An Improvement of Power Demand Prediction Method using Weather Information and Machine Learning: A Case of a Clinic in Japan (II)

Tomoya Inagata
Graduate School of Engineering
Nagasaki Institute of Applied Science
Nagasaki, Japan
s5323001@campus.nias.ac.jp

Keita Matsunaga
Faculty of Engineering
Nagasaki Institute of Applied Science
Nagasaki, Japan
s1819009@campus.nias.ac.jp

Yuji Mizuno
Dept. Medical Science
Osaka Electro-Communication University
Osaka, Japan
y-mizuno@osakac.ac.jp

Fujio Kurokawa
Institute for Innovative Science and Technology
Nagasaki Institute of Applied Science
Nagasaki, Japan
kurokawa_fujio@nias.ac.jp

Masaharu Tanaka
Institute for Innovative Science and Technology
Nagasaki Institute of Applied Science
Nagasaki, Japan
tanaka_masaharu@nias.ac.jp

Nobumasa Matsui
Institute for Innovative Science and Technology
Nagasaki Institute of Applied Science
Nagasaki, Japan
matsui_nobumasa@nias.ac.jp

Abstract—**In Japan, power generation from renewable energy sources has been promoted since the Great East Japan Earthquake in 2011. For this reason, the installing of renewable energy is increasing in Japan. However, the amount of power generated by renewable energies is influenced depending on natural conditions. To ensure a stable supply of electricity, it is important to keep a balance between supply and demand. Therefore, research and development of a demand response is becoming increasingly important. Hospitals and clinics, which are among the most energy-consuming types of medical facilities using renewable energy systems, need to predict an electricity demand to consider carrying out a demand response. This paper proposes a method to improve the accuracy of an electricity demand prediction for a clinic. A neural network is used as a prediction method, and the predictors consist of day of the week and temperature data by Japan Meteorological Agency. As a result, it is clarified that the proposed method is close to $\sqrt{\pi/2} \approx 1.253$, which is the value to be evaluated when the error is normally distributed.**

Keywords—clinic, load prediction, weather information, machine learning

I. Introduction

In Japan, power generation from renewable energy sources has been promoted since the Great East Japan Earthquake in 2011. For this reason, the installing of renewable energy is increasing in Japan. Photovoltaic power generation especially has increased about sevenfold between 2012 and 2020 [1]. However, the amount of power generated by renewable energies is influenced depending on natural conditions. To ensure a stable supply of electricity, it is important to keep a balance between supply and demand. Therefore, research and development of a demand response is becoming increasingly important. Hospitals and clinics, which are among the most energy-consuming types of medical facilities using renewable energy systems, need to predict an electricity demand to consider carrying out a demand response.

Many methods have been proposed for a predicting electricity demand. Reference [2] compares the accuracy of prediction methods based on recurrent neural network (RNN), RNN-based long short-term memory (LSTM), and gated recurrent unit (GRU) using smart meter electricity data. In reference [3], the results of forecasting with RNN-based LSTM using Turkish electricity data show that for a short-term electricity demand predicting, the predicting method with LSTM shows high accuracy. Reference [4-5] shows that combining convolutional neural networks (CNN) and LSTM improves accuracy over using LSTM on their own. In reference [6], a method called multi-channel long short-term memory with time location (TL-MCLSTM) is proposed. This method predicts a power demand by inputting power data and time locations into the LSTM layer at each step. Reference [7] shows that a predicting method using the XGBoost algorithm is more effective than LSTM and a regression method for a short-term electricity demand predicting. In reference [8], it is shown that a hybrid regression and LSTM method is more accurate than a two-stage LSTM method. In reference [9], multilayer perceptron (MLP), gradient boosting regression trees (GBRT) and stacked bidirectional long short-term memory (SB-LSTM) methods are used to forecast an electricity demand from electricity and weather data, and each is compared. In reference [10], an approach to short-term load forecasting (STLF) is proposed using an ensemble prediction network (EPN). It predicts the final electricity demand by using an ensemble consisting of optimized estimates at each node. Comparison with the evaluation indicator root mean square error (RMSE) shows that the EPN method is superior to LSTM, support vector regression (SVR), and MLP. In reference [11], an RNN-based LSTM is used to predict solar power generation. They show that increasing the input variables of the neural network does not increase the evaluation indicator RMSE. Although reference [12] reports that an electricity demand in Japan is highly correlated with temperature, these previous studies have not used temperature to predict an electricity demand. Therefore, reference [13] proposes a forecasting method using weather forecast temperatures. However, renewable energy is increasing the current situation in Japan, it needs to improve prediction of a demand.

This paper proposes an improved method for a predicting electricity demand using machine learning. The structure of the predictors is determined from the correlation between the predictors and an electricity demand. When determining the

This work was supported by JSPS KAKENHI Grant Number JP21K04559.

structure of the predictors, it is necessary to pay attention to the correlation coefficient between predictors.

II. Analysis of a Clinic Demand

Fig. 1 shows an electricity demand of a clinic. The horizontal axis is the months from April to March of the following year. The vertical axis is the total electricity demand [p.u] for each day. Electricity demand data for Sunday and clinic closing days are excluded. An electricity demand data was obtained from the clinic for April 2017 through March 2022. The circle markers show a demand for FY 2017. The cross markers denote a demand for FY 2018. The triangle markers indicate a demand for FY 2019. The square markers are a demand for FY 2020. The diamond markers show a demand for FY 2021. Fig. 1 shows an electricity demand in February is the highest of the year.

Fig. 2 shows the comparison with August in FY 2020 and February in FY 2020 for each day of the week. The horizontal axis of all graphs is time, from 0 to 24 [h]. The vertical axis is an electricity demand [p.u] for each hour. The solid red line indicates a demand for February FY 2020. The solid blue line is a demand for August FY 2020. Demand curves are similar on Sunday since the clinic are closed. The clinic open hours divide it into two groups. Group 1 is Monday-Wednesday-Friday. Group 1 has a similar demand curve. Group 2 is Tuesday-Thursday-Saturday. Group 2 also has a similar demand curve. Both groups have flat demand before clinic open hours. A demand tends to increase during clinic open hours and decrease after clinic open hours. Moreover, since a daily electricity demand changes more rapidly in February than in August, it is important to accurately predict an electricity demand in February.

Fig. 3 plots temperature and an electricity demand in February for five years. Table 1 shows the correlation coefficients for each time period. Temperature data uses the data published by the Japan Meteorological Agency (JMA) for the area around the clinic. The correlation coefficient R is calculated using Eq. (1), where s_{xy} is the covariance between temperature and an electricity demand, σ_x is the standard deviation of temperature, σ_y is the standard deviation of an electricity demand. s_{xy}, σ_x and σ_y are calculated through Eqs. (2) and (4), where x_i is temperature, $\bar{x}$ is the mean of temperature, y_i is an electricity demand, and $\bar{y}$ is the mean of an electricity demand. The correlation coefficient is generally considered to be 0 for no correlation, $0 < R \leq 0.3$ for weak correlation, $0.3 < R \leq 0.6$ for moderate correlation, and $0.6 < R$ for strong correlation [14]. Fig. 4 plots temperature and an electricity demand for Monday in February for five years. Table 2 shows the correlation coefficients for each time period. Tables 1 shows a weak or no correlation between temperature and an electricity demand for each hour. On the other hand, Table 2 shows a high correlation between temperature and an electricity demand for clinic open hours.

$$R = \frac{s_{xy}}{\sigma_x \times \sigma_y} \tag{1}$$

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \tag{2}$$

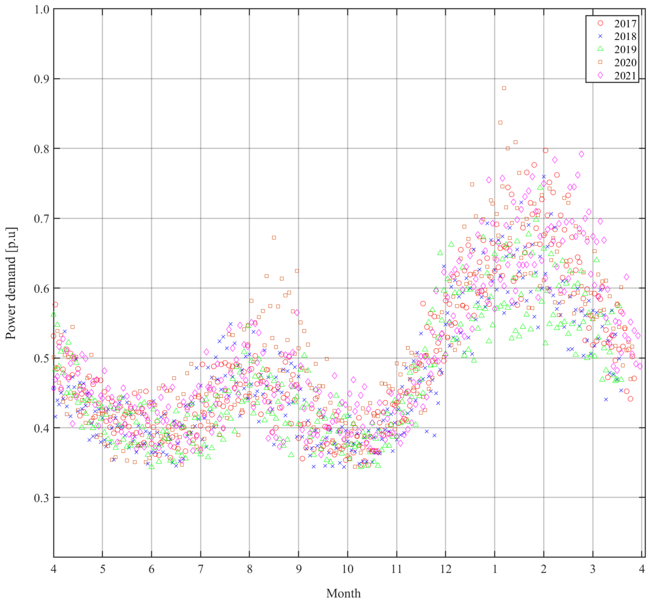

Fig. 1. A clinic power demand for each year.

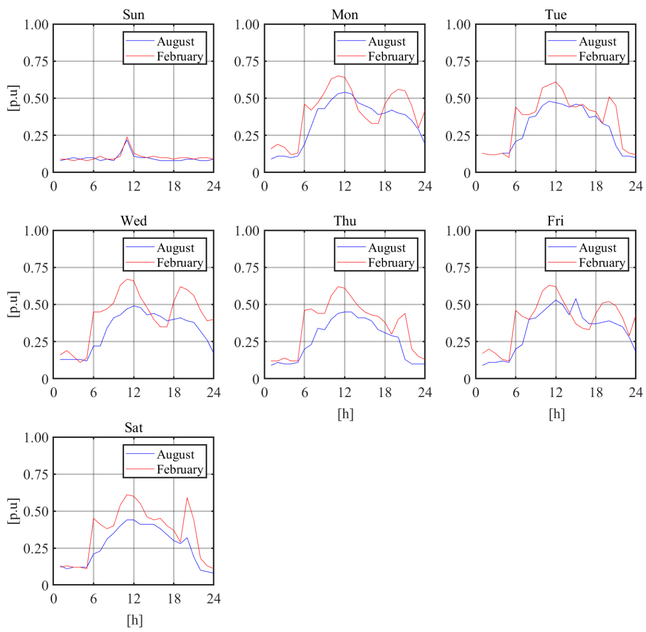

Fig. 2. Comparison with August in FY 2020 and February in FY 2020 for each day of the week.

$$\sigma_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{3}$$

$$\sigma_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

TABLE I. CORRELATION COEFFICIENTS ON ALL DAYS OF THE WEEK

12 midnight	1 a.m.	2 a.m.	3 a.m.	4 a.m.	5 a.m.
0.0123	-0.0127	-0.0675	-0.1327	-0.159	-0.1689
6 a.m.	7 a.m.	8 a.m.	9 a.m.	10 a.m.	11 a.m.
-0.1953	-0.2511	-0.357	-0.3252	-0.2922	-0.2862
12 noon	1 p.m.	2 p.m.	3 p.m.	4 p.m.	5 p.m.
-0.2525	-0.2551	-0.2959	-0.2602	-0.3115	-0.3049
6 p.m.	7 p.m.	8 p.m.	9 p.m.	10 p.m.	11 p.m.
-0.2688	-0.168	-0.1334	-0.1347	-0.0835	-0.1053

TABLE II. CORRELATION COEFFICIENTS ON MONDAY

12 midnight	1 a.m.	2 a.m.	3 a.m.	4 a.m.	5 a.m.
-0.3922	0.3862	0.1443	-0.2326	-0.1715	-0.0522
6 a.m.	7 a.m.	8 a.m.	9 a.m.	10 a.m.	11 a.m.
-0.7835	-0.7476	-0.8958	-0.8957	-0.8871	-0.8604
12 noon	1 p.m.	2 p.m.	3 p.m.	4 p.m.	5 p.m.
-0.818	-0.7468	-0.7733	-0.7304	-0.7117	-0.6583
6 p.m.	7 p.m.	8 p.m.	9 p.m.	10 p.m.	11 p.m.
-0.8091	-0.7854	-0.3839	-0.1745	-0.5401	-0.4748

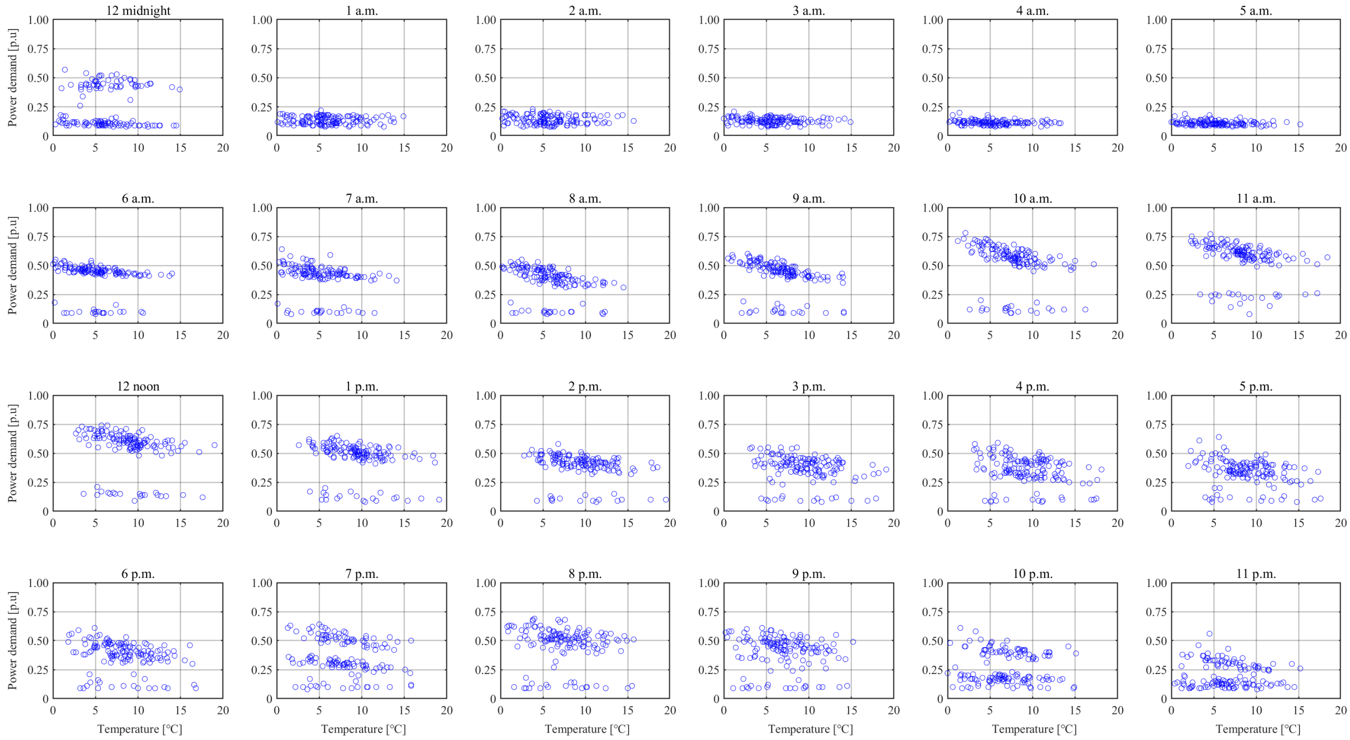

Fig. 3. Relationship between temperature and electricity demand on all days of the week.

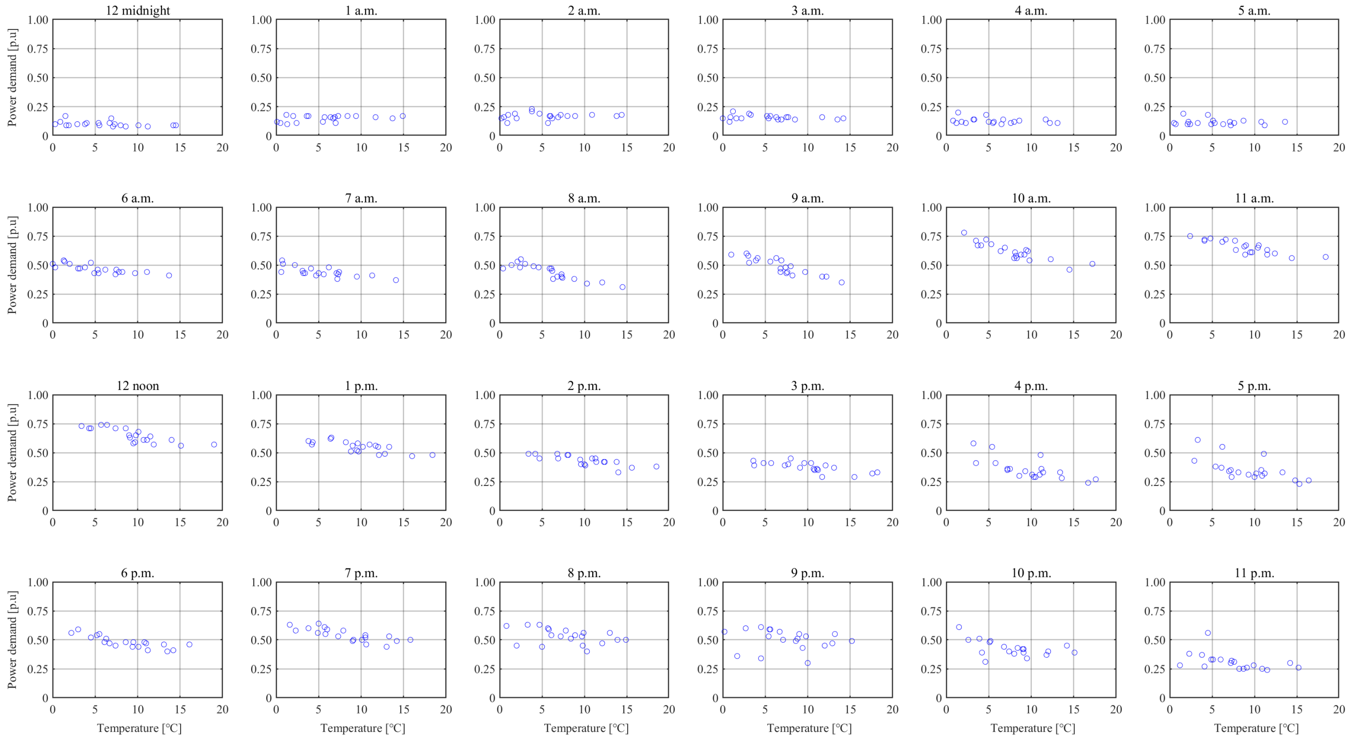

Fig. 4. Relationship between temperature and electricity demand on Monday.

III. Predicting Modelds Using Machine Learning

A. Standardization of an electricity demand

Temperature data published by the JMA is hourly data. On the other hand, an electricity demand data for the clinic is every 30 minutes. So, an hourly electricity demand data is used. In addition, since the temperature data is structured from April to March of the following year, the electricity demand data is also structured accordingly.

The actual temperature data published by the JMA is used as a predictor, and the actual electricity demand of the clinic is used as the training target of a neural network. The predictor and the actual electricity demand data are standardized x_i' using Eq. (5). Where x_i is the value in each unit system, and $\bar{x}$ and σ are the mean and standard deviation.

$$x_i' = \frac{x_i - \bar{x}}{\sigma} \tag{5}$$

B. Improvement of structure of predictors

Fig. 5 shows a structure of the prediction model. The prediction model has an input layer, 3 hidden layers and an output layer. Each hidden layer has 30 nodes. The output layer is an hourly electricity demand.

Table 3 shows the structure of predictors verified in this study. In Case 1, the predictor consists of hourly temperature data and the day of the week. The day of the week is represented by an integer value 1~7. Sunday is 1 and Saturday is 7. In Case 2, the month is added to Case 1. In Case 3, the season is added to Case 1. The season are represented by an integer value 1~4, from December to March as 1, and an integer value given every 3 months thereafter. In Case 4, hourly temperature data and season are added with groupings of days of the week. The group of days of the week is one group of days of the week with the same clinic open hours. Concretely, Monday, Wednesday, and Friday are set as group 1 and given an integer value 1. Tuesday, Thursday, and Saturday are grouped as group 2 and given an integer value 2. Sunday is given an integer value 3. In Case 5, hourly temperature data and grouped days of the week are added to the grouped season. For season, October through March are given an integer value 1 as group 1. April through September are given an integer value 2 as group 2. Temperature data are time series data. When using time-series data consisting of hours, learning is required for each hour. However, if we use data consisting of days, we only need to learn for each day. Therefore, the temperature data is composed of a set of 24-hour data for each day.

Table 4 shows the correlation coefficients between the predictors. Day of the week and temperature data are no correlated in all cases. Month and temperature data are weakly correlated in case 2. Cases 3 through 5 show that there is a moderate or high correlation between temperature data and season. High correlations between predictors means that the number of predictors can be reduced. Therefore, in Cases 3 through 5, the learning accuracy is not expected to increase.

Table 5 shows the regression result for an electricity demand of each case. The data set for the learning, 70% for training, 15% for validation, and 15% for testing. The coefficient of determination R^2 in Eq. (6) is used to evaluate the regression results, where L is the actual value of an electricity demand, $\bar{L}$ is the mean of actual electricity demand, and $\hat{L}$ is the value of electricity demand predicted by NN. The learning accuracy is approximately 96.7% in all cases. The learning accuracy is not much different in all cases.

TABLE III. The Structure of Predictors

Case 1		
Temperature data in hourly	Day of the week	
Case 2		
Temperature data in hourly	Day of the week	Month
Case 3		
Temperature data in hourly	Day of the week	Season
Case 4		
Temperature data in hourly	Group of day of the week	Season
Case 5		
Temperature data in hourly	Group of day of the week	Group of season

TABLE IV. Correlation Coefficient of Predictors

Case 1	Correlation coefficient
Day of the week and temperature	-0.0114
Case 2	
Day and month	-0.0013
Day of the week and temperature	-0.0114
Month and temperature	0.3336
Case 3	
Days of the week and seasons	0.0011
Day of the week and temperature	-0.0114
Season and temperature	0.6518
Case 4	
Days of the week and seasons	0.0014
Day of the week and temperature	-0.0025
Season and temperature	0.6518
Case 5	
Days of the week and seasons	-0.001
Day of the week and temperature	-0.0025
Season and temperature	0.7295

Therefore, in this study, since the learning accuracy is not much different in each case, the predictor in Case 1, which has the least number of predictors, is employed to build the electricity demand predicting model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(L - \hat{L}\right)^2}{\sum_{i=1}^{n} \left(L - \bar{L}\right)^2} \tag{6}$$

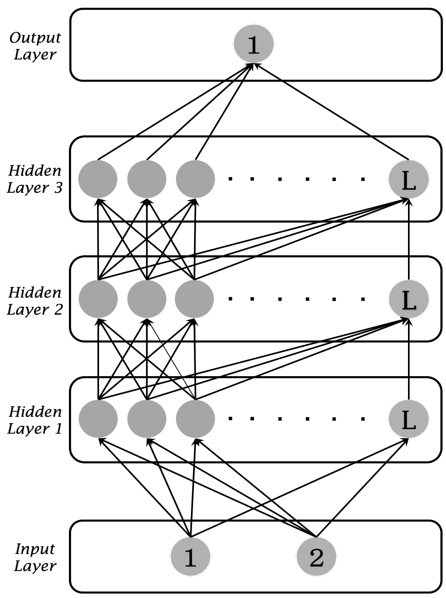

Fig. 5. The prediction model structure.

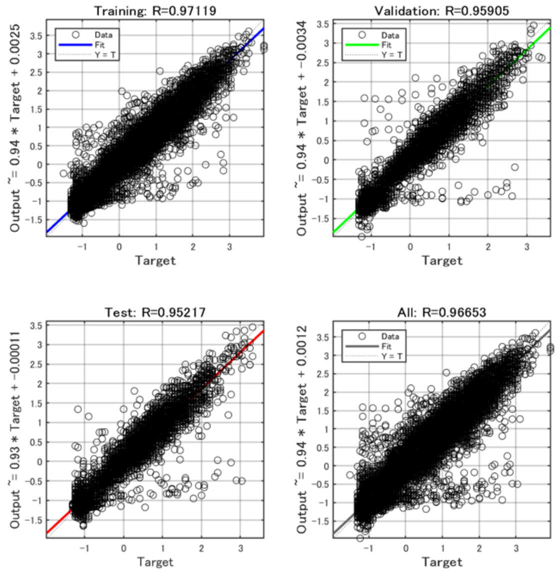

Fig. 6. The regression results for an electricity demand.

TABLE V. THE REGRESSION RESULTS FOR EACH CASE

	Training	Validation	Test	All datasets
Case 1	0.97119	0.95905	0.95217	0.96653
Case 2	0.97332	0.96026	0.94975	0.96794
Case 3	0.97344	0.96023	0.95943	0.96938
Case 4	0.96791	0.96764	0.96500	0.96740
Case 5	0.96904	0.96705	0.96127	0.96757

C. *The Prediction model structure*

The structure of the prediction model is shown in Fig. 5. The inputs of the prediction model consist of temperature data and the day of the week. Each day of the week is represented by converting Sunday through Saturday into integers 1~7. The hidden layer consists of three layers with one layer consisting of 30 nodes. The outputs of the prediction model are an hourly electricity demand. The data set for the learning, 70% for training, 15% for validation, and 15% for testing.

IV. EVALUATION OF THE PREDICTION MODEL

Fig. 6 shows the regression results for an electricity demand. The regression results show a coefficient of determination R^2 of 0.97119 for training, 0.95905 for validation, 0.95217 for testing, and 0.96653 for all datasets.

Fig. 7 shows a comparison of predicted and actual power demand using the prediction model. Data for February FY 2020 were used. The horizontal axis of all graphs is time, from 0 to 24 [h]. The vertical axis is an electricity demand [p.u]. The solid blue line is the actual demand data for a clinic and the solid red line is the prediction demand data.

Table 6 shows the prediction errors for February and August of FY2020. The mean absolute error (MAE), RMSE and the ratio of RMSE/MAE are used as indicators to evaluate the predicting model [15]. MAE and RMSE are calculated using Eq. (7) and (8). The RMSE/MAE ratio of

TABLE VI. THE PREDICTION ERROR

Year	Month	Day of the week	RMSE [kWh]	MAE [kWh]	RMSE/MAE [-]
FY 2020	2	Sunday	1.688	2.115	1.266
		Monday	3.240	3.953	1.227
		Tuesday	3.206	3.942	1.226
		Wednesday	2.734	3.351	1.224
		Thursday	2.741	3.685	1.379
		Friday	2.693	3.282	1.218
		Saturday	3.703	4.832	1.331
	8	Sunday	2.142	2.657	1.286
		Monday	4.267	5.185	1.229
		Tuesday	3.59	4.247	1.186
		Wednesday	4.915	6.045	1.224
		Thursday	2.223	2.734	1.235
		Friday	4.963	6.211	1.253
		Saturday	2.788	3.329	1.191

a good machine learning model generally is known to approach $\sqrt{\pi/2} \approx 1.253$. The average RMSE/MAE ratio was 1.267 in February FY 2020 and 1.229 in August FY 2020. The obtained RMSE/MAE value is close to 1.253.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{L} - L\right)^2} \tag{7}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{L} - L\right| \tag{8}$$

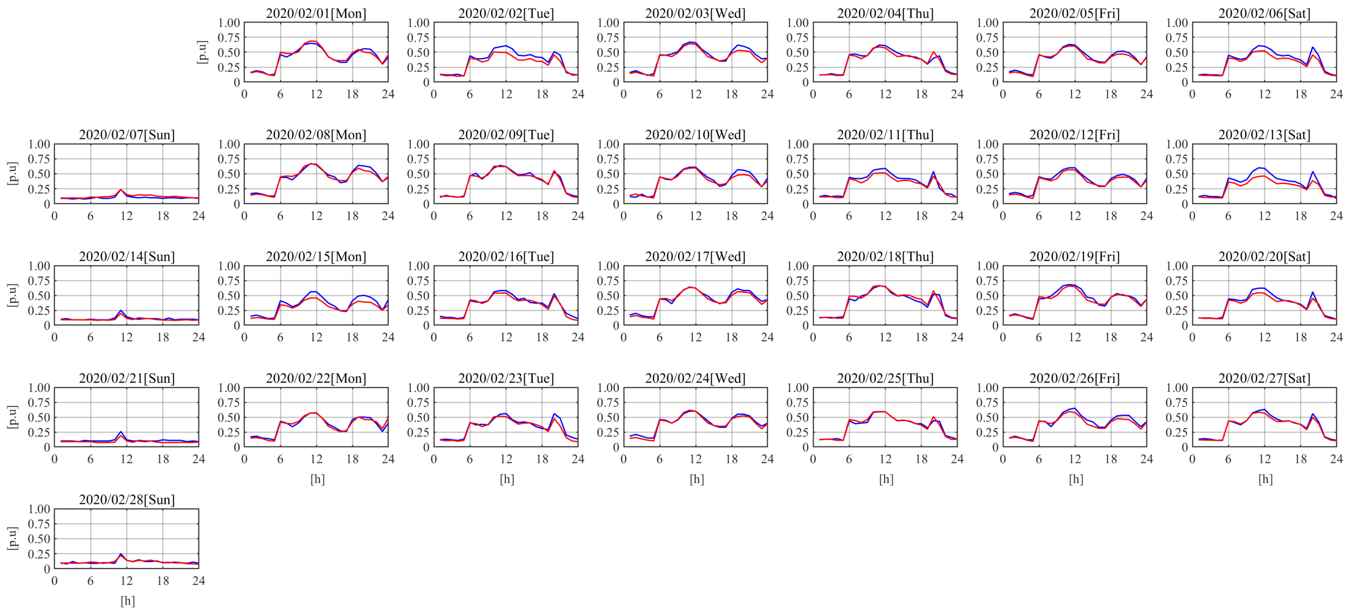

Fig. 7. Comparison of a predicted and actual electricity demand in February FY2020.

V. CONCLUSIONS

This paper proposed a method to improve the predicting electricity demand for a clinic using weather information. The inputs constructure by two pieces of information, day of the week and temperature data by JMA. Temperature data is learned day by day, with each 24-hour period as one data set. The results obtained are as follows:

(1) Prediction accuracy is improved by determining the predictor from its correlation with the output. If the predictors are highly correlated between each other, the predictor structure need to changed.

(2) Prediction accuracy is close to $\sqrt{\pi/2} \approx 1.253$ using a proposed method.

The proposed electricity demand prediction model will be practically implemented in the power system of the clinic.

REFERENCES

[1] Agency for Natural Resources and Energy, "Annual Report on Energy in FY2021," pp. 94-100, 2022.

[2] T. Hossen, A. S. Nair, R. A. Chinnathambi and P. Ranganathan, "Residential Load Forecasting Using Deep Neural Networks (DNN) ," in Proc. North American Power Symposium (NAPS), Fargo, ND, USA, pp. 1-5, 2018.

[3] Idris Bodur, Emre Celik, and Nihat Ozturk, "A Short-Term Load Demand Forecasting based on the Method of LSTM," in Proc. 2021 International Conference on Renewable Energy Research andApplication (ICRERA), pp. 171-174, Sep. 2021.

[4] W. Qian, C. Gu, C. Zhu, Z. Jiang, B. Han and M. Yu, "Short-term Load Forecasting Based on Multi-model Fusion of CNN-LSTM-LGBM," *2021 International Conference on Power System Technology (POWERCON)*, Haikou, China, 2021, pp. 934-939,2021.

[5] S. H. Rafi, Nahid-Al-Masood, S. R. Deeba and E. Hossain, "A Short-Term Load Forecasting Method Using Integrated CNN and LSTM Network," IEEE Access, vol. 9, pp. 32436-32448, 2021.

[6] Xiaorui Shao, Chang Soo Kim, "Multi-Step Short-Term PowerConsumption Forecasting Using Multi-Channel LSTM With Time Location Considering Customer Behavior," IEEE Access, Vol. 8, pp.125263-125273, 2020.

[7] S. H. Rafi, Nahid-Al-Masood and M. M. Mahdi, "A Short-Term Load Forecasting Technique Using Extreme Gradient Boosting Algorithm," 2021 IEEE PES Innovative Smart Grid Technologies - Asia (ISGT Asia), Brisbane, Australia, pp. 1-5,2021.

[8] S. Zhu, M. Zou, S. Z. Djokic, J. Gunda and I. Papic, "Comparison of Single-Stage and Two-Stage LSTM-Based Methods for Forecasting Electricity Demands and Prices," 2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), Novi Sad, Serbia, pp. 1-5, 2022.

[9] M. Zou et al., "Comparison of Three Methods for a Weather Based Day-Ahead Load Forecasting, " 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe), Bucharest, Romania, pp. 1-5, 2019.

[10] M. Z. A. Bhotto, R. Jones, S. Makonin and I. V. Bajić, "Short-Term Demand Prediction Using an Ensemble of Linearly-Constrained Estimators, " in IEEE Transactions on Power Systems, vol. 36, no. 4, pp. 3163-3175, July 2021.

[11] V. De, T. T. Teo, W. L. Woo and T. Logenthiran, "Photovoltaic Power Forecasting using LSTM on Limited Dataset, " 2018 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia), Singapore, pp. 710-715, 2018.

[12] D. Narumi, Y. Niurao, Y. Shimoda and M. Mizuno, "Effects of Temperature Change upon the Regional Electric Power Consumption, " Japan Society of Energy and Resources, vol. 28, no. 3, pp.154-160, 2007, (in Japanese).

[13] Y. Mizuno, M. Tanaka, Y. Tanaka, F. Kurokawa and N. Matsui, "A Prediction of Power Demand using Weather Forecasting and Machine Learning: A Case of a Clinic in Japan," in Proc. 2022 10th International Conference on Smart Grid (icSmartGrid), Istanbul, Turkey, pp. 190-193, 2022.

[14] Haldun Akoglu, "User's guide to correlation coefficients, " Turkish Journal of Emergency Medicine, Volume 18, Issue 3, pp 91-93, 2018.

[15] T. Chai, R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature, " Geosci. Model Dev., 7, pp.1247–1250, 2014.