International Journal of Engineering & Technology, 7 (4.35) (2018) 755-759

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Estimation of Building Energy Efficiency Performance Using Radial Basis Function Neural Network

Che Munira Che Razali *, Shamsul Faisal Mohd Hussein [2], Nolia Harudin[3], Shahrum Shah Abdullah[4]

[1]*Malaysia-Japan International Institute of Technology, UTM*
[2] *Malaysia-Japan International Institute of Technology, UTM*
[3] *Razak Faculty of Technology and Informatic, UTM*
[4] *Malaysia-Japan International Institute of Technology, UTM*
**Corresponding author E-mail: munira2484@gmail.com*

Abstract

Since a pass few decades up to recent, building energy efficiency performance is the top priority due to the sustainability of energy and quality of life. According to recent study related to computer experiment, there are various types of the model has been proposed by the researcher to improve the performance of building energy efficiency. However, there is no empirical evidence to prove the best method in prediction and estimation of energy efficiency that ensure adequate energy to meet todays and future needs. The objective of this paper is to propose Radial Basis Function Neural Network (RBFNN) for estimating the heating load and cooling load of a residential building. This study set out to evaluate different estimation methods of residential building energy efficiency using RBFNN. The data of residential building are obtained from UCI Machine Learning Repository. The dataset of simulation using Ecotect consists of 768 samples with 8 input features and 2 output variables were used to train and test the algorithm of RBFNN. The input variables involved in this experiment are relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area, and glazing area distribution of a building, while the output variables are heating and cooling loads of the building. The analytical result of the proposed method shows that RBFNN produces better result and performance compared with the previous researches.

Keywords: *building energy; radial basis function neural network, estimation, energy consumption*

1. Introduction

Nowadays, the research related to building energy consumption is increasing due to the growing concern about energy efficiency. Based on the World Energy Markets Observatory (WEMO) 2017 report, Malaysia's energy usage is projected to increase up to 4.8% by the year 2030 [1]. Heating load (HL) and cooling load (CL) contribute around 30-40% of building's energy usage. Minimizing the HL and CL plays a major role in ensuring total energy consumption reduction in buildings [2]. Previous research in building energy based on computer experiment is done to optimize parameters for energy efficiency improvement in the residential or commercial buildings.

Recent studies in computer simulation of energy consumption have been conducted using random forest algorithm (RF) [3], support vector machine (SVM), polynomial regression, decision trees, artificial neural network (ANN) [2], genetic programming (GP) [4], support vector regression (SVR) with ANN [5] and multilayer perceptron (MLP) with grid search strategy algorithm [6]. Major study in energy building efficiency by T.Sanas and Xifara (2012) identified that RF hugely outperformed IRLS in finding an accurate functional relationship between the input and output variable. In contrast to earlier findings, M.Y. Cheng (2014) [7] significantly improved the work of T.Sanas and Xifara (2012) in terms of assessment accuracies by predicting both the HL and CL for residential buildings by using evolutionary multivariate adaptive regression splines (EMARS). In the present study, G. Regina (2016) improves the HL and CL predictions using MLP hybrid with grid search strategy [6]. Therefore, the computer experiment method of prediction is a never-ending process and always be improved by researchers from time to time.

According to several studies, B. Yildiz [8] summarized that there are 4 types of regression analysis for prediction of electricity and building energy consumption: (1) ANN with Levenberg Marquardt (LM) and Bayesian regulation (BR) backpropagation; (2) nonlinear autoregressive network with exogenous inputs (NARX) with LM and BR backpropagation; (3) regression trees (RT); and (4) support vector regression (SVR). Along the same line, M. Castelli [4] proved that the overall performance of ANN models relied on the architecture of the neural network that has been designed, which is usually carried out in an ad-hoc and manual way. Furthermore, in his interesting energy consumption analysis, K. Amber et al [9] proposed multiple regression to predict energy consumption in university sector building. In 2014, Y. Zhang [10] demonstrated particle swam optimization (PSO) hybrid with Radial Basis Function Neural Network (RBFNN) to predict the actual energy consumption in building operation to serve the data analysis module of their developed real-time energy monitoring system. In his study, the purpose of using PSO is to optimize the center of basis function and hidden layer nodes in RBF. To further investigate the issue of energy consumption, C. Li [11] proposed new approach, which is ELM with stacked autoencoders (SAEs) to extract the building energy consumption features and predict better result for building performance. Based on comprehensive literature studies that have been done in this study, it is found that there is no published work on

(1)

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

building energy efficiency performance estimation and prediction was conducted by the previous researcher using RBFNN so far. Hence, the specific objective of this paper to propose the RBFNN to predict the HL and CL and improve the estimation of residential building energy efficiency performance. The dataset was first demonstrated experimentally by Tsanas and Xifara [3]. In an analysis using RBFNN and Random Forest, H. Sug [12] found that RBFNN has effective training time which is relatively shorter than other neural network algorithms. Another a good point of using RBFNN is good prediction accuracy with small-sized datasets, which is also true for other types of neural networks. A. Behera et al [13] argued that RBFNN provided important features such as: (1) fast learning; (2) capable for complex linear mapping, and (3) one of the competitive algorithm to solve approximation and classification problems.

2. Methodology

2.1 Radial Basis Function Neural Network

RBFNN is one of the powerful algorithms in ANN has three layers of feed forward fully connected network, which uses RBFNN as the only nonlinearity in the hidden layer neurons. The output layer of RBFNN has no nonlinearity and the connections of the output layer are only weighted. The connections from the input to the hidden layer are not weighted. In addition, RBFNN also has better approximation capabilities, simpler network architecture, and faster learning algorithm[13].

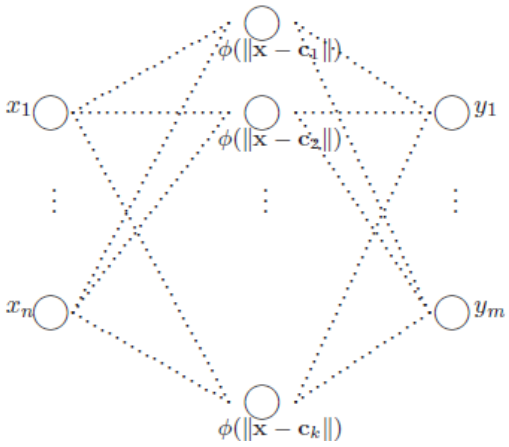


Fig 1:RBF Architecture [14]

The RBFNN model is then expressed as a linear combination of the basis function across all the training points $X_i \in R^n$, i = 1, …, n as given by equation (1).

$$\hat{y}(x) = \sum_{i=1}^{n} w_i \, \emptyset_i(\|x - c_k\|_2) \quad \textbf{(2)}$$

2.1.1 Parameters of RBFNN

In Equation (1), w_i are the generic weights of the basis function. The weight is evaluated by training points x_i and corresponding function values $f(x_i)$. The matrix of the basis function values at the training points is described in equation (2).

$$\emptyset = \begin{bmatrix} \emptyset(\|X_1 - C_1\|) & \emptyset(\|X_1 - C_2\|) & \cdots & \emptyset(\|X_1 - C_k\|) \\ \emptyset(\|X_2 - C_1\|) & \emptyset(\|X_2 - C_2\|) & \cdots & \emptyset(\|X_2 - C_k\|) \\ \vdots & \vdots & \ddots & \vdots \\ \emptyset(\|X_n - C_1\|) & \emptyset(\|X_n - C_2\|) & \cdots & \emptyset(\|X_n - C_k\|) \end{bmatrix}$$

The vector of weight is W and is described in equation (3).

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \quad \textbf{(3)}$$

All the input data are represented by the input matrix X and the output data are represented by the output vector y as below:

$$X = \begin{bmatrix} X^1 \\ X^2 \\ \vdots \\ X^n \end{bmatrix} = \begin{bmatrix} X_1^1 & X_2^1 & \ldots & X_n^1 \\ X_1^2 & X_2^2 & \ldots & X_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ X_1^m & X_1^m & \ldots & X_n^m \end{bmatrix} \; Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \textbf{(4)}$$

Where m is the number of inputs to the RBFNN and n is the number of samples. While target output data are given by the following equation:

$$h = \begin{bmatrix} h_1^1 & h_1^2 & \ldots & h_1^Q \\ h_2^1 & h_2^2 & \ldots & h_2^Q \\ \vdots & \vdots & \ddots & \vdots \\ h_n^1 & h_n^2 & \ldots & h_n^Q \end{bmatrix} \quad \textbf{(5)}$$

In equation (5), Q is the number of outputs and m is the number of data samples. RBFNN also consists of two parameters, which are the spread parameter is scalar while the center parameter is a matrix given by:

$$center = \begin{bmatrix} C_1^1 & C_2^1 & \ldots & XC_s^1 \\ C_1^2 & C_1^2 & \ldots & C_s^2 \\ \vdots & \vdots & \ddots & \vdots \\ C_1^m & C_2^m & \ldots & C_s^m \end{bmatrix} \quad \textbf{(6)}$$

S is the number of RBF center and m is the number of input. Each neuron in the hidden layer consists of an RBFNN centered at a point, depending on the dimensionality of the input or output predictor variables[13]. In comprehensive literature survey of RBFNN, Sanjeev et al also state that the activation functions in hidden layer are given by the basis functions $\emptyset_i$which is depend on parameter center and spread while the input x_i is non-standard manner [13]. Table 1 shows the input activation basis function for RBFNN. In this paper, the Gaussian activation function is used, which is given in Equation (7).

$$\emptyset(\|X - C_k\|) = e^{\frac{-\|X - c_k\|^2}{\beta^2}} \quad \textbf{(7)}$$

In equation (7), x is the input, C_k is the basis function centers, $\|.\|$ denotes that Euclidean Distance and β is the spread parameters. The optimum values for the weight can be found using least square as given by equation (8) :

$$\widehat{W} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = (\emptyset^T \emptyset)^{-1} \emptyset^T \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \textbf{(8)}$$

Finally, the RBFNN's output is computed by using estimated weight based on equation (9) below:

$$Y = \widehat{W} \emptyset \quad \textbf{(9)}$$

2.1.2 RBFNN Activation Function

An activation function in RBFNN is a transfer function which depends on the distance with respect to the center. Selected spread constant can influence the output performance of RBFNN [15]. If the spread value is high, the data points will be scattered to a large distance from the center and hence the maximum function response will be the performance of the model. Thus, the value of the spread must be determined so that the model performance is good both in the calibration period properly for the better performance of the network. In [13], it is found in that the model performance is good with the spread value of 1.0 with 4 function nodes.

Table 1: Transfer function for RBFNN

Equation	Activation Function
$\Phi(r,\sigma) = e^{\frac{-r^2}{\sigma^2}}$	Gaussian
$\Phi(r) = r^3$	Cubic
$\Phi(r) = r^2 \log(r)$	Thin Plate spline
$\Phi(r) = \frac{1}{r+1}$	Cauchy
$\Phi(r,\beta) = \sqrt[2]{r^2+\beta}$	Multiquadratic
$\Phi(r,\beta) = \frac{1}{\sqrt[2]{r^2+\beta}}$	Inverse multi-quadratic
$\Phi(r) = r$	Linear

The RBFNN makes an approximation based on training data, and the Gaussian function is used mostly as the activation function.

2.1.3 K-means clustering

This technique is based on the distance matrix, using Euclidean distance as a criterion. It starts with m initial cluster centers and for all data, the Euclidean distance from each cluster center is calculated, after which the data points are assigned to the closest cluster center. This method is particularly useful in adjusting the spread and center parameter for RBFNN algorithm.

2.1.4 Cross-Validation

Cross-validation is one of the statistical techniques used for separating and partitioning the dataset into training and testing. For each iteration, one segment is used as training dataset and the remaining segments are used as testing and validation dataset. For k-fold cross-validation, the first segment of the dataset is formed based on the value of k-fold. Afterwards, k iterations of training and testing dataset are performed specifically among every iteration. A different fold dataset is held-out for validation whereas the remaining k-1 fold of the dataset is used for testing [16]. Fig. 2 shows how the cross-validation technique separating the dataset into training and testing continuously until the k-fold cross validation is completed.

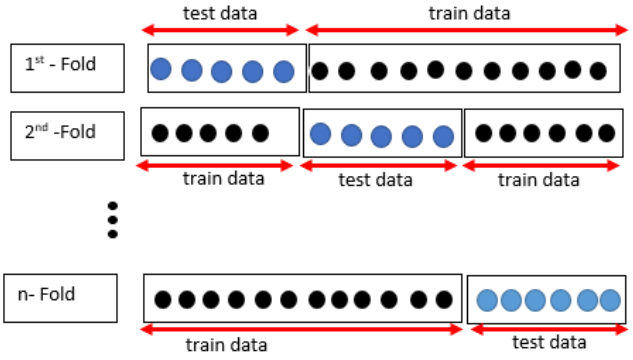


Fig 2: cross-validation technique for partition data

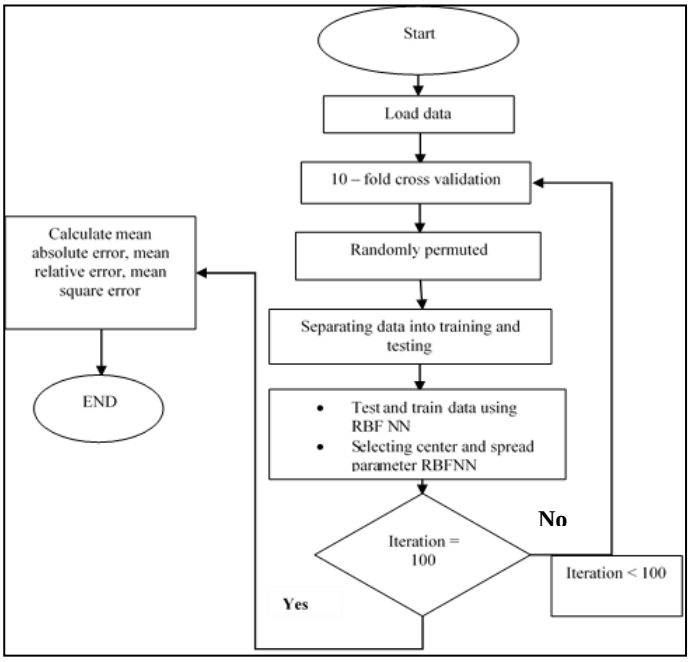


Fig 3:Flowchart for RBFNN estimation algorithm

Referring to Fig. 2, the dataset is partitioned into training and testing using the 10-folds cross-validation technique. Subsequently, the dataset is randomly permuted for each row. Then, the dataset undergoes testing and training phase using RBFNN algorithm. For statistical confidence, the process of training and testing dataset using RBFNN is repeated for 100 iterations. The error for 10 – folds cross-validation with 100 iteration is computed using mean square error (MSE), mean absolute error (MEA) and mean relative error (MRE).

3.3. Error Validation and Performance

The result of the comparison analysis is computed using MSE, MEA, and MRE to compare the performance of RBFNN and others method that proposed by others researcher.

$$\textbf{Mean absolute error} = \frac{1}{N}\sum_{i=1}^{N}\left|Y_i - \hat{Y}_i\right|^2 \quad (10)$$

$$\textbf{Mean square error} = \frac{1}{N}\sum_{i=1}^{N}\left|Y_i - \hat{Y}_i\right|^2 \quad (12)$$

$$\textbf{Mean square error} = 100 \text{ x } \frac{1}{N}\sum_{i=1}^{N}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right|^2 \quad (11)$$

where Y_i is the actual value when the input is X_i, $\hat{Y}_t$ is the predicted value when the input is X_i.

The result for actual and predicted values of HL and CL also calculated using a percentage of fitness. The mathematical formula to calculate the percentage of fitness is shown as below:

$$\% \text{ Fit} = \left(1 - \frac{norm(predicted - actual)}{norm\,(actual - mean(actual))}\right) \times 100 \quad \textbf{(13)}$$

4.1 Experimental studies

In this paper, the residential building energy performance dataset obtained from UCI Machine Learning is used to improve the estimation of energy efficiency performance using different algorithm. The extensive simulations on 768 diverse residential buildings with 8 input variables and 2 output variables are demonstrated using ICOTECT. The first published work using the same dataset is done by Tsanas and Xifara [3] by implementing the Iteratively Reweight Least Square (IRLS) algorithm and improving the estimation result using the RF algorithm. Subsequently, the result of the same work but using RBFNN in this study is also compared with the previously published work using the same dataset to improve the performance of estimation in [3-7]. The description of dataset is shown in detail in Table 2.

Table 2: Input and output variables of dataset

Representative of variables	Variables name
X_1	relative compactness
X_2	surface area
X_3	wall area
X_4	roof area
X_5	overall height
X_6	orientation
X_7	glazing area
X_8	glazing area distribution
Y_1	heating load
Y_2	cooling load

4.2 Experimental Setup

The dataset was simulated using RBFNN algorithm on computer with the following specification: (1) Microsoft Windows 10 Professional Edition; (2) Matlab R2017a; (3) Intel® Core™ i7-6700K Processor. This study describes the comparison of different algorithms using the same dataset that is uploaded on the UCI Machine Learning Repository. In this study, the dataset that consists of 768 was partitioned using 10 folds cross-validation. For the first iteration, 76 samples were used as testing data while the 692 samples used as training data. Subsequently, the next 76 samples were used as a testing data and the remaining samples used for training data. The same process was repeated until the dataset divided into 10 partitions of testing that consist 76 data for each partition and 10 partitions of training data that consist 692 data for each partition. Then, the data was randomly permuted for each row before the analysis, testing, and training process using the RBFNN algorithm.

5. Result and Discussion

Further analysis using the dataset shows an estimation of building energy performance for heating load and cooling load. Figure 3 and Figure 4 represent regression plot and R square value for each HL and CL computed using Matlab 2017a. The results show that RBFNN has a good performance because the predicted and actual values have a high value of R square.

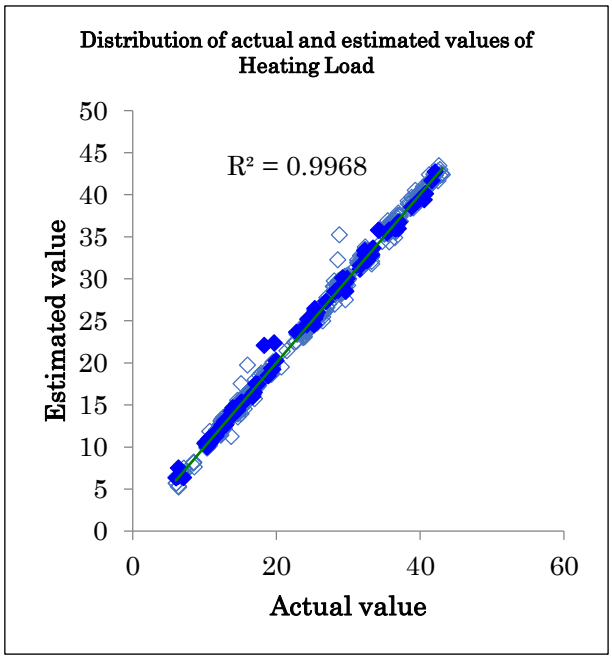


Fig 4: Value of R square for actual and predicted of HL

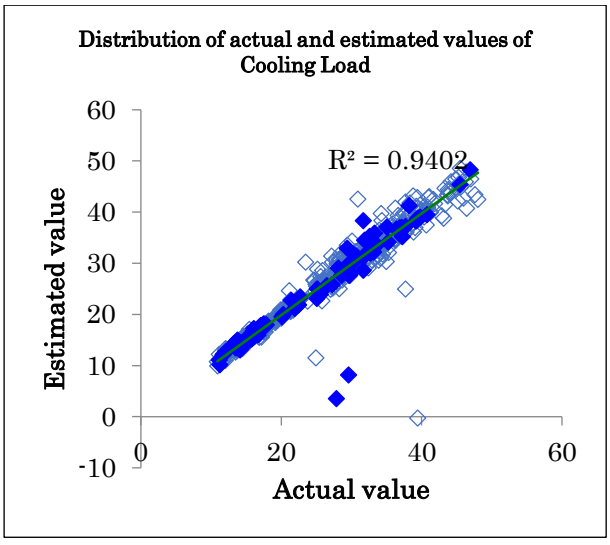


Fig 5: Value of R square for actual and predicted of CL

As mentioned in the literature review, RBFNN is one of the competitive algorithms for estimation. The obtained R square values for HL and CL are each 0.9968 and 0.9402, which indicate a good performance. Fig 3 and Fig 4 show that there are no overfitting occurrences, and the model is significantly satisfied for estimation of energy efficiency performance.

The validation of error for 10 folds cross-validations with 100 repetitions was computed and recorded using MEA, MRE, and MSE. The error of HL and CL are summarizing in Table 3 and Table 4 – in these tables, the value of error obtained in this study is compared with values of error obtained in previous works [3-7].

Table 3: Comparison MEA, MSE, and MRE performance for Heating Load

Reference	Methodology	MEA	MSE	MRE
Tsanas and Xifara [3]	Random Forest	0.51	1.03	2.18
Cheng et al [7]	EMARS	0.35	0.47	-
Chou et al [5]	SVR with ANN	0.24	0.35	-
Casteli et al [4]	GP	0.38	-	0.43
Regina et al [6]	MLP with grid search	0.25	1.04	1.20
This paper	RBF NN	0.32	0.21	1.60

Table 4:Comparison MEA, MSE, and MRE performance for Cooling Load

Reference	Methodology	MEA	MSE	MRE
Tsanas and Xifara [3]	Random Forest	1.42	6.59	4.62
Cheng et al [7]	EMARS	0.71	1.00	-
Chou et al [5]	SVR with ANN	0.97	1.57	-
Casteli et al [4]	GP	0.97	-	3.40
Regina et al [6]	MLP with grid search	0.39	1.28	1.65
This paper	RBF NN	0.89	2.78	3.20

As can be seen from the Table 3 and Table 4, the comparison of error performance between RF [3], EMARS [7], SVR with ANN [5], GP [4], MLP with grid search strategy [6] and RBFNN are shown. Based on the values of MEA, MRE, and MSE for the method proposed by the previous researchers, it is shown that each subsequent work improves previous work in terms of the accuracy of building energy performance estimation. The RBFNN proposed in this study has better performance than RF, SVR with ANN, GP, and EMARS. However, the result obtained by Regina et al [6] is better than RBFNN in this study, except the value of MSE for HL which is 0.21 for RBFNN in this study and 1.04 for MP with grid search strategy in [6]. These findings might encourage the others to develop a hybrid algorithm for building energy efficiency estimation instead of using the stand-alone method of the algorithm.

Table 5: Percentage of fit HL for testing data

Output Variables	**Percentage of the fitness value**
HL	94.33
CL	83.84

Table 5 shows the result of HL and CL in terms of a percentage of fitness for testing data. The percentage fitness is 94.33% for HL and 83.84 % for CL – these shows that both HL and CL value have good performance between actual and predicted.

The findings of this study strongly agree with those in the algorithm literature – the best method for estimation are hybrid algorithm such as MLP with grid search strategy. However, some algorithms show better performance results based on error validation values such as GP and RBFNN. Both MLP and RBFF is the popular architerture in ANN which is widely used for estimation and prediction. In this study, the hybrid method of MLP with grid search obtained in [6] have the best performance compared to the methods proposed by the previous researchers in [3-7]. Although the RBFNN gets slightly higher value of error validation than MLP with grid search, the result produced by the RBFNN method can still be accepted. However, further work for improvement need to be done in future.

Conclusion

Based on the comparison of the results obtained in this study and the previous studies by other researchers in [3-7], it is concluded that the MLP with grid search strategy produces the best estimation result for both HL and CL based on the values of MEA, MSE, and MRE. However, RBFNN perform the best for HL in terms of MSE. For future work, it is suggested to establish the RBFNN algorithm with grid search strategy to estimate the HL and CL for the same dataset.

Acknowledgment

We would like to express our gratitude to the Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM) for the financial support.

References

[1] World Energy Markets Observatory, (2017).
[2] A.G. Alam, C.I. Baek, H. Han, Prediction and Analysis of Building Energy Efficiency Using Artificial Neural Network and Design of Experiments, Appl. Mech. Mater. 819 (2016) 541–545. doi:http://dx.doi.org/10.4028/www.scientific.net/AMM.819.541.
[3] A. Tsanas, A. Xifara, Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools, Energy Build. 49 (2012) 560–567. doi:10.1016/j.enbuild.2012.03.003.
[4] M. Castelli, L. Trujillo, L. Vanneschi, A. Popovič, Prediction of energy performance of residential buildings: A genetic programming approach, Energy Build. 102 (2015) 67–74. doi:10.1016/j.enbuild.2015.05.013.
[5] J.S. Chou, D.K. Bui, Modeling heating and cooling loads by artificial intelligence for energy-efficient building design, Energy Build. 82 (2014) 437–446. doi:10.1016/j.enbuild.2014.07.036.
[6] G. Regina, P. Capriles, Prediction of energy load of buildings using machine learning methods Database and Machine Learnig Methods, (2016).
[7] M.Y. Cheng, M.T. Cao, Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines, Appl. Soft Comput. J. 22 (2014) 178–188. doi:10.1016/j.asoc.2014.05.015.
[8] B. Yildiz, J.I. Bilbao, A.B. Sproul, A review and analysis of regression and machine learning models on commercial building electricity load forecasting, Renew. Sustain. Energy Rev. 73 (2017) 1104–1122. doi:10.1016/j.rser.2017.02.023.
[9] K.P. Amber, M.W. Aslam, A.M. Id, A. Kousar, M.Y. Younis, B. Akbar, G.Q. Chaudhary, S.K. Hussain, Energy Consumption Forecasting for University Sector Buildings, (2017) 1–18. doi:10.3390/en10101579.
[10] Y. Zhang, Q. Chen, Prediction of Building Energy Consumption Based on PSO - RBF Neural Network, (2014) 1–4.
[11] A. Extreme, D. Learning, Building Energy Consumption Prediction : An Extreme Deep Learning Approach, (2017) 1–20. doi:10.3390/en10101525.
[12] H. Sug, A comparison of RBF networks and random forest in forecasting ozone day, Int. J. Math. Comput. Simul. 4 (2010) 59–66.
[13] R. Article, O. Access, A.K. Behera, S. Dehuri, S. Cho, Radial basis function neural networks : a topical state-of-the-art survey RBFNs architecture, (2016) 33–63. doi:10.1515/comp-2016-0005.
[14] M.D. Buhmann, Radial basis functions, Acta Numer. 2000. 9 (2000) S0962492900000015. doi:10.1017/S0962492900000015.
[15] S. Martin, C.T.M. Choi, On the influence of spread constant in radial basis networks for electrical impedance tomography, Physiol. Meas. 37 (2016) 801–819. doi:10.1088/0967-3334/37/6/801.
[16] D. Anguita, L. Ghelardoni, A. Ghio, L. Oneto, S. Ridella, The ‘ K ’ in K-fold Cross Validation, (2012) 25–27.