A Hybrid Approach of Using Both Simulation plus Neural Networks for Window Design Optimization and HVAC Energy Consumption Prediction Modeling

Yesuel. Kim and Seongju Chang
KAIST/Civil and Environmental Engineering, Daejeon, Republic of Korea
Email: yesuel21@kaist.ac.kr, schang@kaist.ac.kr

***Abstract*—Energy use in the building sector accounts for a large percentage of the world's total energy consumption. Specifically, the energy consumption from the whole life cycle perspective of building is 0.4% at design stage, 16% at construction stage, 83.2% at operation stage, and 0.4% at disposal stage. There have been many studies focusing on the design stage to find alternatives to enhance the energy efficiency of buildings. However, there have been few studies considering both of the efficient energy management of the building operation stage for the optimum design model at the same time. As a result of the design phase study, we proposed an improved window design alternative that could save 2736.06 kW of heating and cooling energy per year compared to the base case building. As for optimum window design, we proposed an ANN (Artificial Neural Network) model which predicts the heating and cooling loads. It satisfied the content of ASHRAE (American Society of Heating, Refrigerating, and AirConditioning Engineers) Guideline 14-2002 and IPMVP (International Performance Measurement & Verification Protocol). Based on this study, it would be possible to save energy from the perspective of a building's entire life cycle if window selection options standard that can be referenced at building design stage and heating and cooling system control algorithm applicable to the operation stage are developed together.**

***Index Terms*—heating and cooling load, window performance evaluation, building simulation, ANN**

I. Introduction

Energy use in the building sector accounts for a large percentage of the total energy consumption in the world. Ref. [1] The energy consumption of heating and cooling equipment in the building sector is increasing due to the increase in the size of the building and the improvement of the quality of life. Specifically, the energy consumption from the whole life cycle perspective of the building is 0.4% at design stage, 16% at construction stage, 83.2% at operation stage, and 0.4% at disposal stage. Therefore, for effective energy savings in the building sector, It could be argued that minimizing energy consumption at operational stage is the most important. Ref. [2] However, in general, it becomes reality that the energy saving of operation stage is not fully considered in the construction process of building.

Among building components, windows have insulation properties about 7 times lower than walls, and the heat loss of buildings through windows occupies 10 to 40% of the total heat loss, resulting in excessive heat loss. Since windows are major source of heat loss in terms of energy performance, they lead to an increase in heating and cooling loads. Ref. [3] Therefore, window performance must be considered from the design stage in order to reduce the energy consumption of buildings. In particular, a curtain wall method is widely applied to reduce the weight of the outer shell, ease of construction, aesthetic appearance and outlook. However, curtain wall windows cause a lot of heat loss due to a higher heat conduction rate than walls, and cause excessive solar radiations. Ref. [4] Therefore, if a window decision is made in perspective of energy saving during the design stage, the energy performance of the building can be improved by passive control method.

The Korean government is continuously inducing the reduction of window area ratio by strengthening the external heat pipe rate. Ref. [5] Nevertheless, it is necessary to provide a more detailed guideline considering the fact that design factors(window angle, window type, window to wall area ratio etc.) influence complexly on the heating and cooling load through windows. They implemented the "Window Energy Rating System" (hereinafter referred to as "Window Rating System") from July 2012, which can objectively evaluate the thermal performance of window products. Ref. [6] However, current regulations on window systems are restricted to U-value only, which is vulnerable to summer cooling load. Not only the U-value of the window but also the structural part(area ratio of the building envelope, Number of panes in external window etc.) should be considered for increasing energy efficiency of heating and cooling system. Ref. [7] However, there is no independent guidance on structural aspects of window design.

As such, there are constant problems that humanity has to solve in terms of energy in buildings. Ref. [8] To solve

Manuscript received May 17, 2018; revised July 8, 2019.

doi: 10.18178/ijscer.8.4.300-309

these problems, it is essential to analyze building energy performance to numerically calculate and predict energy consumption of buildings. Energy simulation, a tool for quantitatively analyzing the energy performance of a building, enables energy-efficient alternative analysis at initial design stage. Ref. [9] It is used to predict the heating and cooling load of a building, analyze energy usage, and evaluate the performance of each building element. However, in general, creating a dynamic simulation model has disadvantage such as expertise necessity and considerable time. Ref. [10] Especially, it takes a lot of time to find appropriate U-values for various building components, therefore required time for simulation increases as the building scale increases. Ref. [11] These dynamic simulation model characteristics affect performance evaluation attributes (accuracy, reproducibility, reliability).

To overcome these limitations at operational stage, which occupies a large percentage of the entire life cycle of a building, the concept of 'learning' that can predict the load more intelligently than the dynamic simulation model should be introduced based on the previous building facility operation data for efficient energy management. Ref. [12] Prediction of heating and cooling loads of HVAC systems in buildings is important to improve energy performance for energy conservation and environmental impact reduction. The load prediction in a building is a study that is the basis for optimizing the operation of HVAC system in a building. There are many techniques for finding the solution to predict through learning, but ANN has been developed more progressively and used more widely than other predictive methods.

In this study, we used DesignBuilder, a dynamic simulation tool, to find model alternatives that reduce energy consumption compared to the base case model through repetitive model case studies. The ANN was applied to the generated simulation building data of the model alternative to perform the heating and cooling load prediction. The purpose of this study is to develop a model of building that is energy efficient and ANN prediction algorithm. As mentioned above, among the various components of the building, energy efficiency evaluation was carried out through the dynamic simulation tool used at design stage to find the optimum window design alternative which has a positive influence on the energy consumption. Furthermore, we have developed a heating and cooling load prediction algorithm based on ANN learning concept.

II. Building Simulation Tool

EnergyPlus is a dynamic simulation tool that is widely used around the world. Ref. [13] It was reconstructed by combining the benefits of BLAST and DOE-2 with the US Department of Energy (D.O.E) in 1995. The programming language is FORTRAN90, and it has a modular structure, which is advantageous for enhancing functions. In order to create an EnerygyPlus model accurately, detailed input information (density, specific heat, long-wave absorption rate, short wave absorption rate, etc.) is required as well as simple information(thickness and thermal conductivity provided by drawings and specifications. Ref. [4, 14] Therefore, someone who are designing simulation model should assume the insufficient physical property information by referring to existing general values.

EnergyPlus supports thermal comfort calculations and modeling of radiation systems by applying detailed building physics on heat transfer, moisture and air, including radiation and convection. It also enables the calculation of natural light and shade, visibility measures. In addition, it is possible to design various HVAC system and heat source equipment models. And it is possible to simulate within one-hour time interval (Kim, 2018). EnergyPlus uses the thermal equilibrium equation for all surfaces and air in the room, not the transfer function method that is commonly used for building load calculations. It calculates the response factor(Con-duction Transfer Function) and the indoor air temperature for all the rooms by applying the thermal equilibrium equation. And then it calculates the heating and cooling load according to the difference between the calculated indoor air temperature and the indoor set-point temperature. The thermal equilibrium formulated as Eq. (1).

$$C_z \frac{dT_z}{dt} = \sum_{i=1}^{N_{sl}} Q_i + \sum_{i=1}^{N_{surface}} h_i A_i (T_{si} - T_z) + \sum_{i=1}^{N_{zones}} m_i C_p (T_{zi} - T_z) + \dot{m}_\infty C_p (T_\infty - T_z) + \dot{Q}_{sys} \quad (1)$$

where $\sum_{i=1}^{N_{sl}} Q_i$ means the sum of the convective internal loads, $\sum_{i=1}^{N_{surface}} h_i A_i (T_{si} - T_z)$ means convective heat transfer from the zone surfaces, $\sum_{i=1}^{N_{zones}} m_i C_p (T_{zi} - T_z)$ means heat transfer due to interzone air mixing, $\dot{m}_\infty C_p (T_\infty - T_z)$ means heat transfer due to infiltration of outside air. And $\dot{Q}_{sys}$ means air system output, $C_z \frac{dT_z}{dt}$ means stored energy in zone air, C_z means $\rho_{air} C_p C_T$. where ρ_{air} means zone air density, C_p means zone air specific heat, C_T means sensible heat capacity multiplier.

EnergyPlus, which has such a characteristic, belongs to the simulation engine, and it induces a way for a third developer to develop and provide the user interface so that various levels of users can use it for different purposes and uses. Therefore, EnergyPlus input and output files are configured in text format and other developers can configure these input and output files. Due to the textual input and output, it is very difficult to make the simulation with EnergyPlus alone. Therefore, a third party graphic user interface is needed to compensate for this. EnergyPlus's third party graphical user interface types include DesignBuilder, Openstudio, EBEST, Archism, and gEnergy. DesignBuilder and Openstudio are mainly used in Republic of Korea. Ref. [15] In this study, energy analysis was performed using Design Builder, a program developed in the UK, which is widely known as easy modeling, many templates, and building simulation tools including ASHRAE 90.1.

III. ANN PRINCIPLES

Ref. [16] An ANN mimicking the human nervous system was proposed by Warren McCulloch and Walter Pitts in 1943 based on an algorithm called mathematics and threshold logic. ANN belongs to supervised learning in which learning is performed from a given data set with input and output variable values. In this study, ANN is used as a regression technique. Multilayer neural network, which is a representative model of artificial neural network that imitates biological neural system, consists of input layer, hidden layer, and output layer. Each node composing the neural network receives the external input, calculates the sum by multiplying it by the weight, and processes the input values. Finally, the output value is derived by an activation function. In this case, weights are determined by the learning of the ANN as a part that reflects the influence of each input variable on the output variable. In addition, the activation function determines the result value according to the threshold value. Ref. [1] Typical activation functions are a step function, a sigmoid function, and a linear function. The structure of an ANN with these features can be expressed as shown in Fig. 1.

The 'learning' of ANN refers to the process of determining the weights multiplied by the values of input and output variables in the process of transferring values from neuron to neuron using given learning data and the bias value added to the corresponding values. The method of systematically changing the weights of neural networks according to given learning data is called 'learning rules'. In artificial intelligence, learning methods are divided into supervised learning and unsupervised learning. Supervised learning is a method of learning training data for a result to be derived through a classifier. Ref. [17] Training data sets, validation data sets, evaluation cross-validation is required through testing data sets. Super-vised learning can be divided into 'classification', which is used to determine which category of input data belongs to the use of the model, and 'regression', which is used to find a model that predicts the trend of input data. Ref. [18] Unlike supervised learning, unsupervised learning leads to self-learning without objective function when data is given. There are three representative methods for calculating the error in the learning of neural networks: Stochastic Gradient Descent (SGD), Batch, and Mini Batch. A typical learning rule of a neural network is a delta rule.

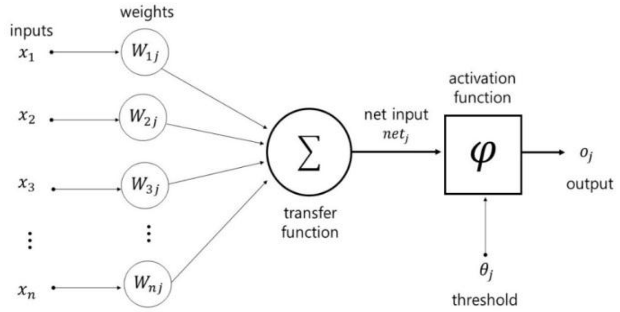

Figure 1. Structure of ANN

The formula of the delta rule is slightly different depending on the active function. The delta rule is formulated as Eq. (2).

$$w_{ij} \leftarrow w_{ij} + \alpha\delta_i x_j \tag{2}$$

In the above equation, w_{ij} represents the connection weights of the output node j and the input node j, and α represents the learning rate from 0 to 1. δ_i is an arbitrary active function which is formulated as Eq. (3).

$$\delta_i = \varphi'(v_i)e_i \tag{3}$$

where e_i is the error of the output node, v_i is the weighted sum of the output nodes, and φ' is the derivative of φ, which the active function of the output node.

This delta rule is a technique to find the correct answer step by step through repetitive tasks. Therefore, the learning data must be relearned repeatedly until the error of the neural network is sufficiently reduced. A monolayer neural network with such a delta rule as a learning rule has a limitation that it can not be applied to linearly inseparable problems and can be applied only to linearly seperable problems. Multilayer neural networks have been developed to overcome these drawbacks. The backpropagation algorithm, which is a typical algorithm for learning the multilayer neural network, propagates the output error of the neural network in reverse order from the output layer, defines the error of the hidden layer, and updates the weight of each layer by the delta rule.

The back propagation algorithm is a method of reducing the error by adjusting the connection weights between the neurons according to the magnitude of error between the output value and the target value derived through the forward propagation process. Thus, the significance of the back propagation algorithm is to provide a systematic way to define the error of the hidden nodes in the multilayer neural network. In the backpropagation algorithm, Tansig (Tangent Sigmoid Function) is used as the activation function of the hidden layer and Purelin (Pure Linear Function) is used as the activation function of the output layer. In addition, the cost function is a function to calculate the output error of the neural network. The cost function has a value proportional to the error from the viewpoint that the larger the error, the more the cost is required. The cost function is formulated as Eq. (4).

$$E(\vec{w}) = \frac{1}{2}\sum_{d=D}(t_d - O_d)^2 \tag{4}$$

In the above equation, D is the number of learning data sets, t_d is the actual value of the output variable of the d-th data in the training data, and O_d is the estimated value of the output variable using the input variable of the dth training data.

IV. Base Case Building Modeling

In this section, we compare and analyze the energy demand of the heating and cooling in the building which depends on the design of the window. In order to apply the physical property information to the actual situation in Republic of Korea, a building in Seocho-gu, which is a case building, was selected as a test bed. For the modeling and energy analysis of selected modeling cases, we used DesignBuilder, a dynamic simulation tool at the design stage. Ref. [4] Simple information(specifications, wall properties, thermal conductivity, etc.) provided by drawings or specifications required in the building modeling process can be obtained easily, but detailed information(density, specific heat, long wave absorption rate, short wave absorption rate etc.) need to be assumed by the simulationist. Therefore the insufficient information was input as general default values provided by the design builder. The typical modeling input information of the test bed building is shown in Table I.

TABLE I. DesignBuilder Model Inputs

Location	Seocho 2-dong, Seocho-gu, Seoul, Republic of Korea		
Building Type	Office Building		
Floor	8 floors		
Building scale	Floor Area	383.49 m^2	
	Depth	1 ~ 7 F	3,300 mm
		8 F	3,450 mm
	Width	North facing wall	27,463 mm
		South facing wall	28,763 mm
	Height	13,800 mm	
External Wall	Configuration	- Stone Granite : 30 mm - PUR Polyurethane : 70 mm - Concrete Block : 200 mm - Cement/Plaster/Mortar : 18 mm	
Window	Type	Double-Pane Color Windows	
	Configuration	External Pane	- 6 mm - Color LowE
		Gass	- 13 mm - Argon Gas
		Internal Pane	- 6 mm - Clear
	SHGC (Solar Heat Gain Coefficient)	0.284	
	U-value	1.33	
Window to Wall ratio (W/W)	0.82		
HVAC System	System Type	VRF (Variable Refrigerant Flow)	
	CoP (Coefficient of Performance)	Heating System	3.5
		Cooling System	3.86
Lighting	Type	LED	
	Luminaire Type	Recessed (Ceiling-mounted lighting)	
Equipment, Device Operation Schedule	6 a.m - 9 p.m		
Occupancy Schedules	6 a.m - 9 p.m		

V. Window Case Study

In this section, a case study was conducted according to 11 types of window designs commonly used in Republic of Korea. The default setting is a double low-e window with argon gas injected between the panes as described in the previous section, and it is named case 1. Ten additional window cases were designed by applying four criteria such as 'window to wall area ratio', 'injection gas', 'low-E coated Yes or No' and 'pane number' to case 1. The diagram of the design process of the window case is shown below. (see Fig. 2).

As shown in Fig. 2, Case 1 corresponds to the base case, and the number of panes is 2, the window to wall area ratio (w/w) is 0.82, the injected gas is argon and the glass is low-E coated. Based on the number of panes, which are the most comprehensive classification criteria, we classified them into two groups consisting of double pane and triple pane window. In case of double pane window group, it was configured with case2, case 3, case 4 and case 5. These were designed by reducing window to wall area ratio to four different proportions in order to investigate the energy saving when the window area is reduced in comparison with the base case. We also designed case 6, case 7 by changing the type of injected gas, changing the low-E coated pane to clear while maintaining the same remaining conditions in the base case. In case of triple pane window group, first we classified the glasses into Clear glass group and low-E coated glass group according to whether they were low-E coated or not. Once again, the final case 8, case 9, case 10, and case 11 alternatives were categorized according to whether the injected gas was argon or air. The specific properties of these generated cases are shown in Table Ⅱ.

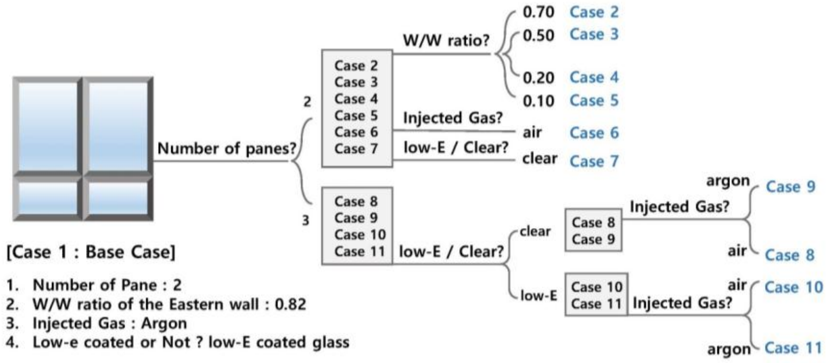

Figure 2. Test cases organization

TABLE II. CASE CLASSIFICATION

No	Type	Configuration	W/W	U-value	SHGC
1	Double lowE	color lowE + argon + clear (6mm + 13mm + 6mm)	0.82	1.363	0.285
2	Double lowE	color lowE + argon + clear (6mm + 13mm + 6mm)	0.70	1.363	0.285
3	Double lowE	color lowE + argon + clear (6mm + 13mm + 6mm)	0.50	1.363	0.285
4	Double lowE	color lowE + argon + clear (6mm + 13mm + 6mm)	0.20	1.363	0.285
5	Double lowE	color lowE + argon + clear (6mm + 13mm + 6mm)	0.10	1.363	0.285
6	Double lowE	color lowE + air + clear (6mm + 13mm + 6mm)	0.82	1.651	0.295
7	Double Clear	clear + argon + clear (6mm + 13mm + 6mm)	0.82	2.511	0.704
8	Triple Clear	clear + air + clear + air + clear (6mm + 13mm + 6mm + 13mm + 6mm)	0.82	1.726	0.612
9	Triple Clear	clear + argon + clear + argon + clear (6mm + 13mm + 6mm + 13mm + 6mm)	0.82	1.593	0.613
10	Triple lowE	color lowE + air + clear + air + lowE (6mm + 13mm + 6mm + 13mm + 6mm)	0.82	0.885	0.234
11	Triple lowE	color lowE + argon + clear + argon + lowE (6mm + 13mm + 6mm + 13mm + 6mm)	0.82	0.304	0.677

VI. BUILDING SIMULATION RESULTS

In this chapter, we have examined how the heating and cooling load changes compared to case 1 for the additional cases defined in the previous section. We also selected the model that showed the greatest amount of energy saving compared to case1. The following is a list of building cases in order of energy savings: case 5, case 11, case 4, case 10, case 3, case 10, case 3, case 2, case 6, case 9, case 8 and case. The Table Ⅲ shows the heating and cooling load, energy saving ranking for each building case. According to the table, two of the most energy saving alternatives are case 5, which reduces the window to wall area ratio of the eastside external wall to 0.1, and case 11, which is a Argon injected triple low-E coated window. It is shown that energy savings for case 5 and case 11 amounted to -3044.72 kW and -2736.06 kW, respectively.

Through a comprehensive analysis of these, it was confirmed that the reduction of window to wall area ratio at the design stage positively affects energy saving. However, due to the site conditions of the base case building(case1) surrounded by other buildings in the south, north, and west, it is necessary to guarantee the natural lights and views of the east side wall windows. Therefore, it was concluded that the application of triple low-E coated window rather than reducing the window to wall area ratio of the east side wall would be advantageous both in terms of the building energy as well as the occupant comfort. Finally case 11, the second most energy saving case, was selected as an optimal alternative for this study. The building-related data of case 11 was used for energy prediction in the next chapter.

TABLE III. ENERGY SIMULATION RESULTS

case no.	Heating [kW]	Cooling [kW]	Total H+C [kW]	Energy Savings [kW]	ranking
1	49889.70	41454.21	91343.82	-	7
2	49948.10	40906.00	90854.11	-489.71	6
3	50045.14	39967.72	90012.86	-1330.96	5
4	50121.78	38601.86	88723.64	-2620.18	3
5	50146.12	38152.98	88299.10	-3044.72	1
6	51503.37	41246.35	92749.72	+1408.90	8
7	49260.38	49628.12	98888.50	+7544.68	11
8	52109.43	43954.45	96063.88	+4720.06	10
9	46519.23	48785.15	95304.38	+3960.56	9
10	49081.62	40573.61	89655.23	-1688.59	4
11	47823.51	40784.26	88607.76	-2736.06	2

VII. ANN Input Variables Selection

In this chapter, we have developed the energy prediction ANN model through the data generated in the previous chapter. First, input variable selection process is necessary. This process is essential to avoid using irrelevant parameters in predictions and to construct a set of input variables based on those that are most relevant to target data [12]. There are many ways to select input variables. In this study, we analyzed the correlation between input and output variables through regression analysis and used the coefficient of determination (R^2) as an index. Regression analysis is a statistical technique that measures the correlation between one or more input variables and output variables. R^2 is a measure of how well the regression line describes the actual data, and can be formulated as Eq. (5).

$$\text{Coefficient of Determination}(\mathrm{R}^2) = \frac{(\mathrm{S_{xy}})^2}{\mathrm{S_{xx}} \times \mathrm{S_{yy}}} \quad (5)$$

where

$$S_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \quad (6)$$

$$S_{xx} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (7)$$

$$S_{yy} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \quad (8)$$

The coefficient of determination is expressed as a value between 0 and 1, and the closer to 1, the closer the points of the scatter plot in the graph are concentrated around the regression line. That is, it means that the input variable has a large influence on the output variable. Among the many building data generated from the simulation in the previous section, we firstly extracted a list of internal and external variables related to the window, which is the building element covered in previous chapter. The relationship between the extracted input parameters and the heating and cooling load, which is an output parameter, was analyzed. The Table Ⅳ summarizes the results. The results show that Mech Vent/Net Vent/Infiltration and the heating and cooling load has the greatest correlation which R^2 was 0.3638. The R^2 of solar altitude variable was 0.1706, followed by 0.0851, the R^2 of solar gains exterior windows. On the other hand, the lowest correlated variable was solar azimuth with 0.0001 R^2 value. The correlation coefficients are variously derived when correlation analysis of input parameter and output parameter is performed. Therefore, it is essential to select optimal input variables through correlation analysis. In this study, it was concluded that the variables ranging from 1 to 6 have a relatively good correlation with the output parameter. As a result, 6 variables(mech vent/net vent/infiltration, solar altitude, solar gains exterior windows, air temperature, relative humidity, wind speed) were selected as the final input parameters.

TABLE IV. Relationship between Input and Output Variables

input parameters	R^2	ranking
Air Temperature	0.042400	4
Outside Dry-Bulb Temperature	0.000761	9
Solar Gains Exterior Windows	0.085100	3
Relative Humidity	0.031950	5
Mech Vent/Net Vent/Infiltration	0.363800	1
Wind Speed	0.018670	6
Wind Direction	0.001973	8
Solar Altitude	0.170600	2
Solar Azimuth	0.000100	10
Atmospheric Pressure	0.002440	7

VIII. Data Preprocessing

Before determining the structure of the ANN model, it is essential to normalize the input and output datasets to ensure the accuracy of the model [19]. Normalization is a preprocessing technique that allows data of each dimension to have values within the same range. This process is essential because each data item has different units or ranges. The input and output sample data are normalized from 0 to 1. The data normalization process can be formulated as Eq. (9).

$$X = \frac{x_i - x_{min}}{x_{max} - x_{min}} \times 0.8 + 0.1 \quad (9)$$

In the above equation, x_i means value for normalization, x_{min} means minimum value, and x_{max} means maximum value.

IX. Initial ANN Model Development

Ref. [20, 21] An initial model based on an artificial neural network was developed through the MATLAB neuron network toolbox. The data used in the ANN development is 1-year simulation data composed by 8760 points with variable building parameters generated at 1-hour intervals, 24h a day. More specifically the data set consists with case 11 building selected as the optimal alternative among the 11 building cases in the previous chapter. In addition the simulation data is split into three different datasets, which is for training, validation and testing the ANN model. The training data set is used for obtaining the ANNs parameters. The validation data set is necessary in the training and post training processes in order to prevent the over-training in the ANNs. The test data set is used to provide an independent data set for benchmarking and testing purposes. The distribution ratios of the training, validation, and testing data sets to the total data are 0.75, 0.15, and 0.15. The number of neurons in the output layer of the model is one, and the heating and cooling load, which is the final value to be predicted, is determined as the output value. The number of input neurons is 6, and the input layer of the model is composed of the six variables that have the highest correlation with the output variable in the input variable selection process. The conditions and structure of the initial model of artificial neural network designed with these considerations are shown in the following table. Ref. [22, 23] As shown in Table Ⅴ, the number of hidden neurons in the initial artificial neural network was set to 1, and the number of initial neurons in the hidden layer was selected to be 13 according to Eq. (10). The hidden and output layer activation functions are applied as Tangent Sigmoid, Pure Linear which is given as a basic function in MATLAB Toolbox. We use the most widely used feedforward backpropagation network as a model learning algorithm. The learning function is the LMA (Levenberg-Marquardt Algorithm, TRAINLM) [24], which has proven to be the fastest in various studies. Tansig was used as the activation function of the hidden layer and purelin was used as the activation function of the output layer. The goal is arbitrarily set to 0.0 in the toolbox, which means that learning continues until the calculated output value and the ideal output value are equal. However, since the number of learning epochs is set to 1000, it is automatically completed when the learning process is repeated 1000 times. After setting the condition and structure of the model, we created the initial mod-el as a computer language with the neural network toolbox of MATLAB.

TABLE V. INITIAL ANN MODEL

Model Components		Contents
Structure	Input Layer	Number of Neurons : 6 1) Indoor Air Temperature 2) Solar Gains Exterior Windows 3) Relative Humidity 4) Mech Vent/Nat Vent/Infiltration 5) Wind Speed 6) Solar Altitude
	Hidden Layer	Number of layers : 1 Number of Neurons : 13 $N_h = 2N_i + 1$ (10) N_i means the number of input nodes.
	Output Layer	Number of Neurons : 1 1) Energy(Heating/Cooling Load)
Transfer Function	Hidden Neurons	Tangent Sigmoid
	Output Neurons	Pure Linear
Training Method	Epoch	1000 times
	Algorithm	Levenberg-Marquardt
	Number of datasets	Total 8760 datasets (1year, 1hour interval) - 5836 : training (70%) - 1459 : validation (15%) - 1465 : Testing (15%)

X. ANN MODEL OPTIMIZATION PROCESS

The parameter optimization process of the initial model was developed. In the ANN model, the optimization process is essential for improving the prediction performance of the model. In general, the parameters considered in the Levenberg-Marquardt learning are the number of epoch, the goal, the number of hidden layers, and the number of hidden neurons. As a result of repetitive experiments with different epochs and goals, the parameters do not have a significant effect on the data learning of the present study. So we set the epoch to 1000 and goal to 0.0 which is given as default values in the neural network toolbox of MATLAB. Generally, as the number of hidden layers and the number of hidden neurons increase, the connection weight increases and the prediction performance can be improved. However, it takes a long time to calculate and there is a disadvantage that the overfitting problem may occur. Therefore, it is important to select the optimal number of these in the ANN model. In this chapter, the final model is derived by analyzing how the performance of the model is changed while increasing the number of hidden neurons from 1 to 20 and the number of hidden layers from 1 to 10. The MBE (Mean Biased Error, %) value between the predicted value and the input value was calculated according to the change of each parameter value. The parameter condition for deriving the smallest MBE was determined to be the optimal value. The MBE indicates the bias of the predicted value relative to the measured value, and the closer to zero the result is, the closer to the measured value. The indicator MBE is formulated as Eq. (11).

$$MBE = \frac{\sum_{i=1}^{n} |S_i - M_i|}{\sum_{i=1}^{n} M_i} \times 100 \tag{11}$$

In the above equation, S_i represents the predicted heating and cooling load data generated through the prediction process of the ANN model, and M_i represents the initial heating and cooling load data of the target building generated by the design builder as data before the prediction.

XI. ANN MODEL OPTIMIZATION RESULTS

The results of the parameter optimization process show MBE of 32.282 for 1 hidden neuron and 13.1661 for 15 MBE. The MBE result analysis graph of the number of hidden neurons 1 to 20 is as follows (see Fig. 3).

Next, the number of hidden layer neurons was fixed to the optimum number of 15, and the process of optimizing the number of hidden layers was performed. In the case of the number of hidden layers, the MBE value is 13.1661 when the number is 1, and the lowest MBE is 7.4115 when the number is 5. The graph of MBE result analysis for the number of hidden layers 1 to 10 is as follows (see Fig. 4).

Therefore, the optimized ANN model developed in this study showed the smallest MBE value of 7.4115 when the number of hidden neurons and hidden layers were 15 and 5, respectively.

XII. MODEL PERFORMANCE EVALUATION CRITIA

Ref. [25] The ASHRAE Guideline 14-2002 provides statistical criteria for comparing and evaluating the measured data and simulation results. Predictive performance evaluation uses MBE, Cv(RMSE), and R^2as a measure. R^2 is a measure of how close the predicted value of the proposed model is to the regression line of the measured value. According to ASHRAE Guideline 14, the correlation is considered appropriate when the R^2value is 0.75 or more. In addition, the MBE, which is a measure for confirming the normality of the model, has good prediction performance when the value is within ± 10%. Cv(RMSE), which is a measure of the accuracy of the model, defines good predictive performance when the value is within 30%. Ref. [26] Similarly IPMVP also

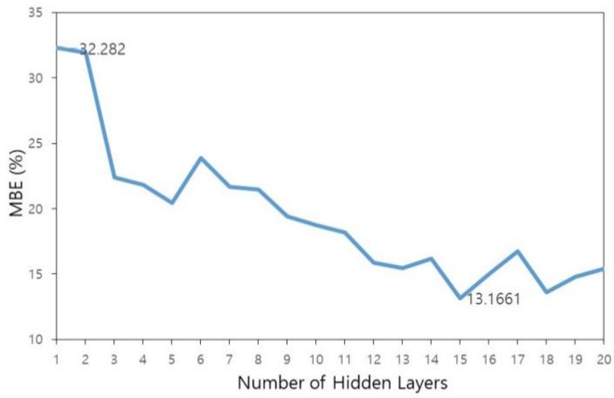

Figure 3. Optimization of the number of hidden neurons

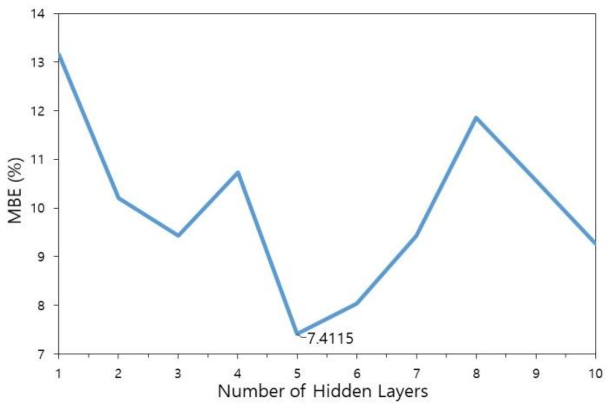

Figure 4. Optimization of the number of hidden layers

provides the criteria for performance evaluation of predictive models. Details of these performance criteria are shown in Table VI.

TABLE VI. PREDICT PERFORMANCE EVALUATION CRITIA

	MBE	Cv(RMSE)	R^2
ASHRAE	smaller than ±10%	smaller than ±30%	bigger than 0.75
IPMVP	smaller than ±20%	smaller than ±20%	bigger than 0.75

Cv(RMSE), which is one of the measures used in the above performance criteria, is a method of obtaining the Cv(coefficient of variation) of the RMSE(root mean square error). This is an error analysis method that grasps the error of the model considering the variance. The result is expressed as the error rate(%) and is generally calculated as a positive number. Also, the closer to 0%, the more accurate the predicted value, which is formulated as Eq. (12). The formulas for $R^2$and MBE are expressed in Eq. (5) and Eq. (11), respectively.

$$Cv(RMSE) = \frac{RMSE}{A} \times 100 \tag{12}$$

where

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(S_i - M_i)^2}{n}} \tag{13}$$

In the above equations, S_i generated through the ANN model prediction process means predicted heating and cooling load data, and M_i means initial heating and cooling load data of the target building generated by the design builder as the data before the prediction. A represents the average value of M_i , and n represents the total number of data, which is 8760 in this case.

XIII. Model Performance Analysis

In order to evaluate the prediction performance of the finally developed artificial neural network model, the data generated by the design builder simulation and the developed model predictive data were compared and analyzed using the three performance evaluation scales R^2, MBE, and Cv(RMSE). As a result, R^2 was 0.9592, MBE was 7.4115, and Cv(RMSE) was 15.18. Therefore, it satisfies ASHRAE Guideline 14-2002, which is a criterion that the MBE should be smaller than 10, Cv(RMSE) should be smaller than 30, and that it should be larger than 0.75 in the building performance evaluation criteria presented above. It also satisfied IPMVP, which is a standard for the requirement that hidden neuron should be less than 20, Cv(RMSE) should be less than 20, and R^2 should be greater than 0.75. As a result, this ANN model has proven to be suitable as a load prediction model. The scatter plot and the regression plot were used to analyze how well the output of the model fit with the actual data. It can be seen from the scatter plot below that the densities are concentrated close to the diagonal line of the upper right, meaning the fitting line. In addition, it is confirmed that the trend of the line representing the predicted data (blue line) and the line representing the existing design builder data (gray line) are almost similar. We demonstrated the regression fitting and accuracy of this ANN result model according to the numerical result analysis and the analysis result on the graph (see Fig. 5).

XIV. Conclusions

This study proposes a method to increase energy efficiency of heating and cooling by establishing one test bed building in two phases: design and operation. The process and conclusion of this study are summarized as follows.

For design stage study, the window element, which is a building element that has a great influence on the efficiency of the heating and cooling system, is set as a case study building element. And we additionally designed 10 kinds of window cases generally used in Republic of Korea. After applying the designed case to the base case building, we conducted the analysis of the energy consumption of the heating and cooling system. The case of saving 2736.06 kW of heating and cooling load per year compared with the base case was decided as the final model alternative (building case 11). For building modeling and energy analysis, a representative building simulation tool, Design Builder, was used. It has been confirmed that the appropriate window design of the building alone can save a considerable amount of energy annually during the building life cycle in a passive manner.

To propose energy efficiency enhancement measures for the operational stage study, heating and cooling load prediction algorithm was developed with ANN, which is one of the widely used machine learning techniques. This study can be considered as an extension of the design stage study which suggests new energy efficient building design. The building-related data for the final model (building case 11) selected in Chapter 3 was created with

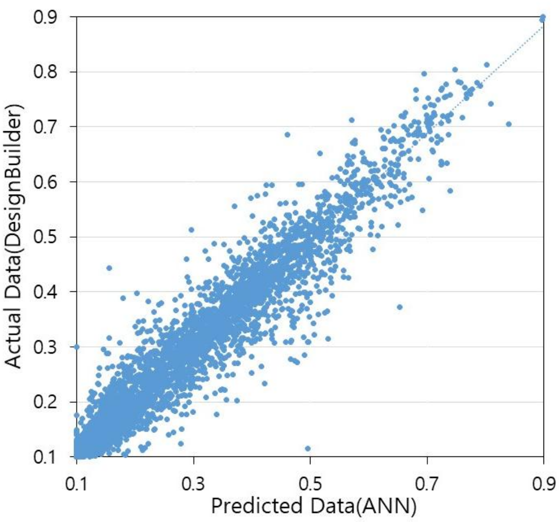

Figure 5. Correlation of Simulation and ANN learning

design builder, which is dynamic building simulation tool. The generated building data was used in the ANN model development process in Chapter 4. R^2, which is a measure of the correlation between data, was used to select ANN input variables. The data set consists of a total of 8760 data generated by design builder at 1 hour intervals including 6 input variables and 1 output variable. In addition, for the data preprocessing process the data was normalized to a value between 0 and 1 to improve the computational efficiency of the ANN model. Next, for the model optimization, repeated error analysis was performed by changing the number of hidden neurons and hidden layers in the basic model. The number of hidden neurons and hidden layers for the smallest error were defined as the optimal model condition. MBE was used as a measure of the error, and it was confirmed that the smallest MBE value was 7.4115 when the number of hidden neurons and hidden layers were 15 and 5, respectively. In order to evaluate the prediction performance of the finally developed ANN model, the data generated by the design builder simulation and the output data of ANN model were compared and analyzed using the three performance evaluation scales R^2, MBE, and Cv(RMSE). As a result, R^2 was 0.9592, MBE was 7.4115, and Cv(RMSE) was 15.18. As a result, it was confirmed that the proposed method meets the ASHRAE Guideline 14-2002 and IPMVP, which is the evaluation criterion of building performance. With the development of this ANN algorithm, it is possible to predict the accurate load even with a small amount of building information. The model can be used for hourly heating and cooling load prediction of HVAC systems.

The contents of this study can contribute to the optimal window design process at the design stage of the building. It will also contribute to the development of an efficient energy management system at the operational stage. As a result, it can contribute to improving energy efficiency from the perspective of the entire building life cycle. The one limitation of this study are that the design stage only considers general window cases of Republic of Korea and does not consider more special window cases because of complexity and time consuming of the work. In fact, the Republic of Korea government is continuously inducing the reduction of window to wall area ratio by strengthening the external wall heat transfer rate in the domestic market. Ref. [12] However, it is necessary to provide a more detailed guideline considering overall design factors such as window angle, window type, window area ratio. If more specific case are considered in the scope of the study additionally in the future, it will be possible to form reference guidelines for selecting appropriate windows in the design stage of the building. The other limitation is that the ANN algorithm developed in the operation phase study did not apply to actual building data. Therefore, it is necessary to expand the research scope to the optimization control that can confirm the suitability of prediction and the energy reduction performance by utilizing BEMS (Building Energy Management System) data of actual building.

Based on the process of this study it will be possible to increase the energy efficiency throughout the entire life cycle of the building, including both design stage and operation stage. Therefore, it is expected that the design of the building will be done in a comprehensive and efficient manner considering not only the energy saving in the passive method at the design stage but also the active system energy control at the operating stage of the building.

ACKNOWLEDGMENT

This paper is based on a research which has been conducted as part of the Climate Change Research Hub of KAIST (Grant No. N11180107)

REFERENCES

[1] S. Kang, “Development of ANN (Artificial Neural Network) based predictive model for energy consumption of HVAC system in office building,” M.S. thesis, Dept. Archi. Eng., Chung-ang Univ., Seoul, Korea, Aug. 2017.

[2] H. G. Jeong, "Geonmul-eseoui BEMS doib-eul wihan gisulgaebal salye(Examples of technology development for introduction of BEMS in buildings)," *Korea Green Building Council*, vol. 16, no. 2, pp. 38-47. June 2015.
[3] H. J. Kwak, H. I. Jang, C. Y. Lee, and S. J. Seo, "Comparison on the energy consumption of building according to fenestration energy performance rating system," presented at the SAREK Conference 2012, vol. 11, pp. 113-116, Nov. 2012.
[4] Y. S. Kim, "Empirical comparison of energy performance assessment by ISO 13790 and EnergyPlus," M.S. thesis, Dept. Future Cit., Sungkyun-kwan Univ., Suwon, Korea, Aug. 2016.
[5] S. H. Kang, H. J. Yu, B. H. Lee, Y. H. Jo, S. S. Kim, and K. W. Kim, "A comparison of energy consumption on facade design factors in office building," presented at the Korean Institute of Architectural Sustainable Environment and Building Systems Spring Conference, vol. 3, pp. 145-148, Mar. 2012.
[6] W. J. Hwang, K. J. Kim, and W. K. Choi, "The analysis on energy performance according to characteristics of glazing in high-rise office buildings," *Korean Solar Energy Society Spring Conference*, vol. 4, pp. 156-161, Apr. 2011.
[7] H. C. Yu, Y. H. Oh, and S. K. Park, "The optimal window system of office buildings considering energy efficiency," *Journal of Korean Solar Energy Society*, vol. 25, no. 4, pp. 53-60, Dec. 2005.
[8] Study for standardization of quantitative evaluation method of building energy performance, Ministry of Land Infrastructure and Transport Republic of Korea, Dec. 2014.
[9] S. B. Kim, "A study on the uncertainty of domestic building energy simulation," M.S. thesis, Dept. Archi. Eng., Daejeon University, Daejeon, Korea, Aug. 2016.
[10] S. A. Kalogirou, C. C. Neocleous, and C. N. Schizas, "Building heating load estimation using artificial neural networks," in *Proc. the 17th International Conference on Parallel Architectures and Compilation Techniques*, vol. 8, pp. 14, Sep. 1997.
[11] C. S. Park, "Normative assessment of technical building performance," *Journal of Architecture Institute (Planning)*, vol. 22, no. 11, pp. 337-344, Nov. 2006.
[12] H. Chaowen and W. Dong, "Prediction on hourly cooling load of buildings based on neural networks," *International Journal of Smart Home*, vol. 9, no. 2, pp. 35-52, Jan. 2015.
[13] D. B. Crawley, K. L. Linda, O. P. Curtis, and C. W. Frederick "EnergyPlus: New, capable and linked," *World Renewable Energy Congress VIII*, At Denver, Colorado, pp. 292-302, Nov. 2004
[14] K. U. Ahn, Y. J. Kim, D. W. Kim, S. H. Yoon, and C. S. Park, "Difficulties and issues in simulation of a high-rise office Building," in *Proc. the 13th Conference of IBPSA*, Chambery, France, vol. 2628, pp. 824-831, Aug. 2013.
[15] Kharn homepage. [Online]. Available: http://www.kharn.kr/news/article.html?no=3961, last accessed 2018/08/03.
[16] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115-133, Dec. 1943.
[17] H. J. Lee, "A study on the target group selection method in supervised learning model," M.S. thesis, Dept. Statistics, Chung-ang University, Seoul, Korea, Dec. 2008.
[18] C. P. Kim, S. W. Noh, and S. L. Park, "Analysis on the distribution of RF threats using unsupervised learning techniques," Journal of KIMST, vol. 19, no. 3, pp. 346-355, Jun. 2016.
[19] H. K. Ghritlahre and R. K. Prasad, "Development of optimal ANN model to estimate the thermal performance of roughened solar air heater using two different learning algorithms," *Ann. Data. Sci*, vol. 5, no. 3, pp. 453-467, Jan. 2018.
[20] H. Demuth and M. Beale, "Neural Network Toolbox for use with Matlab--User'S Guide Verion 3.0," Apr. 1993.
[21] H. B. Demuth and M. H. Beale, "Neural network toolbox; for use with MATLAB; Computation, Visualization, Programming; User's Guide, Version 4," Math Works, 2000.
[22] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Neural Networks for Perception*, pp. 65-93 1992.
[23] D. Datta, S. A. Tassou, and D. Marriott, "Application of neural networks for the prediction of the energy consumption in a supermarket," in *Proc. the International Conference CLIMA*, pp. 98-107, 2000.
[24] J. J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory," in *Numerical Analysis*, pp. 105-116. Springer, Berlin, Heidelberg, 1978.
[25] Guideline, A. S. H. R. A. E., Guideline 14-2002, Measurement of Energy and Demand Savings. American Society of Heating, Ventilating, and Air Conditioning Engineers, Atlanta, Georgia, 2002.
[26] IPMVP. International Performance Measurement and Verification Protocol, 1997.

Yesuel Kim received the B.Sc. degree in engineering(architectural) from Inha University, Incheon, Republic of Korea, from 2011 to 2016. After completing her B.Sc., she is attending a M.Sc. in Korean Advanced Institute of Science and Technology(KAIST) since 2016. Her research interest focuses on intelligent systems such as Artificial Neural Network and their applications on buildings, energy management systems. She is also interested in research that finds building alternatives that show optimal energy efficiency using dynamic simulation tools in the design phase of a building.

Seongju Chang is an expert in the field of smart & green building systems. He is currently an invited professor of the department of Civil & Environmental Engineering at KAIST, Korea. Prof. Seongju Chang got his bachelor's and master's degrees at Seoul National University in Korea majoring in architecture and received Ph.D. degree in Building Performance & Diagnostics at Carnegie Mellon University, Pittsburgh, USA in 1999. His current research interest includes various smart and sustainable architectural/urban systems including smart building components, energy efficient and resource circulative sustainable buildings and cities. He is also listed on the Marquis Who's Who as one of the potential contributors for resolving World's problems.