# Application of Machine Learning in Forecasting Energy Usage of Building Design

Truong Xuan Dan, Phan Nguyen Ky Phuc, *Member, IEEE*

**Abstract** – ***As rapid movement of Information technology, the amount of data and unexpected factors impact on forecasting are gradually uncontrollable, thus, traditional method may be not enough efficient to deal with this issue. Therefore, the development of AI is advantageous for forecasting when data extends significantly. In this paper, machine learning model was developed to estimate energy load based on the characteristics of building design. This determination helps business and engineer reduce energy consumption cost and environmental impact. The data set was collected from energy simulation in the study of Tsanas, A., & Xifara, A. (2012) including 768 observations with 8 inputs and 2 outputs (heating load and cooling load). The energy loads were achieved through innovative methods such as Artificial Neural Network, Support Vector Machine or Random Forest (nonlinear) and Multilinear Regression (linear) with the support of Interactions. The performance of result from neural network technique was quite overfitting with dataset and better than linear as 20% in RMSE. So, it is proposed that using ANN forecasting combine with the predictor variable interaction of Multilinear Regression helps the user analyze the predictive value coordinate with input adjustment. Generally, the research supports the feasibility of machine learning in building energy forecast based on historical data and the building design parameter as well as the possibility to apply to another dataset for prediction purpose.***

**Keywords:** ***forecast, energy load, building design, energy cost, environmental impact, Artificial Intelligent, ANN, Multilinear regression.***

## I. INTRODUCTION

Thermal load estimation based on building design is significantly essential to improve the energy performance. In general, energy optimization plays a vital role in civil planning as well as infrastructure and facility planning in industrial engineering system. Energy planning helps business and engineer not only save cost but also minimize environmental impact. In order to control the engineer consumption, heating load and cooling load need to be determined among to the feature of building design. Heating load and cooling load are the amount of energy that building needs to maintain thermal comfort of inner temperature. The efficiency of energy load is defined based on the design of building which contains two main types: passive design and design for climate. Design for climate approach is modifying the building to ensure the thermal comfort with minimal auxiliary heating and cooling in location climate, whereas passive design takes the advantages of outside condition to heating and cooling system for reducing the energy consumption as attempt to control comfort.

Thermal comfort, energy consumption and environmental impact are considered as indicators to evaluate the energy system. It was reported that about 35 to 40% of total energy was consumed in buildings in the developed countries with 50 to 65% of electricity consumption [1]. Specifically, In the UK, energy consumption for space heating contributed to about 50% of the service sector energy consumption as reported by IEA [2]. According to the statistic of Vietnam Academy of Science and Technology in 2017, the energy usage of industrial and commercial building in Vietnam accounted for over 50% of total energy consumption and it was estimated that this amount would be extended to 65% till 2030. Additionally, in Viet Nam, there are still many plants and warehouses in conventional industrial park are not satisfied with the thermal standard - TCVN 5508:2009 for working environment. Thus, analysis and energy assessment are a crucial part for improving Industrial planning.

The objective of this paper is using AI for energy estimation analysis as well as making the comparison with linear forecasting. The research contains five main parts. The previous one is introducing the general problems and proposing solution for this area. The following section is reviewing the past work relevant to this area containing predictive methods and energy analysis. Then, Section 3 illustrates the implementation of data collection and study. Section 4 describes the methodology used for mining and developing the energy load. Section 5 specifies the modelling process and result analysis.

## II. LITERATURE REVIEW

Forecasting is a core-business tool to make predictions of the future based on past and present data. Conventionally, there are several quantitative approaches to forecast including time series method, causal forecasting as well as qualitative approaches like judgmental method. Generally, those methods are mostly based on the analysis of previous data and additional information to get the behavior of future data. Mentioned to the problem of energy load forecasting based on the building design, the research of Hong, T., Gui, M., Baran, M. E., & Willis, H. L. (2010) developed statistical model for electricity forecast which built the Multi Linear Regression model. In specific, the energy load was estimated by regression model with the support of diagnostics statistics to evaluate and continuously improve the Linear Model. They chose this traditional method despite the advance of machine learning approach due to its simplicity and explanatory analysis. Additionally, Deng, J., & Jirutitijaroen, P. (2010) also applied time series analysis for short-term

Truong Xuan Dan is senior of Industrial Engineering Department, International University – Vietnam National University (corresponding author to provide phone: +8490-6963-942; e-mail: dantruong.iuerd@gmail.com).

Phan Nguyen Ky Phuc is the lecturer of Industrial Engineering Department, International University – Vietnam National University (corresponding author to provide phone: +84965865512; email: pnkphuc@hcmiu.edu.vn).

978-1-5386-5126-1/18/$31.00 ©2018 IEEE

Singapore electricity demand forecasting. Two time-series models were proposed which were multiplicative decomposition model and seasonal ARIMA model. They compared two models by conducting the analysis of trend, seasonal, cyclic, and random component from historical data into the multiplicative model as well as time series analysis by ARIMA.

With the extension of information technology, machine learning is gradually applied in forecasting. Soon, the study was promoted by Hong, W. C. (2009) using data mining technique to build predictive model for Taiwan regional electric load. In detail, Support Vector Regression with Immune Algorithm was created due to the comparison of many machine learning methods in order to get the better performance in the prediction of nonlinear data pattern. The result shown that SVRIA had better result than Support Vector Machine (SVM), Regression model and Artificial Neural Network (ANN) model with the evaluation of generalization error. Furthermore, Tsanas, A., & Xifara, A. (2012) built a statistical machine learning framework to study the effect of industrial building design on the energy load. They investigated the association strength of building characteristics with the energy consumption using various non-parametric statistical analysis. Then, they made the comparison between classical linear regression method, nonlinear non-parametric method and random forest classification to estimate the energy load. Meanwhile, Bonetto, R., & Rossi, M. (2017) extended artificial intelligent application including SVM and Nonlinear Auto-Regressive Neural Network (NAR) to perform multi-step ahead forecasting [3]. They generated a parallel and efficient training framework, using the historical data and demand traces from real deployments to ensure the accuracy of considered techniques. And the result indicated that machine learning approach achieved smaller prediction error than conventional ARMA but it was still confusion in choosing training algorithm. In the advance of this research, multiple data mining technique were applied by Chou, J. S., & Bui, D. K. (2014) to consider which one had better performance in order to conduct the model combination. Specifically, ensemble learning was considered for this combination to avoid 'overtraining' issue as using artificial intelligent approach to develop predictive model.

ANN is a basic predictive method of machine learning due to the widen use and uncomplicated process to get result. . Application of ANN in forecasting area was clearly described on the study of Agatonovic-Kustrin, S., & Beresford, R. (2000). While it can get the result with lowest error indicator , it cannot explicit the relationship of input variables and forecasting value. In addition, there are many training algorithms for preprocessing as applied to ANN which were mentioned in the theory of Hagan, M.T., H.B. Demuth, and M.H. Beale (1996). Obviously, the study of Sari, Y. (2014) used several above training algorithms to analyze the effect of them on the learning performance of neural networks on the inverse kinematics model learning of a seven-joint redundant robotic manipulator is investigated.

Multiple linear regression is the most common form using for prediction in regression term. As analysis, it is used for investigating the relationship of the estimated value and predictor variables. The key theory of this approach is the mathematic formulation for the linear equation of forecasting which estimates the multiple linear regression model. The advance of multiple linear regression with interactions compare to the conventional one is the synthesis of interaction term between quantitative predictors. Hagan, M. T., Demuth, H. B., & Beale, M. H. (2010) proposed several transformations of basic linear regression model among the interactive effect in econometrics. In specific, linear equation was enhanced with interaction such as robustness and panel data.

The above studies used not only machine learning but also conventional methods in energy estimation. Each past work was logically persuasive for the method selection either nonlinear or linear, hence, this paper would make a comparative result of single model in order to point out clearly the disadvantages and advantages of them.

## III. ENERGY DATASET

In order to implement machine learning in prediction, this paper used the dataset of building design to forecast the energy load. Specifically, the dataset was created by Angeliki Xifara (Civil Engineer) and was processed by Athanasios Tsanas (Oxford Centre for Industrial and Applied Mathematics, University of Oxford, UK) [4]. The energy load were analyzed by using 12 different building shapes simulated in Ecotect software. All the buildings have the same volume which is 771.75 m3, the material used for each 18 elements are the same for all building forms. The simulation was implemented under these conditions [5]:

- Internal design: clothing: 0.6 clo, humidity: 60%, air speed: 0.30 m/s, lighting level: 300 Lux. The internal gains were set to sensible (5) and latent (2 W/m2), while the infiltration rate was set to 0.5 for air change rate with wind sensitivity 0.25 air changer per hour.
- Thermal properties: 95% efficiency, thermostat range 19–24 ◦C, with 15–20 h of operation on weekdays and 10–20 h on weekends

The buildings differ with respect to surface area, wall area, glazing area, orientation, and so on. These characteristics are considered as the attributes of learning model and the outcome would be heating load and cooling load. Although the data used in this study obtained by the simulation could be biased in some way, they represent actual real data with high probability and as such will be considered as ground truth. Moreover, any inconsistency in the simulated data and actual real-world data does not affect whatsoever the methodology developed in this study [5].

In general, the dataset contains 768 samples with 8 attributes predict two outcomes. The table summarizes the inputs and outputs in this study and mathematically represents the typical value for each variable.

TABLE I. ENERGY DATA COLLECTION

| Variable | Name | Description | Data summary |
|---|---|---|---|
| X1 | Relative Compactness (RC) | Measurement of the closure compact in building design. The more RC the more energy load. | 12 types: 0.98, 0.90, 0.86, 0.82, 0.79, 0.76, 0.74, 0.71, 0.69, 0.66, 0.64, 0.62. |
| X2 | Surface Area | Surface area of building | (Sqft) |
| X3 | Wall Area | Area covered by width of the wall | (Sqft) |
| X4 | Roof Area | Area coved under roof. Actual area where cooling or heating would be required | 4 types: 110.25, 122.5, 147, 220.5 |
| X5 | Overall Height | Overall height of building | (ft) |
| X6 | Orientation | Building orientation with respect to sun and wind direction | 4 orientations: 2, 3, 4, 5 representing North, South, East, West |
| X7 | Glazing Area | Glass area in on wall (window, etc). Proportion of floor area which is covered by windows, glass walls, glass roofs etc. | 4 types: 0.00, 0.10, 0.25, 0.40 |
| X8 | Glazing Area Distribution | How glazing area distributed in building | 5 distribution scenarios for each glazing area: (1) uniform: 25% glazing on each side, (2) north: 55% on the north side and 15% on each of the other sides, (3) east: 55% on the east side and 15% on each of the other sides, (4) south: 55% on the south side and 15% on each of the other sides, and (5) west: 55% on the west side and 15% on each of the other sides |

Table 2 conducts that there are 4 main influential factors can affect the output: Relative compactness, Surface Area, Glazing Area and Orientation.

## IV. METHODOLOGY

### A. Artificial Neural Network

Similar to working principle of brain-inspired system, ANN tool in machine learning uses interconnection of neurons as processing system consisting of multiple input, weight, bias, and the output activation. The link of neurons in the network figures their interaction and each link be associated with weight. Specifically, the ANN implements the external data into one or several hidden layers and transforms them into the output.

TABLE II. DATA ANALYSIS

| Predictor variable | Mean | Min | Max | Standard deviation | Correlation | |
|---|---|---|---|---|---|---|
| | | | | | *Heating Load* | *Cooling Load* |
| Relative Compactness (RC) | 0.764 | 0.62 | 0.98 | 0.106 | 0.622 | 0.634 |
| Surface Area | 671.708 | 514.5 | 808.5 | 88.086 | -0.658 | -0.673 |
| Wall Area | 318.5 | 245 | 416.5 | 43.626 | 0.456 | 0.427 |
| Roof Area | 176.604 | 110.25 | 220.5 | 45.166 | -0.862 | -0.863 |
| Overall Height | 5.25 | 3.5 | 7 | 1.751 | 0.889 | 0.896 |
| Orientation | 3.5 | 2 | 5 | 1.119 | -0.0026 | 0.0143 |
| Glazing Area | 0.234 | 0 | 0.4 | 0.133 | 0.270 | 0.208 |
| Glazing Area Distribution | 2.812 | 0 | 5 | 1.551 | 0.0873 | 0.0505 |
| Heating load | 22.307 | 6.01 | 43.1 | 10.090 | | |
| Cooling load | 24.587 | 10.9 | 48.03 | 9.513 | | |

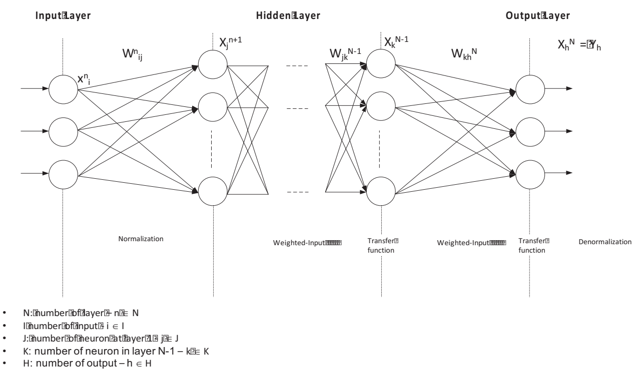



Figure 1. *Artificial Neural Network*

Feed forward neural network is the simplest and effective type of artificial neural network. The information in this network moves forward from the input layer, through the hidden layers and to the output without cycle or loop.

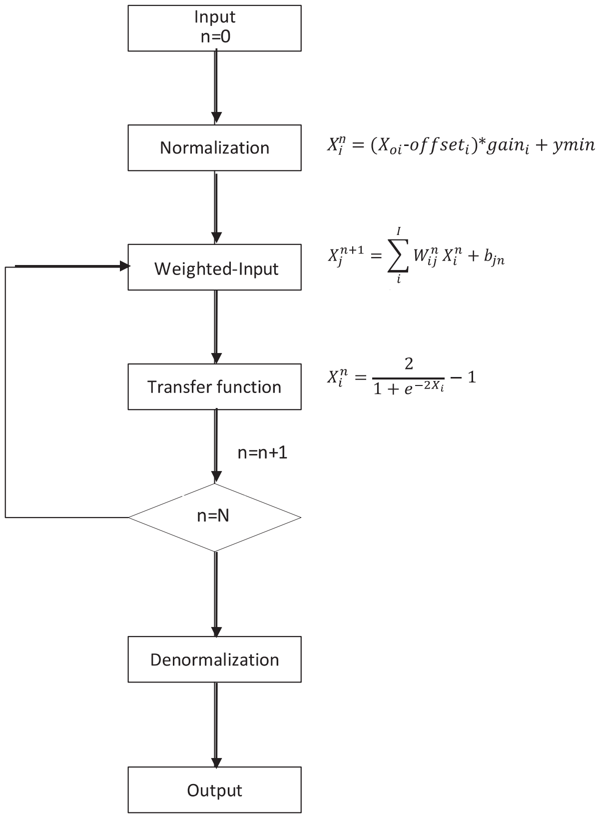



Figure 2. *ANN Process*

As generating several training algorithms were named in [6], Bayes Regularization was the best one for this network. BR training in ANN is based on the Bayes' theorem to determine the probability of weight distribution among to each element in each layer. In general, BR algorithm normalizes data and updates the value of weight and bias in each neuron depend on loss function minimization and overfitting elimination [7].

$$S(W) = \beta \sum_{h=1}^{No} [\, y_{0h} - f(X_i)]^2 + \alpha \sum_{j=1}^{Nw} W_j^2 \qquad (1)$$

where $N_0$ is the number of outputs, $y_{0h}$ is the observed value of output, $f(X_i)$ is the result of BRANN, and $N_w$ is the number of weights. Given the initial value of the hyper parameters α and β, S(W) is the loss function of BR training which is minimized with respect to the weight W. The value of α and β is updated by maximizing the evidence.

### B. Multilinear Regression with Interactions

With the aim of finding the appropriate Linear Regression Model for energy data, this research used 'fitlm' in Matlab. The principle working of this function is finding the fitted Linear model for greater accuracy on low- through medium-dimensional data sets. It is mainly based on the diagnostic value, model criterion, and the algorithm of Robust fitting and QR decomposition to generate the model.

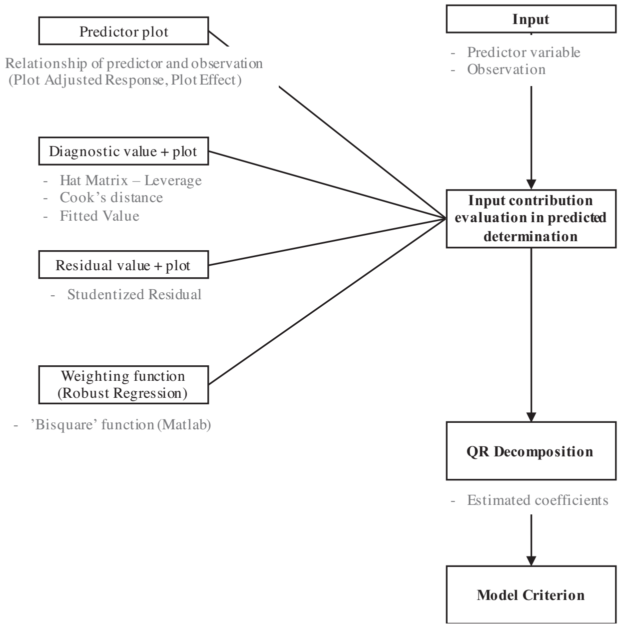



Figure 3. *Multilinear Regression with interactions*

Difference from traditional multilinear regression, the Interaction allows user to estimate the additional coefficients indicating the influential level among each variable [8]. The predictive function is described in detail:

$$Yi= a+\sum_{j=1}^{8} b_j X_{ij} + Interaction \quad (2)$$

$$Interaction=b_{12}X_1X_2+ b_{13}X_1X_3+b_{14}X_1X_4+\ldots+b_{78}X_7X_8 \quad (3)$$

### *C. Support Vector Regression:*

Support Vector Machine is a branch of supervised learning as using learning algorithm for data analysis in the aspect of classification and regression. In this case, Support Vector Regression (SVR) with Nonlinear Regression was used to support the energy prediction. Instead of using the error as loss function, SVR uses the distance measurement for this function [9].

Nonlinear approach has to map the data into high dimensional feature space, which is similar to linear through the Gram matrix of Kernel function. Then, ε-insensitive loss function for its optimization model and predictive model determination. ε -insensitive loss function can be known as Dual formulation in nonlinear SVR. The objective function is maximizing the margin regression similar to minimizing the Euclidean norm of flatness after mapping by Kernel algorithm.

$$L(\alpha)=\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)G(x_i,x_j)+\varepsilon\sum_{i=1}^{N}(\alpha_i-\alpha_i^*)-\sum_{i=1}^{N}y_i(\alpha_i-\alpha_i^*) \quad (4)$$

where

- $\alpha_n, \alpha_{n*}$ :nonlinear multiplier
- $G(x_i, x_j)$: Gram matrix contains element $g_{i,j}$ which is equal to inner predictor as transformation at attribute j of observation i.
- ε: margin tolerance or epsilon margin
- $y_i$: observation $i^{th}$
- $\xi_n, \xi_{n*}$: slack variable for soft margin for each point.
- C: box constraint which is a parameter controlling the maximum penalty impose on margin epsilon as well as preventing overfitting problem.
- $f(x_n)$: predictive model

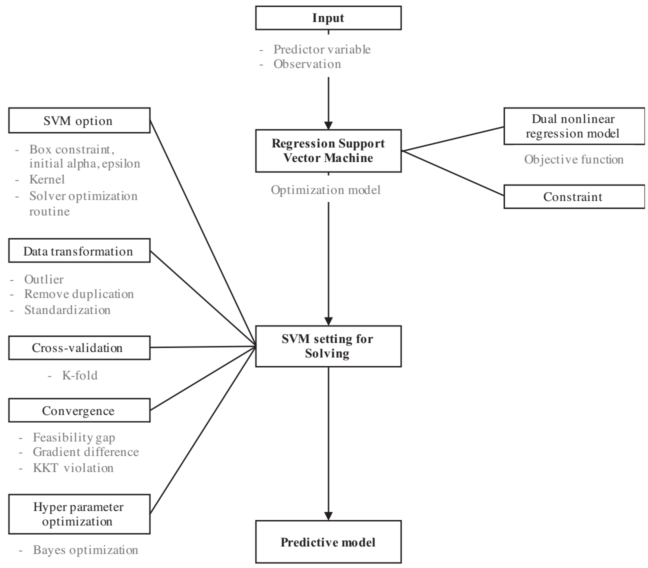



Figure 4. *Support Vector Regression*

### *D. Regression Tree*

The prediction tree is an alternative approach to nonlinear regression using the recursive partitioning to simulate the forecast value. Particularly, regression tree is built through iterative process that splits the data into many partitions or branches, then continuously splits each branch into smaller groups. The algorithm for grouping and splitting data is based on the minimization of the sum of square deviations from the mean in the separate partitions. For forecasting energy usage, ensemble method and random forest algorithm are chosen.

Ensemble approach combined multiple weak regression trees to form a new regression model which was stronger and more accurate than single tree. Particularly, Bootstrap Aggregation was selected to predict the energy consumption. The main reason of this selection is to reduce the variance of a decision tree and it is appropriate with the dataset having many observations. The principle behind this technique is using the collection of Bootstrap set to train the based decision trees and end up with the ensemble of the various tree models through the combination among their votes.

Beside Bagging, Random forest was also chosen to observe and analyze the result of energy load. This method is an innovation of bagging in decision trees. While Bagging trains the data with all of the predictor variables, random forest chooses random training set for each base learner and node of these trees uses the selected feature for splitting. At that point, the tree in next level will be continuously grown. This technique trains faster than Bagging method because the number of feature choosing for splitting is certainly smaller than all original features. Specifically, the feature training for each node was randomly selected and the rest of them would be used for splitting other branch of decision tree.

## V. FORECASTING RESULT

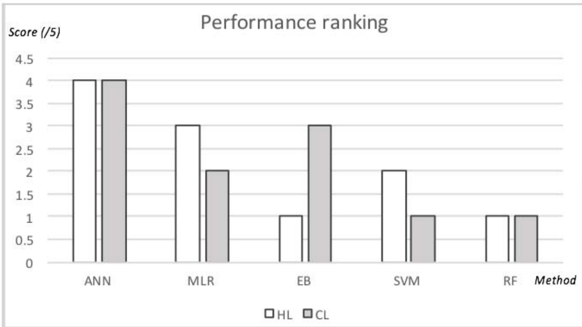
Before applying the predictive model, the energy dataset had to be experienced in preprocessing. Defining the training subset, a random no stratified partition for k-fold cross-validation was applied with k=5. The training set was used for generating the forecasting model and the testing test was used for model evaluation. The following table gives the information about model parameter conducted from training set.

TABLE III. MODEL PARAMETER

| Single model | Model information | |
|---|---|---|
| ANN | Hidden Layers | 1 |
| | Neuron in each layer | 4 |
| | Transfer function | Tangential function |
| | Training function | Bayes Regularization |
| | Maximum Epochs | 1000 |
| | Maximum Training time | Inf |
| | Performance Goal | 0 |
| | Minimum Gradien | 0 |
| | Maximum Validation Checks | 0 |
| | Mu | 0.005 |
| | Mu Decrease Ratio | 0.1 |
| | Mu Increase Ratio | 10 |
| | Maximum Mu | 10000000000 |
| MLR | Number of Interactionns | 29 |
| | Number of parameter | 37 |
| | Leverage evaluation | 0 and 1 |
| | Cook distance | 3*average |
| SVR | Box Constraint | 13.8653 |
| | Margin Tolerance | 1.3866 |
| | Delta Gradient Tolerance | 0.001 |
| | Gap Tolerance | 0 |
| | KKT Tolerance | 0 |
| | Kernel function | Polynomial (p=3) |
| | Kernel scale | 0.806 |
| | Bias | -56.6211 |
| | Weght | 0.00187 |
| | Solver method | SMO |
| Regression tree | Number of trained | 30 |
| | Number of tree | 100 |
| | Number of sample predictor | 3 |
| | Min leaf size | 5 |
| | Number of nodes | {95,129} |

Noted that number of hidden layers and neurons in each layer were determined through numerous training times based on the performance of cross-validated dataset, and such sizes were as low as possible in order to avoid overtraining in ANN. Also, it was achieved depend on the several experiences related to "rules of thumb" in ANN [10].

The data and model was run in Matlab 2016 through several fitting model functions. The running time was quite fast compare to other software which was approximately 12 second for getting the result of all methods.




MAPE (%)

Figure 5. *Performance ranking of each model*

In order to interpret the performance of each model based on error indicator, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) were obtained from testing set. Also, as defining the score of each model, the average ranking of energy load among those was determined. The higher score, the lower error each approach performed. The above chart indicates that ANN technique provides the better result than others.

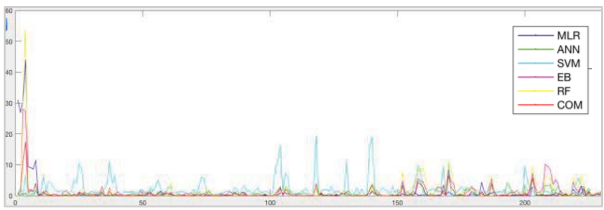
TABLE IV. ERROR INDICATOR OF EACH MODEL.

| Method / Error Indicator | *Heating load* | | | *Cooling load* | | |
|---|---|---|---|---|---|---|
| | RMSE | MAE | MAPE | RMSE | MAE | MAPE |
| ANN | 0.8192 | 0.6934 | 3.0572 | 1.8033 | 1.3166 | 4.6978 |
| SVR | 1.0122 | 0.5561 | 2.909 | 2.4548 | 1.396 | 4.8101 |
| MLR | 1.4212 | 0.12243 | 6.0788 | 2.0934 | 1.6422 | 6.9376 |
| Bagging | 1.1772 | 0.8233 | 3.8359 | 1.9024 | 1.3534 | 4.7333 |
| Random Forest | 1.2288 | 0.8399 | 3.7788 | 1.9488 | 1.4026 | 4.9004 |

Overall, it can be seen that ANN has the highest performance in term of all indicators which is nearly 40% in heating load and 13% in cooling load estimation after applying prediction model to testing data. Mentioned to Ensemble Bagging and Random Forest, the running principle of them were nearly similar, so their result were not significantly different. Particularly, in heating load estimation, those techniques did not have a good perform as cooling load. Moreover, Multilinear Regression is ranked as the second-best method when generating thermal energy, while cooling load is Ensemble Bagging. Although the Multilinear Regression approach did not get the highest performance when forecasting energy, but the result of this technique was quite reasonable which could discover the behavior of data set as well as the correlation of building feature and energy consumption through the Interaction parameters. Hence, choosing the forecast model with error evaluation is not efficient as its ability when extending more data.

RMSE (BTU)

MAE (BTU)




*Testing set*

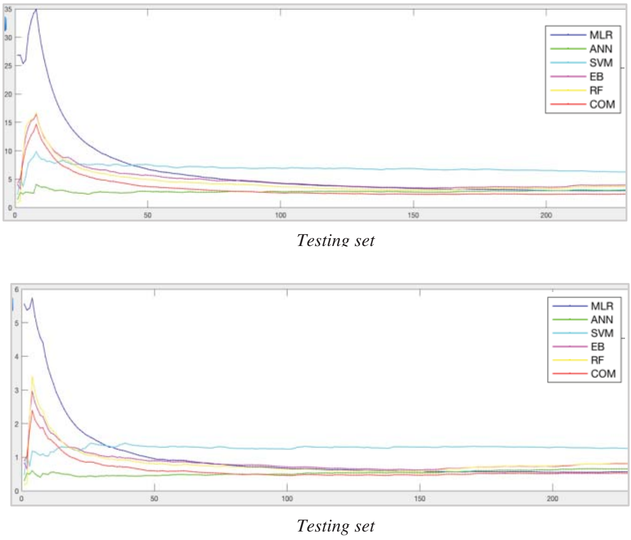



Figure 6. *RMSE, MAE, MAPE of each model in heating load*

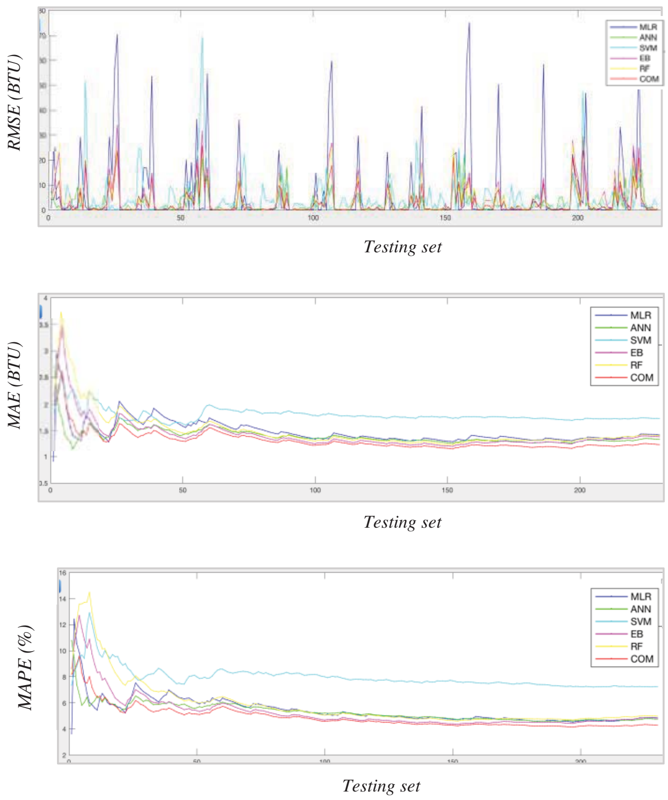



Figure 7. *RMSE, MAE, MAPE of each model in cooling load*

When evaluating based on the stability generation, ANN also had the most stable performance as predicting on testing set of both heating load and cooling load. The figures demonstrate the persistence of this method on all of error indicators. Fig.7 illustrates that Bagging method has nearly equivalent to the performance of ANN in estimating cooling load. Conversely, in heating load, SVR is ranked as second-best with respect to MAE and MAPE, illustrated in Fig.6. Even though Multilinear Regression was chosen as the most reasonable method for forecasting, the stationary evaluation of this technique was not good as error measurement. Thus, so as to maintain the prediction perforamance in terms of error indicator as well as consistent procedure, the machine learning application with the support of explanatory indication of linear method is proposed to implement.

## VI. CONCLUSION

Generally, according to the evaluation of several machine learning techniques and conventional methods in energy prediction with regard to error indicator and steady generation, it could be conducted that ANN was the best technique. Although the number of observation was 768, the process for cleansing data was not complicated due to the amount of predictor variable. It contained 8 factors for building with non-missing value. So, the result was quite good and the best performance approach, however, could lead to overtraining issue when setting the parameter be quite appropriate with current amount of given data set.

In particular, ANN cannot produce the explanatory estimation because its generation is learning from data through several study algorithms. Meanwhile, Multilinear Regression helps the user in identifying which building design factor has high contribution in energy estimation. This work produces the information that RC, surface, glazing and roof area were 3 elements had high interaction and could be changeable as controlling the energy consumption level. The interactive level of each feature was measured through the correlation and model coefficient in interaction parts as can be figured in Table 5 in Appendix. Thus, the ANN generation and explanatory result of multilinear regression produced the reliable forecast of energy consumption. The advance of AI can deal with the issue of large data and regression result allows user to control the energy usage as adjusting on predictor feature.

In advance, to extend this model for various applications, it requires the intervention of another methods to give the reasonable result as well as maintain the further forecasting. Hence, the combination model is proposed to create the hybrid between linear and non-linear model to deal with the data confusion and also the overfitting problem.

## APPENDIX

TABLE 5. CORRELATION AMONG EACH INPUT OF ENERGY DATASET

| | x1 | x2 | x3 | x4 | x5 | x6 | x7 | x8 |
|---|---|---|---|---|---|---|---|---|
| x1 | 1 | | | | | | | |
| x2 | -0.9919 | 1 | | | | | | |
| x3 | -0.2037 | 0.1955 | 1 | | | | | |
| x4 | -0.8688 | 0.8807 | -0.2923 | 1 | | | | |
| x5 | 0.8277 | -0.8581 | 0.2809 | -0.9725 | 1 | | | |
| x6 | 0 | 0 | 0 | 0 | 0 | 1 | | |
| x7 | 1.2E-17 | 1.3E-16 | -7.9E-19 | -1.38E-16 | 1.86E-18 | 0 | 1 | |
| x8 | 1.7E-17 | -3.5E-16 | 0 | -1.07E-16 | 0 | 0 | 0.2129 | 1 |
| y1 | 0.622 | -0.658 | 0.456 | -0.862 | 0.889 | -0.003 | 0.270 | 0.087 |
| y2 | 0.63434 | -0.6730 | 0.4271 | -0.8625 | 0.8957 | 0.0142 | 0.2075 | 0.0505 |

## REFERENCE

[1] Chou, J. S., & Bui, D. K. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. Energy and Buildings, 82, 437–44, 2014.

[2] EIA. International Energy Outlook 2017 Overview. U.S. Energy Information Administration, IEO2017(2017), 143, 2017.

[3] Bonetto, R., & Rossi, M. Machine Learning Approaches to Energy Consumption Forecasting in Households, 6–9, 2017.

[4] Tsanas, A., & Xifara, A. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy and Buildings, 49, 560–567, 2012

[5] Zangheri, P., Armani, R., Pietrobon, M., Pagliano, L., Fernandez Boneta, M., & Müller, A. Heating and cooling energy demand and loads for building types in different countries of the EU, (March), 86, 2014.

[6] Sari, Y. Performance evaluation of the various training algorithms and network topologies in a neural-Network-Based inverse kinematics solution for robots. International Journal of Advanced Robotic Systems, 11(1), 2014.

[7] Burden, F., & Winkler, D.. Bayesian regularization of neural networks. Methods in Molecular Biology, 458, 25–44., 2008.

[8] Hong, T., Gui, M., Baran, M. E., & Willis, H. L. Modeling and forecasting hourly electric load by multiple linear regression with interactions. IEEE PES General Meeting, PES 2010, 1–8, 2010.

[9] Hong, W. C. Electric load forecasting by support vector model. Applied Mathematical Modelling, 33(5), 2444–2454, 2009.

[10] Faqs.org. (2018). comp.ai.neural-nets FAQ, Part 3 of 7: Generalization Section - How many hidden units should I use?. [online] Available at: http://www.faqs.org/faqs/ai-faq/neural-nets/part3/section-10.html [Accessed 16 Mar. 2018].