found by taking the square root of MSE. Root mean squared error is the standard deviation of the difference between actual and predicted values. The closer the RMSE value obtained by the model is to zero, the closer the values predicted by the model will be to the real values. Since RMSE penalizes large errors more, it may be more appropriate in some situations. RMSE avoids the undesirable use of absolute values in many mathematical calculations [74].

Mean Absolute Error (MAE): Mean absolute error (MAE) is frequently used in regression problems because it is easy to interpret. The MAE is the measure of the difference between two continuous variables, i.e., the sum of the absolute values of the differences between the actual and predicted values [74]. An MAE value close to 0 indicates a successful analysis.

3. Results

In this study, predictions were made for cooling load. The inputs are the total floor area, aspect ratio, ceiling height, window material, external wall material, roof material, window wall ratio north faced, window wall ratio south faced, horizontal shading, and orientation of the building. The output is cooling load (Figure 9).

In this study, five foundational regression algorithms and five ensemble algorithms were used for the generation, training, and validation of the models. The Voting algorithm has been implemented with different combinations of base learners (lr: linear regression, knr: k-neighbor regression, ent: elastic net regression, dtr: decision tree regression, svr: support vector regression, rfr: random forest regression, gbr: gradient boosting regression, hgbr: a histogram-based gradient boosting regression). The Stacking algorithm has been implemented by utilizing different combinations of base learners and Gradient Boosting Regressor as the final estimator. The accurate metric values obtained as a result of the predictions made using each algorithm and each implementation of Voting and Stacking algorithm are provided in Table 5.

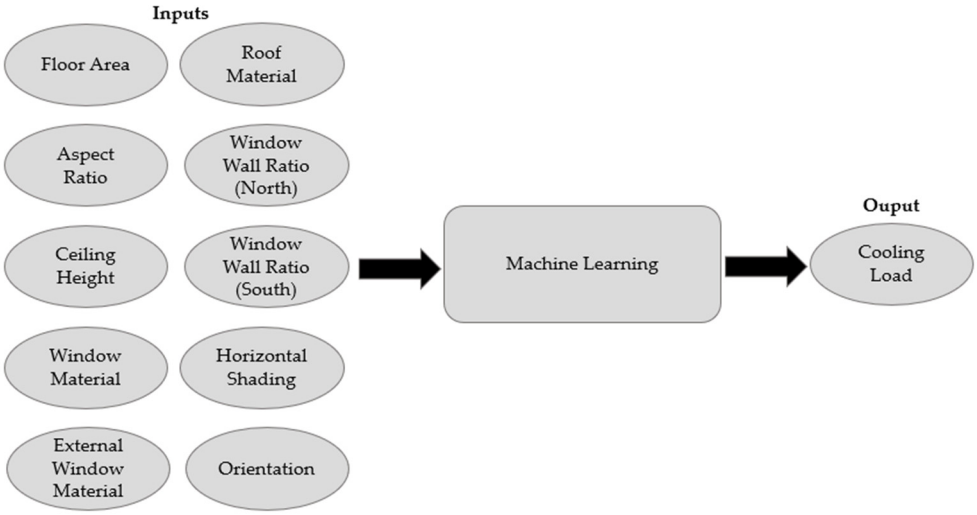

Figure 9. Inputs and output used in machine learning.

Table 5. Accurate metrics of models.

Algorithm	R^2	NMSE	NRMSE	NMAE
Foundational Models				
Linear Regression	0.8656	−237.8091	−15.4138	−10.9330
Decision Tree Regressor	0.9569	−75.7774	−8.7182	−5.1185
Elastic Net	0.8015	−351.5527	−18.7423	−13.1531
K Neighbors Regressor	0.8447	−275.0733	−16.5700	−10.6251
Support Vector Regressor	0.7341	−471.1797	−21.6914	−14.1668
Ensemble Models				
Random Forest Regressor	0.9835	−29.1874	−5.3989	−3.3252
Gradient Boosting Regressor	0.9857	−25.2759	−5.0185	−3.3374
Hist. Gradient Boosting Regressor	0.9949	−8.9500	−2.9849	−1.7973
Voting				
lr, rfr, gbr	0.9717	−50.1243	−7.0727	−4.6916
knr, dtr, hgbr	0.9748	−45.2701	−6.7476	−4.2715
knr, ent, rfr	0.9238	−132.6369	−11.5088	−7.6540
svr, rfr, gbr	0.9564	−77.2832	−8.7821	−5.5664
rfr, gbr, hgbr	0.9921	−13.9571	−3.7286	−2.2663
lr, dtr, rfr, gbr	0.9766	−41.0442	−6.4132	−3.4676
lr, dtr, ent, knr, svr	0.9001	−176.9352	−13.2938	−8.5382
lr, dtr, rfr, gbr, hgbr	0.9835	−29.1855	−5.4077	−3.4638
Stacking				
Final Estimator = Gradient Boosting Regressor				
lr, rfr, gbr	0.9889	−19.5029	−4.4094	−2.6754
knr, dtr, hgbr	0.9948	−9.0843	−2.9966	−1.8139
knr, ent, rfr	0.9849	−26.6835	−5.1600	−3.1831
svr, rfr, gbr	0.9888	−19.7641	−4.4362	−2.6954
rfr, gbr, hgbr	0.9949	−8.9284	−2.9811	−1.7841
lr, dtr, rfr, gbr	0.9890	−19.2676	−4.3964	−2.6668
lr, dtr,et, knr, svr	0.9687	−57.0651	−7.5189	−4.5797
lr, dtr, rfr, gbr, hgbr	0.9949	−8.9397	−2.9843	−1.7669

As mentioned previously, the R^2 value approaching one indicates that the model has a high success rate. When the results of all models are compared, the lowest R^2 value (0.7341) was obtained in the SVR implementation, while the highest R^2 score (0.9949) was obtained with two models, (i) the histogram gradient boosting regression algorithm and (ii) in two of the stacking implementations, (a) the combination of random forest

regression, gradient boosting regression and histogram-based gradient boosting regression with final estimator being the gradient boosting regression and (b) the combination on linear regression, decision tree regression, random forest regression, gradient boosting regression, and histogram-based gradient boosting regression with final estimator being the gradient boosting regression. Among the foundational models, the model with the highest R^2 score (0.9569) was the decision tree. The model with the lowest R^2 score (0.7341) is SVR (i.e., this was also the lowest overall R^2 score).

In ensemble methods, the training/validation with histogram gradient boosting algorithm and stacking resulted in the highest R^2 value (0.9949). Combinations of which the highest R^2 value was obtained in the stacking model, were the combinations in which multiple ensemble methods were used together as base learners. As the low error rate is an indication of better success in regression algorithms, among all models, the success of the histogram gradient boosting algorithm and stacking models can also be confirmed through negative MSE/RMSE/MAE values for i, ii.a and ii.b (NMSE = {−8.93; −8.93; −8.94}, NMAE = {−1.8; −1.78; −1.77}). When negative values of MSE/RMSE/MAE are considered, the smaller number denotes a higher error. The model with the highest error rate (NMSE = −471.17, NMAE = −14.16) was SVR and the results complied with low R^2 scores obtained by the same model. Among the foundational models, the decision tree model has the lowest error rate (NMSE = −75.77, NMAE = −5.11).

In ensemble methods, the stacking models have the lowest error rates. The combination of which the lowest MSE is obtained (NMSE = −8.93) was the Stacking model with the combination of random forest regression, gradient boosting regression, and histogram-based gradient boosting regression with the final estimator being the gradient boosting regression. The combination in which the lowest MAE is obtained (NMAE = −1.77) was the combination of linear regression, decision tree regression, random forest regression, gradient boosting regression, and histogram-based gradient boosting regression with the final estimator being the gradient boosting regression.

The analyses were performed with a PC equipped with IntelI CoreI I3-5005U CPU @ 2.00 GHz (4Cores). When the time performance of the models was evaluated, the SVR model was the model with the worst time performance (54 s for 10-fold cv process) among the foundational models and is the model with the lowest overall R^2 score. Among the models with the best performance, the time performance of the Histogram Gradient Boosting model (24 s for 10-fold cv process) was much better than the best-performing Stacking model (178 s for 10-fold cv process). A comparison of the models is shown in Table 6.

Table 6. Comparison of Models.

Algorithm	Computational Speed	Success
Foundational Models		
Linear Regression	Fast	Medium
Decision Tree Regressor	Fast	High
Elastic Net	Fast	Medium
K Neighbors Regressor	Fast	Medium
Support Vector Regressor	Low	Low
Ensemble Models		
Random Forest Regressor	Fast	High
Gradient Boosting Regressor	Fast	Very High
Hist. Gradient Boosting Regressor	Fast	High
Voting	Low	High
Stacking	Low	Very High

4. Discussion and Conclusions

Due to the rapid population growth throughout the world, the energy demand is increasing day by day. The built environment is one of the key consumers of energy. Especially for buildings in tropical climates, the energy load required for cooling is very

high. Therefore, it is necessary to know the energy load required for the cooling of the building to develop building designs with a focus on energy efficiency.

In the study, various machine learning algorithms are implemented to predict the cooling load of a tropical building based on its architectural attributes such as floor area, aspect ratio, ceiling height, window material, external wall material, roof material, window wall ratio north faced, window wall ratio south faced, horizontal shading, and orientation. The main findings of this study are as follows:

(1) Ensemble learning algorithms/models are superior to foundational algorithms models in the prediction of the cooling load of the building through regression. Among the ensemble models, stacking-based models were found to be most successful when compared to others. Ensemble models have been more successful (high R^2, low error) than base models as they combine decisions from multiple models to improve their overall performance.
(2) It is observed that Support Vector Regression was the least efficient model among all foundational and ensemble models, not only in terms of performance/accuracy but also in terms of time performance in the training/validation stages.
(3) When only the foundational algorithms were compared, Decision Tree Regression was the model with the best performance. This indicates that Tree Based approaches can be efficient in the prediction of the cooling load of buildings based on their architectural properties.
(4) In a similar study, Guo et al. [75] predicted heating and cooling loads based on light gradient boosting machine algorithms. Common models in our study and [75] are Random Forest and SVR. The same R^2 values were obtained for Random Forest in both studies, but SVR had a higher R^2 value in Guo et al. [75]. This indicates (a) that based on the nature of the dataset, SVR can also provide accurate results, so tests with SVR should not be neglected in studies for developing cooling load prediction models, (b) that Tree Based approaches and Ensemble models are very promising in cooling load prediction.
(5) When the time performance of the models is taken into account, the Histogram Gradient Boosting algorithm appears as the optimal model, as it also provides a good prediction performance.

In summary, the results of the study have demonstrated that Ensemble Learning algorithms can be successfully used to establish the relationship between the architectural properties of tropical buildings and their cooling load because ensemble methods come to a conclusion by using more than one predictor in the same prediction task. In this method, the results of predictors with different metric scores are combined with different methods (voting, stacking, etc.). Thus, more successful performance is achieved. Furthermore, the cooling load of tropical buildings can be accurately predicted through the use of ensemble learning algorithms. Future research will focus on how hyperparameter optimization would enhance the performance of the provided models. The accuracy of the prediction model provided in this paper can be further enhanced through the addition of other predictor variables such as the occupancy status of the rooms, occupancy schedule, space usage conditions, and characteristics.

Author Contributions: A.N.S. and Ü.I. generated the data; Y.A. and Ü.I. generated the analysis codes; Y.A., Ü.I. and G.B. developed the theory, background, and formulations of the problem; Verification of the results was performed by Y.A. and Ü.I.; The text of the paper was written by Y.A., Ü.I. and G.B.; The figures were drawn by Y.A.; Ü.I., S.K. and Z.W.G. edited the paper; G.B. and Z.W.G. supervised the research direction. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be made available upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. United Nations Brundtland Commission Report. 1987. Available online: http://www.un-documents.net/our-common-future.pdf (accessed on 20 April 2023).
2. Houghton, J. Global warming. *Rep. Prog. Phys.* **2005**, *68*, 1343. [CrossRef]
3. Pérez-Lombard, L.; Ortiz, J.; Pout, C. A review on buildings energy consumption information. *Energy Build.* **2008**, *40*, 394–398. [CrossRef]
4. Erkmen, F.İ.; Gedik, G.Z.; Sözen, M.Ş. Comparison of Cooling Loads in Hot Climates. *Megaron* **2006**, *1*, 112.
5. Tıkansak, T.E. Energy Efficiency in Housing. *ICONARP Int. J. Archit. Plan.* **2013**, *1*, 189–200.
6. Guan, L. Sensitivity of building cooling loads to future weather predictions. *Archit. Sci. Rev.* **2011**, *54*, 178–191. [CrossRef]
7. Annual Energy Review. Available online: https://www.eia.gov/totalenergy/data/annual/ (accessed on 18 April 2023).
8. Amasyali, K.; El-Gohary, N.M. A review of data-driven building energy consumption prediction studies. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1192–1205. [CrossRef]
9. Cao, X.; Dai, X.; Liu, J. Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade. *Energy Build.* **2016**, *128*, 198–213. [CrossRef]
10. Malaysia. Available online: https://www.wikipedia.org/ (accessed on 20 April 2023).
11. Omar, E.O.H.; Endut, E.; Saruwono, M. Adapting by altering: Spatial modifications of terraced houses in the Klang Valley area. *Asian J. Environ.-Behav. Stud.* **2016**, *2*, 1–10. [CrossRef]
12. NAPIC. *Property Stock Report: Residential Property Stock Table Q4 2018*; National Property Information Centre (NAPIC): Putrajaya, Malaysia, 2018.
13. Hashim, A.H.; Rahim, Z.A. The influence of privacy regulation on Urban Malay families living in terrace housing. *Int. J. Archit. Res.* **2008**, *2*, 94–102.
14. Khan, T.H. Is Malaysian Terrace Housing an outdated planning concept. *Scott. J. Arts Soc. Sci. Sci. Stud.* **2012**, *3*, 114–128.
15. Ju, S.-R.; Omar, S.B. A typology of modern housing in Malaysia. *Int. J. Hum. Ecol.* **2010**, *11*, 109–119.
16. Sadeghifam, A.N. Development of Cooling Load Prediction Tool for Low-Rise Residential Buildings. Ph.D. Thesis, University of Technology Malaysia, Johor Bahru, Malaysia, 2019.
17. What Is Cooling Load? Purpose & Calculation. Available online: https://electricalworkbook.com/cooling-load/ (accessed on 17 April 2023).
18. Summary Cooling Design Table. Available online: https://designbuilder.co.uk/helpv7.0/Content/Summary_Cooling_Design_Table.html (accessed on 17 April 2023).
19. McKeen, P.; Fung, A.S. The effect of building aspect ratio on energy efficiency: A case study for multi-unit residential buildings in Canada. *Buildings* **2014**, *4*, 336–354. [CrossRef]
20. How High Ceilings Can Be Affecting the Efficiency of Your HVAC. Available online: https://makeitmowery.com/how-high-ceilings-can-be-affecting-the-efficiency-of-your-hvac/ (accessed on 17 April 2023).
21. Yıldız, Y.; Göksal Özbalta, T.; Arsan, Z.D. *Impact of Window-to-Wall Surface Area for Different Window Glass Types and Wall Orientations on Building Energy Performance: A Case Study for a School Building Located in Izmir*; Megaron: İzmir, Turkey, 2011.
22. Sarıkçıoğlu, N. Thermal and Economical Analysis of Commonly Used Building Walls for Cooling Applications. Master's Thesis, Gaziantep University, Gaziantep, Turkey, 2011.
23. Yao, L.; Chini, A.; Zeng, R. Integrating cost-benefits analysis and life cycle assessment of green roofs: A case study in Florida. *Hum. Ecol. Risk Assess. Int. J.* **2020**, *26*, 443–458. [CrossRef]
24. Gonca, Ö. The Effect of Buildings' Window/Wall Ratio and Orientation Parameters on Solar Energy Gain. *Sci. Eng. Des. J. East. Anatolia Reg.* **2021**, *3*, 425–441.
25. Kim, S.H.; Shin, K.J.; Choi, B.E.; Jo, J.H.; Cho, S.; Cho, Y.H. A study on the variation of heating and cooling load according to the use of horizontal shading and venetian blinds in office buildings in Korea. *Energies* **2015**, *8*, 1487–1504. [CrossRef]
26. Sun Control and Shading Devices. Available online: https://www.wbdg.org/resources/sun-control-and-shading-devices (accessed on 17 April 2023).
27. Lim, T.; Yim, W.S.; Kim, D.D. Analysis of the thermal and cooling energy performance of the perimeter zones in an office building. *Buildings* **2022**, *12*, 141. [CrossRef]
28. Köroğlu Isin, N.; Alaloğlu, M.; Erdoğan, A.; Acar, L. Hourly Analysis Program. *HVAC Refrig. Fire Fight. Sanit. J.* **2011**, *73*, 1–8.
29. Maçka Kalfa, S. The Method Using in Determination of Heating and Cooling Loads for Residential Buildings in Turkish Climate Regions. Ph.D. Thesis, Karadeniz Technical University, Trabzon, Turkey, 2014.
30. Stephens, B. Illinois Institute of Technology-Civil, Architectural and Environmental Engineering, Building Science [Powerpoint Slides]. 2017. Available online: https://www.built-envi.com/wp-content/uploads/cae331_513_lecture22_cooling-load-calcs-part1.pdf (accessed on 17 April 2023).
31. Li, X.; Lu, J.-H.; Ding, L.; Xu, G.; Li, J. Building cooling load forecasting model based on LS-SVM. In Proceedings of the 2009 Asia-Pacific Conference on Information Processing, Shenzhen, China, 18–19 July 2009; IEEE: Piscataway, NJ, USA, 2009; Volume 1, pp. 55–58.

32. Li, Q.; Meng, Q.; Cai, J.; Yoshino, H.; Mochida, A. Applying support vector machine to predict hourly cooling load in the building. *Appl. Energy* **2009**, *86*, 2249–2256. [CrossRef]
33. Gao, Z.; Yu, J.; Zhao, A.; Hu, Q.; Yang, S. A hybrid method of cooling load forecasting for large commercial building based on extreme learning machine. *Energy* **2022**, *238*, 122073. [CrossRef]
34. Sha, H.; Moujahed, M.; Qi, D. Machine learning-based cooling load prediction and optimal control for mechanical ventilative cooling in high-rise buildings. *Energy Build.* **2021**, *242*, 110980. [CrossRef]
35. Ngo, N.T. Early predicting cooling loads for energy-efficient design in office buildings by machine learning. *Energy Build.* **2019**, *182*, 264–273. [CrossRef]
36. Rana, M.; Sethuvenkatraman, S.; Goldsworthy, M. A data-driven approach based on quantile regression forest to forecast cooling load for commercial buildings. *Sustain. Cities Soc.* **2022**, *76*, 103511. [CrossRef]
37. Xuan, Z.; Xuehui, Z.; Liequan, L.; Zubing, F.; Junwei, Y.; Dongmei, P. Forecasting performance comparison of two hybrid machine learning models for cooling load of a large-scale commercial building. *J. Build. Eng.* **2019**, *21*, 64–73. [CrossRef]
38. Zingre, K.; Srinivasan, S.; Marzband, M. Cooling load estimation using machine learning techniques. In Proceedings of the ECOS 2019—32nd International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems, Wroclaw, Poland, 23–28 June 2019.
39. Python (3.6) [Computer Software]. Available online: http://python.org (accessed on 19 April 2023).
40. Anaconda3 [Computer Software]. Available online: https://anaconda.org/ (accessed on 19 April 2023).
41. NumPy. Available online: https://numpy.org/ (accessed on 19 April 2023).
42. Pandas—Python Data Analysis Library. Available online: https://pandas.pydata.org/ (accessed on 19 April 2023).
43. Scikit-Learn: Machine Learning in Python. Available online: https://scikit-learn.org/stable/ (accessed on 19 April 2023).
44. Akyiğit, H.E.; Taşcı, T. Customer Churn Analysis in the Insurance Sector Using Machine Learning. *J. Des. Archit. Eng.* **2021**, *2*, 66–79.
45. Aydın, Y.; Işıkdağ, Ü.; Bekdaş, G.; Nigdeli, S.M.; Geem, Z.W. Use of Machine Learning Techniques in Soil Classification. *Sustainability* **2023**, *15*, 2374. [CrossRef]
46. Cakiroglu, C.; Islam, K.; Bekdaş, G.; Nehdi, M.L. Data-driven ensemble learning approach for optimal design of cantilever soldier pile retaining walls. In *Structures*; Elsevier: Amsterdam, The Netherlands, 2023; Volume 51, pp. 1268–1280.
47. Aydın, Y.; Bekdaş, G.; Nigdeli, S.M.; Isıkdağ, Ü.; Kim, S.; Geem, Z.W. Machine Learning Models for Ecofriendly Optimum Design of Reinforced Concrete Columns. *Appl. Sci.* **2023**, *13*, 4117. [CrossRef]
48. Cakiroglu, C.; Bekdaş, G. Predictive Modeling of Recycled Aggregate Concrete Beam Shear Strength Using Explainable Ensemble Learning Methods. *Sustainability* **2023**, *15*, 4957. [CrossRef]
49. Bekdaş, G.; Cakiroglu, C.; Kim, S.; Geem, Z.W. Optimal dimensioning of retaining walls using explainable ensemble learning algorithms. *Materials* **2022**, *15*, 4993. [CrossRef] [PubMed]
50. Yücel, M.; Nigdeli, S.M.; Bekdaş, G. Machine Learning-Based Model for Optimum Design of TMDs by Using Artificial Neural Networks. In *Optimization of Tuned Mass Dampers: Using Active and Passive Control*; Springer International Publishing: Cham, Switzerland, 2022; pp. 175–187.
51. Mohan, S.; Thirumalai, C.; Srivastava, G. Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access* **2019**, *7*, 81542–81554. [CrossRef]
52. Massaro, A.; Magaletti, N.; Cosoli, G.; Leogrande, A.; Cannone, F. Use of machine learning to predict the glycemic status of patients with diabetes. *Med. Sci. Forum* **2022**, *10*, 11.
53. Casson, R.J.; Farmer, L.D. Understanding and checking the assumptions of linear regression: A primer for medical researchers. *Clin. Exp. Ophthalmol.* **2014**, *42*, 590–596. [CrossRef] [PubMed]
54. Luu, Q.H.; Lau, M.F.; Ng, S.P.; Chen, T.Y. Testing multiple linear regression systems with metamorphic testing. *J. Syst. Softw.* **2021**, *182*, 111062. [CrossRef]
55. Rath, S.; Tripathy, A.; Tripathy, A.R. Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. *Diabetes Metab. Syndr. Clin. Res. Rev.* **2020**, *14*, 1467–1474. [CrossRef] [PubMed]
56. Rokach, L.; Maimon, O. Decision trees. In *Data Mining and Knowledge Discovery Handbook*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 165–192.
57. Safavian, S.R.; Landgrebe, D. A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* **1991**, *21*, 660–674. [CrossRef]
58. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* **2005**, *67*, 301–320. [CrossRef]
59. Taneja, S.; Gupta, C.; Goyal, K.; Gureja, D. An enhanced k-nearest neighbor algorithm using information gain and clustering. In Proceedings of the 2014 Fourth International Conference on Advanced Computing & Communication Technologies, Rohtak, India, 8–9 February 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 325–329.
60. Ayhan, S.; Erdoğmuş, Ş. Kernel Function Selection for the Solution of Classification Problems via Support Vector Machines. *Eskişehir Osman. Univ. J. Econ. Adm. Sci.* **2014**, *9*, 175–201.
61. Dietterich, T.G. Ensemble learning. *Handb. Brain Theory Neural Netw.* **2002**, *2*, 110–125.
62. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
63. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

64. Nhat-Duc, H.; Van-Duc, T. Comparison of histogram-based gradient boosting classification machine, random Forest, and deep convolutional neural network for pavement raveling severity classification. *Autom. Constr.* **2023**, *148*, 104767. [CrossRef]
65. Histogram-Based Gradient Boosting Ensembles in Python. Available online: https://machinelearningmastery.com/histogram-based-gradient-boosting-ensembles/ (accessed on 18 April 2023).
66. Džeroski, S.; Ženko, B. Is combining classifiers with stacking better than selecting the best one? *Mach. Learn.* **2004**, *54*, 255–273. [CrossRef]
67. VotingRegressor. Available online: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingRegressor.html (accessed on 20 April 2023).
68. Akram, V.K.; Taşer, P.Y. Ensemble Learning-based Method for Detection of Byzantine Attacks in Wireless Sensor Networks. *Dokuz Eylul Univ. Fac. Eng. J. Sci. Eng.* **2020**, *22*, 905–918.
69. Wolpert, D.H. Stacked generalization. *Neural Netw.* **1992**, *5*, 241–259. [CrossRef]
70. Alpaydin, E. *Introduction to Machine Learning*, 3rd ed.; MIT Press: Cambridge, MA, USA, 2014; Volume 67, pp. 69–72.
71. Stacking Classifier Approach for a Multi-Classification Problem. Available online: https://towardsdatascience.com/stacking-classifier-approach-for-a-multi-classification-problem-56f3d5e120c8#:~:text=Just%20like%20other%20ensemble%20techniques,Figure%20%2D1 (accessed on 20 April 2023).
72. Niu, M.; Li, Y.; Wang, C.; Han, K. RFAmyloid: A web server for predicting amyloid proteins. *Int. J. Mol. Sci.* **2018**, *19*, 2071. [CrossRef]
73. Saigal, S.; Mehrotra, D. Performance comparison of time series data using predictive data mining techniques. *Adv. Inf. Min.* **2012**, *4*, 57–66.
74. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [CrossRef]
75. Guo, J.; Yun, S.; Meng, Y.; He, N.; Ye, D.; Zhao, Z.; Jia, L.; Yang, L. Prediction of heating and cooling loads based on light gradient boosting machine algorithms. *Build. Environ.* **2023**, *236*, 110252. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.