Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2020, 7(9):61-75**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

A Holistic Hybrid Entropy-TOPSIS-based Method for the prioritization of Machine Learning Models in Energy Efficient buildings

Eslam Mohammed Abdelkader*[1], Nehal Elshaboury[2], Abobakr Al-Sakkaf[3,4]

[1]Assistant Lecturer, Structural Engineering Department, Faculty of Engineering, Cairo University, Giza, Egypt
[2]Assistant Lecturer, Construction and Project Management Research Institute, Housing and Building National Research Center, Giza, Egypt
[3]Ph.D. candidate, Department of Building, Civil, and Environmental Engineering, Concordia University, Montreal, QC, Canada. E-mail: alsakkaf.abobakr@gmail.com
[4]Assistant lecturer, Department of Architecture & Environmental Planning, College of Engineering & Petroleum, Ḥadhramout University, Mukalla, Yemen
*Corresponding author, E-mail: eslam_ahmed1990@hotmail.com

Abstract Heating and cooling loads account for a considerable proportion of energy consumption in buildings, which contributes significantly in amplifying environmental emissions and global warming. The characteristics of the buildings play a monumental role in their energy performance. In this context, it is essential to build an efficient and robust machine learning models to simulate energy consumption of buildings. This research introduces a hybrid entropy-TOPSIS based method for the prioritization of machine learning models based on a set of performance indicators. Six types of machine learning models are investigated namely, Elman neural network, linear regression, K-nearest neighbor, Gaussian process regression, gradient boosted decision trees and relevance vector machine. Shannon entropy is applied to compute the relative weights of the performance metrics. Eventually, Technique for Order of Preference by Similarity to Ideal Solution is utilized to obtain a full ranking of the prediction models.

Keywords Heating and cooling; energy consumption; machine learning; K-nearest neighbor; Shannon entropy; TOPSIS

1. Introduction

The building sector is responsible for more than 30% of energy consumption all over the world [1]. The thermal energy in any building involves heating and cooling loads, and these loads are regulated by heating ventilation and air conditioning (HVAC) system [2]. The HVAC systems provide a desirable indoor air condition based on the computed heating and cooling loads [3]. The required heating and cooling loads are estimated based on building properties, usage, and climate conditions. Therefore, the proper design of HVAC systems ensures enhanced energy performance of buildings [4]. The level of energy consumption is increasing globally because of rising living standards. For instance, the building construction industry consumes about 40% of the total energy in the United States and the European Union [5].

The aforementioned reasons have pushed researchers to develop predictive tools that model building energy consumption [6-7]. In general, there are two common methods for examining the energy performance of buildings namely, forward modeling and inverse modeling [8]. In the forward modeling, the tools used to simulate the energy performance are based on the actual attributes of the building. There are some disadvantages associated with the application of such simulators because of the difficulty of setting certain parameters in practice. Moreover, it is not suitable for appraising the energy consumption of occupied buildings [9]. Besides,

the accuracy of this method is dependent on the level of precision of the included parameters as well as being a time-consuming method. This can be seen by the varying accuracies obtained from different simulation programs [10].

In order to overcome the limitations and drawbacks associated with the simulation packages, many researchers adopt inverse modeling to evaluate building energy consumption [11]. This method allows evaluating the impact of significant factors on building energy consumption [12-13]. It can be seen that the primary advantages of machine learning models are their ease of implementation and high speed of performance. Besides, the structure of machine learning models could handle the changes in the input variables [4]. In this regard, the next section presents an extensive literature review on the applicability and feasibility of machine learning models in building energy consumption.

Literature Review

Poel *et al* [14] defined the energy performance of buildings as the amount of energy that is consumed to satisfy the different needs associated with the use of the building. The authors mentioned that the energy performance could be triggered by many characteristics in any building such as insulation, installation, design, positioning, solar exposure, energy production, indoor climate, influence of neighboring structures, and many other factors. Fahmy *et al* [15] assessed the energy consumption, energy cost, and thermal comfort of three different external walls specifications in Egypt. The results yielded that the 10cm GRC (C2) wall specification is more energy-efficient compared to the single wall of half red-brick –Ct. Cuadrado *et al* [16] applied integrated value model for the sustainability assessment concept to compute the environmental sustainability index for timber structures. The proposed model could assist decision-makers to identify the environmentally friendly options that must be undertaken at the design phase. Mousa *et al* [17] applied building information modeling to visualize carbon emissions in any building. The proposed methodology acts as a tool that aids facility managers in making informative decisions. Marzouk and Mohammed Abdlekader [18] introduced a hybrid fuzzy multi-objective non-dominated sorting genetic algorithm II model to select the most sustainable materials of building components. The model accounted for the time, environmental, and cost constraints. Finally, a technique for order preference by similarity to ideal solution decision-making technique was applied to select the most feasible solution among the Pareto-optimal solutions.

The application of artificial intelligence has attracted great attention in the field of energy consumption of buildings. In particular, machine learning models have emerged as powerful tools and techniques in building energy management [4]. For example, Jihad and Tahiri [19] forecasted the heating and cooling loads of residential buildings using an artificial neural network (ANN) learning algorithm. It was proven that ANN is a powerful tool that could be useful for architects and designers in building design. Kavaklioglu [20] modeled the heating and cooling loads in residential buildings using the partial least squares method. The developed model was assessed against the ordinary least squares model. The results indicated that the application of partial least squares regression is feasible in modeling heating and cooling loads in buildings. Srihari and Santhi [21] built prediction of heating and cooling load to improve the energy efficiency of buildings using four regression and three gradient boosting models. It was concluded that machine learning models can predict heating and cooling loads with high accuracy. The obtained loads could be utilized to install efficient heating, cooling, and ventilation systems and thus result in economic and environmental gains. Bui *et al* [4] predicted the heating and cooling loads in energy-efficient buildings using genetic algorithms (GA) and the imperialist competition algorithm (ICA). The aforementioned artificial intelligent models have been applied to optimize the weights and biases of the ANN. The authors concluded that the ICA-ANN model exhibited superior performance when compared to the GA-ANN and ANN models.

Moayedi *et al* [22] employed six machine learning techniques, namely multi-layer perceptron regressor (MLPr), lazy locally weighted learning (LLWL), alternating model tree (AMT), RF, ElasticNet (ENet), and radial basis function regression (RBFr) for the prediction of heating load in energy-efficient buildings. The outcomes of the above-mentioned models were assessed using root relative squared error (RRSE), root mean square error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), and relative absolute error (RAE) statistical indexes. The results showed the superiority of the RF model in estimating the heating load in energy-

efficient buildings. Mohammed Abdelkader *et al* [23] performed a comprehensive analysis to predict heating and cooling loads in residential buildings using several machine learning models. These models are; back-propagation artificial neural network, generalized regression neural network, radial basis neural network, radial kernel support vector machine (SVM), and analysis of variance kernel SVM. The comparisons were conducted using the mean absolute percentage error (MAPE), MAE, and RMSE. Furthermore, the significance of the aforementioned models was evaluated using the student's t-test. The results yielded that the radial basis function network outperformed the remaining models. Roy *et al* [24] proposed a deep neural network (DNN) machine learning technique to estimate the heating load and cooling load of residential buildings. The output of DNN has been compared with gradient boosted machine (GBM), Gaussian process regression (GPR), and minimax probability machine regression (MPMR). The performance of the proposed model was assessed using statistical performance metrics such as variance accounted for (VAF), relative average absolute error (RAAE), root means absolute error (RMAE), R^2, standard deviation ratio (RSR), MAPE, Nash–Sutcliffe coefficient (NS), RMSE, weighted MAPE (WMAPE), and MAPE. It was proved that DNN and GPR models have yielded the best predictions for heating and cooling loads.

Proposed Method

The ultimate objective of the present study is to construct a multi-criteria decision making model for the prioritization of machine learning models in energy efficient buildings. As shown in Figure 1, the proposed method is composed of four main modules discussed in the following lines. The dataset utilized in the present study is composed of 768 data points of energy simulation generated using Ecotect energy analysis software. . The dataset was published by UCI machine learning repository [25] based on the work published by Tsanas and Xifara [26]. The simulated energy is produced from different buildings of different surface areas and dimensions. The input variables incorporated to predict the heating and cooling loads of buildings encompass relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area and glazing area distribution. The number of possible values of relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area and glazing area distribution are 12, 12, 7, 4, 2, 4, 4 and 6, respectively.

Six machine learning models are designed to build heating and cooling loads such that a separate prediction model is built for each of the heating and cooling loads. These models are Elman neural network (ENN), linear regression (LR), K-nearest neighbor (KNN), Gaussian process regression (GP), Gradient boosted decision trees (GBT) and relevance vector machine (RVM). Their accuracies are analyzed assessed as per five performance metrics, namely mean absolute percentage error (MAPE), root-mean squared error (RMSE), mean absolute error (MAE), normalized absolute error (NAE) and root relative squared error (RRSE). As a result of the presence of numerous performance indicators, it is very hard to find a machine learning model which behaves efficiently with respect to the different performance metrics. Thus, this study applies Shannon entropy to objectively compute the relative importance of the attributes, which are the performance metrics in the present study. Then, Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is applied to sort the machine learning models based on the afore-mentioned performance metrics.

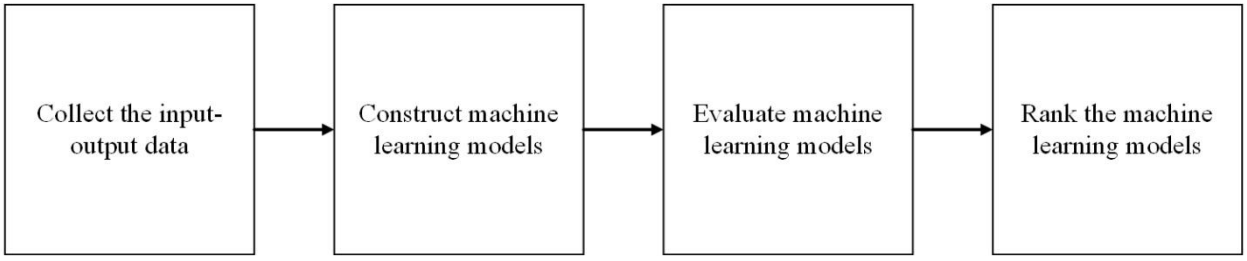

Figure 1: Framework of the proposed method

Model Development

This section describes the machine learning models and multi-criteria decision making techniques delineated in the previous section.

Elman Neural Network

Elman neural network, proposed by Elman [27], is a recurrent-based neural network with internal time-delay feedback [28]. The major advantages of recurrent neural networks are the prediction capabilities of time series and nonlinear data, fast convergence, and accurate mapping abilities [29]. The architecture of the Elman network is composed of an input layer, a hidden layer, an output layer, and a context layer. The connections among the input layer, the hidden layer, and the output layer can be considered as a multi-layer feed-forward neural network. The output values from the input layer and the context layer are used to activate the hidden layer. Then, the hidden layer is used to activate the output layer. Besides, the outputs from the hidden layer are used as inputs to the context layer. The network also supports several user-based options such as the training functions and the number of nodes in each layer. For illustration purposes, the Elman network is associated with feedback connections from the output layers to the input one. These connections take into consideration the current and previous state of neurons from the signals collected from the input layer. In other words, the context layer, addressed for information recording, is conducted for connecting the previous iterations with the next ones. ENN is trained using back propagation algorithm, based on the inputs and targets given to the network [28]. Besides, ENN is characterized by better learning efficiency, approximation ability, and memory ability [30]. Therefore, the Elman network is preferable for time series data forecasting and system identification and prediction [31]. On the other side, the major disadvantages of ENN are summarized as follows: low convergence rate, ease of being trapped at the local minimum, and lack of theory to determine the initial weights and threshold of the network [30].

Linear Regression

Regression analysis is applied to investigate and model the relationship between a set of dependent and independent variables [32]. Furthermore, it predicts the magnitude of the dependent variable(s) when the independent variable(s) are given [33]. There are two categories of regression analysis namely, single and multiple regression. Single linear/univariate regression analyzes the relationship between one dependent variable and one independent variable. On the other hand, multiple linear/multivariate regression is concerned with analyzing the relationship between one dependent variable and more than one independent variable [34]. The relationship between the two variables can be represented by a single line. When there are three and four variables, the relationship is a plane and a body, respectively. The major advantages of analyzing data using regression models are: a) determining the degree of influence of the predictor independent variable(s) on the dependent variable(s), b) identification of outliers and anomalies, c) fast modeling speed when the amount of data is large, and d) non-requirement of complicated calculations. On the other hand, the application of an incomplete or false data leads to an incorrect causal correlation between the dependent and independent variables. Besides, the non-linear data cannot be fitted using linear regression models [35].

K-nearest Neighbor

The k-nearest neighbor algorithm is one of the oldest and simplest non-parametric methods used for classification and regression problems [36]. Besides, it is considered as one of the top ten most influential data mining algorithms [37]. The KNN regression algorithm predicts the values of new data points based on the degree of closeness between test points and points in the training data sets. There are various methods for calculating the distance between training and test data, which are; Euclidian, Manhattan and Minkowski (for continuous variables) and Hamming distance (for categorical variables). Choosing the optimal value for K is essential to the implementation of this algorithm [38]. This is done by running the KNN algorithm several times with different K values and selecting the K value that yields the lowest error while maintaining the algorithm's ability to accurately make predictions [39]. The major advantages of KNN are simplicity and ease of implementation [40]. Besides, it yields competitive results without the need to tune several parameters and make additional assumptions. This versatile algorithm can be used for multiple purposes (i.e. classification, regression, and search) [39]. However, its major drawbacks involve slow learning and sensitivity to the local structure of the data [36].

Gaussian Process Regression

Gaussian process regression (GPR) is a nonparametric model that can be utilized in exploration and exploitation scenarios [41]. It could be used for both regression and classification tasks [42]. It has proven its prediction capabilities in cases where there are not a large number of samples [43]. It works by extending the idea of a probability distribution of numbers to a probability distribution of functions. The probability distribution of functions could be determined by mean function and covariance function. The covariance function in GPR is determined by a chosen kernel function that describes how much influence one point has on another. This effectively determines the smoothness of the function in the distribution. Given a set of function values, a probability distribution of functions could be fit to the given function values. Considering the whole fit distribution of functions, the mean and confidence interval could be determined. This results in obtaining not only the regression function but also the probabilistic bounds on the prediction [44]. The major advantages of GPR can be summarized as follows: a) it makes few assumptions about the shape of the estimator function and the choice of covariance function, b) it can be constructed to change the width of the local weighting functions separately for each known input dimension, and c) it returns both point predictions and confidence intervals around those predictions [45].

Gradient Boosted Decision Trees

Gradient boosting decision tree [46] is one of the most powerful techniques for building predictive models, due to its efficiency, accuracy, and interpretability [47]. It has been widely applied in classification, regression, and ranking problems [48]. It creates the model in a stage-wise fashion by building a series of trees where each tree is trained so that it attempts to correct the mistakes of the previous trees in the series. It typically uses lots of shallow trees (weak learners) to build a model that creates fewer mistakes as more trees are added. The number of estimators in GBDT is an important parameter in controlling the model's complexity. Besides, the learning rate controls how hard each new tree tries to correct mistakes from previous rounds. When the learning rate is high, each successive tree emphasizes correcting the mistakes of the predecessors and thus resulting in more complex trees and vice versa [49]. The major advantages of GBDT are summarized as follows: fast training and prediction time, high accuracy, and small memory footprint [48]. Besides, it handles a mixture of feature types and it does not require normalization of features to produce good results. However, GBDT has several downsides: a) it is very difficult for humans to interpret, b) it requires careful tuning of the model's parameters, and c) it is facing challenges when dealing with problems with high dimensional sparse features for accuracy, efficiency and computational problems [47].

Relevance Vector Machines

Relevance vector machine (RVM), introduced by Tipping [50], can be used as an alternative to support vector machine (SVM) for both regression and classification problems. In other words, RVM is divided into two categories, namely relevance vector regression (RVR) and relevance vector classification (RVC) [51]. RVM represents a Bayesian formulation of a linear model that results in a sparse representation than that achieved by SVM [52]. The major advantages of RVM over the SVM are: a) ability to make probabilistic predictions with fewer relevance vectors for a given dataset, b) reduced sensitivity to the hyper parameter settings, c) ability to use non-Mercer kernels, and d) non-requirement to define the error/margin trade-off parameter [53]. However, RVM has not found widespread application because of its slow training procedure [54].

Shannon Entropy

The concept of Shannon entropy is utilized in order to calculate the weights of the performance metrics. Shannon introduced the concept of information entropy in 1948, which can be defined as average amount of information. It assigns a smaller weight to the attribute if this attribute has similar values across the alternatives, i.e., if the measures of performance of the alternatives of a given attribute are relatively equal, therefore this attribute is considered as relatively unimportant by the decision maker.

The first stage is to calculate the Weight (P_{ij}) which is calculated using Equation (1).

$$P_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}} \quad (1 \le i \le m, 1 \le j \le n) \tag{1}$$

Where;

P_{ij} represents the weight of the i-th alternative with respect to j-th attribute. x_{ij} represents measure of performance of the i-th alternative with respect to j-th attribute.

The second stage involves computing the Entropy value and it is calculated using Equation (2).

$$e_j = -k * \sum_{j=1}^{n} P_{ij} \times \ln P_{ij} \quad (1 \le i \le m, 1 \le j \le n) \tag{2}$$

Where;

$$k = \frac{1}{\ln(m)} \tag{3}$$

Where;

e_j refers to the Entropy value of j-th attribute.

The third stage encompasses calculating variation coefficient for different attributes and this is done via Equation (4).

$$d_j = 1 - e_j \tag{4}$$

Where;

d_j represents variation coefficient of j-th attribute.

The fourth stage is to weight for each attribute and it is calculated using Equation (5).

$$w_j = \frac{d_j}{\sum_{j=1}^{n} d_j} \tag{5}$$

Where;

w_j represents weight of each attribute.

TOPSIS Technique

TOPSIS refers for Technique for Order Preference by Similarity to Ideal Solution. TOPSIS utilizes the Euclidean distances to compare between the alternatives using the positive and negative ideal solutions as a reference. TOPSIS decision making technique is divided into five main steps:

The decision matrix is normalized where the purpose of this step is to convert performance attributes into non-dimensional ones. The normalized decision matrix is computed using Equation (6).

$$r_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x^2_{ij}} \tag{6}$$

The weighted normalized matrix is generated using Equation (7).

$$v_{ij} = r_{ij} \times w_j \tag{7}$$

The ideal and negative ideal solutions are determined. A^* indicates the most preferable alternative or ideal solution. On the contrary, A^- indicates the least preferable alternative or negative ideal solution. For benefit criteria, decision maker wants to obtain the maximum value among all alternatives. On the other hand, the decision maker wants to obtain minimum value among all alternatives for cost criteria.

$$A* = \{(\max v_{ij} | j \in J), (\min v_{ij} | j \in J'), i = 1,2,3 \ldots . M\} = \{v*_1, v*_2 \ldots \ldots \ldots . . v*_N\} \quad (8)$$

$$A- = \{(minv_{ij} | j \in J), (maxv_{ij} | j \in J'), i = 1,2,3 \ldots . M\} = \{v-_1, v-_2 \ldots \ldots \ldots . . v-_N\} \quad (9)$$

Where;

$$J = \{, j = 1,2,3, \ldots \ldots \ldots . N | jassociatedwithbenefitcriteria\} \quad (10)$$

$$J' = \{, j = 1,2,3, \ldots \ldots \ldots . N | jassociatedwithcostcriteria\} \quad (11)$$

The fourth step constitutes computing the separation distance of each alternative to the ideal and negative ideal solutions. $s*_i$ represents the separation distance of each alternative in the Euclidean way from the ideal solution. On the contrary, $s-_i$ represents the separation distance of each alternative in the Euclidean way from the negative ideal solution.

$$s*_i = (\sum_{j=1}^{n} (v_{ij} - v*_j))^{\frac{1}{2}} \quad (12)$$

$$s-_i = (\sum_{j=1}^{n} (v_{ij} - v-_j))^{\frac{1}{2}} \quad (13)$$

The fifth step involves calculating the relative closeness of an alternative A_i to the ideal solution $A*$. The relative closeness is calculated using Equation (14). The larger $c*_i$ the closer to the ideal solution. Alternatives are ranked in descending order.

$$c*_i = \frac{s-_i}{s*_i + s-_i} \quad (14)$$

Performance Metrics

The present study adopts five performance indicators to compare between the six machine learning models. The three performance indicators are: mean absolute percentage error, root-mean squared error, mean absolute error, normalized absolute error and root relative squared error. MAPE, RMSE, MAE, NAE and RRSE can be computed using Equations (15), (16), (17), (18) and (19), respectively [55-57].

$$RMSE = \sqrt{\frac{1}{K}\sum_{i=1}^{K} (O_i - P_i)^2} \quad (15)$$

$$MAE = \frac{1}{K}\sum_{i=1}^{K} |(O_i - P_i)| \quad (16)$$

$$MAPE = \frac{100}{k} \times \sum_{i=1}^{K} \frac{|P_i - O_i|}{O_i} \quad (17)$$

$$NAE = \sum_{i=1}^{K} \frac{|(O_i - P_i)|}{O_i} \quad (18)$$

$$RRSE = \sum_{i=1}^{K} \frac{(O_i - P_i)^2}{(O_i - O^{\wedge})^2} \quad (18)$$

Where;
O_i and P_i stand for the observed and predicted heating or cooling loads, respectively. O^ denotes the average of the actual values. K indicates number of observations.

Model Implementation

The dataset is composed of 768 instances, such that614 data points are used for training purpose while the remaining 154observations are used for testing purposes. A sample of the data set required to build the heating and coolingloads prediction models is presented in Table 1. The terms "X1", "X2", "X3" "X4", "X5", "X6", "X7" and "X8" denoterelative compactness, surface area, wall area, roof area, overall height, orientation, glazing area and glazing area distribution, respectively. The terms "Y1" and "Y2 stand for the heating load and cooling load, respectively.

Table 1: Sample of data set for the prediction of heating and cooling loads [26]

X1	X2	X3	X4	X5	X6	X7	X8	Y1	Y2
0.76	661.50	416.50	122.50	7.00	2	0.40	3	39.32	38.17
0.76	661.50	416.50	122.50	7.00	3	0.40	3	39.84	38.48
0.76	661.50	416.50	122.50	7.00	4	0.40	3	38.89	39.66
0.76	661.50	416.50	122.50	7.00	5	0.40	3	39.68	40.10
0.71	710.50	269.50	220.50	3.50	2	0.40	3	14.07	16.11
0.71	710.50	269.50	220.50	3.50	3	0.40	3	14.03	15.47
0.71	710.50	269.50	220.50	3.50	4	0.40	3	13.94	16.70
0.71	710.50	269.50	220.50	3.50	5	0.40	3	13.86	16.10
0.66	759.50	318.50	220.50	3.50	2	0.40	3	15.16	17.04
0.66	759.50	318.50	220.50	3.50	3	0.40	3	15.18	17.63
0.66	759.50	318.50	220.50	3.50	4	0.40	3	14.72	18.10
0.66	759.50	318.50	220.50	3.50	5	0.40	3	14.90	18.22
0.86	588.00	294.00	147.00	7.00	2	0.40	4	32.38	31.53
0.86	588.00	294.00	147.00	7.00	3	0.40	4	31.66	36.20
0.86	588.00	294.00	147.00	7.00	4	0.40	4	32.15	36.21
0.86	588.00	294.00	147.00	7.00	5	0.40	4	32.75	31.00

As stated earlier, six machine learning models are developed to predict the heating and cooling loads. For the Elman neural network, the number of hidden and context layers are two while the number of hidden and context neurons are four. In the K-nearest neighbor algorithm, the number of neighbors is assumed two. For the Gaussian process regression, radial basis function is the kernel function and the kernel length scale is three. In the gradient boosted decision trees, the number of trees is assumed 100. The maximum depth, learning rate and sample rate are 10, 0.01 and 1, respectively. In the relevance vector machines, the kernel function is radial basis, and the kernel length scale is three. An illustration of the performances of the K-nearest neighbor, gradient boosted decision trees and relevance vector machines in the prediction of heating loads are depicted in Figures 2, 3 and 4, respectively.

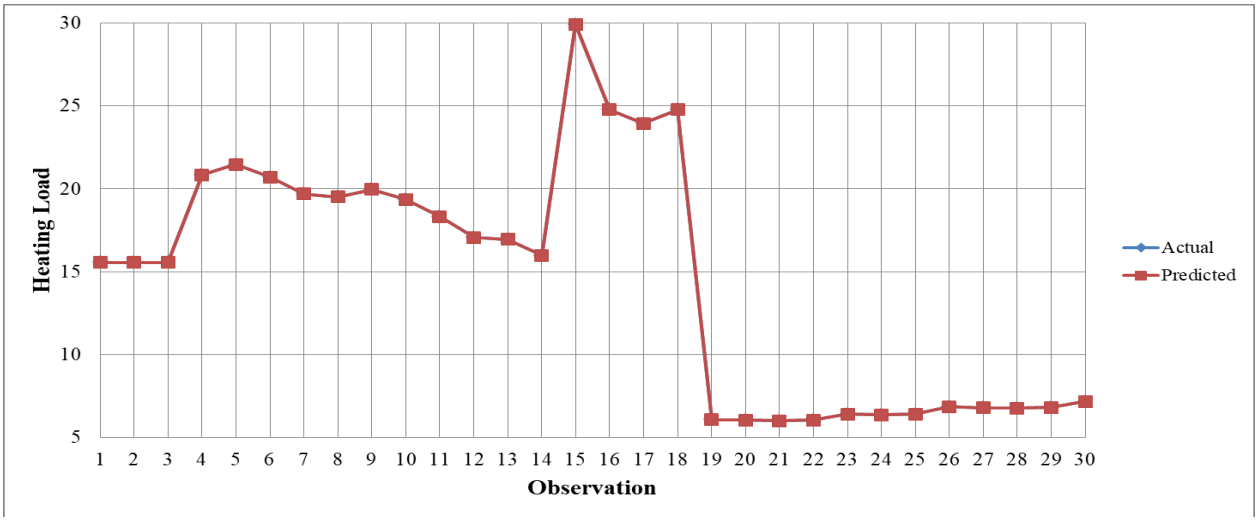

Figure 2: Actual and predicted heating loads using K-nearest neighbor

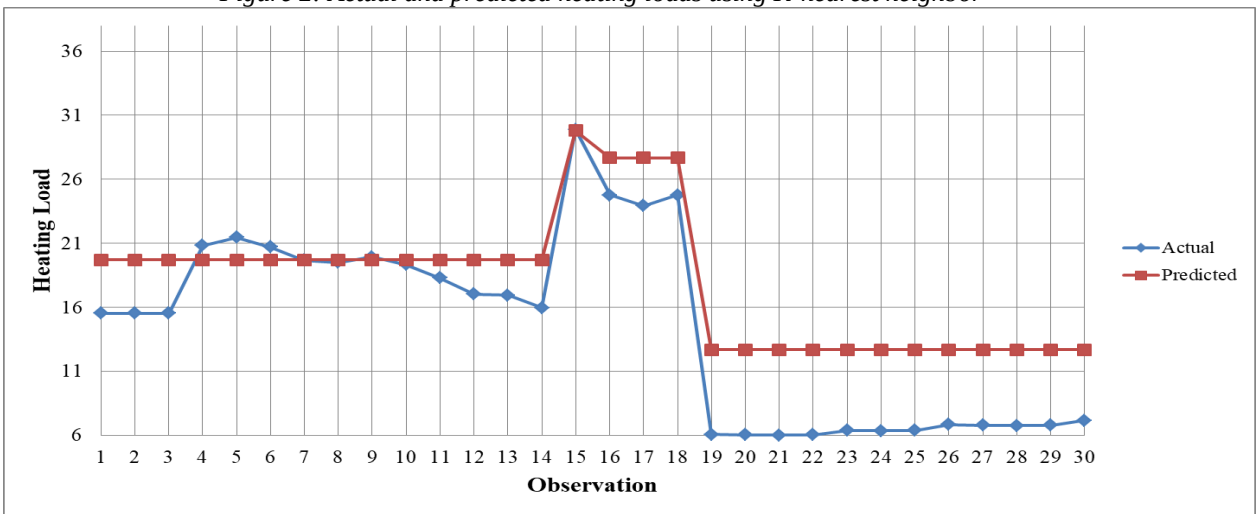

Figure 3: Actual and predicted heating loads using gradient boosted decision trees

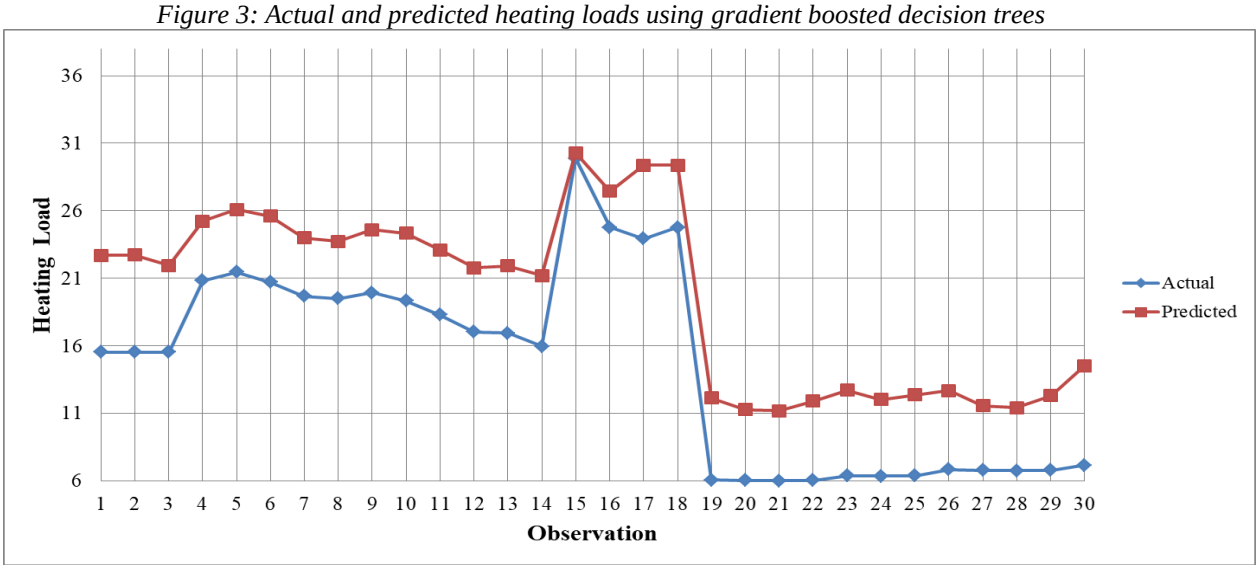

Figure 4: Actual and predicted heating loads using relevance vector machines

As shown in the previous figures, K-nearest neighbor succeeded in simulating the heating loads while gradient boosted decision trees and relevance vector machines failed to model the heating loads. Samples of the predicted values of the heating and cooling loads based on the six machine learning models are presented in Tables 2 and 3, respectively.

Table 2: Predicted heating loads using the six machine learning models

ID	ENN	LR	KNN	GP	GBT	RVM
1	18.33	27.76	15.55	16.79	21.29	22.71
2	18.08	27.76	15.55	16.94	22.16	22.73
3	17.78	29.07	15.55	17.09	22.30	21.95
4	17.48	29.07	20.84	23.25	27.29	25.24
5	21.83	29.07	21.46	22.31	26.16	26.10
6	21.52	33.66	20.71	22.05	26.64	25.60
7	21.16	33.66	19.68	22.86	29.51	24.00
8	20.83	33.66	19.5	21.90	25.41	23.72
9	19.37	7.77	19.95	22.73	24.02	24.59
10	19.17	7.77	19.34	20.61	24.71	24.35
11	18.89	7.77	18.31	16.35	27.78	23.12
12	18.58	9.08	17.05	19.86	22.21	21.79
13	22.75	9.08	16.95	18.75	22.47	21.93
14	22.46	9.08	15.98	16.77	24.57	21.20
15	22.11	9.72	29.9	32.52	30.54	30.29
16	21.76	9.72	24.77	23.16	28.65	27.45
17	25.60	9.72	23.93	25.06	32.51	29.37
18	25.23	9.72	24.77	26.69	32.73	29.36
19	24.84	11.03	6.07	6.46	10.10	12.15
20	24.48	11.03	6.05	6.98	10.93	11.30
21	29.76	11.03	6.01	8.27	11.76	11.20
22	29.24	11.68	6.04	16.79	11.99	11.91

Table 3: Predicted cooling loads using the six machine learning models

ID	ENN	LR	KNN	GP	GBT	RVM
1	29.52	49.38	21.33	21.29	24.46	25.17
2	30.06	106.82	21.33	22.16	24.46	25.32
3	30.91	164.26	21.33	22.30	24.46	25.18
4	32.03	26.20	28.28	27.29	24.72	30.39
5	39.54	31.24	25.38	26.16	24.72	31.32
6	39.68	88.69	25.16	26.64	24.72	31.07
7	39.29	146.13	29.6	29.51	25.00	29.76
8	37.58	52.22	27.3	25.41	24.69	25.74
9	28.20	5.22	21.97	24.02	24.69	26.42
10	24.80	62.66	23.49	24.71	24.69	26.70
11	20.48	120.11	27.87	27.78	24.97	26.44
12	16.94	55.23	23.77	22.21	24.51	24.93
13	24.00	59.66	21.16	22.47	24.51	25.12
14	20.49	117.10	24.93	24.57	24.51	24.84
15	17.60	0.00	31.27	30.54	32.44	27.60
16	15.61	36.69	29.79	28.65	29.86	25.13
17	18.33	20.75	29.68	32.51	29.86	25.51
18	17.17	78.20	29.79	32.73	29.86	25.61
19	16.04	46.73	10.9	10.10	16.31	21.85
20	15.01	10.71	11.19	10.93	16.31	21.53
21	20.41	68.15	10.94	11.76	16.31	21.43
22	18.36	125.60	11.17	11.99	16.31	21.61

A performance comparison between the different machine learning models is described in Table 4 and 5, respectively. The comparison is conducted as per split validation, whereas 80% of the dataset are used for training and the remainder of the dataset is used for testing. As shown in Table 4, K-nearest neighbor achieved the best performance, whereas it achieved MAPE, RMSE, MAE, NAE and RRSE of 2.329%, 1.332, 0.479, 0.023 and 0.123, respectively. Multiple linear regression provided the least prediction accuracies, such that it achieved MAPE, RMSE, MAE, NAE and RRSE of 23.654%, 6.914, 3.411, 0.237 and 0.686, respectively. With respect to the cooling loads, K-nearest neighbor yielded the best performance, whereas it achieved MAPE, RMSE, MAE, NAE and RRSE of 1.408%, 0.97, 0.34, 0.014 and 0.022, respectively. Multiple linear regression yielded the least prediction accuracies such that it achieved MAPE, RMSE, MAE, NAE and RRSE of 26.169%, 7.506, 3.701, 0.262 and 0.744, respectively.

Table 4: Performance comparison of the six machine learning models for predicting heating loads

Model	MAPE	RMSE	MAE	NAE	RRSE
ENN	7.303%	2.511	1.670	0.073	0.249
LR	23.654%	6.914	3.411	0.237	0.686
KNN	2.329%	1.332	0.479	0.023	0.132
GP	8.622%	1.993	1.617	0.086	0.198
GBT	18.989%	3.761	3.386	0.190	0.373
RVM	12.677%	3.048	2.382	0.127	0.302

Table 5: Performance comparison of the six machine learning models for predicting cooling loads

Model	MAPE	RMSE	MAE	NAE	RRSE
ENN	8.256%	2.314	1.682	0.083	0.229
LR	26.169%	7.506	3.701	0.262	0.744
KNN	1.408%	0.970	0.340	0.014	0.022
GP	6.273%	1.923	1.488	0.063	0.202
GBT	14.976%	3.829	3.250	0.150	0.403
RVM	23.724%	6.160	5.152	0.237	0.648

Shannon entropy is applied to compute the weights of the performance indicators. The weights of attributes are illustrated in Table 6. It is revealed thatNAE had the largest weight followed by MAPE while RMSE had the lowest weight among the attributes. The rankings of the machine learning models are displayed in Table 7. As can be seen, KNN achieved the highest relative closeness of 99.189 while regression analysis provided the lowest relative closeness of 22.869. As such, K-nearest neighbor yielded the highest ranking among the six machine learning models in the prediction of heating and cooling loads followed by Gaussian process regression and then Elman neural network.

Table 6: Entropy value, variation coefficient and weight of the performance indicators

Terms	MAPE	RMSE	MAE	NAE	RRSE
e_j	0.911	0.931	0.928	0.911	0.916
d_j	0.089	0.069	0.072	0.089	0.084
w_j	22.062%	17.116%	17.893%	22.098%	20.832%

Table 7: Ranking of the machine learning models based on TOPSIS

Terms	$s*_i$	$s-_i$	$c*_i$	Ranking
ENN	0.004	0.036	90.606	2
LR	0.034	0.010	22.869	6
KNN	0	0.040	99.189	1
GP	0.003	0.037	92.339	2
GBT	0.015	0.026	63.679	4
RVM	0.021	0.020	50.095	5

Conclusion

The magnitude of heating and cooling energy consumption in the building sector is significant. In this regard, it is decisive to build an automated machine learning model for mimicking the energy consumption in buildings, which results in highly energy-efficient buildings. This research introduces a holistic hybrid entropy-TOPSIS-based method for sorting machine learning models in the energy-efficient buildings. The present study analyzes the behavior of six types of machine learning models, namely Elman neural network, linear regression, K-nearest neighbor, Gaussian process regression, gradient boosted decision trees and relevance vector machine. Their prediction capacities are investigated relying on five performance indicators namely,mean absolute percentage error, root-mean squared error, mean absolute error, normalized absolute error and root relative squared error. Machine learning models behave differently with respect to the performance indicators. Thus, it is essential to design a holistic method for ranking them based on the different metrics. The present study utilizes Shannon entropy algorithm to derive the weighting vector of the attributes, which resembles the performance criteria in the present study then TOPSIS was used to rank the machine learning models. Normalized absolute error had the highest importance weighting of 22.098% while root mean-squared error had the least importance weighting of 17.116%. It was concluded that K-nearest neighbor ($c*_i = 99.189$) achieved the highest ranking followed by Gaussian process ($c*_i = 92.339$) while the multiple linear regression analysis ($c*_i = 22.869$) attained the least ranking across the different machine learning models.

References

[1]. International Energy Agency (IEA), (2015). Key World Energy Statistics, International Energy Agency: Paris, France.

[2]. McQuiston, F. C., & Parker, J. D. (1982). Heating, Ventilating, and Air Conditioning: Analysis and Design, John Wiley & Sons Inc.: New York, USA.

[3]. Castaldo, V. L., Pigliautile, I., Rosso, F., Cotana, F., De Giorgio, F., & P isello, A. L. (2018). How Subjective and Non-Physical Parameters Affect Occupants' Environmental Comfort Perception, *Energy and Buildings*, Vol. 178, pp. 107-129.

[4]. Bui, D. T., Moayedi, H., Anastasios, D., & Foong, L. K. (2019). Predicting Heating and Cooling Loads in Energy-Efficient Buildings Using Two Hybrid Intelligent Models, *Applied Sciences*, Vol. 9, No. 17, pp. 3543.

[5]. Cao, X., Dai, X., & Liu, J. (2016). Building Energy Consumption Status Worldwide and the State-of-the-Art Technologies for Zero Energy Buildings during the Past Decade, *Energy and Buildings*, Vol. 128, pp. 198-213.

[6]. Egan, J., Finn, D., Soares, P. H., Baumann, V. A., Aghamolaei, R., Beagon, P., Neu, O., Pallonetto, F., & O'Donnell, J. (2018). Definition of a Useful Minimal-Set of Accurately Specified Input Data for Building Energy Performance Simulation, *Energy and Buildings*, Vol. 165, pp. 172-183.

[7]. Fanti, M. P., Mangini, A. M., & Roccotelli, M. (2018). A Simulation and Control Model for Building Energy Management, *Control Engineering Practice*, Vol. 72, pp. 192-205.

[8]. Zhao, H. X., & Magoulès, F. (2012). A Review on the Prediction of Building Energy Consumption, *Renewable and Sustainable Energy Reviews*, Vol. 16, No. 6, pp. 3586-3592.

[9]. Park, J., Lee, S. J., Kim, K. H., Kwon, K. W., & Jeong, J. W. (2016). Estimating Thermal Performance and Energy Saving Potential of Residential Buildings Using Utility Bills, *Energy and Buildings*, Vol. 110, pp. 23-30.

[10]. Yezioro, A., Dong, B., & Leite, F. (2008). An Applied Artificial Intelligence Approach towards Assessing Building Performance Simulation Tools, *Energy and Buildings*, Vol. 40, No. 4, pp. 612-620.

[11]. O'Neill, Z., & O'Neill, C. (2016). Development of a Probabilistic Graphical Model for Predicting Building Energy Performance, *Applied Energy*, Vol. 164, pp. 650-658.

[12]. Catalina, T., Virgone, J., & Blanco, E. (2008). Development and Validation of Regression Models to Predict Monthly Heating Demand for Residential Buildings, *Energy and Buildings*, Vol. 40, No. 10, pp. 1825-1832.

[13]. Yu, Z., Haghighat, F., Fung, B. C., & Yoshino, H. (2010). A Decision Tree Method for Building

Energy Demand Modeling, *Energy and Buildings*, Vol. 42, No. 10, pp. 1637-1646.
[14]. Poel, B., Van Cruchten, G., & Balaras, C. A. (2007). Energy Performance Assessment of Existing Dwellings, *Energy and Buildings*, Vol. 39, No. 4, pp. 393-403.
[15]. Fahmy, M., Mahdy, M. M., & Nikolopoulou, M. (2014). Prediction of Future Energy Consumption Reduction Using GRC Envelope Optimization for Residential Buildings in Egypt, *Energy and Buildings*, Vol. 70, pp. 186-193.
[16]. Cuadrado, J., Zubizarreta, M., Pelaz, B., & Marcos, I. (2015). Methodology to Assess the Environmental Sustainability of Timber Structures, *Construction and Building Materials*, Vol. 86, pp. 149-158.
[17]. Mousa, M., Luo, X., & McCabe, B. (2016). Utilizing BIM and Carbon Estimating Methods for Meaningful Data Representation, *Procedia Engineering*, Vol. 145, pp. 1242-1249.
[18]. Marzouk, M., & Mohammed Abdelkader, E. (2019). A Hybrid Fuzzy-Optimization Method for Modeling Construction Emissions, *Decision Science Letters*, Vol. 9, No. 1, pp. 1-20.
[19]. Jihad, A. S., &Tahiri, M. (2018). Forecasting the Heating and Cooling Load of Residential Buildings by Using a Learning Algorithm "Gradient Descent", Morocco, *Case Studies in Thermal Engineering*, Vol. 12, pp. 85-93.
[20]. Kavaklioglu, K. (2018). Robust Modeling of Heating and Cooling Loads Using Partial Least Squares Towards Efficient Residential Building Design, *Journal of Building Engineering*, Vol. 18, pp. 467-475.
[21]. Srihari, J., &Santhi, B. (2018). Prediction of Heating and Cooling Load to Improve Energy Efficiency of Buildings Using Machine Learning Techniques, *Journal of Mechanics of Continua and Mathematical Sciences*, Vol. 13, No. 5, pp. 97-113.
[22]. Moayedi, H., Bui, D. T., Dounis, A., Lyu, Z., & Foong, L. K. (2019). Predicting Heating Load in Energy-Efficient Buildings through Machine Learning Techniques, *Applied Sciences*, Vol. 9, No. 20, pp. 4338.
[23]. Mohammed Abdelkader, E., Al-Sakkaf, A., & Ahmed, R. (2020). A Comprehensive Comparative Analysis of Machine Learning Models for Predicting Heating and Cooling Loads, *Decision Science Letters*, Vol. 9, pp. 1-12.
[24]. Roy, S. S., Samui, P., Nagtode, I., Jain, H., Shivaramakrishnan, V., & Mohammadi-Ivatloo, B. (2019). Forecasting Heating and Cooling Loads of Buildings: A Comparative Performance Analysis, *Journal of Ambient Intelligence and Humanized Computing*, Vol. 11, No. 3, pp. 1253-1264.
[25]. Asuncion, A. &Newman, D. J. (2007). UCI Machine Learning Repository. *University of California, Irvine*. Retrieved from https://archive.ics.uci.edu/ml/index.php.
[26]. Tsanas, A. & Xifara, A. (2012).Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings*, Vol. 49, pp. 560-567.
[27]. Elman, J. (1990). Finding Structure in Time, *Cognitive Science*, Vol. 14, No. 2, pp. 179-211.
[28]. Ren, G., Cao, Y., Wen, S., Huang, T., & Zeng, Z. (2018). A Modified Elman Neural Network with a New Learning Rate Scheme, *Neurocomputing*, Vol. 286, pp. 11-18.
[29]. Wang, J., Wang, J., Fang, W., & Niu, H. (2016). Financial Time Series Prediction Using Elman Recurrent Random Neural Networks, *Computational Intelligence and Neuroscience*, Vol. 2016, pp.14.
[30]. Xu, L., & Zhang, Y. (2019). Quality Prediction Model Based on Novel Elman Neural Network Ensemble, *Complexity*, Vol. 2019, pp. 11.
[31]. Li, C., Zhu, L., He, Z., Gao, H., Yang, Y., Yao, D., & Qu, X. (2019). Runoff Prediction Method Based on Adaptive Elman Neural Network, *Water*, Vol. 11, No. 6, pp. 1113.
[32]. Montgomery, D., Peck, E., & Vinning, G. (2012). Introduction to Linear Regression Analysis (5th edition). John Wiley & Sons: New York, USA.
[33]. Quraishi, M., & Mouazen, A. (2013). Development of a Methodology for In Situ Assessment of Topsoil Dry Bulk Density, *Soil and Tillage Research*, Vol. 126, pp. 229-237.
[34]. Uyanik, G., & Guller, N. (2013). A Study on Multiple Linear Regression Analysis, *Procedia - Social and Behavioral Sciences*, Vol. 106, pp. 234-240.
[35]. Weedmark, D. (2018). The Advantages & Disadvantages of a Multiple Regression Model. Accessed

April 17, 2020. https://sciencing.com/advantages-disadvantages-multiple-regression-model-12070171.html.

[36]. Novakovic, J., Veljovic, A., Ilic, S., & Papic, M. (2016). Experimental Study of Using the K-Nearest Neighbour Classifier with Filter Methods, Proceedings of the Computer Science and Technology Conference, Varna, Bulgaria.

[37]. Wu, X., Kumar, V., Quinlan, R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z. H., Steinbach, M., Hand, D., & Steinberg, D. (2007). Top 10 Algorithms in Data Mining, *Knowledge and Information Systems*, Vol. 14, No. 1, pp. 1-37.

[38]. Singh, A. (2018). A Practical Introduction to K-Nearest Neighbors Algorithm for Regression (with Python code). Accessed April 17, 2020. https://www.analyticsvidhya.com/blog/2018/08/k-nearest-neighbor-introduction-regression-python/.

[39]. Harrison, O. (2018). Machine Learning Basics with the K-Nearest Neighbors Algorithm. Accessed April 17, 2020. https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761.

[40]. Dubey, H. (2013). Efficient and Accurate KNN based Classification and Regression. MSc Thesis, International Institute of Information Technology Hyderabad, India.

[41]. Schulz, E., Speekenbrink, M., & Krause, A. (2018). A Tutorial on Gaussian Process Regression: Modelling, Exploring, and Exploiting Functions, *Journal of Mathematical Psychology*, Vol. 85, pp. 1-16.

[42]. Kapoor, A., Grauman, K., Urtasun, R., & Darrell, T. (2010). Gaussian Processes for Object Categorization, *International Journal of Computer Vision*, Vol. 88, No. 2, pp. 169-188.

[43]. Lee, K., Cho, H., & Lee, I. (2019). Variable Selection Using Gaussian Process Regression-Based Metrics for High-Dimensional Model Approximation with Limited Data, *Structural and Multidisciplinary Optimization*, Vol. 59, pp.1439-1454.

[44]. Görtler, J., Kehlbeck, R., & Deussen, O. (2019). A Visual Exploration of Gaussian Processes. Accessed April 2, 2020. https://distill.pub/2019/visual-exploration-gaussian-processes/.

[45]. Caywood, M., Roberts, D., Colombe, J., Greenwald, H., & Weiland, M. (2017). Gaussian Process Regression for Predictive but Interpretable Machine Learning Models: An Example of Predicting Mental Workload across Tasks, *Frontiers in Human Neuroscience*, Vol. 10, pp. 647.

[46]. Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine, *Annals of statistics*, Vol. 29, No. 5, pp. 1189-1232.

[47]. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). Light GBM: A Highly Efficient Gradient Boosting Decision Tree, Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), California, USA.

[48]. Si, S., Zhang, H., Keerthi, S., Mahajan, D., Dhillon, I., & Hsieh, C. (2017). Gradient Boosted Decision Trees for High Dimensional Sparse Output, Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia.

[49]. Brownlee, J. (2016). A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning. Accessed April 18, 2020. https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/.

[50]. Tipping, M. 2001. Sparse Bayesian Learning and the Relevance Vector Machine, *Journal of Machine Learning Research*, Vol. 1, pp. 211-244.

[51]. Liu, L., Wang, J., & Liu, R. (2017). Infrared Small Target Detection Based on Relevance Vector Regression, *Procedia Computer Science*, Vol. 111, pp. 315-322.

[52]. Caesarendra, W., Widodo, A., & Yang, B. S. (2010). Application of Relevance Vector Machine and Logistic Regression for Machine Degradation Assessment, *Mechanical Systems and Signal Processing*, Vol. 24, pp. 1161-1171.

[53]. Pal, M. (2008). Support Vector Machines/Relevance Vector Machine for Remote Sensing Classification: A Review, Proceedings of the Brainstorming Workshop on Application of Advanced Soft Computing Techniques in Geospatial Data Analysis, Bombay, India.

[54]. Dong, C., & Tian, L. (2012). Accelerating Relevance-Vector-Machine-Based Classification of Hyperspectral Image with Parallel Computing, *Mathematical Problems in Engineering*, Vol. 2012, pp. 13.

[55]. Mohammed Abdelkader, E., Zayed, T., & Marzouk, M.: A Computerized Hybrid Bayesian-Based Approach for Modeling the Deterioration of Concrete Bridge Decks, *Structure and Infrastructure Engineering*, Vol. 25, No. 19, pp. 1178-1199.

[56]. Nazari, A., Rajeev, P., and Sanjayan, J. G. (2015). Offshore pipeline performance evaluation by different artificial neural networks approaches. *Measurement*, Vol.76, pp.117–128.

[57]. Ranjith, S., Setunge, S., Gravina, R., & Venkatesan, S. (2013). Deterioration Prediction of Timber Bridge Elements Using the Markov Chain. *Journal of Performance of Constructed Facilities*, Vol. 27, No. , pp. 319–325.