Article

Deep Neural Network Approach for Prediction of Heating Energy Consumption in Old Houses

Sungjin Lee [1,2,*] **, Soo Cho** [1] **, Seo-Hoon Kim** [1] **, Jonghun Kim** [1] **, Suyong Chae** [1] **, Hakgeun Jeong** [1] **and Taeyeon Kim** [2]

1   Korea Institute of Energy Research, 152, Gajeong-ro, Yuseong-gu, Daejeon 34129, Korea; scho@kier.re.kr (S.C.); takesky@kier.re.kr (S.-H.K.); jonghun@kier.re.kr (J.K.); sychae@kier.re.kr (S.C.); hgjeong@kier.re.kr (H.J.)
2   Department of Architecture and Architectural Engineering, Yonsei University, 50, Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea; tkim@yonsei.ac.kr
*   Correspondence: sjleeki@kier.re.kr; Tel.: +82-42-860-3172

Abstract: Neural network models are data-driven and are effective for predicting and interpreting nonlinear or unexplainable physical phenomena. This study collected building information and heating energy consumption data from 16,158 old houses, selected key input variables that affect the heating energy consumption based on the collected datasets, and developed a deep neural network (DNN) model that showed the highest accuracy for the prediction of heating energy consumption in an old house. As a result, 11 key input variables were selected, and an optimal DNN model was developed. This optimal DNN model showed the highest prediction accuracy (R^2 = 0.961) when the number of hidden layers was five and the number of neurons was 22. When the optimal DNN model was applied for the standard model of low-income detached houses, the prediction accuracy ($Cv(RMSE)$) of the optimal DNN model, compared to the EnergyPlus calculation result, was 8.74%, which satisfied the ASHRAE standard sufficiently.

Keywords: old detached house; prediction of heating energy consumption; deep neural network; data-driven model approach

1. Introduction

According to the 2018 National Housing Information Survey of the Korean Statistics Information Service (KOSIS), there are about 17.63 million houses in Korea, 9% of which were built before 1979 when housing insulation standards were enacted. Of those buildings, 47% are more than 20 years old [1]. Old houses are vulnerable to a lack of insulation and poor airtightness, which may cause heat loss and thus excessive energy consumption. For this reason, the government announced the 3rd National Energy Plan and has made efforts to improve the energy welfare system and to reduce the energy consumption of old houses. Old houses, however, are smaller in size and more numerous than general buildings. In addition, due to lack of diagnostic equipment, manpower and the long diagnostic time required, it is difficult to measure all parameters affecting energy consumption and to predict energy consumption by the parameters.

There are two model approaches to predict building energy consumption: physics-based models and data-driven models [2]. Physics-based models are based on the laws of thermodynamics and physics. EnergyPlus, eQuest, and Trnsys are representative physics-based energy-simulation software tools developed on this model approach [3]. They calculate building energy consumption using building parameters, air-conditioning and heating equipment system parameters, as well as environmental parameters such as building construction details, operation schedules, HVAC (Heating, Ventilation, and Air Conditioning) design information and climate, sky, and solar/shading information. However, the physics-based model requires a lot of computation time and resources for simulation [4] and, in many cases, it does not accurately reflect the thermal performance of the actual

building because it simplifies the model to account for the lack of available detailed building information at the time of simulation [5]. On the other hand, the data-driven model is an effective method for modeling physical phenomena for which the theory is unknown or unexplainable. This approach has recently attracted attention from researchers [6–10] because it can simulate energy consumption from available building information and energy data without the need for detailed modeling and the numerous input parameters required by the physics-based approach. The data-driven approach, instead, requires enough data to obtain accurate simulation results, and requires insight into the appropriate preprocessing of datasets and the understanding of simulation results.

Neural network modeling is a data-driven approach used widely in the prediction of building energy consumption [11–20]. Matlab Neural Network Toolbox, TensorFlow/Keras, and PyTorch are representative energy-simulation software tools widely used for neural network modeling. ANN is a learning algorithm created by simulating human neural networks. Early ANN started with a single hidden layer in the neural network structure. This structure had been applied to nonlinear regression analysis. However, with growing inputs dimensions and interference components, the shallow layer ANN model cannot fit these situations, while the deep neural network (DNN) is capable of meeting the requirements such as better accuracy, deceased computational time and noise robustness.

The neural network model has been used by many researchers for building energy simulation. González and Zamarreño [16] proposed the use of a data-driven approach as a building energy-consumption prediction method. Their use of an ANN model to predict building energy consumption produced more accurate calculations with fewer input data, and more quickly than conventional physics-based models. Huang et al. [17] and Biswas et al. [18] applied it to residential buildings and heating systems. Tardioli et al. [19] used an ANN model to predict energy demand on the urban level, rather than for individual buildings. Mohandes et al. [20] used a DNN model to predict energy consumption of a commercial building, and Luo et al. [21] applied DNN model for predicting electricity consumption of an office building. They extracted features of weather data and DNN could provide accurate week-ahead energy consumption. In summary, the use of a DNN model for prediction of building energy consumption has been mostly for commercial or residential buildings, but less for old houses.

This study aimed to develop a DNN model for predicting the heating energy consumption for 16,158 old houses in Korea. The key input variables that affect the heating energy consumption were selected, and based on these input variables, an optimal DNN model was proposed by determining the structural parameters (the number of hidden layers and the number of their neurons) with the highest prediction accuracy. In addition, we evaluated the applicability of the optimal DNN model to low-income standard house.

2. Characteristics of Old Houses

The Korea Institute of Energy Research (KIER) has been carrying out an energy efficiency improvement project for old houses since 2014. The energy efficiency improvement project is conducted to support fuel poverty (Public Aid Recipients) based on the national Energy Law with the aim of preventing energy inequality and social polarization and improving energy efficiency. People in the fuel poverty group have lower-income and spend more than 10% of their ordinary income for energy purchases. The government collects building information through diagnosticians for about 20,000 households each year and, with an on-site inspection, supports repair/replacement work for walls, windows, doors, airtightness and boilers to improve energy efficiency. The building information includes architecture scheme (householder, address, region, building structure, building orientation, number of residents, year of completion, building use, floor plan) and energy performance (area of building envelope, U-value, ACH (Air Changes per Hour), heating equipment) before and after the repair/replacement work [22]. Table 1 shows a sample portion of the collected building information.

Table 1. An example of datasets that have collected building information and heating energy consumption of old houses.

| Number | General | | | | | | Area | | | | | U-Value | | | | | | HVAC | | Heating Energy Consumption |
	City	Building Orientation	Structure	Region	Year of Completion	Heating Space	Wall	Window	Door	Roof	Floor	Wall	Window	Door	Roof	Floor	ACH	Boiler	Efficiency	
1	Daegu	East	Heavy	Gyeongsang	25-Jan-90	16.30	34.22	0.86	1.44	16.32	16.32	0.76	3.90	2.70	0.52	0.76	1.00	Oil	90.00	2985
2	Busan	South	Heavy	Busan	01-Jan-70	40.00	56.88	4.35	1.89	0.00	40.08	1.54	3.90	2.70	0.00	1.54	1.00	Gas	83.50	6823
3	Wando	South	Heavy	Jeollanam-do	07-May-64	60.00	53.88	0.24	2.88	60.00	60.00	1.54	5.29	2.40	1.54	1.54	1.00	Oil	80.00	14,975
4	Wando	South	Heavy	Jeollanam-do	03-Jul-80	60.00	76.81	3.08	7.28	60.00	60.00	1.05	5.30	2.40	1.05	1.05	1.00	Oil	80.00	12,701
5	Wando	South	Heavy	Jeollanam-do	08-Aug-70	53.00	54.98	5.24	4.00	53.00	53.00	1.54	5.30	2.50	1.54	1.54	1.00	Oil	80.00	14,486
6	Wando	South	Heavy	Jeollanam-do	11-Sep-65	51.00	57.54	3.08	5.82	51.00	51.00	1.54	5.30	2.40	1.54	1.54	1.00	Oil	80.00	14,243
7	Wando	South	Heavy	Jeollanam-do	29-May-85	16.00	33.37	1.54	1.89	16.00	16.00	0.58	2.80	2.70	0.58	0.58	1.00	Oil	80.00	2524
8	Wando	East	Heavy	Jeollanam-do	07-Jul-87	50.00	74.38	3.79	3.46	50.00	50.00	0.58	4.47	2.58	0.58	0.58	1.00	Oil	80.00	6172
9	Wando	North	Heavy	Jeollanam-do	07-Jul-83	25.00	30.40	5.44	1.36	25.00	25.00	0.58	6.60	2.40	0.58	1.16	1.00	Oil	80.00	3553
10	Wando	South	Heavy	Jeollanam-do	08-Jun-68	37.00	49.98	8.50	1.36	37.00	37.00	1.54	5.30	2.40	1.54	1.54	1.00	Oil	80.00	10,489
11	Wando	North	Heavy	Jeollanam-do	06-Apr-85	50.00	69.73	5.88	1.89	50.00	50.00	0.58	2.80	2.70	0.58	0.58	1.00	Oil	80.00	5502
12	Wando	North-East	Heavy	Jeollanam-do	02-Apr-95	21.00	41.28	2.42	0.00	21.00	21.00	0.76	6.60	0.00	0.52	0.76	1.00	Oil	80.00	3148
13	Wando	South	Heavy	Jeollanam-do	07-May-03	53.00	67.66	8.00	3.78	53.00	53.00	0.58	6.60	2.55	0.35	0.41	1.00	Briquette	70.00	4213
14	Wando	South	Heavy	Jeollanam-do	18-Apr-64	100.00	96.76	8.40	0.00	100.00	100.00	1.54	3.87	0.00	1.54	1.54	1.00	Oil	80.00	24,453
15	Wando	South	Heavy	Jeollanam-do	03-Jul-90	30.00	62.74	0.00	1.89	30.00	30.00	0.76	0.00	2.70	0.52	0.76	1.00	Oil	80.00	5495
16	Wando	South	Heavy	Jeollanam-do	06-Jul-89	45.00	57.52	5.04	3.69	45.00	45.00	0.76	3.90	2.55	0.52	0.76	1.00	Oil	80.00	6165
17	Wando	South	Heavy	Jeollanam-do	07-Jun-86	50.00	65.58	6.45	1.89	50.00	50.00	0.58	5.30	2.70	0.58	0.58	1.00	Oil	80.00	6533
18	Wando	South	Heavy	Jeollanam-do	20-Aug-88	50.00	65.47	3.41	1.89	50.00	50.00	0.76	3.90	2.70	0.52	0.76	1.00	Oil	85.00	5382
19	Wando	North	Heavy	Jeollanam-do	04-Aug-88	86.00	81.90	8.96	1.89	172.00	0.00	0.76	3.90	2.70	0.52	0.00	1.00	Oil	80.00	12,023
20	Wando	South	Heavy	Jeollanam-do	16-Sep-00	60.00	68.08	6.16	1.89	60.00	60.00	0.76	5.30	2.40	0.52	0.76	1.00	Oil	80.00	8290
21	Wando	South	Heavy	Jeollanam-do	22-May-70	38.00	55.63	3.49	1.36	38.00	38.00	1.54	5.30	2.70	1.54	1.54	1.00	Oil	80.00	10,057
22	Wando	North-East	Heavy	Jeollanam-do	23-Sep-77	28.00	44.41	1.76	1.44	28.00	28.00	1.54	6.60	2.70	1.54	1.54	1.00	Oil	80.00	9014
23	Busan	South	Heavy	Gyeongsang	08-Apr-71	16.00	16.60	1.08	0.00	17.00	17.00	1.54	3.90	0.00	1.54	1.54	1.00	-	50.00	4232
24	Wando	South	Heavy	Jeollanam-do	13-Jul-78	50.00	60.06	5.94	0.00	50.00	50.00	1.54	6.60	0.00	1.54	1.54	1.00	Oil	80.00	15076
25	Wando	South	Heavy	Jeollanam-do	07-Jul-85	45.00	49.58	13.34	0.00	45.00	45.00	0.58	5.11	0.00	0.58	0.58	1.00	Oil	85.00	6360
26	Wando	South	Heavy	Jeollanam-do	07-Jun-65	45.00	75.76	9.88	5.00	45.00	45.00	1.54	5.30	2.40	1.54	1.54	1.00	Oil	80.00	14,025
27	Wando	South	Heavy	Jeollanam-do	20-Jun-59	25.00	43.18	10.26	1.12	25.00	25.00	1.54	5.30	2.70	1.54	1.54	1.00	Oil	80.00	8623
28	Wando	South	Heavy	Jeollanam-do	08-Jun-83	40.00	64.73	5.07	2.88	40.00	40.00	0.58	5.61	2.40	0.58	1.16	1.00	Oil	80.00	6932
29	Wando	South	Heavy	Jeollanam-do	08-Mar-86	65.00	68.71	3.92	1.89	65.00	65.00	0.58	3.90	2.70	0.58	0.58	1.00	Oil	90.00	6329
30	Wando	South	Heavy	Jeollanam-do	10-Aug-63	40.00	68.64	2.64	3.52	40.00	40.00	1.54	5.30	2.40	1.54	1.54	1.00	Oil	80.00	13,233
31	Wando	North-East	Heavy	Jeollanam-do	05-Jun-85	28.00	53.55	3.52	1.89	28.00	28.00	0.58	6.60	2.70	0.58	0.58	1.00	Oil	80.00	4378
32	Wando	South	Heavy	Jeollanam-do	18-Jun-83	45.00	57.36	8.64	0.00	45.00	45.00	0.58	6.60	0.00	0.58	1.16	1.00	Electricity	100.00	5939
33	Wando	West	Heavy	Jeollanam-do	08-May-65	48.00	59.17	0.00	9.91	48.00	48.00	1.54	0.00	2.44	1.54	1.54	1.00	Oil	80.00	14,901
34	Wando	South	Heavy	Jeollanam-do	18-Jun-77	45.00	57.70	13.83	1.95	45.00	45.00	1.54	5.30	2.40	1.54	1.54	1.00	Oil	80.00	14,866
35	Wando	South-East	Heavy	Jeollanam-do	14-Sep-02	40.00	60.74	7.42	5.76	40.00	40.00	0.58	6.18	2.40	0.35	0.41	1.00	Oil	80.00	5843

This study targeted 16,158 old houses among the 20,000 households for which building information was collected. The 16,158 old houses were completed more than 20 years, and were deteriorated in structure and function due to aging, and had high energy-consuming houses with low insulation performance [23]. The old houses had either a light-weight structure (panel, wood, and prefabricated structure) or a heavy-weight structure (steel, concrete, and masonry). The average heating space area was 42.89 m^2, and the average areas of envelope elements including the roof, walls, floors, windows and doors, was 42.66, 47.2, 40.90, 6.57 and 2.27 m^2, respectively. The U-value was measured using a heat-flux meter according to the ISO 9869-1. An average analysis method was used and each measurement took 72 h. The U-value was measured by the spot measurement method. The average U-value for the roof, walls, floors, windows and doors was 0.97, 0.91, 1.01, 4.28 and 2.78 W/(m^2·K), respectively. ACH was applied at 1-h, which was suggested in the results of a detailed survey of low-income houses in Korea [22]. The heating equipment used individual boilers (type of boiler: gas, oil and briquettes). The efficiency of the boilers used the value from their name plates which the diagnostician investigated in the field. The average boiler efficiency was 84.4%. Heating energy consumption was calculated according to ISO 52016 [24]. ISO 52016 is an international standard for procedures for calculating the heating and cooling energy consumption of buildings. It contains calculation methods for the assessment of sensible energy needs for heating and cooling, latent energy needs for dehumidification, design sensible heating/cooling loads, design latent loads, and internal temperatures. The average energy consumption was 279.42 kWh/(m^2·a), which is higher than the minimum energy consumption (173.20 kWh/(m^2·a)) to escape from fuel poverty [25].

3. Input Data and Configuration of DNN Model

3.1. Preparation of Input Data

The input data for the DNN model used in this study came from the building information for 16,158 old houses surveyed in 2019 (1 January 2019−31 December 2019). This input data was prepared for modeling through a preprocess of error elimination–missing value elimination–outlier elimination–normalization.

1. Error elimination: this is the process of eliminating inaccurate information from the collected input data. For example, address or meteorological data, etc., that do not match or contradict the actual data for the old, detached houses of this study, were removed.
2. Missing data elimination: collected data may contain missing values, and statistical analysis using these datasets does not produce the desired results. Therefore, the data were checked for missing values and the data set containing missing values was removed or replaced with one containing the correct values.
3. Outlier elimination: an outlier, in statistics, is a data point that differs significantly from other observations. The outlier may be due to variability in the measurement, or it may indicate experimental error; the latter are sometimes excluded from the data set. An outlier can cause serious distortion of the analytical results. The outlier value is determined to be greater than 97.5% or less than 2.5% in the normal distribution. This study removed outliers using the Mahalanobis distance method [26]. The Mahalanobis distance (MD) is the distance between two points in multivariate space. The MD calculation—Mahalanobis Score (Probability)—*p*-value test was conducted using IBM SPSS statistic software. Through this, 16,158 data were extracted from 17,008 data.
4. Normalization: the data range distribution is adjusted by changing the range of data with different scales to 0 and 1. In this study, data normalization was performed using the following Equation (1):

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

3.2. Neural Network Model

The DNN model had input layers, hidden layers, and an output layer (Figure 1), similar to the human neuron structure which has a dendrite, an axon, and cell bodies. Each node of the model received an external input and adjusted its influence on the output by weight and bias; the output value was derived by an activation function. The representative activation functions were step function, sigmoid function and linear function. The optimal weights in the learning process were adjusted using back propagation [27]. The DNN model is often used to solve nontheoretical problems such as pattern recognition and classification, as well as prediction in place of mathematical models. Back propagation is a representative algorithm for learning neural networks based on supervised learning, the signals of which are forward and backward propagated. As a learning function of the DNN model, the LMA (Levenberg Marquardt Algorithm) is widely used. The LMA is an algorithm for solving the nonlinear least squares problem [28]. It updates the link weight and bias values according to optimization [29]. This study proposes a DNN model as an alternative for estimating the heating energy consumption of old houses.
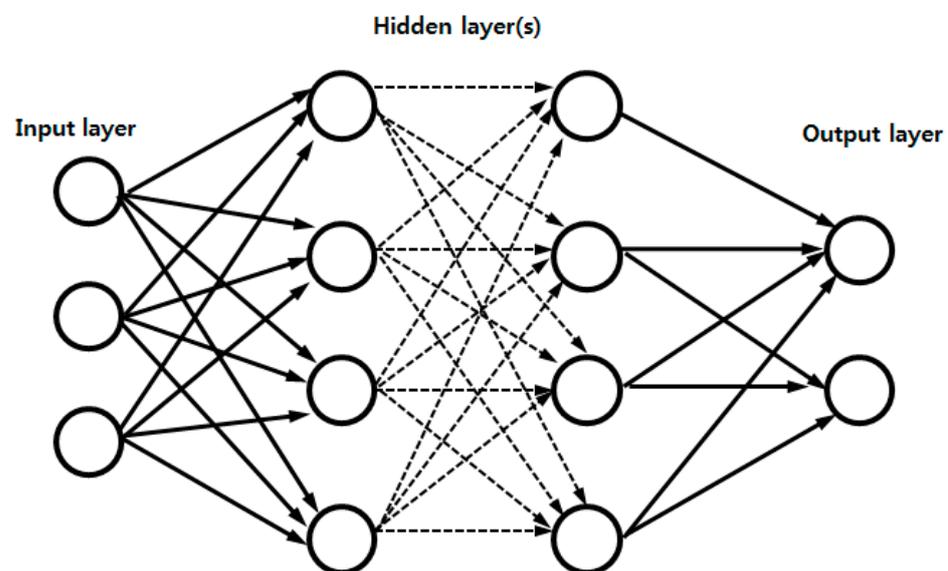

Figure 1. Conceptual structure of the deep neural network (DNN) model [22].

3.3. Modeling Approach

In this study, a DNN model for the prediction of heating energy consumption in old houses was developed through the process of Selection of key input variables–setting initial model conditions–optimization. The selection of key input variables affects the prediction accuracy of the model. If the input variables are incorrectly selected, the model prediction accuracy may drop significantly. Therefore, it is very important to find the key input variables that are most relevant to the target variables. The key input variables in this study were selected through correlation analysis of statistical techniques [30] for the input variables and target variables, as shown in Table 2, and then through prediction accuracy analysis [31] using an initial DNN model as shown in Table 3.

1. Correlation analysis: the Pearson correlation coefficients between the input variable and the target variable were calculated using the IBM Statistical Package for the Social Sciences (SPSS) [32]. From these, the input variables with a strong correlation were included as the key input variables.
2. Prediction accuracy analysis using an initial DNN model: the following stepwise method was applied. Possible combination cases were created, based on the key input variables selected from the correlation analysis above, and the coefficient of determination (R^2) was calculated for each combination case. The final key input

variables were determined by a stepwise method that excludes the combination case of the input variable with the lowest coefficient of determination.

Table 2. Input variables and target variables considered for the DNN modelling of old houses.

Type	Input Variable	Target Variable
General details	Region (32 cities) Building orientation (E, W, S, N, NE, NW, SE, SW) Structure (heavy, light) Year of completion ACH Type of boiler Boiler efficiency	Energy consumption [kWh/a]
Area [m^2]	Heating space area Wall Roof Floor Window Door	
U-value [W/(m^2·K)]	Wall Roof Floor Window Door	

Table 3. Feature and setting value of the structural and learning parameters of initial DNN model.

Parameters		Feature	Value
Model		Back propagation efficiently computes the gradient of the loss function with respect to the weights of the network for a single input-output example	LMA (Levenberg Marquardt Algorithm) [29]
Data division (%)		Training:Validation:Testing	70:15:15
Structural Parameters	Hidden layer	The increase in the number of hidden layers/neurons, the predict performance is improved, but the calculation takes a long time	1
	Neuron of hidden layer		10
Learning parameters	Learning rate	The smaller value has improved predict performance, but the longer it takes to learn	0.2
	Momentum	Basically, it starts with 0.1, and the bigger value it is, the faster the learning speed	0.6
	Epochs	Maximum number of learning	1000
	Goal	Target error ratio between actual and predicted values	0.01

In this study, modeling was performed by using the 'MATLAB Neural Network Toolbox' of The MathWorks, Inc. [33]. The initial DNN model used the values of the parameters in Table 3. Optimization of the model is the process of selecting the most suitable structural variables (number of hidden layers, number of their neurons) and parameters affecting the learning speed from the initial neural network model, to improve the predictive performance of the model. The number of hidden layers and the number of neurons in the hidden layers which have the highest prediction accuracy (R^2-value) are determined by simulation in the range of the number of hidden layers = 1–10 and the number of neurons = 10–30, for the selected key input variables. This study did not consider the influence of parameters that affect the learning speed of the simulation.

4. Results and Discussion

4.1. Selection of Key Input Variables

The analysis of the correlation between the input variable and the target variable was performed by using SPSS, and the results are summarized as the Pearson correlation coefficient in Table 4. The Pearson correlation coefficient [34] is the quantification index of the linear correlation between X and Y. This is the covariance between two variables divided by the product of their standard deviations, which has a value between +1 and −1. A perfect positive linear correlation is measured as +1, 0 means no linear correlation and −1 means a perfect negative linear correlation. As shown in Table 4, the Pearson correlation coefficients are in the order of roof U-value > roof area > wall U-value > floor area > floor U-value > year of completion > wall area > heating space area > boiler efficiency > door area > window U-value > window area > region > structure > door U-value > ACH > building orientation. The boiler type was not considered in the present analysis, because it does not affect heating energy consumption. As a result of the correlation analysis, roof U-value (0.636), roof area (0.617), wall U-value (0.604), floor area (0.557), floor U-value (0.550), year of completion (−0.539), wall area (0.4888), and heating space area (0.430) were selected as the primary key input variables, because of high correlation with the target variable. Boiler efficiency (−0.276), door area (0.275), window heat permeability (0.269), and window area (0.252), with relatively low Pearson correlation coefficients, were selected as variables to be considered again in the analysis of the prediction accuracy of the initial DNN model.

Table 4. Results of Pearson correlation analysis using IBM SPSS Statistics Software.

Type	Input Variable	Pearson Correlation Coefficient	Rank
General details	Region	0.162	13
Area [m^2]	Building orientation	−0.012	17
	Structure	0.159	14
	Year of completion	−0.539	6
	ACH	0.012	16
	Type of boiler	N/A	N/A
	Boiler efficiency	−0.276	9
	Heating space area	0.430	8
	Wall	0.488	7
U-value (Heat transmission coefficient)	Roof	0.617	2
[W/(m^2·K)]	Floor	0.557	4
	Window	0.252	12
	Door	0.275	10
	Wall	0.604	3
	Roof	0.636	1
	Floor	0.550	5
	Window	0.269	11
	Door	−0.051	15

Table 5 shows the coefficient of determination, R^2, which represents the prediction accuracy of the initial DNN model for each combination case. The values of R^2 from Case 1 to Case 16 ranged from 0.890 to 0.936. The coefficient of determination is a measure of explanation of the model for the data set, which means that the higher this value the higher the prediction accuracy of the model. As seen in Table 5, Case 13 shows the highest prediction accuracy. When the initial DNN was modeled with 11 input variables (excluding door area), it had the highest prediction accuracy (R^2 = 0.936). Based on these results, we selected 11 input variables (roof U-value, roof area, wall U-value, floor area, floor U-value, year of completion, wall area, heating space area, boiler efficiency, window area, and window U-value) as the final key input variables for the DNN model for predicting energy consumption of old houses.

Table 5. Cases of available combinations of the input variables selected from the SPSS correlation analysis and performance results (R^2-values) for each case predicted from the initial DNN model.

Parameter	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10	Case 11	Case 12	Case 13	Case 14	Case 15	Case 16
Window U-value [W/(m²·K)]	×	○	×	×	×	○	×	×	○	○	×	○	○	○	×	○
Window area [m²]	×	×	○	×	×	○	○	×	×	×	○	○	○	×	○	○
Door area [m²]	×	×	×	○	×	×	○	○	○	×	×	○	×	○	○	○
Boiler efficiency [%]	×	×	×	×	○	×	×	○	×	○	○	×	○	○	○	○
Heating space area [m²]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Wall area [m²]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Year of completion	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Floor U-value [W/(m²·K)]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Floor area [m²]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Wall U-value [W/(m²·K)]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Roof area [m²]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Roof U-value [W/(m²·K)]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
R^2-value	0.893	0.890	0.896	0.895	0.931	0.896	0.896	0.933	0.896	0.929	0.933	0.898	0.936	0.935	0.935	0.935

Note: O and X are input variables that were considered and not considered in the case analysis, respectively.

4.2. Optimal DNN Model

The optimization of the DNN model was performed by calculating R^2 while changing the structural variables (the number of hidden layers and the number of their neurons) for the selected key input variables, and then identifying the structural variables with the highest R^2 (i.e., best prediction accuracy) for use in the DNN model. In the optimization process, the number of hidden layers was increased from 1 to 10 and the number of neurons in the hidden layer was increased from 10 to 30. R^2 was then calculated for all possible combinations of the structure variables. In these calculations, since the value of R^2 showed a slight deviation for every time of training with random sampling, it was trained 30 times and the average of the results was used.

Table 6 shows the R^2 calculation results when the number of neurons was fixed at 10 and the number of hidden layers was varied from 1 to 10. As shown in this table, when the number of hidden layers was five, the prediction accuracy was the highest, with an R^2 of 0.954. Table 7 shows the results obtained by fixing the number of hidden layers to five (as determined in the previous Table 6) and changing the number of neurons from 10 to 30. When the number of neurons was 22, the prediction accuracy was the highest, with an R^2 of 0.961. The results of the regression analysis (R^2) for the training, validation, test and total datasets are plotted in Figure 2.

Table 6. Predict performance results (R^2-values) for the change of hidden layer number when the number of neurons is fixed to 10.

Number of Hidden Layer	1	2	3	4	5	6	7	8	9	10
R^2-value	0.936	0.945	0.949	0.951	0.954	0.953	0.952	0.953	0.952	0.950

Table 7. Predicted performance results (R^2-values) for the change of neuron number when the number of hidden layers is fixed to five.

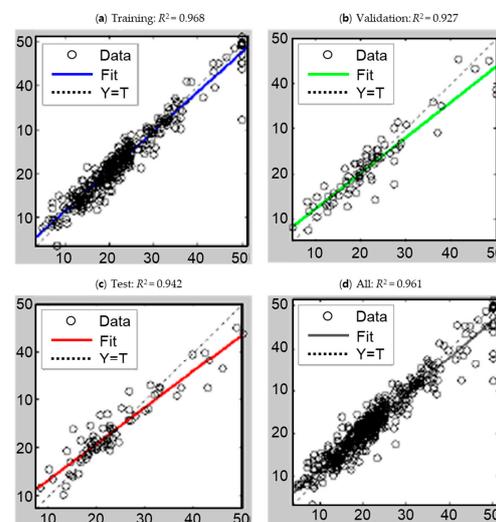| Number of Neurons in Hidden Layer	10	11	12	13	14	15	16
R^2-value	0.954	0.955	0.955	0.954	0.956	0.958	0.954
Number of Neurons in Hidden Layer	**17**	**18**	**19**	**20**	**21**	**22**	**23**
R^2-value	0.958	0.956	0.956	0.959	0.958	0.961	0.958
Number of Neurons in Hidden Layer	**24**	**25**	**26**	**27**	**28**	**29**	**30**
R^2-value	0.959	0.958	0.959	0.958	0.960	0.959	0.958

Figure 2. Regression results of an optimized DNN model for: (**a**) training; (**b**) validation; (**c**) test; and (**d**) all datasets.

4.3. Summury and Discussion

As shown in Table 4 above, the results of the correlation analysis and prediction accuracy analysis showed that the U-values and envelope areas of the roof, wall, and floor had a high correlation, which influences the heating energy consumption through envelope heat loss proportional to the U-value and envelope area [24]. The completion year of the building was not used in physics-based model (e.g., E+), but in the DNN model, it was identified as a major variable that influences the prediction results. This variable is also related to the U-value standard of the national building code, and the standard has been strengthened recently. The heating space area has a proportional relationship with the occupied space and the building volume and affects the heating load through heat loss by infiltration. Although the boiler efficiency is an important variable in heating energy consumption, it was placed in a relatively low rank in this study. This seems to have a low correlation due to pipe heat loss and efficiency decline due to aging. The area and U-value of the window, like the roof, wall, and floor described above, are variables that contribute to the envelope heat loss, but their area is small and thus the influence is small. Other variables (door area and U-value, region, structure, ACH, building orientation) did not significantly affect the prediction of heating energy consumption of old houses by DNN model in this study.

The structure of the neural network varies depending on the complexity or hyper-parameter setting, and if an appropriate model is not selected, it may cause underfitting or overfitting. Underfitting refers to a state in which learning is not performed because there is too little data to approach the decision boundary, while overfitting refers to a case where the variance is high due to overtraining. To avoid these problems, regularization, hyperparameter selection, and supply of sufficient amount of learning/verification data are necessary [28,35]. This study performed learning by dividing 16,158 data sets into training, validation, and testing. As a result of learning, the neural network structure of hidden layers 5 and neurons 22 had the highest prediction accuracy (R^2 = 0.961), which shows that this optimized DNN model is suitable for predicting the heating energy consumption of old houses.

The developed optimal DNN model showed high predictive accuracy even when applied to Korean standard house that were not used for the above training. The Korean standard house is a house selected by extracting features that can have representativeness from architectural information, envelope performance, and floor plan obtained by surveying about 3000 low-income detached houses [23]. The floor plan of the standard house is shown as Figure 3. Table 8 shows the building information used as input for the model simulation and the heating energy consumption of the standard house as calculated by EnergyPlus. A description of EnergyPlus is summarized in Appendix A, and the details are contained in the User's Manual from the Department of Energy [36]. Table 9 compares the result of the EnergyPlus calculation (in Table 8) with the result of heating energy consumption predicted by the optimal DNN model for the standard house. The annual heating energy consumption for the standard model house, as calculated by EnergyPlus, was 12,143 kWh/a (annual heating energy consumption per unit area: 272.75 kWh/(m^2·a)). As predicted by the optimal DNN model from this study, the annual heating energy consumption was 13,307 kWh/a (annual heating energy consumption per unit area: 298.80 kWh/(m^2·a)). The *Cv*(*RMSE*) of heating energy consumption predicted by the optimal DNN model was 8.74%. The *Cv*(*RMSE*) is a coefficient of variation (*Cv*) of a root mean square error (*RMSE*) and is a measure of the difference between an actual and a predicted value. The closer the *Cv*(*RMSE*) is to 0%, the more accurate the prediction. ASHRAE (The American Society of Heating, Refrigerating and Air-Conditioning Engineers) suggests that the acceptable value of *Cv*(*RMSE*) is 15% on a monthly, and 30% on an hourly, scale basis, as shown in Table 10 [37]. The *Cv*(*RMSE*) result of 8.74% indicates that the optimal DNN model can be used to evaluate the heating energy consumption of low-income houses. In addition, the DNN model developed in this study will be applicable to energy prediction of residential buildings in other countries with similar climate and

building types. This result will be useful in setting the standard of energy consumption and reduction targets in large regional units such as old housing communities and cities.
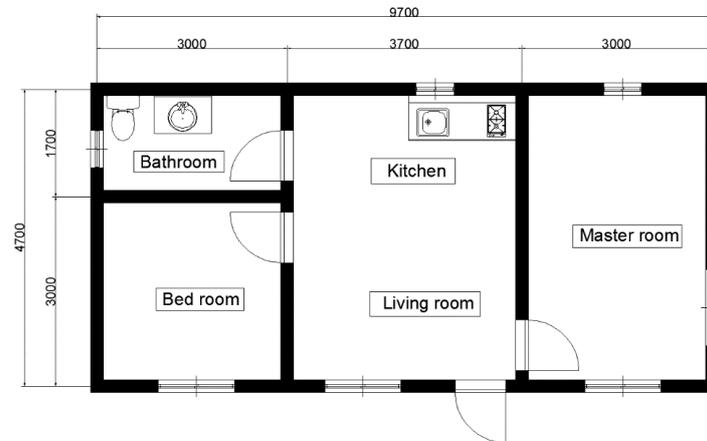

Figure 3. Floor plans of Korean standard house.

Table 8. Building Information and heating energy consumption (calculated by EnergyPlus) of standard house.

Division	Values
Window U-value [W/(m^2·K)]	5.84
Window area [m^2]	8.05
Heating space area [m^2]	44.52
Wall area [m^2]	56.09
Year of completion	1980
Floor U-value [W/(m^2·K)]	1.05
Floor area [m^2]	44.52
Wall U-value [W/(m^2·K)]	1.05
Roof area [m^2]	44.52
Roof U-value [W/(m^2·K)]	1.05
Boiler efficiency [%]	80
Energy consumption(Calculated) [kWh/yr]	12143
Energy consumption (Calculated/m^2) [kWh/m^2·yr]	272.75

Table 9. *Cv(RMSE)* of heating energy consumptions predicted by the optimal DNN model to those calculated by EnergyPlus for the standard house.

Division	Values
Optimal DNN model-predicted annual heating consumption [kWh/yr]	13,307
Optimal DNN model- predicted annual heating consumption per unit area [kWh/m^2·yr]	298.90
EnergyPlus-calculated annual heating consumption [kWh/yr]	12,143
EnergyPlus-calculated annual heating consumption per unit area [kWh/m^2·yr]	272.75
Cv(RMSE) [%]	8.74

Table 10. ASHRAE guideline of tolerance limits for the *Cv*(*RMSE*) of building energy modelling.

Division	Monthly	Hourly
Tolerance limits (*Cv(RMSE)*)	15%	30%

5. Conclusions

This study developed an optimized DNN model for predicting heating energy consumption by selecting key input variables for 16,158 old houses and determining the structural variables (number of hidden layers and their neurons) with the highest prediction accuracy. As a result of correlation and prediction-accuracy analyses, 11 key input variables (roof U-value, roof area, wall U-value, floor area, floor U-value, year of completion, wall area, heating space area, boiler efficiency, window area, and window U-value) were selected. The optimal DNN model for predicting the heating energy consumption of old houses showed the highest accuracy ($R^2 = 0.961$) when the number of hidden layers and neurons in hidden layers was five and 22, respectively. The developed optimal DNN model showed high accuracy (*Cv*(*RMSE*) = 8.74%) in the applicability evaluation for heating energy consumption of the standard house and satisfied the ASHRAE standard sufficiently. Moreover, the optimal DNN model, compared to the physics-based model (EnergyPlus), can have high energy prediction accuracy with fewer input variables, and can reduce modeling and simulation time. These advantages will make it easier for field engineers to use with less expertise and experience in building energy modeling. We believe that our study makes a significant contribute to energy welfare by improving the energy efficiency, diagnosis and prediction for low-income houses.

Author Contributions: Design, modeling, simulation and writing, S.L.; conceptualization and methodology, S.C. (Soo Cho); data collection and analysis, S.-H.K.; project administration, J.K., S.C. (Suyong Chae) and H.J.; supervision, T.K. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

Appendix A

The EnergyPlus program has its roots in both the BLAST and DOE-2 programs. BLAST (Building Loads Analysis and System Thermodynamics) and DOE-2 were both developed and released in the late 1970s and early 1980s, as energy and load simulation tools. EnergyPlus is an energy analysis and thermal load simulation program. The program can be detailed for internal conditions; can consider the interactions between factors in each hourly load calculation; and calculates energy consumption through integrated simulations between buildings, systems, and plants. In addition, the module-type structure provides flexibility, and the calculation results can be accurately reproduced in the actual building, over time. See the User's Manual of the Department of Energy [36] for details.

References

1. Korean Statistical Information Service (KSIS). Population and Housing Census. 2018. Available online: http://kosis.kr/index/index.do (accessed on 4 September 2020).
2. Bourdeau, M.; Zhai, X.Q.; Nefzaoui, E.; Guo, X.; Chatellier, P. Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustain. Cities Soc.* **2019**, *48*, 101533. [CrossRef]
3. O'Neill, Z.; Narayanan, S.; Brahme, R. Model-based thermal load estimation in buildings. *Proc. Simbuild* **2010**, *4*, 474–481.
4. Naganathan, H.; Chong, W.O.; Chen, X. Building energy modeling (BEM) using clustering algorithms and semi-supervised machine learning approaches. *Autom. Constr.* **2019**, *72*, 187–194. [CrossRef]
5. Ryan, E.M.; Sanquist, T.F. Validation of building energy modeling tools under idealized and realistic conditions. *Energy Build.* **2012**, *47*, 375–382. [CrossRef]
6. Esen, H.; Escen, M.; Ozsolak, O. Modelling and experimental performance analysis of solar-assisted ground source heat pump system. *J. Exp. Theor. Artif. Intell.* **2017**, *29*, 1–17. [CrossRef]
7. Kang, I. Development of ANN (Artificial Neural Network) Based Predictive Model for Energy Consumption of HVAC System in Office Building. Master's Thesis, Chung-Ang University, Seoul, Korea, 2017.
8. Robinson, C.; Dilkina, B.; Hubbs, J.; Zhang, W.; Guhathakurta, S.; Brown, M.A.; Pendyala, R.M. Machine learning approaches for estimating commercial building energy consumption. *Appl. Energy* **2017**, *208*, 889–904. [CrossRef]
9. Ahmad, T.; Chen, H.; Huang, R.; Yabin, G.; Wang, J.; Shair, J.; Akram, H.M.A.; Mohsan, S.A.H.; Kazime, M. Supervised based machine learning models for short, medium and long-term energy prediction in distinct building environment. *Energy* **2018**, *158*, 17–32. [CrossRef]
10. Chou, J.-S.; Tran, D.-S. Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. *Energy* **2018**, *165 Pt B*, 709–726. [CrossRef]
11. D'Amico, A.; Ciulla, G.; Traverso, M.; Brano, V.L.; Palumbo, E. Artificial Neural Networks to assess energy and environmental performance of buildings: An Italian case study. *J. Clean. Prod.* **2019**, *239*, 117993. [CrossRef]
12. Ciulla, G.; D'Amico, A.; Brano, V.L.; Traverso, M. Application of optimized artificial intelligence algorithm to evaluate the heating energy demand of non-residential buildings at European level. *Energy* **2019**, *176*, 380–391. [CrossRef]
13. Katsatos, A.L.; Moustris, K.P. Application of Artificial Neuron Networks as energy consumption forecasting tool in the building of Regulatory Authority of Energy, Athens, Greece. *Energy Procedia* **2019**, *157*, 851–861. [CrossRef]
14. Satrio, P.; Mahlia, T.M.I.; Giannetti, N.; Saito, K. Optimization of HVAC system energy consumption in a building using artificial neural network and multi-objective genetic algorithm. *Sustain. Energy Technol. Assess.* **2019**, *35*, 48–57.
15. Deb, C.; Lee, S.E.; Santamouris, M. Using artificial neural networks to assess HVAC related energy saving in retrofitted office buildings. *Sol. Energy* **2018**, *163*, 32–44. [CrossRef]
16. González, P.A.; Zamarreño, J.M. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy Build.* **2005**, *37*, 595–601. [CrossRef]
17. Huang, Y.; Yuan, Y.; Chen, H.; Wang, J.; Guo, Y.; Ahmad, T. A novel energy demand prediction strategy for residential buildings based on ensemble learning. *Energy Procedia* **2019**, *158*, 3411–3416. [CrossRef]
18. Biswas, M.A.R.; Robinson, M.D.; Fumo, N. Prediction of residential building energy consumption: A neural network approach. *Energy* **2016**, *117 Pt 1*, 84–92. [CrossRef]
19. Tardioli, G.; Kerrigan, R.; Oates, M.; O'Donnell, J.; Finn, D. Data Driven Approaches for Prediction of Building Energy Consumption at Urban Level. *Energy Procedia* **2015**, *78*, 3378–3383. [CrossRef]
20. Mohandes, S.R.; Zhang, X.; Mahdiyar, A. A comprehensive review on the application of artificial neural networks in building energy analysis. *Neurocomputing* **2019**, *340*, 55–75. [CrossRef]
21. Luo, X.J.; Oyedele, L.O.; Ajayi, A.O.; Akinade, O.O.; Owolabi, H.A.; Ahmed, A. Feature extraction and genetic algorithm enhanced adaptive deep neural network for energy consumption prediction in buildings. *Renew. Sustain. Energy Rev.* **2020**, *131*, 109980. [CrossRef]
22. Korea Institute of Energy Research (KIER). *Low-Income Household Energy Efficiency Improvement Project Report*; Korea Institute of Energy Research (KIER): Daejeon, Korea, 2018.
23. Kim, J. Heating Energy Baseline and Saving Model Development of Detached Houses for Low-Income Households. Master's Thesis, University of Science and Technology, Daejeon, Korea, 2015.
24. International Organization for Standardization (ISO). *Energy Performance of Buildings—Energy Needs for Heating and Cooling, Internal Temperatures and Sensible and Latent Heat Loads—Part 1: Calculation Procedures*; Standard No. 52016-1; ISO: Geneva, Switzerland, 2017.
25. Lee, S.-J.; Kim, J.; Jeong, H.; Yoo, S.; Lee, S. Heating Energy Efficiency Improvement Analysis of Low-income Houses. *J. KIAEBS* **2017**, *11*, 212–218.
26. Hazewinkel, M. *Encyclopedia of Mathematics*; Springer: Berlin, Germany, 2001.
27. Aggarwal, C.C. *Neural Networks and Deep Learning: A Textbook*, 1st ed.; Springer: Berlin, Germany, 2018.
28. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; The MIT Press: Cambridge, MA, USA, 2016.
29. Ye, Z.; Kim, M.K. Predicting electricity consumption in a building using an optimized back-propagation and Levenberg-Marquardt back-propagation neural network: Case study of a shopping mall in China. *Sustain. Cities Soc.* **2018**, *42*, 176–183. [CrossRef]
30. Hinton, P.R.; McMurray, I.; Brownlow, C. *SPSS Explained*, 2nd ed.; Routledge: London, UK, 2014.

31. Ahmad, M.W.; Mourshed, M.; Rezgui, Y. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy Build.* **2017**, *147*, 77–89. [CrossRef]
32. IBM. IBM SPSS Software. Available online: https://www.ibm.com/kr-ko/analytics/spss-statistics-software (accessed on 4 September 2020).
33. Mathworks. Matlab Software. Available online: https://www.mathworks.com/products/matlab.html (accessed on 4 September 2020).
34. Wang, J.C. A study on the energy performance of hotel buildings in Taiwan. *Energy Build.* **2012**, *49*, 268–275. [CrossRef]
35. Reed, R.; MarksII, R.J. *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*; A Bradford Book: Cambridge, MA, USA, 1999.
36. Department of Energy (DOE). EnergyPlus Software. Available online: https://energyplus.net/ (accessed on 4 September 2020).
37. Department of Energy (DOE). *M&V Guidelines: Measurement and Verification for Federal Energy Projects, Version 3.0*; Department of Energy (DOE): Washington, DC, USA, 2008; pp. 21–22.