



Article

# Forecasting Electricity Consumption in Commercial Buildings Using a Machine Learning Approach

Junhwa Hwang <sup>1</sup>, Dongjun Suh <sup>1,\*</sup> and Marc-Oliver Otto <sup>2</sup>

<sup>1</sup> Department of Convergence and Fusion System Engineering, Kyungpook National University, Sangju 37224, Korea; jhhwang1907@knu.ac.kr

<sup>2</sup> Department of Mathematics, Natural and Economic Sciences, Ulm University of Applied Sciences, Prittwitzstr, 10, 89075 Ulm, Germany; Marc-Oliver.Otto@thu.de

\* Correspondence: dongjunsuh@knu.ac.kr; Tel.: +82-54-530-1482

Received: 1 October 2020; Accepted: 9 November 2020; Published: 11 November 2020



**Abstract:** Prediction of electricity consumption is a key research area for efficient power grid operation. Accurate electricity consumption predictions of buildings can prevent power shortages in modern cities, reduce social costs caused by unnecessary energy supply, and support stable and efficient power grid operation. In this study, an electricity consumption prediction model is proposed using open-access data for the monthly and daily electricity consumption of 28 commercial buildings in Seo-gu, Gwangju, South Korea. In the case of the electricity consumption prediction of a building, information about specific parameters that affect energy consumption in target buildings is required. However, inappropriate parameter selection of the prediction model can lead to decreased prediction accuracy. Therefore, we propose a two-step approach to develop a highly accurate electricity consumption prediction model by overcoming the limitations of insufficient information. In the first step, the electricity consumption model of the building is derived by reflecting the characteristics of an individual building that constitutes a building community. In the second step, we use additional information, including the specific building's features, as well as the energy facility types of the building. Using dynamic-time-warping-based clustering classification, we could infer the energy equipment information of the buildings. We apply the two-step method to develop a prediction model using machine learning methods. In addition, we propose an optimal prediction model by comparing the performance of a traditional time-series analysis technique and machine learning techniques. In this study, the proposed model performs >27.5% better than the existing model. Using the proposed model, it will be possible to accurately predict electricity consumption of commercial buildings, and it can be used as a major guideline for the power supply and demand of buildings and cities.

**Keywords:** LSTM; DNN; demand response; machine learning; commercial building

## 1. Introduction

Climate change and the occurrence of anomalies have been a global and consistent problem. Indiscriminate development is the major cause of climate change and various anomalies. As such, carbon-emission regulations were negotiated internationally through the Kyoto Protocol (Kyoto Conference on Prevention of Global Warming) and the Paris Agreement [1,2]. International regulations have had a major impact on carbon-emission reductions by countries and companies. Countries and companies are implementing methods, such as the use of renewable energy, carbon recycling, construction efficient power grids, and operations to reduce carbon emissions [3]. More than 40% of the world's energy consumption is used within buildings. In addition, carbon emissions from buildings account for one-third of the global total [4,5], and building energy consumption continues

to increase every year. Therefore, energy savings in buildings is a critical component for addressing climate change. The energy demand forecasting method is an effective way to conserve energy in buildings.

Demand response is important for the construction of an efficient power grid and operation [6]. Accurate electricity consumption prediction is important for establishing plans for electricity supply and demand at the national or provincial level and for power system facilities. Accurate electricity consumption predictions can prevent unnecessary power supply and power shortages. For example, when excessive electricity demand is predicted, unnecessary spending may occur due to the development of excess power generation facilities. On the other hand, insufficient electricity demand prediction may cause a power shortage phenomenon due to insufficient power generation facilities and lead to blackouts. In fact, a blackout phenomenon occurred due to a power shortage in September 2011 in Korea. Therefore, a stable plan for power supply and demand should be established with accurate demand prediction, minimized unnecessary costs, and stable power grid operation through securing a stable power reserve rate [7,8].

Demand response is another important research topic in the smart grid field. A smart grid is an intelligent power grid system capable of efficient operation and reduction of energy, where energy can be used efficiently through energy storage management, user energy demand management, and peak load management. The smart grid environment exchanges information about power generation and consumption through two-way communication [7,9,10]. Therefore, it is possible to appropriately adjust the amount of energy generated according to the amount of future electricity consumption predicted via consumer electricity consumption information. This enables rational and efficient electricity use by reducing unnecessary power supply.

Buildings can be classified into structure types based on volume, shape, and purpose. In terms of the purpose, buildings can also be further classified into several types, such as detached houses, apartments, commercial buildings, and medical and public buildings [11]. In Korea, commercial buildings account for 17.9% of all buildings and account for a large proportion of total building energy consumption [12]. Buildings have different electricity consumption patterns depending on their usage; thus, it is necessary to predict electricity demand for each. Specifically, commercial buildings show similar patterns with respect to occupancy because of similar commercial economic activities on weekdays and public holidays [7]. Therefore, in this study, we propose a daily and monthly electricity consumption prediction model for commercial buildings located in Seo-gu, Gwangju.

This study aims to develop an electricity consumption prediction model for commercial buildings and proposes a method to improve model performance through a two-step approach to compensate for insufficient information. In the first step, the electricity consumption prediction model of multiple buildings is developed by reflecting the features of individual buildings constituting the building community. The electricity consumption prediction model for multiple buildings might fail to highlight the degree of influence of features of individual buildings on the electricity consumption pattern. Therefore, when the input parameters are applied in consideration of the differences in power consumption level according to the total building area, the prediction performance of the proposed model can be improved.

In the second step, we infer the information of the building's heating and cooling facilities by classifying the electricity consumption patterns of buildings through the dynamic time warping (DTW)-based clustering method. We utilize the inferred facility information as an input parameter to improve the prediction accuracy of the electricity consumption prediction model.

To develop an electricity consumption prediction model with high accuracy, information on heating and cooling facilities of individual buildings is required. However, it is impossible to gather all of the information related to cooling and heating facility types of the target buildings due to the Personal Information Protection Act.

As the electricity consumption patterns of buildings vary according to the features of the heating and cooling facilities of each building, we analyze the information of representative cooling and heating facilities constituting commercial buildings in Korea and conduct a scenario-based analysis to derive information about building energy equipment. Therefore, the facility information inferred from scenario-based analysis and DTW-based clustering method enable improved accuracy of the proposed model.

In the last part, we conduct performance evaluation using machine learning models artificial neural network (ANN), deep neural network (DNN), long short-term memory (LSTM), support vector regression (SVR), and the time-series-based method seasonal autoregressive integrated moving average-X (SARIMAX) is optimize using the design of experiments. Moreover, we present accurate daily and monthly electricity consumption of the commercial buildings.

## 2. Research Framework

As shown in Figure 1, the study was conducted in three steps. In Step 1, we performed data collection and pre-processing to train the model properly. In Step 2, for the development of the electricity consumption prediction model, input parameters were selected through correlation analysis between factors that influence electricity consumption and potential energy consumption. The model performance was evaluated using five different methods related to machine learning and time series to develop the electricity consumption prediction model using the derived input parameters. In Step 3, the accuracy of the electricity consumption prediction was improved by inferring information on heating and cooling facilities of individual buildings. We identified the types of electricity consumption patterns of commercial buildings through scenario-based analyses and classified the buildings based on their main heating and cooling facilities. The accuracy of the prediction model could be improved using the inferred facility information as an additional input parameter.

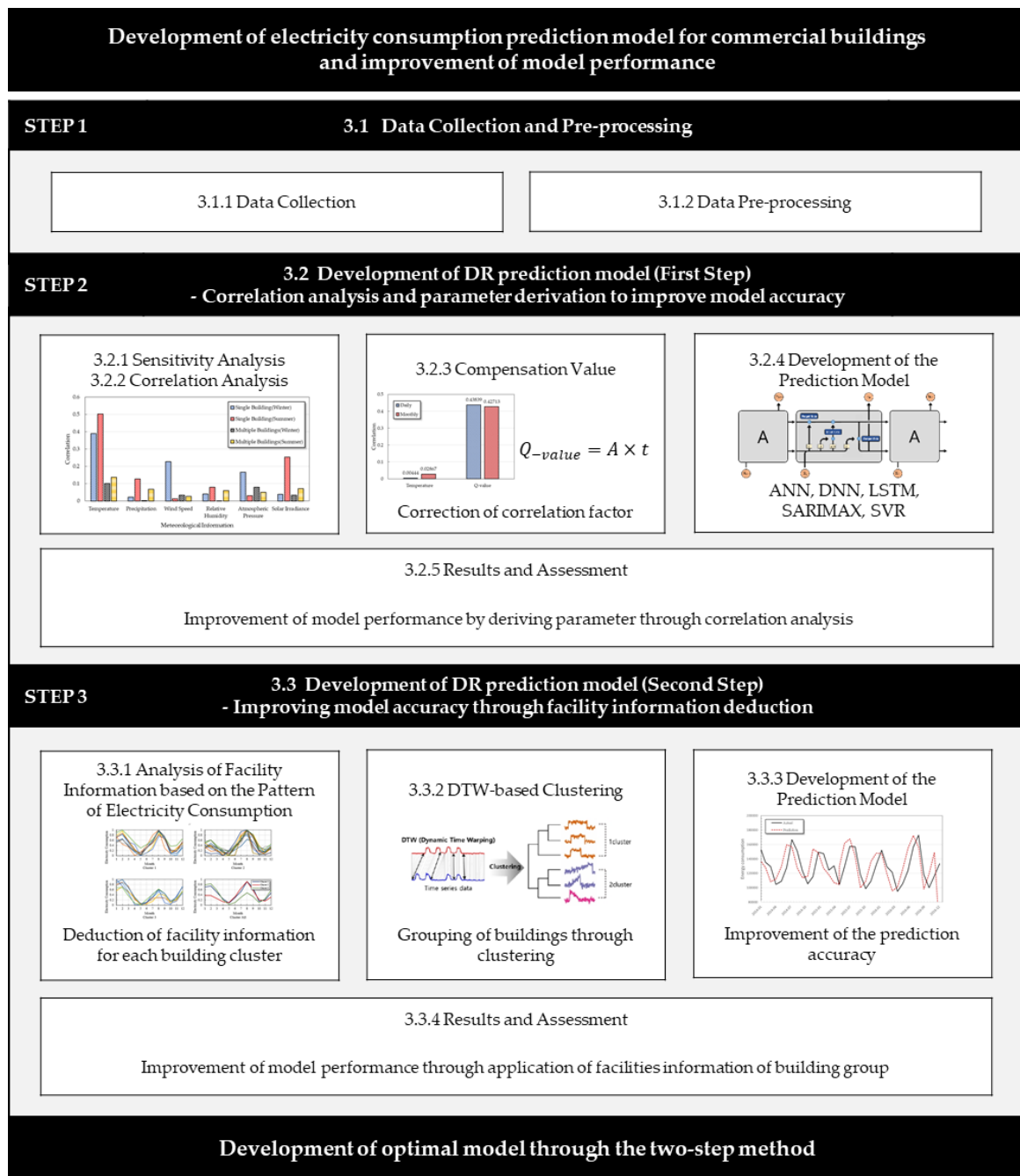


Figure 1. Research framework.

*Literature Review*

In the field of energy demand response prediction, various studies have been conducted according to various prediction targets, prediction periods, and model techniques. Specifically, a variety of machine learning and statistical methods based on time series have been widely applied. In this section, previous studies on prediction models are discussed considering the time scale given in Table 1.

**Table 1.** Summary of the studies on prediction model analysis.

Time Scale	Research	Object of Prediction	Features	Model	Evaluation
Hourly	Walker et al. [13]	Commercial Buildings Electricity Consumption	Meteorological Parameters, day of the week, hour of the day, month of the year, seasons, working day, Autoregressive parameters	Boosted-tree/Random forest, SVM, ANN,	MAPE, CV-RMSE, R <sup>2</sup> , Theil U-Statistics,
	Chae et al. [14]	Commercial Buildings Electricity Consumption	Meteorological Parameters, Time indicator, Operational condition	ANN	CV-RMSE, MBE, APE, MSE
	Ryu et al. [15]	Various Building type Load Consumption	Weather features, day of the week, weekday indicator, date information of label	DNN (RBM, ReLU), SNN, DSHW, ARIMA	MAPE, RRMSE
	Rahman et al. [16]	Commercial Buildings Electricity Consumption	HVAC Critical, HVAC Normal, Convenience, Critical, Convenience Power Normal, CRAC Critical, CRAC Norma	DNN, LSTM, MLP, NN, RNN	RMS, RMSE, Pearson Coefficient
	Jain et al. [17]	Electricity consumption of multi-family residential buildings	Meteorological Parameters, Spatial granularity	SVR	CV
Daily	Song et al. [18]	Oil production	Pressure, Temperature, Permeability, Porosity, Well length	LSTM	MAPE, MAE, RMSE
	Shao et al. [19]	Hotel Building Electricity Consumption.	Meteorological Parameters, Building measurement information, Building information.	SVR	MSE, R <sup>2</sup>
	Bouktif et al. [20]	Electric Load	Meteorological Parameters	RNN, LSTM, NN, Extra Trees, Random Forest	CV, RMSE, MAE
	Ngo et al. [21]	Cooling load in office buildings	Building information, building envelopes, Internal loads	ANN, SVR, CART, LR, Ensemble	R, RMSE, MAE, MAPE, SI, Computing time
Monthly	Jeong et al. [8]	Educational Building Electricity Consumption	Characteristics by educational facility	SARIMA, ANN, Hybrid (SARIMA, ANN)	MAPE, RMSE, MAE
	Choi et al. [22]	Gas consumption in Building	Building information Date, Temperature	ANN	R <sup>2</sup> , Pearson Coefficient
	Lee et al. [23]	Gas consumption	Gas consumption	DTW Clustering	-

First, in the case of hourly predictions, Walker et al. predicted the electricity consumption by hour in a small area of the smart grid environment for 47 commercial buildings, and used Boosted tree/Random forest, support vector machine (SVM), and ANN [13]. Chae et al. predicted electricity consumption in 15 min for commercial buildings [14]. Variable importance analysis was used to select the important variables regarding electricity consumption; building information and electricity consumption data were obtained through a building management system, and a predictive model was developed using an ANN. Ryu et al. predicted hourly load for various buildings [15]. As a prediction technique, they used a rectified linear unit without pre-training and a DNN based on a pre-training restricted Boltzmann machine. Rahman et al. used DNN, LSTM, MLP, NN, and RNN to predict hourly electricity consumption for commercial and residential buildings [16]. Jain et al. used SVR to predict hourly electricity consumption in multi-family residential buildings and made sensor-based predictions by investigating the effects of time and spatial granularity on model accuracy [17].

Related to daily models, Song et al. proposed a model for predicting daily oil production [18]. LSTM was used as a prediction technique, and the PSO algorithm was used to optimize the LSTM. Shao et al. predicted the daily electricity consumption for hotel buildings [19]. A predictive model was developed using the SVR model technique; meteorological parameters and air conditioning unit operation data were used as input parameters. The accuracy was improved after data quality evaluation and pretreatment. Bouktif et al. predicted the daily electricity consumption data for France's metropolitan areas [20]. RNN, LSTM, NN, Extra Trees, and Random Forest were used as model techniques, and the LSTM was optimized using genetic algorithm (GA). Ngo et al. developed a cooling load prediction model for 243 office buildings, and used ANN, SVR, LR, CART, and ensemble as prediction models [21].

In case of the monthly prediction, Jeong et al. developed a monthly prediction model to determine the annual energy cost budget for educational facilities, and used SARIMA, ANN, and Hybrid (SARIMA, ANN) [8]. The problems associated with the existing model were identified and a new model was proposed through the predictive model. Choi et al. and Lee et al. predicted the monthly gas consumption of buildings through DTW-based clustering classifications [22,23].

### 3. Method

As introduced in the research framework, steps 1, 2, and 3 were performed to develop the electricity consumption prediction model.

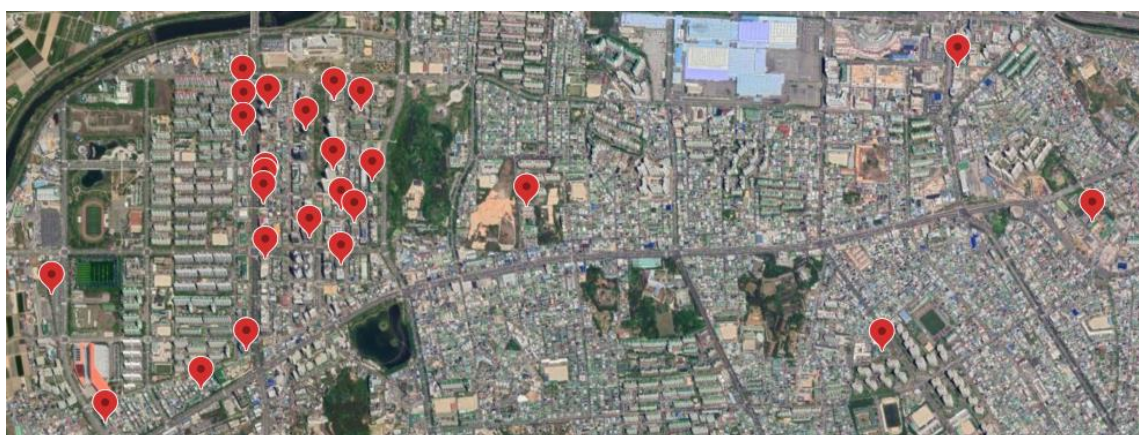
#### 3.1. STEP 1: Data Collection and Pre-Processing

To develop accurate and proper prediction models, it is very important to collect and pre-process data in an appropriate form [24]. Improper data collection and inadequate pre-processing inhibits model training and reduces the model's performance. Therefore, it is necessary to perform data pre-processing in terms of handling missing data as well as underfitting and overfitting issues [25].

##### 3.1.1. Data Collection

All the data used in this study were collected from the Public Open Data [26]. We collected 15-min and monthly electricity consumption data (i-SMART, the Power Portal Service) of 28 commercial buildings located in Seo-gu, Gwangju, Korea. In addition, we collected the meteorological data and building information of the target buildings from the KMA (Korea Meteorological Administration) [27,28]. Figure 2 shows commercial buildings' location map in Gwangju, South Korea.





**Figure 2.** Commercial buildings' location map in Gwangju, South Korea.

First, the electricity consumption data of the target buildings were collected from January to December 2016. Collected electricity consumption data was with 15-min resolution; therefore, we converted data into daily and monthly time scales. In addition, we collected data for three years from January 2014 to December 2016 to develop a monthly power consumption prediction model. Second, meteorological and building information were gathered as explanatory variables for input parameters. Meteorological data consisted of temperature, precipitation, wind speed, atmospheric pressure, relative humidity, and solar irradiance, and building information comprised the total area, number of floors, and underground floor. In addition, data including energy equipment of the buildings related to heating and cooling systems were also necessary accurately to predict the amount of energy consumption; however, it is impossible to collect the majority of the information of the building energy facilities. Therefore, this study tries to compensate for insufficient facility information through the inference method of the scenario-based analysis, which is the analysis of the electricity consumption pattern of the building and clustering classification. Furthermore, we also conducted data analysis in consideration of holidays and weekends, which greatly affect the power consumption. The collected input and target values are described in Table 2.

**Table 2.** Input parameter and target value information.

	<b>Input Parameter</b>	<b>Target Value</b>
Daily	Temperature (°C), Precipitation (mm), Wind Speed (m/s), Atmospheric Pressure (hPa), Relative Humidity (%), Solar Irradiance (MJ/m <sup>2</sup> ), Total Area (m <sup>2</sup> ), Number of Floors, Underground Floors, Day of the Week, <i>Q-value</i> , Facility type	January–December 2016 Daily electricity consumption for 1 year (kW)
Monthly	Temperature (°C), Wind Speed (m/s), Atmospheric Pressure (hPa), Relative Humidity (%), Solar Irradiance (MJ/m <sup>2</sup> ), Total Area (m <sup>2</sup> ), Number of Floors, Underground Floors, <i>Q-value</i> , Facility type	January 2014–December 2016 Monthly electricity consumption for 3 years (kW)

### 3.1.2. Pre-Processing

In the pre-processing process, a few pieces of information regarding daily precipitation were missed, so the Lagrange interpolation method was applied. If the time series are presented as  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , the Lagrange interpolation can be formulated as presented in Equation (1).

$$L(x) = \sum_{i=0}^n y_i \prod_{j=0, j \neq i}^n \frac{(x - x_j)}{(x_i - x_j)} \quad (1)$$

where  $L(x)$  is the Lagrange polynomial,  $y$  represents the interpolation value, and  $n$  is the size of the data used for interpolation [29]. Thus, some missing values were interpolated through Lagrange interpolation and used as input parameters.

The electricity consumption data of the buildings were also pre-processed to a suitable form toward a proper time scale. The collected electricity consumption data with 15-min resolution was summed and converted into daily electricity consumption. Monthly data was collected on a monthly basis. Therefore, as shown in Table 3, the input parameters consist of buildings information, meteorological information, and day of week information; a new  $Q$ -value can be used as a correction factor for analyzing the correlation between the input parameters and electricity consumption. The electricity consumption is used as a dependent variable.

To adjust the numerical differences among the datasets, min-max normalization was conducted. The normalization technique prevents the performance deterioration of the model training process due to the significant differences among target data; this helps to improve the speed of data learning process [30]. The min-max normalization is formulated as presented in Equation (2)

$$z = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2)$$

where  $x$  is an original value and  $z$  is the normalized value. Tables 3 and 4 summarize the training data for the development of the model by dividing it into daily and monthly data. To develop the prediction model for commercial buildings, target data of the 28 buildings were processed as training.





Table 3. Cont.

Building Number	Date	Temperature (°C)	Precipitation (mm)	Wind Speed (m/s)	Relative Humidity (%)	Atmospheric Pressure (hPa)	Solar Irradiance (MJ/m <sup>2</sup> )	Total Area (m <sup>2</sup> )	Number of Floors	Underground Floors	Day of the Week	Electricity Consumption (kW)
BN #26	1 January 2016	7.1	0.7	1.6	71.1	1018.7	9.29	54,304.04	17	3	1	12,819.84
	... 31	...	...	...	...	...	...	...	...	...	...	...
	December 2016	1.9	0	0.6	77.5	1021.8	6.62	54,304.04	17	3	1	13,757.76
BN #27	1 January 2016	7.1	0.7	1.6	71.1	1018.7	9.29	7279.9	9	2	1	1201.32
	... 31	...	...	...	...	...	...	...	...	...	...	...
	December 2016	1.9	0	0.6	77.5	1021.8	6.62	7279.9	9	2	1	1768.92
BN #28	1 January 2016	7.1	0.7	1.6	71.1	1018.7	9.29	2191.04	10	1	1	577.6
	... 31	...	...	...	...	...	...	...	...	...	...	...
	December 2016	1.9	0	0.6	77.5	1021.8	6.62	2191.04	10	1	1	574.39

Table 4. Example dataset of multiple buildings (monthly).

Building Number	Date	Temperature (°C)	Wind Speed (m/s)	Relative Humidity (%)	Atmospheric Pressure (hPa)	Solar Irradiance (MJ/m <sup>2</sup> )	Total Area (m <sup>2</sup> )	Number of Floors	Underground Floors	Electricity Consumption (kW)
BN #1	January 2014	2.1	1.8	58	1016.5	285.64	8238.56	9	2	96,194
	February 2014	4.2	2.2	57	1015.4	295.26	8238.56	9	2	94,178
	December 2016	4.7	1.5	69	1016.1	241.72	8238.56	9	2	69,539
BN #2	January 2014	2.1	1.8	58	1016.5	285.64	21,935.23	8	1	156,744
	December 2016	4.7	1.5	69	1016.1	241.72	21,935.23	8	1	133,456
BN #3	January 2014	2.1	1.8	58	1016.5	285.64	6401.65	5	1	31,872
	December 2016	4.7	1.5	69	1016.1	241.72	6401.65	5	1	29,750
BN #4	January 2014	2.1	1.8	58	1016.5	285.64	9939.17	5	1	81,523
	December 2016	4.7	1.5	69	1016.1	241.72	9939.17	5	1	61,373
BN #26	January 2014	2.1	1.8	58	1016.5	285.64	54,304.04	17	3	546,912
	December 2016	4.7	1.5	69	1016.1	241.72	54,304.04	17	3	461,424
	January 2014	2.1	1.8	58	1016.5	285.64	7279.9	9	2	66,898
BN #27	December 2016	4.7	1.5	69	1016.1	241.72	7279.9	9	2	50,179
	January 2014	2.1	1.8	58	1016.5	285.64	2191.04	10	1	27,079
BN #28	December 2016	4.7	1.5	69	1016.1	241.72	2191.04	10	1	23,605

### 3.2. STEP 2: Development of DR Prediction Model (First Step)

Correlation analysis and parameter derivation to improve model accuracy.

#### 3.2.1. Sensitivity Analysis

Sensitivity analysis (SA) is the study of how a mathematical model or parameter variation of a system affects the output or performance of a system [31–34]. In this section, we analyzed the effect of each meteorological parameter on energy demand through SA. Table 5 shows the SA of the electricity consumption and meteorological information.  $\delta$ (delta) is SA indices for individual inputs, and S1 is first-order global SA indices or main effect indices. Among the meteorological variables, the both delta and S1 indices of the temperature parameter shows the highest value. We recognized that the temperature factor has the highest influence on the energy consumption.

**Table 5.** Sensitivity analysis of the electricity consumption and meteorological information in a single building.

Season	Winter		Summer	
	$\delta$ (Delta)	S1	$\delta$ (Delta)	S1
Temperature	0.181880	0.246228	0.212678	0.243081
Precipitation	0.103899	0.041697	0.179938	0.085850
Wind Speed	0.085606	0.063668	0.048359	0.013573
Relative Humidity	0.056584	0.015088	0.012891	0.028091
Atmospheric Pressure	0.063352	0.031969	0.070607	0.024481
Solar Irradiance	0.030567	0.025631	0.048970	0.081328

#### 3.2.2. Correlation Analysis

Correlation analysis is a method of confirming the degree of correlation between variables. The Pearson correlation method, widely applied in data analysis, was employed in this model. The closer the correlation value is to 0, the weaker the linear correlation, where +1 and –1 indicate strong positive and negative linear correlations, respectively [35]. We investigated the correlation analysis with respect to the meteorological information and features of the building regarding the amount of electricity consumption of the building to derive the relationship between the consumption of electricity and various features of the building.

Table 6 represents the results of the correlation analysis for both summer and winter season in consideration of seasonal features for a building in terms of energy consumption. Among them, the correlation between electricity consumption and temperature variation is relatively high, with a negative high value of –0.38969 in winter, and a positive high value of 0.502439 in summer.

**Table 6.** Correlation of the electricity consumption and meteorological information from a single building.

Season	Temperature	Precipitation	Wind Speed	Relative Humidity	Atmospheric Pressure	Solar Irradiance
Winter	–0.38969	0.022229	0.227532	–0.04063	0.166891	–0.03883
Summer	0.502439	–0.12635	0.012127	–0.08027	0.030038	0.252418

On the contrary, Figure 3 shows the results of the correlation analysis for a single building as well as for 28 buildings. In case of a single building, the correlation between temperature and electricity consumption is high, but it is low in the case of multiple buildings; –0.101294 in winter and 0.137393 in summer.

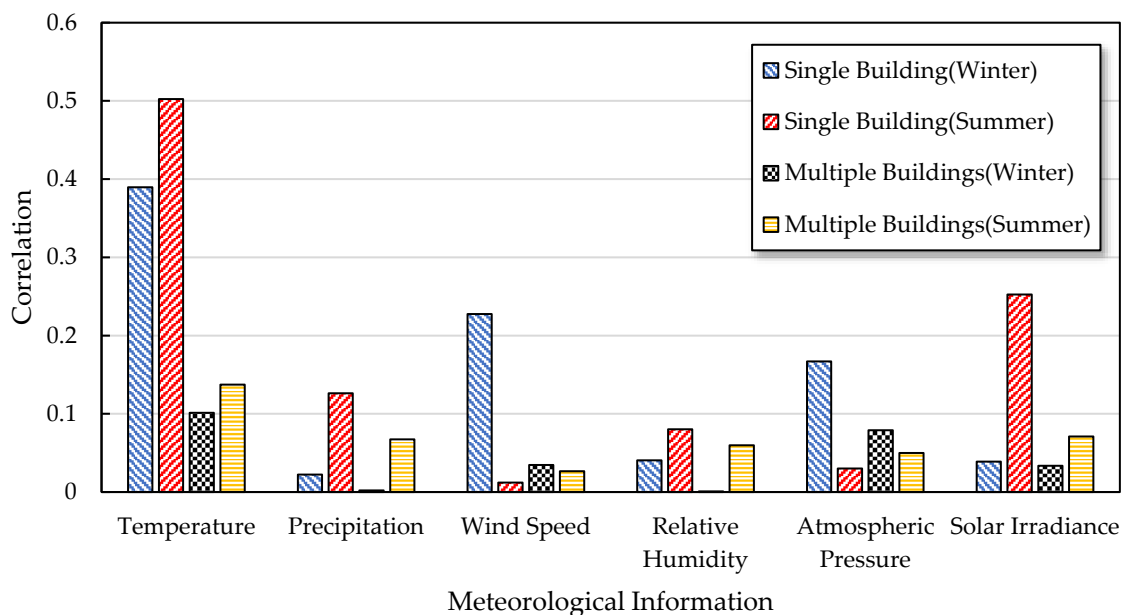


Figure 3. Correlation between electricity consumption and meteorological information.

In case of the correlation between temperature and electricity consumption, the single building case has a high correlation whereas the multiple building case shows a low correlation.

According to Tables 6 and 7, the volume of the building, including the total area, has a high correlation value between energy consumption; therefore, we considered the compensation values for data analysis because the different sizes of the buildings have significant influence on the electricity consumption variation.

Table 7. Correlation of the electricity consumption and meteorological and building information from multiple buildings.

Season	Meteorological Information					
	Temperature	Precipitation	Wind Speed	Relative Humidity	Atmospheric Pressure	Solar Irradiance
Winter	-0.101294	0.002157	0.034499	0.000760	0.078926	-0.073448
Summer	0.137393	-0.067337	-0.026587	-0.059833	0.049873	0.095713
Season	Building Information					
	Total Area	Number of Floor	Underground Floor			
Winter	0.610065	0.404868	0.254268			
Summer	0.638708	0.527150	0.445445			

### 3.2.3. Compensation Value

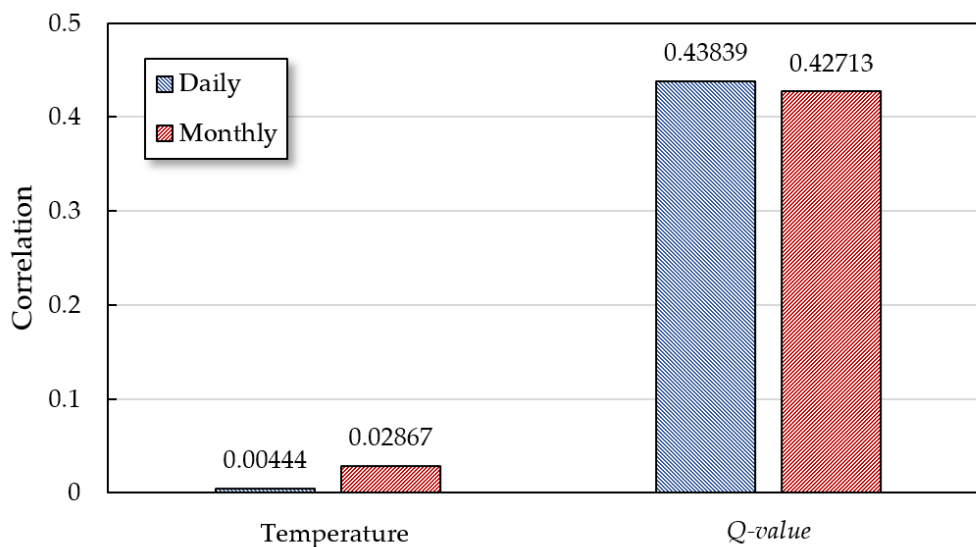
In this study, we derived a new parameter to explain the degree of electricity consumption according to the building size and variation in temperature. Therefore, a new variable was derived by combining the total area and temperature variables regarding building energy consumption.

The correction factor was derived by referring to the thermal conductivity formula to compensate for the correlation according to different total areas and temperature changes of the 28 buildings.

$$Q_{-value} = A \times t \tag{3}$$

where *A* is the total area and *t* is the temperature. A new parameter, *Q-value* was derived through the multiplication of the total area and the temperature parameter. Figure 4 shows the correlation of the electricity consumption and the results of the both temperature and *Q-value*. The correlation of the

temperature with the electricity consumption is as low as 0.00444 and 0.02867, whereas the correlation of *Q-value* with electricity consumption is as high as 0.43839 per day and 0.42713 per month.



**Figure 4.** Correlation with electricity consumption in multiple buildings.

#### 3.2.4. Development of the Prediction Model (First Step)

For the first step, we developed the prediction model, such as a traditional time-series-based model (SARIMAX) and a machine learning model (ANN, DNN, LSTM, SVR) through the correlation analysis between electricity consumption and the input parameter to the independent variable. Test Case 1 adopts monthly and daily electricity consumption data as a target value with input parameters such as meteorological data, building information, and day of the week information. In Test Case 2, *Q-value* is additionally considered an input parameter. Therefore, we evaluated five different model techniques with two test cases (Test Cases 1 and 2).

- ANN

Artificial neural networks (ANNs) are most often implemented in predictive models in machine learning. The ANN is a network model resembling functions similar to the human brain. There are numerous neurons in the human brain to process and collect information through neurons. The mathematical model was introduced in 1943 by McCulloch and Pitts [36]. The ANN's nonlinear approach is suitable for solving complicated relationships between input and output data. The ANN consists of an input layer, hidden layer, and an output layer and receives data through the input layer, processes it in the hidden layer, and obtains the result in the output layer. The operation in the hidden layer is as follows:

$$f(x) = \sum_i W_i X_i + \beta_i \quad (4)$$

$$g(f(x)) = \frac{\sinh(f(x))}{\cosh(f(x))} = \frac{e^{f(x)} - e^{-f(x)}}{e^{f(x)} + e^{-f(x)}} \quad (5)$$

where  $f(x)$  represents a combination function.  $W$  represents the weight between layers.  $X$  represents the  $i$ th neuron in the input layer.  $\beta$  represents the bias of the  $i$ th neuron. The  $g$  function is a transfer function, and Equation (5) represents the hyperbolic tangent function [37].



- SVR

SVM is one of the most popular artificial intelligence methods as a machine learning model. SVM is widely used for a classification and regression methods, and the SVM-based regression method is called the SVR. SVM was first introduced by Vapnik (1995) as a robust learning algorithm for solving nonlinear problems [38]. SVM is used to determine the optimal decision hyperplane classifying the classes by maximum margin. SVR is a learning method that includes as much data as possible in a certain margin on the same principle as SVM. The sensitivity of the model can be controlled by adjusting the margin [19,39].

- SARIMAX

Autoregressive integrated moving average (ARIMA) is widely utilized in traditional statistical time-series analysis, and is a model obtained by mixing the autoregressive (AR) model and moving average (MA) model. Seasonal autoregressive integrated moving average (SARIMA) is the model used when the data contains seasonal characteristics. In addition, seasonal autoregressive integrated moving average-X (SARIMAX) is the model that considers both seasonal characteristics and external variables. In this study, SARIMAX is used to apply the electricity consumption data type and external variables with seasonal characteristics. In this study, SARIMAX is used to apply the electricity consumption data type and external variables with seasonal characteristics [40,41].

- DNN

Deep neural network (DNN) is an ANN with multiple hidden layers. DNN solves the overfitting problem, which is a disadvantage of the ANN, and reduces the learning time. In addition, DNN can handle complicated models with fewer nodes than ANN [42]. DNN is widely used in demand forecasting and big data fields because it can model complicated nonlinear relationships. Equation (6) represents the ability for DNN to determine the weighted value.

$$\Delta w_{ij}(t+1) = \Delta w_{ij}(t) + \eta \frac{\partial C}{\partial w_{ij}}. \quad (6)$$

where  $w$  is weight,  $\eta$  is learning rate,  $C$  is cost function. The choice of cost function is determined by factors such as the type of learning (supervised learning, self-learning (machine learning), reinforcement learning) and the activation function [43].

- LSTM

Long short-term memory (LSTM) is suitable for a prediction model based on time-series analysis. LSTM was first proposed by Sepp Hochreiter and Jürgen Schmidhuber and was developed to solve the vanishing gradient problem of RNN [44].

The basic structure of LSTM contains of cells, which consist of an input gate, an output gate, and a forget gate, which stores and transmits data from the cell as shown in Figure 5. The role at the gate is as follows. The forget gate is the process of deciding whether to discard past data, the input gate determines whether to store the current information, and the output gate is the process of deciding which output value to output. The LSTM structure solves the problem of long-term prediction of RNN [18].

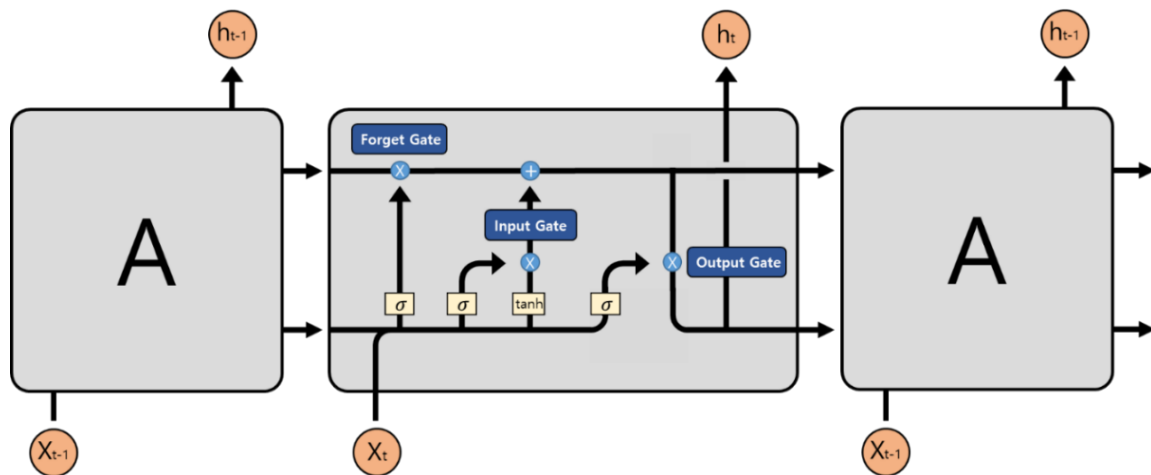


Figure 5. Structure of LSTM (Long Short-Term Memory).

In this study, to predict the electricity consumption at time  $t + 1$ , the model was developed by adding the electricity consumption, the target value at time  $t$ , to the input parameter at time  $t + 1$ .

### 3.2.5. Results and assessment (First Step)

- Error calculation

To evaluate the electricity consumption prediction performance for each model, we use mean absolute percentage error (MAPE), root mean square error (RMSE), mean bias error (MBE), and coefficient of variation (CV) evaluation. Equations (7)–(10) describe the equations of the performance evaluation.

$$\text{MAPE} = \frac{100}{n} \sum_{t=1}^n \left| \frac{A_t - P_t}{A_t} \right| \quad (7)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (A_t - P_t)^2} \quad (8)$$

$$\text{MBE} = \frac{\sum_{t=1}^n (A_t - P_t)}{\sum_{t=1}^n A_t} \quad (9)$$

$$\text{CV} = \frac{\text{RMSE}}{\frac{1}{n} \sum_{t=1}^n A_t} = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (A_t - P_t)^2}}{\frac{1}{n} \sum_{t=1}^n A_t} \quad (10)$$

where  $A_t$  is actual data,  $P_t$  is target data, and  $n$  is number of samples. MAPE is a value expressed as a percentage of the error between the measured value and the predicted value and is generally used to predict the error [45]. RMSE is a value representing the error between the measured value and the predicted value as standard deviation [46]. The RMSE is a relative value and is affected by the size of the sample. Therefore, the RMSE can be evaluated by relative comparison on the same data set. MBE is a value representing the average of the deviation between the measured value and the predicted value. The closer to 0, the better regardless of the sign [47]. CV is expressed as the square root of the mean squared error, and the model with the minimum value or 0 is the optimal model [14].

- Assessment of trained models

Figure 6 and Table 8 show the results of the performance evaluation with the *Q-value* for commercial multiple buildings. Test Case 2 shows the performance results of applying the *Q-value* as one of the input parameters, whereas Test Case 1 considers meteorological data, building information, and day of weekend or not as input values. Besides, the performance results of the SARIMAX ANN, DNN, and LSTM show that the overall result of Test Case 2 has good performance compared to that of Test Case 1.

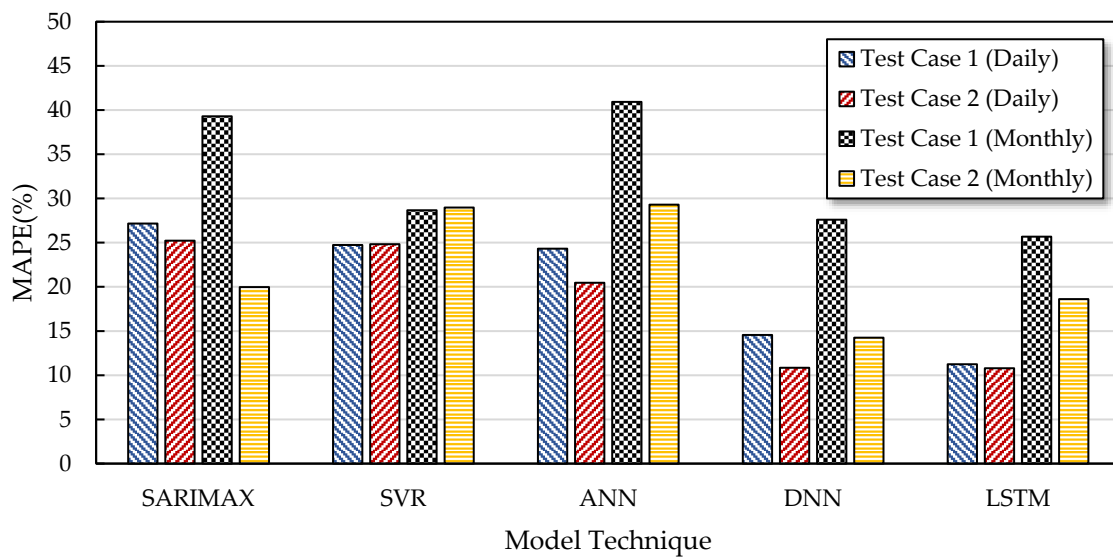


Figure 6. Individual assessment graph for trained models (MAPE).

Table 8. Individual evaluation summary for trained models in multiple buildings (First Step).

Test Case	Performance Evaluation	SARIMAX	SVR	ANN	DNN	LSTM	
Daily	Test Case 1	MAPE (%)	27.15991	24.74141	24.32531	14.54982	11.23937
		RMSE (kW)	557.6002	711.1884	571.3286	406.2006	579.5171
		MBE (%)	-1.18098	-1.22898	2.777059	-0.76317	-0.31540
		CV (%)	18.21689	23.17986	18.66540	13.27064	18.13398
	Test Case 2	MAPE (%)	25.22460	24.83279	20.45893	10.83902	10.78970
		RMSE (kW)	669.5132	701.3550	452.3658	382.9493	389.8103
		MBE (%)	-1.07525	-1.21958	2.63918	1.12560	0.258391
		CV (%)	21.42904	22.85935	14.77887	12.51102	12.73517
Monthly	Test Case 1	MAPE (%)	39.30651	28.67256	40.92174	27.59819	25.68580
		RMSE (kW)	29795.64	41911.26	28848.64	15503.43	32260.52
		MBE (%)	2.531365	-8.84265	-1.27958	-4.62698	-4.156262
		CV (%)	34.81914	44.91909	33.71248	18.11729	37.547024
	Test Case 2	MAPE (%)	19.96770	28.96614	29.30571	14.24719	18.61913
		RMSE (kW)	27423.98	41840.32	17719.81	13946.75	26755.26
		MBE (%)	-1.94524	-9.77504	-1.43820	-0.47899	2.69541
		CV (%)	29.00761	44.84306	20.70734	14.94766	31.13961

Most of the performance results considering *Q-values* show that the accuracy has improved except for the SVR case. The performance result of the SVR shows a similar outcome whether considering *Q-value* or not. In addition, the DNN and LSTM have the highest accuracy compared to the performance results of the SARIMAX, SVR, ANN, DNN, and LSTM.

However, to improve the accuracy of the demand prediction model for multiple buildings, which has been improved with the addition of  $Q$ -value, it is necessary to analyze more various influencing factors affecting electricity consumption.

Therefore, in this study, to improve the performance of the predictive model, we try to use the information related to heating and cooling facilities that affect the electricity consumption of the building for analysis. Also, we intend to infer facility information through a clustering technique for using equipment information that is difficult to collect.

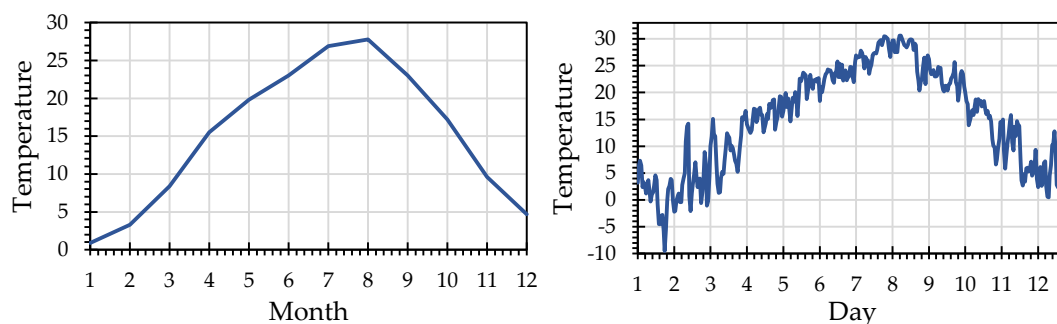
### 3.3. STEP 3: Development of DR Prediction Model (Second Step)

Improving model accuracy through facility information deduction.

#### 3.3.1. Analysis of Facility Information Based on the Pattern of Electricity Consumption

Energy consumption in commercial building consists of heating and cooling, lighting, and office equipment, and heating and cooling accounts for 48.9% of the total [48]. Energy consumption of lighting and office equipment account for a certain level of energy regardless of seasonal changes; however, energy demand for heating and cooling facilities tends to vary according to the season. Therefore, it is crucial to utilize information on heating and cooling facilities of the buildings in terms of the development of electricity demand response prediction models for multiple buildings.

As shown in Figure 7, as South Korea has four distinct seasons in spring, summer, autumn, and winter according to temperature changes, electricity consumption varies by season. The amount of energy consumption for heating and cooling occupies a large portion of total energy, especially summer and winter season. We found that it is possible to derive facility information of the buildings by analyzing power consumption pattern with regard to seasonal features.



**Figure 7.** Daily and monthly temperature for January–December 2016 in Gwangju, South Korea.

Generally, building energy consumption for heating and cooling includes gas, district heating, and solar energy as well as electricity in Korea. Therefore, the energy used for heating and cooling facilities is inferred based on the pattern of electric energy consumption. In this step, we performed a classification process based on the fact that the heating and cooling system of the commercial building can be categorized in accordance with the literature survey in Korea [49–52].

On the basis of summer and winter, information on heating and cooling facilities for the three scenarios are inferred. The information on heating and cooling facilities is inferred by classifying them into three cases based on summer and winter. The following seasonal features can be derived from the electricity usage pattern for cooling of buildings. The average temperature is 25 °C during July and August in Korea, and due to the hot weather, most buildings need to use air conditioning to maintain building temperature. Accordingly, most of the buildings consume a lot of electricity. Meanwhile, in the case of public buildings, the energy saving policy requires restricted use of air conditioners to maintain the indoor temperature between 26 °C and 28 °C. Due to the temperature restrictions imposed for public buildings, the electricity consumption is lower than that of other

buildings, and consumption patterns are consistent. Thus, two features of buildings can be considered: unlimited electricity consumption and limited and consistent electricity consumption in the summer.

In Seo-gu, Gwangju, Korea, the average annual temperature in winter is 5 °C. When the ambient temperature is lower than a certain temperature, the building uses more energy for heating. Heating systems used in Korean commercial buildings is classified into heating types using electricity and gas (city gas and a system using district heating).

Most of the commercial buildings use electric heat pump (EHP) systems, whereas some commercial buildings use mixed type the heating system adopting electricity and gas together. Therefore, for commercial buildings, heating systems can be divided into two types; EHP system and mixed heating.

Therefore, buildings can be divided into three cases according to the type of heating and cooling energy source used in summer and winter. Table 9 shows three clusters classified by facility types.

**Table 9.** The number of three clusters classified by facility types.

Case	Summer	Winter
Case 1	Electricity as a cooling facility	Electricity as a heating facility
Case 2	Electricity as a cooling facility	Mixed energy as a heating facility
Case 3	Electricity as a cooling facility (Electricity usage restrictions)	Electricity as a heating facility

Case 1 involves a building that uses electric heating and cooling facilities in both summer and winter. Case 2 involves a building that uses electric cooling in summer and mixed energy heating in winter. Case 3 involves a building that adopts temperature restriction in summer and electric heating in winter. These three cases are used as criteria for clustering classification; 28 buildings were classified into these clusters through clustering analysis (Section 3.3.2).

### 3.3.2. DTW-Based Clustering

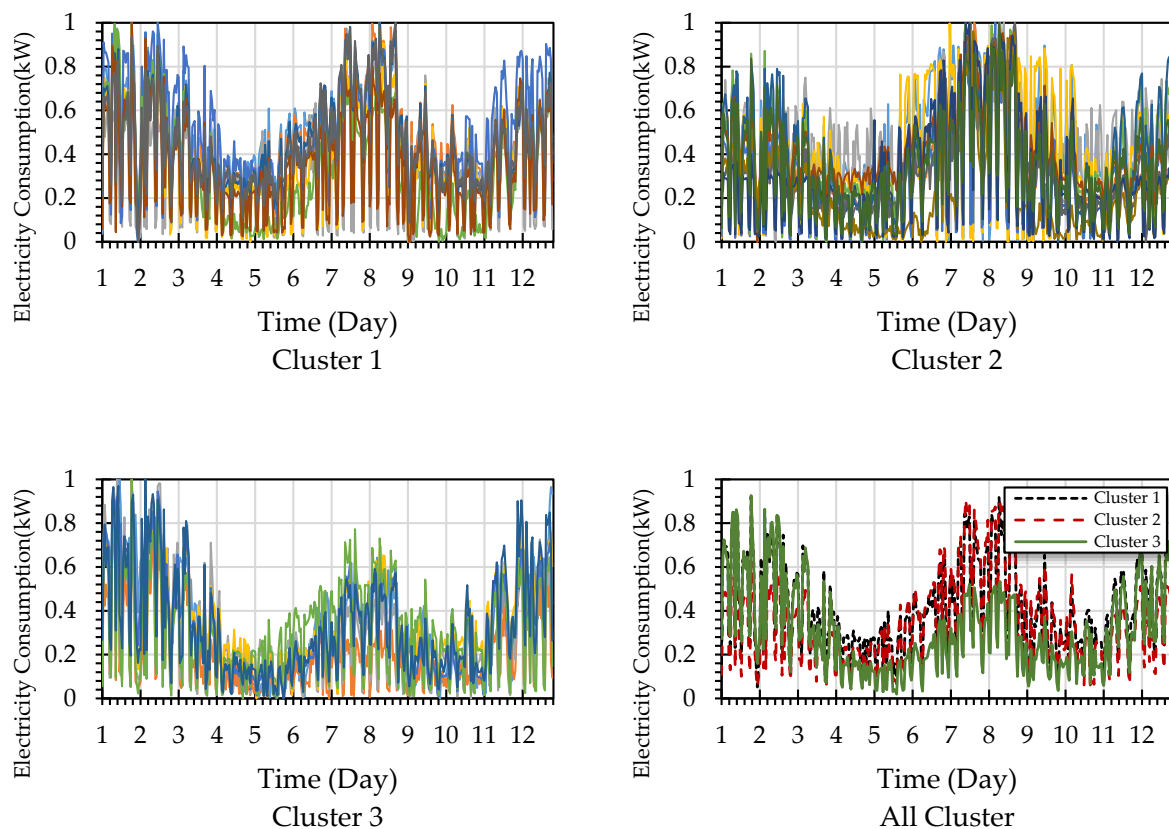
Each building exhibits different electricity consumption patterns depending on the heating and cooling facility types or purpose of use. In particular, the pattern of electricity consumption differs according to the type of heating and cooling facility, and the pattern of electricity consumption can be identified through the inference of the facility types. Therefore, as in Section 3.3.1, the facility type of the buildings is estimated, and then we can classify the building clusters according to the electricity consumption patterns of various buildings.

The hierarchical clustering of 28 buildings is required based on the results of facility information analysis. The clustering technique is a method of dividing data having similar patterns into the similar group. By calculating the distance between the data, the data of adjacent distances belong to a group. Representatively, clustering based on the Euclidean distance algorithm can be used, and similarity between entities can be found by calculating the shortest distance between entities. However, the Euclidean distance algorithm has a limitation in reflecting time series. The electricity consumption has time-series information, and continuous change over time is a key feature. Thus, DTW algorithm is able to measure similarity of time series and suitable for data changing non-linearly in the time dimension [53]. Therefore, a hierarchical clustering analysis was performed using the DTW algorithm.

Prior to clustering classification, electricity consumption data of each building was normalized to reduce clustering classification errors.

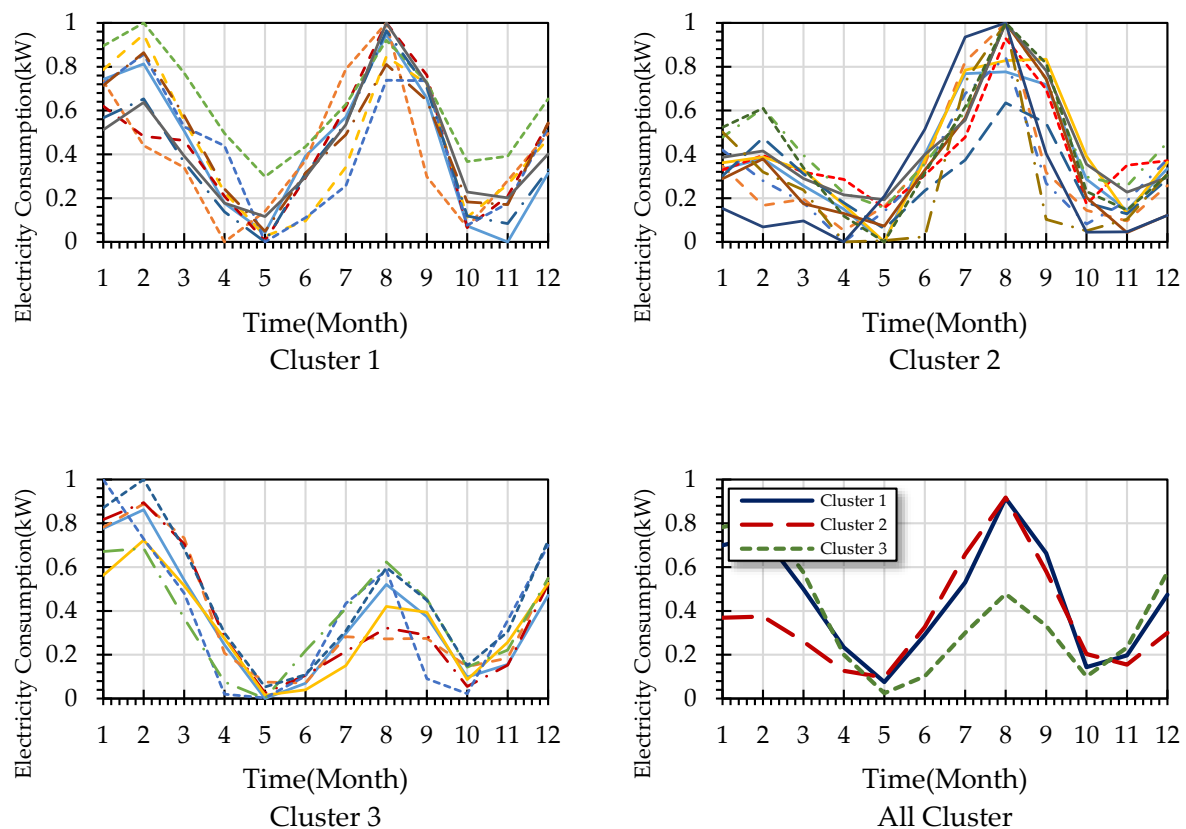
Figures 8 and 9 show the electricity consumption patterns of the buildings according to each cluster. The colored lines on each graph represent each building. Clusters are classified based on three cases, which are the results of heating and cooling facility information analyses. We can see that buildings having similar patterns of electricity consumption in a formed cluster. Cluster 1 is composed of buildings that consume a lot of electricity in summer and winter. Cluster 2 is composed of buildings with high power consumption in summer and relatively low power consumption in winter. Cluster 3 is composed of buildings that consume more electricity in winter than in summer. All of the 28 buildings, the number of buildings in Cluster 1 is 8, the number of buildings in Cluster 2 is 12, and the number of buildings in cluster is 7 buildings. The graph of the all the clusters present an average value of the electricity consumption of each cluster.

Case 1, described in Table 8, is equal to the power consumption pattern of Cluster 1, Case 2 equals Cluster 2, and Case 3 equivalent Cluster 3. Accordingly, we can see that the result of the clustering and building's energy consumption pattern inferred from the energy facility types of the buildings are identical. Therefore, the features of heating and cooling systems affecting energy consumption can be derived, and we can use this inferred value to predict future energy consumption accurately.



**Figure 8.** Daily electricity consumption of each building cluster.





**Figure 9.** Monthly electricity consumption of each buildings cluster.

### 3.3.3. Development of a Prediction Model (Second Step)

The clustering analysis through the estimation of the building's facility was classified into three buildings with similar patterns of electricity consumption. Twenty-eight buildings classified into three clusters were applied to the demand prediction model.

To design an optimal prediction model and evaluate the performance, a proposed model was evaluated using three different datasets.

Table 10 shows the description of the experimental datasets.

**Table 10.** Description of the experimental datasets.

Dataset Name	Group Description
Test Case 1	Original dataset in building group.
Test Case 2	Dataset using <i>Q-value</i> (First Step in Section 3.2)
Test Case 3	Dataset using <i>Q-value</i> (First Step) and facility information (Second Step in Section 3.3)

Test Case 1 was evaluated using an original dataset, Test Case 2 was evaluated using a dataset that used *Q-value* (described in Section 3.2), and Test Case 3 was evaluated using a dataset that used *Q-value* and facility information obtained via clustering analysis.

These three test cases were evaluated at different time scales (both daily and monthly); same input parameters were utilized, including meteorological information, building information, and *Q-value*. As shown in the previous experiment, the performance evaluation was conducted using machine learning and time series-based statistical models: ANN, DNN, LSTM, SVR and SARIMAX.

In the case of ANN, DNN, and LSTM, hyperparameters have a significant influence on the performance of model training; in particular, the difference between accuracy and learning speed can

be controlled via hyperparameter tuning. The machine learning model can be evaluated according to the number of nodes and batch size. Therefore, we conducted experiments by adjusting the number of nodes and batch size to find the optimal parameter selection. Batch sizes of 5, 10, 20, and 50 were tested, the number of nodes was evaluated from 10 to 400, and the selected number of hidden layers was three. The optimal hyperparameter values were selected based on reference [54]. In all experiments, the same number of nodes and batch sizes were applied. Finally, the hyperparameters of the test model with the highest accuracy were selected.

From the experimental result, we found that the error rate was low when most of the values in the daily and monthly models were  $\geq 100$  or more nodes; we can get similar accuracy rates with the number of nodes over 100. In this study, we evaluated the models considering various batch sizes and number of nodes to obtain high accuracy.

### 3.3.4. Results and Assessment (Second Step)

Figure 10 and Table 11 show the results of performance evaluation for each test case. Test Case 3 evaluated by dataset using *Q-value* and facility information shows better performance in the MAPE, RMSE, MBE and CV compared to Test Case 2 and Test Case 3 in all five methods with every time scale. In the first step, Test Case 3 shows better performance than Test Case 2, and we could confirm that the derived value of the heating and cooling type renders improved performance.

A short-term electricity demand response prediction model was evaluated by each time period. The LSTM exhibited an error rate of 8.97% for MAPE and 12.70% for CV in daily prediction. In case of the monthly prediction, the DNN exhibited an error rate of 10.85% for MAPE and 13.58% for CV, the daily prediction showed a lower error rate than the monthly prediction.

The proposed method was assessed according to five machine learning and time series estimation techniques to determine the optimal model. In the daily case, the LSTM showed the best performance, whereas the DNN presented the best performance in the monthly case. As there was insufficient data in the case of the monthly evaluation compared to the daily case, the influence of the time-series effect of the LSTM technique was less reflected.

Summarizing the abovementioned results, the LSTM showed the best performance in Test Case 3 in the daily case. In the daily case, the LSTM showed the best performance of MAPE 8.97%, RMSE 388.67 (kW), MBE 0.18%, and CV 12.7%. In the second step, wherein *Q-value* and facility information, were used we are able to obtain improved performance in every evaluation technique in daily and monthly cases.

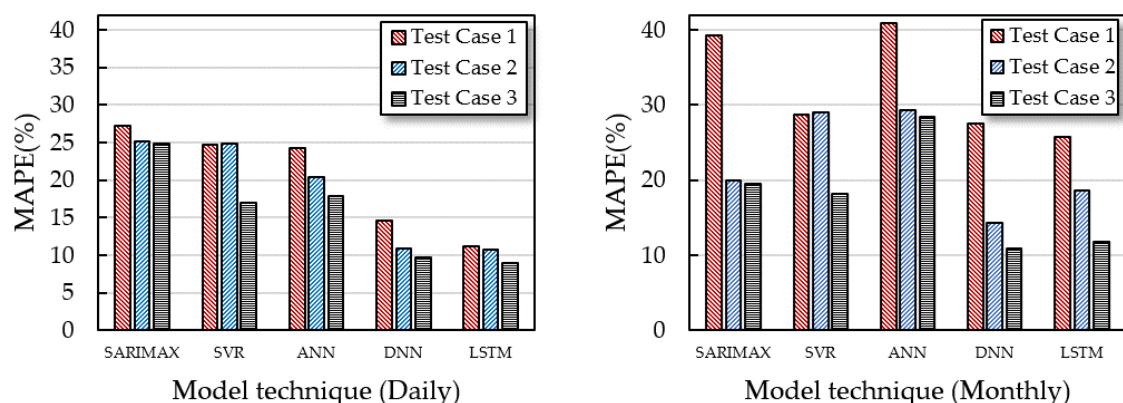


Figure 10. Assessment of individual model (Steps 2 and 3).

**Table 11.** Comparison of the predicted results from Test Cases 1, 2, and 3 (Second Step).

	Test Case	Performance Evaluation	SARIMAX	SVR	ANN	DNN	LSTM
Daily	Test Case 1	MAPE (%)	27.15991	24.74141	24.32531	14.54982	11.23937
		RMSE (kW)	557.6002	711.1884	571.3286	406.2006	579.5171
		MBE (%)	-1.18098	-1.22898	2.777059	-0.76317	-0.31540
		CV (%)	18.21689	23.17986	18.66540	13.27064	18.13398
	Test Case 2	MAPE (%)	25.22460	24.83279	20.45893	10.83902	10.78970
		RMSE (kW)	669.5132	701.3550	452.3658	382.9493	389.8103
		MBE (%)	-1.07525	-1.21958	2.63918	1.12560	0.258391
		CV (%)	21.42904	22.85935	14.77887	12.51102	12.73517
	Test Case 3	MAPE (%)	24.84949	16.95384	17.88299	9.77652	8.96914
		RMSE (kW)	653.5616	414.9893	439.2258	426.7818	388.6730
		MBE (%)	-0.37903	0.646133	-0.17665	-0.10945	0.18410
		CV (%)	20.91847	13.5258	14.3496	13.9101	12.6980
Monthly	Test Case 1	MAPE (%)	39.30651	28.67256	40.92174	27.59819	25.68580
		RMSE (kW)	29795.64	41911.26	28848.64	15503.43	32260.52
		MBE (%)	2.531365	-8.84265	-1.27958	-4.62698	-4.156262
		CV (%)	34.81914	44.91909	33.71248	18.11729	37.547024
	Test Case 2	MAPE (%)	19.96770	28.96614	29.30571	14.24719	18.61913
		RMSE (kW)	27423.98	41840.32	17719.81	13946.75	26755.26
		MBE (%)	-1.94524	-9.77504	-1.43820	-0.47899	2.69541
		CV (%)	29.00761	44.84306	20.70734	14.94766	31.13961
	Test Case 3	MAPE (%)	19.50041	18.11309	28.35975	10.84625	11.79058
		RMSE (kW)	26215.69	14633.71	17113.59	12667.49	12205.81
		MBE (%)	-1.93994	-0.75605	-1.32507	0.17049	-1.28465
		CV (%)	27.72955	15.68392	19.99891	13.57659	13.08178

- Comparison of actual and predicted values

Figures 11 and 12 show the actual and predicted values for one of the target buildings (building number 1) based on the results listed in Table 11. The figures show the monthly and daily electricity consumption prediction results generated using five predictive model techniques. Test Case 3 using facility information and the *Q-value* as an input parameter showed a high degree of agreement between the actual and predicted values in every evaluation model. Overall, the DNN and LSTM had the highest agreement between the predicted and actual values compared to other model techniques, and the highest accuracy was observed in Test Case 3.

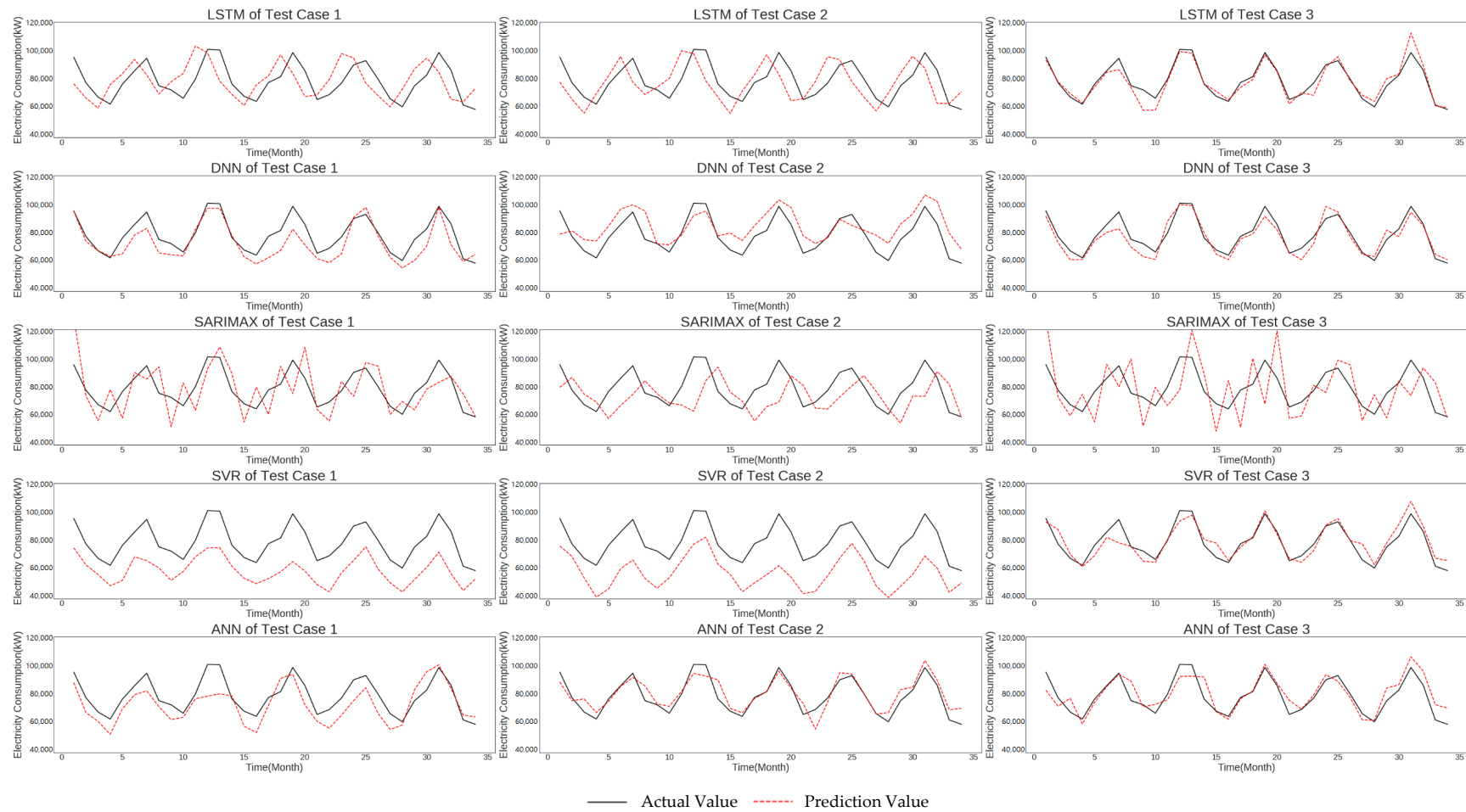


Figure 11. Graph comparing actual and predicted values in monthly cases.

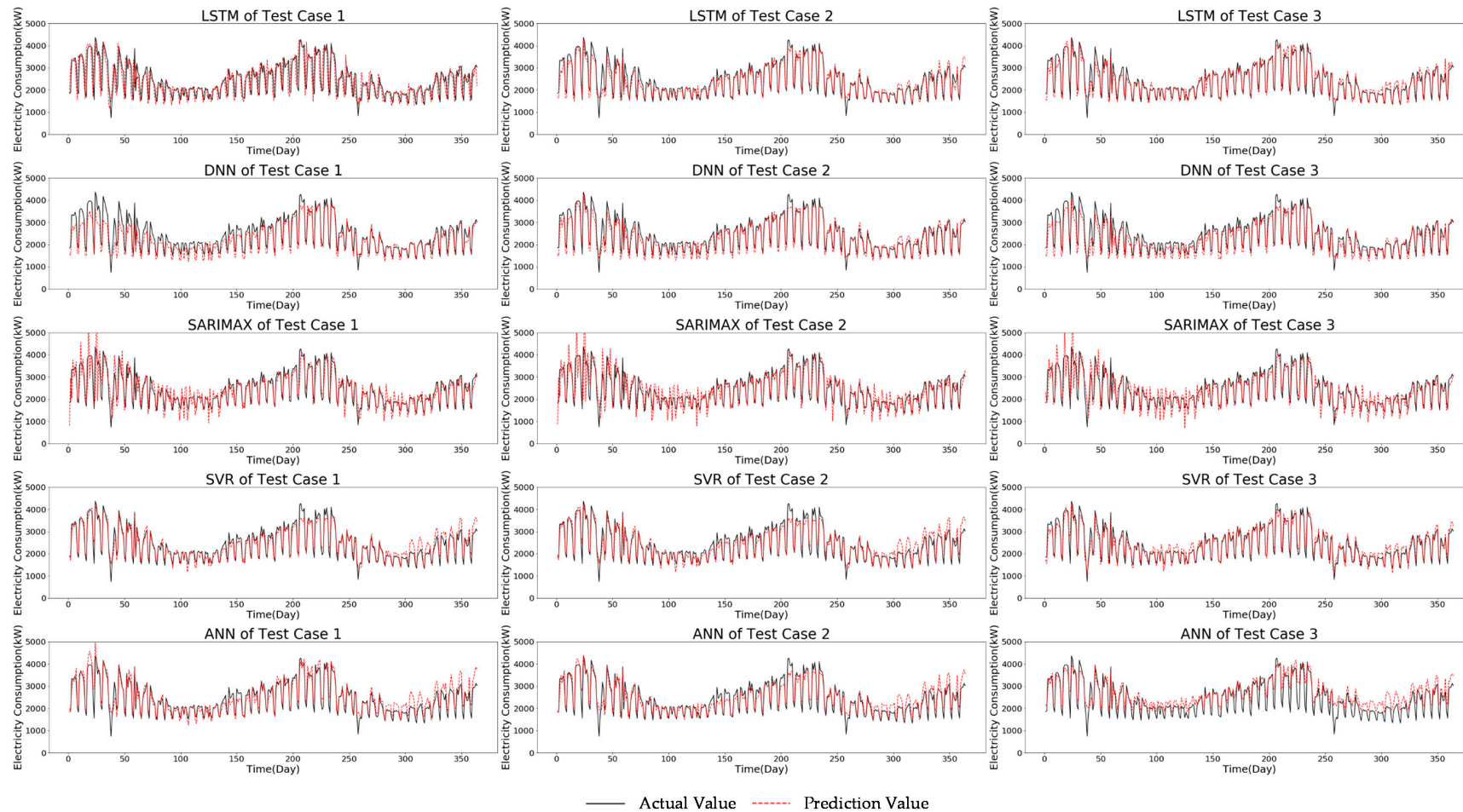


Figure 12. Graph comparing actual and predicted values in daily cases.



#### 4. Discussion and Conclusions

In this study, a daily and monthly electricity consumption prediction model for 28 commercial buildings was proposed using open-access data in Seo-gu, Gwangju, South Korea. Performing energy research of multiple buildings in metropolitan city is difficult as it is cumbersome to gather all of the specific information of target buildings, such as heating and cooling system types, facility information, and the prediction accuracy is decreased when developing a prediction algorithm using insufficient information. Therefore, we proposed a two-step approach to overcome the limitations of data collection, and we designed machine learning and statistics-based prediction models considering daily and monthly cases, belonging to the range of short-term electricity demand prediction. For the development of the prediction algorithm, five different techniques—machine learning techniques (ANN, DNN, LSTM, and SVR) and the traditional time-series analysis technique (SARIMAX)—were evaluated with respect to various test conditions.

In the case of the multiple buildings, in the first step, there is a problem that the features of individual buildings are not reflected. Therefore, to apply the features of individual buildings, a predictive model was developed by analyzing the correlation of electricity consumption with meteorological and building information. Through correlation analysis, the new parameter was developed by combining the total area and temperature variables that explain the effects of the features of individual buildings. As a result of developing the daily monthly prediction model, performance improvement was achieved in all models except for the SVR; in particular, significantly improved results were obtained in the monthly model.

In the second step, as the electricity consumption patterns vary for buildings according to the heating and cooling equipment, building information and building facility information are crucial in developing a building's electricity demand prediction model. However, collection of building facility information is limited due to various problems. Therefore, we inferred building energy facility types and propose a predictive model by performing electricity consumption pattern analysis and DTW-based clustering classification. Based on the inferred information of the heating and cooling types, buildings were divided into three categories and evaluated.

In the experiment, the Test Case 3 model, which reflected the characteristics of individual buildings and the energy consumption pattern in the prediction model, showed the highest performance by applying the *Q-value* and information on the heating and cooling facilities. It was confirmed that the prediction model performed 27.5% better than the existing model (Table 11). The degree of agreement between the actual and predicted values was confirmed through Figures 11 and 12.

In this study, an improved building electricity consumption prediction model was developed based on building and facility information using a two-step approach. Building and facility information are essential in developing a highly accurate electricity consumption prediction model. However, inappropriate parameter selection of the prediction model can lead to decreased prediction accuracy.

With the approach proposed in this study, we prove that the problems associated with insufficient data needed for the prediction can be resolved. In addition, we show the importance of parameter selection for model prediction affecting energy consumption as well as verifying which model would be a suitable machine learning and time-series algorithm for short-term energy prediction. By enabling stable power supply through accurate electricity demand response prediction, efficient power operation in specific regions and cities will be possible.

**Author Contributions:** Conceptualization and methodology were conducted by J.H. and D.S. Writing of the original draft was accomplished by J.H. and D.S. Writing, including review and editing, was performed by D.S. and M.-O.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by “Human Resources Program in Energy Technology” of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and was granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea. (No. 2019401000040) and Korea Electric Power Corporation (grant number R19X001-04).

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



## References

1. Cullen, D. Climate change. *Nature* **2011**, *479*, 267–268. [CrossRef]
2. Horowitz, C.A. Paris Agreement. *Int. Leg. Mater.* **2016**, *55*, 740–755. [CrossRef]
3. Bae, K.Y.; Jang, H.S.; Jung, B.C.; Sung, D.K. Effect of Prediction Error of Machine Learning Schemes on Photovoltaic Power Trading Based on Energy Storage Systems. *Energies* **2019**, *12*, 1249. [CrossRef]
4. Kneifel, J.; Webb, D. Predicting energy performance of a net-zero energy building: A statistical approach. *Appl. Energy* **2016**, *178*, 468–483. [CrossRef]
5. Walker, S.; Labeodan, T.; Boxem, G.; Maassen, W.; Zeiler, W. An assessment methodology of sustainable energy transition scenarios for realizing energy neutral neighborhoods. *Appl. Energy* **2018**, *228*, 2346–2360. [CrossRef]
6. Kwon, O.S.; Bin Song, K. Development of Short-Term Load Forecasting Method by Analysis of Load Characteristics during Chuseok Holiday. *Trans. Korean Inst. Electr. Eng.* **2011**, *60*, 2215–2220. [CrossRef]
7. Saleh, M.S.; Althaibani, A.; Esa, Y.; Mhandi, Y.; Mohamed, A.A. Impact of clustering microgrids on their stability and resilience during blackouts. In Proceedings of the 2015 International Conference on Smart Grid and Clean Energy Technologies (ICSGCE), Offenburg, Germany, 20–23 October 2015; pp. 195–200.
8. Jeong, K.; Koo, C.; Hong, T. An estimation model for determining the annual energy cost budget in educational facilities using SARIMA (seasonal autoregressive integrated moving average) and ANN (artificial neural network). *Energy* **2014**, *71*, 71–79. [CrossRef]
9. Muralitharan, K.; Sakthivel, R.; Vishnuvarthan, R. Neural network based optimization approach for energy demand prediction in smart grid. *Neurocomputing* **2018**, *273*, 199–208. [CrossRef]
10. Ahmad, T.; Chen, H. Potential of three variant machine-learning models for forecasting district level medium-term and long-term energy demand in smart grid environment. *Energy* **2018**, *160*, 1008–1020. [CrossRef]
11. Building Act. Korea Law Information Center. Available online: <http://www.law.go.kr/법령/건축법> (accessed on 20 September 2020).
12. Statistics on Buildings. Molit Statistics System. 2019. Available online: <https://stat.molit.go.kr/> (accessed on 20 September 2020).
13. Walker, S.; Khan, W.; Katic, K.; Maassen, W.; Zeiler, W. Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings. *Energy Build.* **2020**, *209*, 109705. [CrossRef]
14. Chae, Y.T.; Horesh, R.; Hwang, Y.; Lee, Y.M. Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. *Energy Build.* **2016**, *111*, 184–194. [CrossRef]
15. Ryu, S.; Noh, J.; Kim, H. Deep Neural Network Based Demand Side Short Term Load Forecasting. *Energies* **2016**, *10*, 3. [CrossRef]
16. Rahman, A.; Srikumar, V.; Smith, A.D. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. *Appl. Energy* **2018**, *212*, 372–385. [CrossRef]
17. Jain, R.; Smith, K.M.; Culligan, P.J.; Taylor, J.E. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Appl. Energy* **2014**, *123*, 168–178. [CrossRef]
18. Song, X.; Liu, Y.; Xue, L.; Wang, J.; Zhang, J.; Wang, J.; Jiang, L.; Cheng, Z. Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model. *J. Pet. Sci. Eng.* **2020**, *186*, 106682. [CrossRef]
19. Shao, M.; Wang, X.; Bu, Z.; Chen, X.; Wang, Y. Prediction of energy consumption in hotel buildings via support vector machines. *Sustain. Cities Soc.* **2020**, *57*, 102128. [CrossRef]
20. Bouktif, S.; Fiaz, A.; Ouni, A.; Serhani, M.A. Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches. *Energies* **2018**, *11*, 1636. [CrossRef]
21. Ngo, N.-T. Early predicting cooling loads for energy-efficient design in office buildings by machine learning. *Energy Build.* **2019**, *182*, 264–273. [CrossRef]
22. Choi, D.; Lee, Y.; Koh, M. The Prediction and Valuation of Gas Consumption in Building using Artificial Neural Networks Based on Clustering Method. *KIEAE J.* **2018**, *18*, 69–74. [CrossRef]

23. Lee, Y.-J.; Ko, M.-J.; Choi, D.-S. Pattern and Energy Intensity Analysis of Monthly Gas Energy Consumption in Apartment Using Dynamic Time Warping Hierarchical Clustering. *J. Korean Soc. Living Environ. Syst.* **2019**, *26*, 134–139. [[CrossRef](#)]
24. Pyle, D. *Data Preparation for Data Mining*; Morgan Kaufmann Publishers: Burlington, MA, USA, 1999.
25. Bourdeau, M.; Zhai, X.; Nefzaoui, E.; Guo, X.; Chatellier, P. Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustain. Cities Soc.* **2019**, *48*. [[CrossRef](#)]
26. Public Open Data. Available online: <https://www.data.go.kr/> (accessed on 20 September 2020).
27. Korea Meteorological Administration (KMA). Available online: <https://data.kma.go.kr> (accessed on 20 September 2020).
28. A Legister of Building. Saeumteo. Available online: <https://cloud.eais.go.kr> (accessed on 20 September 2020).
29. Meijering, E. A chronology of interpolation: From ancient astronomy to modern signal and image processing. *Proc. IEEE* **2002**, *90*, 319–342. [[CrossRef](#)]
30. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Available online: <https://arxiv.org/abs/1502.03167> (accessed on 20 September 2020).
31. Sobol, I. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* **2001**, *55*, 271–280. [[CrossRef](#)]
32. Wei, P.; Lu, Z.; Yuan, X. Monte Carlo simulation for moment-independent sensitivity analysis. *Reliab. Eng. Syst. Saf.* **2013**, *110*, 60–67. [[CrossRef](#)]
33. Wei, P.; Lu, Z.; Song, J. Moment-Independent Sensitivity Analysis Using Copula. *Risk Anal.* **2013**, *34*, 210–222. [[CrossRef](#)]
34. Borgonovo, E. A new uncertainty importance measure. *Reliab. Eng. Syst. Saf.* **2007**, *92*, 771–784. [[CrossRef](#)]
35. Adler, J.; Parmryd, I. Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander’s overlap coefficient. *Cytom. Part A* **2010**, *77A*, 733–742. [[CrossRef](#)]
36. McCulloch, W.S.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. 1943. *Assoc. Symb. Log* **1990**, *52*, 99.
37. Gurney, K. *An Introduction to Neural Networks*; CRC Press: Boca Raton, FL, USA, 2018.
38. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
39. Dong, B.; Cao, C.; Lee, S.E. Applying support vector machines to predict building energy consumption in tropical region. *Energy Build.* **2005**, *37*, 545–553. [[CrossRef](#)]
40. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; Wiley: Hoboken, NJ, USA, 2016.
41. Contreras, J.; Espínola, R.; Nogales, F.J.; Conejo, A.J. ARIMA models to predict next-day electricity prices. *IEEE Trans. Power Syst.* **2003**, *18*, 1014–1020. [[CrossRef](#)]
42. Bengio, Y.; Courville, A.; Vincent, P. Representation Learning: A Review and New Perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [[CrossRef](#)] [[PubMed](#)]
43. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.E.; Mohamed, A.-R.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.N.; et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Process. Mag.* **2012**, *29*, 82–97. [[CrossRef](#)]
44. Greff, K.; Srivastava, R.K.; Koutnik, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A Search Space Odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *28*, 2222–2232. [[CrossRef](#)]
45. De Myttenaere, A.; Golden, B.; Le Grand, B.; Rossi, F. Mean Absolute Percentage Error for regression models. *Neurocomputing* **2016**, *192*, 38–48. [[CrossRef](#)]
46. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [[CrossRef](#)]
47. Stone, R. Improved statistical procedure for the evaluation of solar radiation estimation models. *Sol. Energy* **1993**, *51*, 289–291. [[CrossRef](#)]
48. Son, C.-H.; Yang, I.-H. A Study on the Effect of Envelope Factors on Cooling, Heating and Lighting Energy Consumption in Office Building. *J. Korean Inst. Illum. Electr. Install. Eng.* **2012**, *26*, 8–17.
49. Jung, S. A Study on the Comparison of Maximum Power Demand in Building’s Heating and Cooling systems: The Case of EHP, GHP and absorption chiller-heater system+CAV. *Resid. Environ. Inst. Korea* **2012**, *10*, 303–311.
50. Kim, Y.; Lee, T. Analysis of Energy Consumption Characteristics of a Medium Sized Office Building. *Korea Facil. Manag. Assoc.* **2012**, *9*, 41–49.

51. Cho, M.S.; Le, D.Y. An Analysis of Residential Building Energy Consumption Using Building Energy Integrated Database—Focused on Building Uses, Regions, Scale and the Year of Construction Completion. *J. Real Estate Anal.* **2017**, *3*, 101–118. [[CrossRef](#)]
52. Building Cooling Facility Survey Report. October 2014. Available online: <http://www.prism.go.kr> (accessed on 20 September 2020).
53. Mishra, S.; Shafi, Z.; Pathak, S. Time series event correlation with DTW and Hierarchical Clustering methods. *PeerJ Prepr.* **2019**, 1–12. [[CrossRef](#)]
54. Heaton, J. *Artificial Intelligence for Humans, Volume 3: Deep Learning and Neural Networks*; CreateSpace Independent Publishing Platform: Scotts Valley, CA, USA, 2015.

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).