Article

Assessment of ANN Algorithms for the Concentration Prediction of Indoor Air Pollutants in Child Daycare Centers

Jeeheon Kim [1], Yongsug Hong [2], Namchul Seong [3,*] and Daeung Danny Kim [4,*]

1   Eco-System Research Center, Gachon University, Seongnam 13120, Korea; jlaw80@naver.com
2   Division of Human-Architectural Engineering, Daejin University, 1007 Hoguk-Ro, Pocheon 11159, Korea; yshong@daejin.ac.kr
3   Department of Architectural Engineering Kangwon National University, Samcheok-si 25913, Korea
4   Architectural Engineering Department, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran 31261, Saudi Arabia
*   Correspondence: inamchul@kangwon.ac.kr (N.S.); dkim@kfupm.edu.sa (D.D.K.)

Abstract: As the time spent by people indoors continues to significantly increase, much attention has been paid to indoor air quality. While many IAQ studies have been conducted through field measurements, the use of data-driven techniques such as machine learning has been increasingly used for the prediction of indoor air pollutants. For the present study, the concentrations of indoor air pollutants such as CO_2, $PM_{2.5}$, and VOCs in child daycare centers were predicted by using an artificial neural network model with three different training algorithms including Levenberg–Marquardt, Bayesian regularization, and Broyden–Fletcher–Goldfarb–Shanno quasi-Newton methods. For training and validation, data of indoor pollutants measured in child daycare facilities over a 1-month period were used. The results showed all the models produced a good performance for the prediction of indoor pollutants compared with the measured data. Among the models, the prediction by the LM model met the acceptable criteria of ASHRAE guideline 14 under all conditions. It was observed that the prediction performance decreased as the number of hidden layers increased. Moreover, the prediction performance was differed by the type of indoor pollutant. This was caused by patterns observed in the measured data. Considering the outcomes of the study, better prediction results can be obtained through the selection of suitable prediction models for time series data as well as the adjustment of training algorithms.

Keywords: indoor air pollutants; ANN model; training algorithm; child daycare center

1. Introduction

People generally spend most of their time in buildings, with this time rapidly increasing due to the situation caused by the SARS-CoV-2 virus [1,2]. Thus, much attention has been paid to the improvement of indoor environmental quality including indoor air quality, thermal parameters, etc. [3–5]. Normally, the quality of indoor air is highly influenced by outdoor air pollution and other indoor sources [1,6]. Specifically, indoor sources originating from building materials, appliances, human activities, etc. have produced indoor air contaminants [7–9]. Regarding indoor pollutants, many studies have performed investigations aiming to reduce the concentration or prevent their occurrence [10–15]. While most studies have focused on indoor pollutants in residential and non-residential buildings, people have started to notice the importance of indoor pollutants in certain facilities such as child daycare centers.

Several studies have observed severe indoor pollutants in child daycare centers through field measurements. According to the study of Oh et al., high levels of $PM_{2.5}$ and PM_{10} were found in ten child daycare centers in South Korea, which were highly influenced by traffic conditions and vehicles through various openings [16]. In the case

of child daycare centers in Paris, Roda et al. measured biological and chemical pollutants [17]. In their findings, some chemical pollutants were above the acceptable indoor levels in 28 child daycare centers. A similar result was observed in the findings of a study by Hwang et al. [18]. Both biological and chemical pollutants were measured in 25 child daycare centers. Specifically, the concentration of VOCs was the highest among the measurements in their data. In the field measurements performed by Madureira et al., severe concentrations of biological pollutants were found in nine child daycare centers [19]. Other studies observed high concentrations of indoor pollutants through their measurements in various child daycare centers [20–23].

To improve indoor air quality and prevent indoor pollution, the prediction of concentrations of various indoor pollutants affecting occupants' health is essential. Even though severe indoor air quality conditions in child daycare centers were reported, most IAQ studies have focused on indoor pollutants in residential or commercial buildings. In addition, most of these studies collected data on indoor pollutants through measurements, which is time-consuming and costly. Another technique for predicting indoor air pollutants is the use of simulations. In a study by Heibat et al., the researchers used a coupling method including CONTAM and WUFI for the prediction of CO_2, $PM_{2.5}$, and VOCs [24]. However, the accuracy of the simulation results can be highly dependent on the user's experience.

Recently, a prediction made by utilizing advanced computer applications was significantly recognized with the development of data-driven methods such as machine learning techniques [19,20]. Machine learning techniques have been used for extracting data patterns and quantifying the impacts of various design parameters [25]. The data-driven methods have been widely used in applications of engineering, medicine, and economics. Among various data-driven methods, artificial neural network (ANN) models, support vector machine regression (SVR), random forest, XGBoost, etc. have been applied for various purposes such as energy consumption predictions, thermal performance quantification, mechanical system diagnostics, and so on [26–30]. For the present study, the ANN model was chosen to predict indoor air quality in a child daycare center due to its high prediction accuracy [27]. According to several studies, the ANN model showed the best performance among the machine learning methods [31–33]. In general, different learning algorithms were used with a regular ANN model. However, the prediction results can become unstable caused by fluctuations in these training algorithms [34]. To provide more reliable prediction results, the performance of several training algorithms was tested. While the predictions made by using machine learning techniques have been widely employed in various fields, there were few studies available for indoor air pollutant prediction. In addition, the predictive performance by different training algorithms was rarely investigated. This study presents the difference in the predictive performance of indoor air pollutants by applying different training algorithms in the ANN model.

2. Machine Learning Applications for the Prediction of Indoor Air Quality

For the prediction of indoor air quality, Jeong et al. predicted indoor environmental parameters such as temperature, humidity, and CO_2 by using a machine learning technique [35]. By comparing the collected data, a highly correlated relationship between the data and prediction results was achieved. In addition, the IAQ management was conducted remotely by using IoT systems [36]. Through cloud data analysis, the comparison of measurement data with simulations was implemented to improve the mechanical exhaust systems. Li et al. used a machine learning method such as the random forest algorithm to predict $PM_{2.5}$ concentrations in residential buildings [37]. In their study, the random forest model showed excellent performance for the prediction of $PM_{2.5}$ concentration levels. Kallio et al. investigated the performance of four machine learning methods: Ridge regression, decision tree, random forest, and multilayer perceptron to predict indoor CO_2 concentrations [38]. The abovementioned machine learning models showed a better performance than statistical methods regarding indoor CO_2 concentration predictions. Another study by Taheri et al. performed comparisons of several machine learning algorithms

including support vector machines, AdaBoost, random forest, gradient boosting, logistic regression, and multilayer perceptron [39]. In addition, Sassi et al. utilized a deep learning technique for data analysis and augmented reality (AR) to predict indoor air quality based on the monitoring by an IoT system [40].

Regarding the ANN model application, Saad et al. proposed an IAQ monitoring system to identify sources affecting IAQ levels. By utilizing ANN techniques, their study recognized the patterns of measured data and proved that the proposed system was able to measure indoor air quality levels. In addition, the sources affecting indoor air quality such as ambient air, presence of chemicals and fragrances, food and beverages, and human activity were classified successfully [41]. In the case of the study performed by Putra et al., ANN models were used to predict indoor air quality with the data, which were measured 8 h a day for several months. For this study, the authors utilized the Levenberg–Marquardt training method and proved that this training method produced good prediction results [42]. Moreover, Dai et al. constructed an ANN model by using indoor CO_2 concentration data sets in a residential building to predict indoor air quality with ventilation rates [43]. About 80% of the overall accuracy levels by the constructed ANN model were achieved and the authors proved that the indoor CO_2 concentration predicted by the ANN model was highly influenced by locations and outdoor air temperatures. According to the study of Egala et al., a practical approach was presented to train ANN models regarding the prediction of indoor CO_2 concentration [44]. By training the model with collected data for a month, computational errors were reduced and high predictive accuracy was achieved. To control HVAC (Heating, ventilation, and air conditioning) systems for improving indoor air quality, Tagliabue et al. used the ANN model [33]. Moreover, Amuthadevial implemented machine learning methods including nonlinear ANN models, statistical multi-level regression, neural purge, and deep learning short and long-term memory (DL-LSTM) to predict concentration levels of SO_2, CO, NOx, and O_3 [32]. Regarding indoor $PM_{2.5}$, PM_{10}, and NO_2 concentrations, Zhang et al. also used several machine learning techniques such as multiple linear regression (MLR), time series regression (TSR), and ANN models [31]. In their study, the ANN model showed the best performance among the machine learning methods.

3. Methodology

Figure 1 presents the research process for the present study. Due to high concentrations of CO_2, $PM_{2.5}$, and VOCs in child daycare centers, the measurement data of CO_2, $PM_{2.5}$, and volatile organic compounds (VOCs) provided by the Big Data Environment Platform were chosen and these were converted to input data for the ANN model. By using the input data, the concentrations of CO_2, $PM_{2.5}$, and VOCs were predicted with different training algorithms of the ANN model. The prediction results of different training algorithms were evaluated by CV (RMSE) (coefficient of variation of the root mean square error) and MBE (mean bias error). In addition, the suitability of each model was assessed by R^2 (coefficient of determination).

3.1. Collection of Training Data Set

The dataset used for the present study was composed of the measurement data provided by the Big Data Environment Platform [45]. For input data, major indoor pollutants such as CO_2, $PM_{2.5}$, and VOCs were measured at 5-min intervals in the child daycare facilities during the month of May 2021. The data consisted of 8929 sets for each pollutant of which 80% of these data were used for training and 20% for testing.
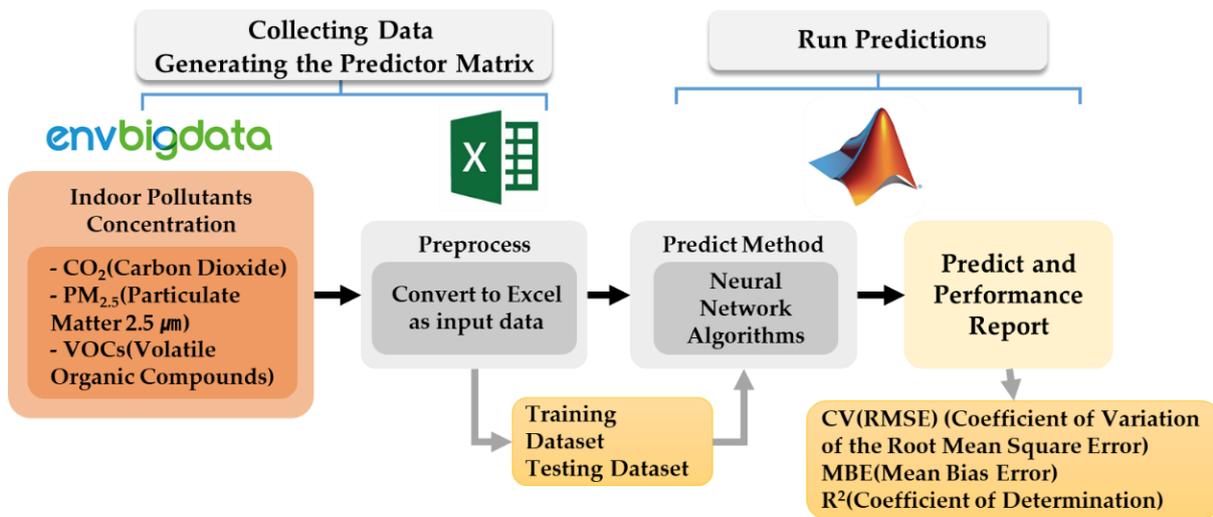
Figure 1. Schematic diagram of the process of predicting indoor air pollution concentrations.

3.2. Indoor Pollutant Concentration Prediction Model

The concentration of indoor pollutants was predicted by using the learning neural network models in the Neural Networks Toolbox of MATLAB [46]. A data multilayer neural network training shows excellent performance when this is optimized by using the slope of neural network performance against neural network weights and the Jacoby matrix of a neural network error. The slope and Jacoby matrix were calculated using the feed-forward back-propagation algorithms. As commonly used in ANN models, feed-forward networks can produce output data quickly avoiding delays [47]. In addition, the feed-forward back-propagation algorithms perform the iteration of updating weights and biases values of network parameters and back-propagate the error for training ANN models [34,48]. For the present study, three different feed-forward back-propagation algorithms were used, which were Levenberg–Marquardt (LM), Bayesian regularization (BR), and Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton (BFG). The performance for the prediction of indoor air pollution was evaluated.

The feed-forward neural network model is generally composed of an input layer, a hidden layer, and an output layer [49]. The hidden layer forms the structure of the ANN and the neuron exists in each layer. Since the number of neurons in the hidden layers mainly influences the calculation prediction and time, the number of neurons was fixed as 20. When the number of the hidden layers was changed to 1, 3, and 5, the predicted performance was compared (Figure 2). As one of the learning parameters, the number of epochs was 100. The total number of data were 8927, in which 80% and 20% of datasets were used for training and testing, respectively. Detailed conditions are summarized in Table 1.
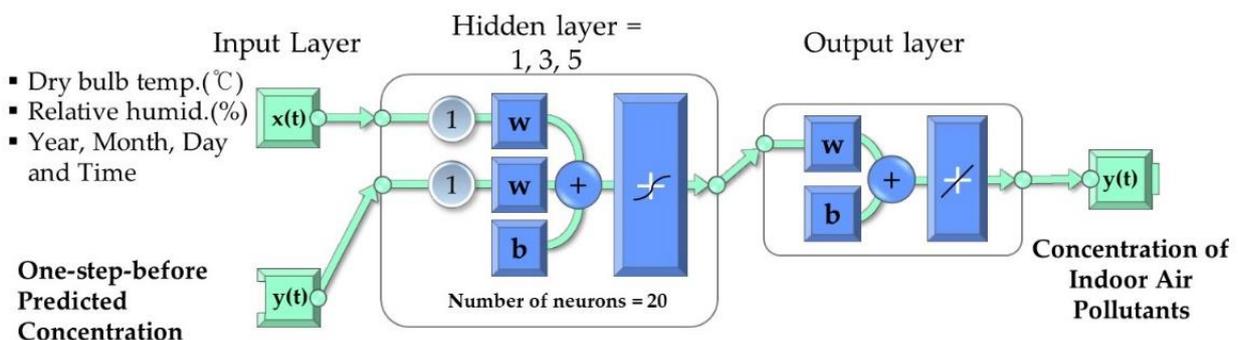

Figure 2. Schematic of the multilayer shallow neural network training algorithms for predicting indoor air pollution concentrations.

Table 1. Training parameters.

Parameter		Value
Number of hidden layers		1, 3, 5
Number of neurons		20
Epochs		100
Data	Training	7142 (80%)
	Testing	1785 (20%)

For input data, the measurement date and time, indoor thermal parameters such as temperature and humidity, and the prediction results of each pollutant fed back from the output layers were used. In the hidden layer, measurement data at the 5-min interval were received as an input signal from the input layer. The feed-forward neural network operations were performed through the internal neurons. The output layer predicted the indoor air pollutant concentrations after 5-min of the input signal point based on the hidden layer calculation result.

By using CV(RMSE) and MBE, the performance evaluation indicators of the predictive model were validated. CV(RMSE) refers to the degree of scattering of estimated values in consideration of variance, and MBE is an error analysis index that identifies errors by tracking how close estimates form clusters through data bias. The models will be declared to be calibrated if they are within the acceptable values of ASHRAE Guideline 14 (Table 2) [50]. The equation for obtaining CV(RMSE) and MBE are as follows.

$$\text{MBE} = \sum^n (y_i - \hat{y}_l) / [(n - p) \times \overline{y}] \cdot 100 \tag{1}$$

$$\text{Cv(RMSE)} = 100 \cdot \left[\sum (y_i - \overline{y_l})^2 / (n - p) \right]^{1/2} / \overline{y}, \tag{2}$$

where n is the number of data points, p is the number of parameters, y_i is the utility data used for calibration, $\hat{y}_i$ is the simulation predicted data, and $\overline{y}$ is the arithmetic mean of the sample of n observations. In addition, the suitability of the model was evaluated by using R^2.

Table 2. Acceptable Calibration Tolerances in building energy performance prediction.

Calibration Type	Index	ASHRAE Guideline 14 [50]
Monthly	MBE_monthly	±5%
	CvRMSE_monthly	15%
Hourly	MBE_hourly	±10%
	CvRMSE_hourly	30%

4. Results

For the present study, the concentrations of CO_2, $PM_{2.5}$, and VOCs in child daycare centers were predicted by using three different feed-forward back-propagation algorithms including LM, BR, and BGF. The performance for the prediction of indoor air pollution was evaluated by CV(RMSE) and MBE. Moreover, the suitability of each model was assessed by R^2.

4.1. CO₂

The prediction results of CO_2 are summarized in Figure 3. The indoor CO_2 concentration is greatly affected by human breathing. In addition, it can be seen that the concentration is altered by the number and activities of occupants. As shown in all graphs in Figure 3, the CO_2 concentration gradually increased to 1500–2000 ppm and then, decreased to 500 ppm. This trend regarding the CO_2 concentration repeated during the measurement period. In addition, a similar trend was observed in all prediction methods with three different

training algorithms. Moreover, the values of training and testing predicted by the ANN model were close to the measurement data.
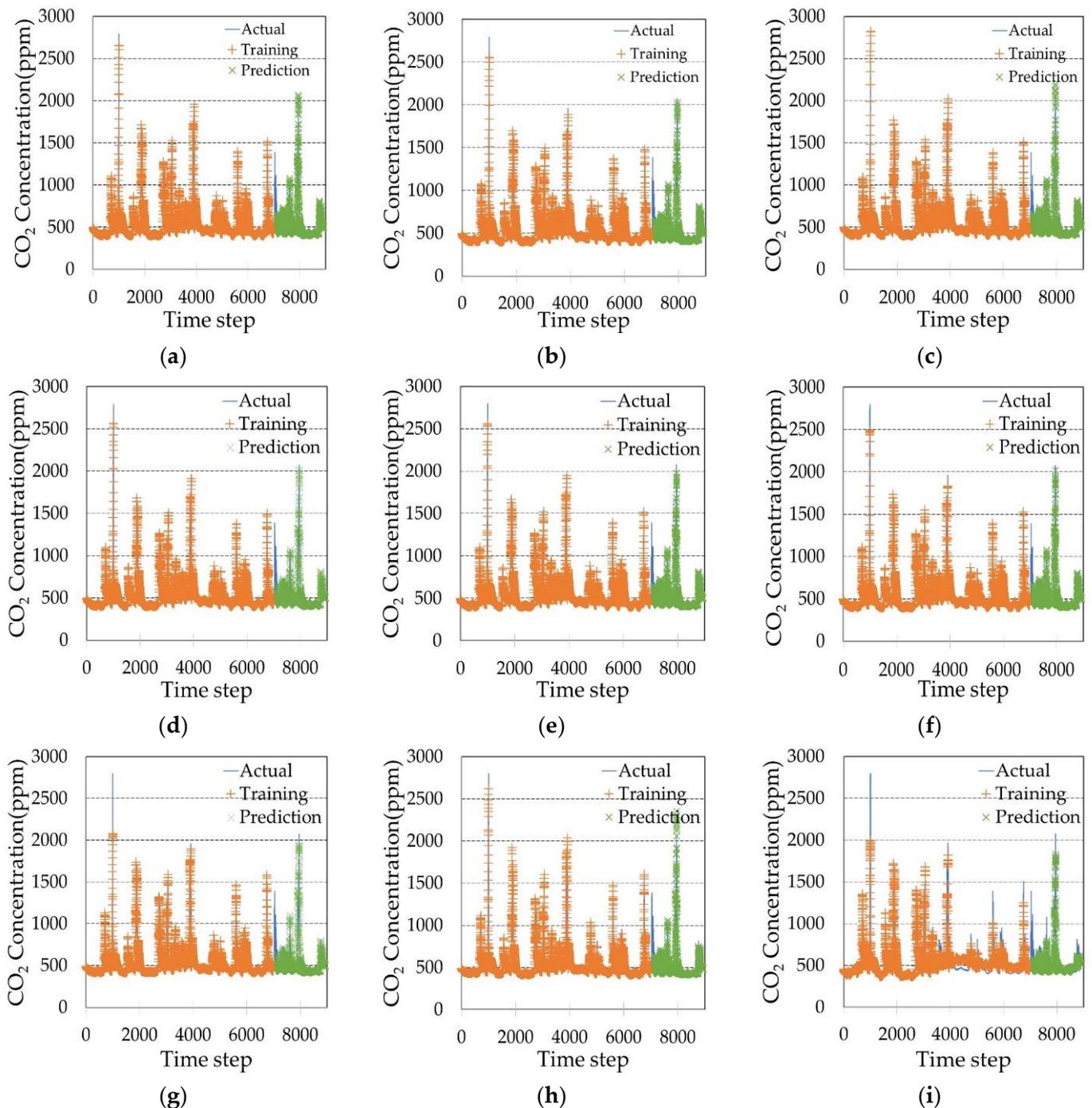

Figure 3. Prediction results of CO_2 concentration. (**a**) LM-Hidden layer-1; (**b**) LM-Hidden layer-3; (**c**) LM-Hidden layer-5; (**d**) BR-Hidden layer-1; (**e**) BR-Hidden layer-3; (**f**) BR-Hidden layer-5; (**g**) BFG-Hidden layer-1; (**h**) BFG-Hidden layer-3; (**i**) BFG-Hidden layer-5.

Table 3 presents CV(RMSE) and MBE of CO_2 prediction results. The CV(RMSE) of the LM model ranges from 4.04% to 4.47% and shows a low degree of dispersion under all conditions. The MBE ranges from 7.53% to 8.56%, which shows an excellent predictive performance due to bias reduction. For the BR model, CV(RMSE) and MBE show ranges of 4.06–4.21% and 7.53–8.06%, respectively. This also shows good predictive performance

satisfying the acceptable range of the ASHRAE guideline 14. For both the LM model and BR model, the values of CV(RMSE) and MBE increased as the number of hidden layers increased. However, a slight increase in those values was observed. In the case of the BFG model, a somewhat larger increase in the CV(RMSE) and MBE was observed than that predicted by using the LM model and BR model. Moreover, the MBE for the BFG model with more than three hidden layers could not satisfy the acceptance criteria of the ASHRAE guidelines 14.

Table 3. CV(RMSE) and MBE of CO_2 Concentration Prediction Result.

Training Algorithm	Hidden Layers-1		Hidden Layers-3		Hidden Layers-5	
	CV(RMSE) (%)	MBE (%)	CV(RMSE) (%)	MBE (%)	CV(RMSE) (%)	MBE (%)
LM	4.04	7.53	4.06	8.03	4.47	8.56
BR	4.06	7.53	4.15	7.98	4.21	8.06
BFG	5.26	9.86	7.65	10.16	12.54	10.86

Figure 4 presents the R^2 of the prediction results. As shown, the high suitability values of 0.9999 and 0.9998 were observed for both the LM model and BR model due to a low degree of dispersion. In addition, a better prediction result was shown when the number of the hidden layer was one. Because of a large degree of dispersion generated by the BFG model, R^2 of the BFG model was large and the suitability was low. Specifically, R^2 measured 0.9393 when the number of the hidden layer was five.

4.2. PM$_{2.5}$

Figure 5 shows the prediction results of the indoor $PM_{2.5}$ concentration. For all models, $PM_{2.5}$ concentration is largely increased to about 160 μg/m^3 and a slight difference was observed in the training. In the prediction phase, the $PM_{2.5}$ concentration is maintained under 30–40 μg/m^3, which is close to the measurement data.

The CV(RMSE) and MBE of $PM_{2.5}$ concentration are shown in Table 4. For both LM and BR models, the range of CV(RMSE) and MBE are within the acceptable criteria of the ASHRAE guidelines 14 in all three hidden layers. Between the two models, the best performance of the prediction is observed when the BR model was used. However, in the case of the BFG model, the prediction performance decreased as the number of the hidden layers increased, even though the range of CV(RMSE) and MBE is within the acceptable criteria of the ASHRAE guidelines 14 when the hidden layer is one. Moreover, the CV(RMSE) is 27.90% when the number of the hidden layer is five for the BFG model.
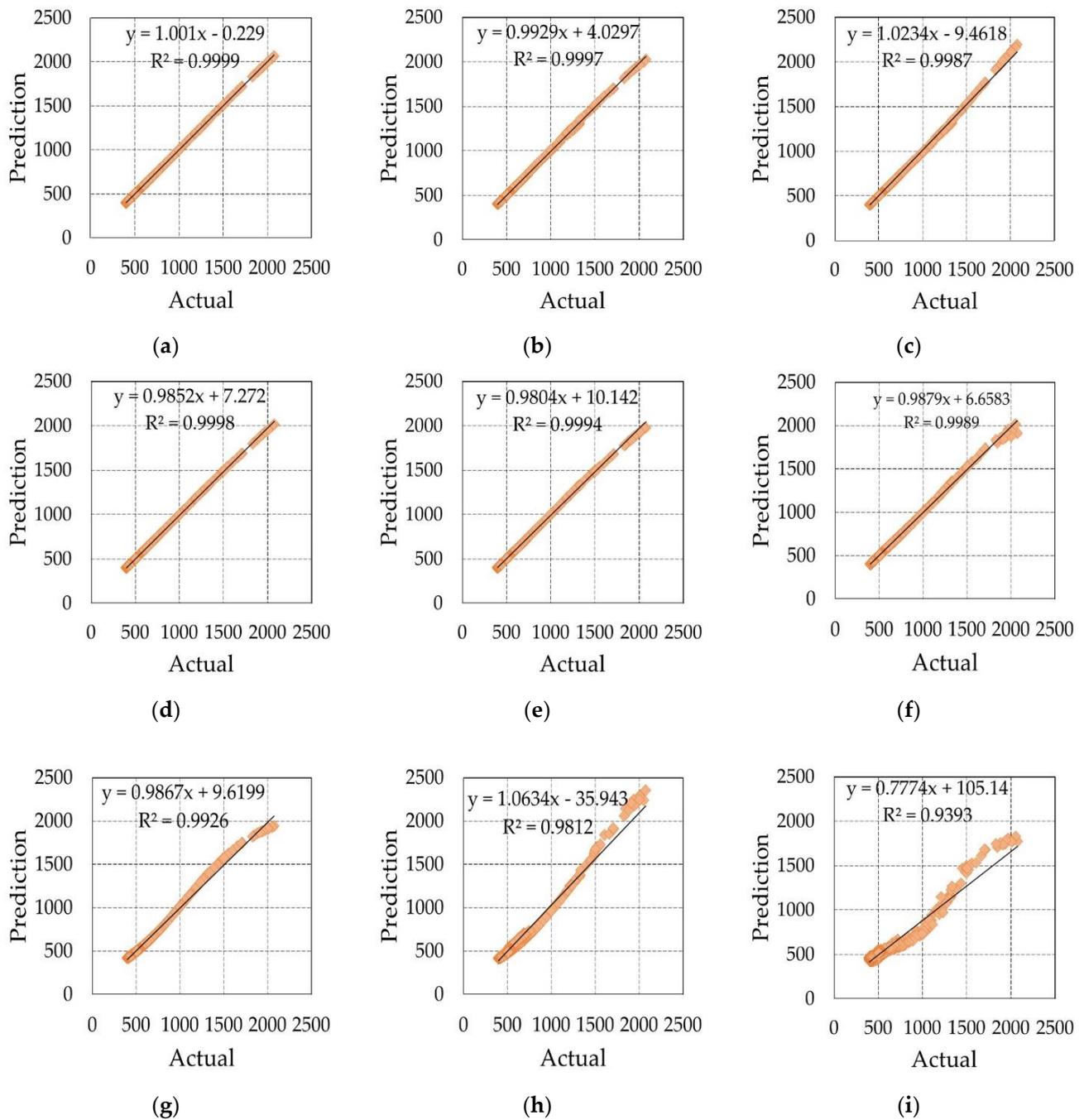
Figure 4. R^2 of prediction results of CO_2 concentration. (**a**) LM-Hidden layer-1; (**b**) LM-Hidden layer-3; (**c**) LM-Hidden layer-5; (**d**) BR-Hidden layer-1; (**e**) BR-Hidden layer-3; (**f**) BR-Hidden layer-5; (**g**) BFG-Hidden layer-1; (**h**) BFG-Hidden layer-3; (**i**) BFG-Hidden layer-5.
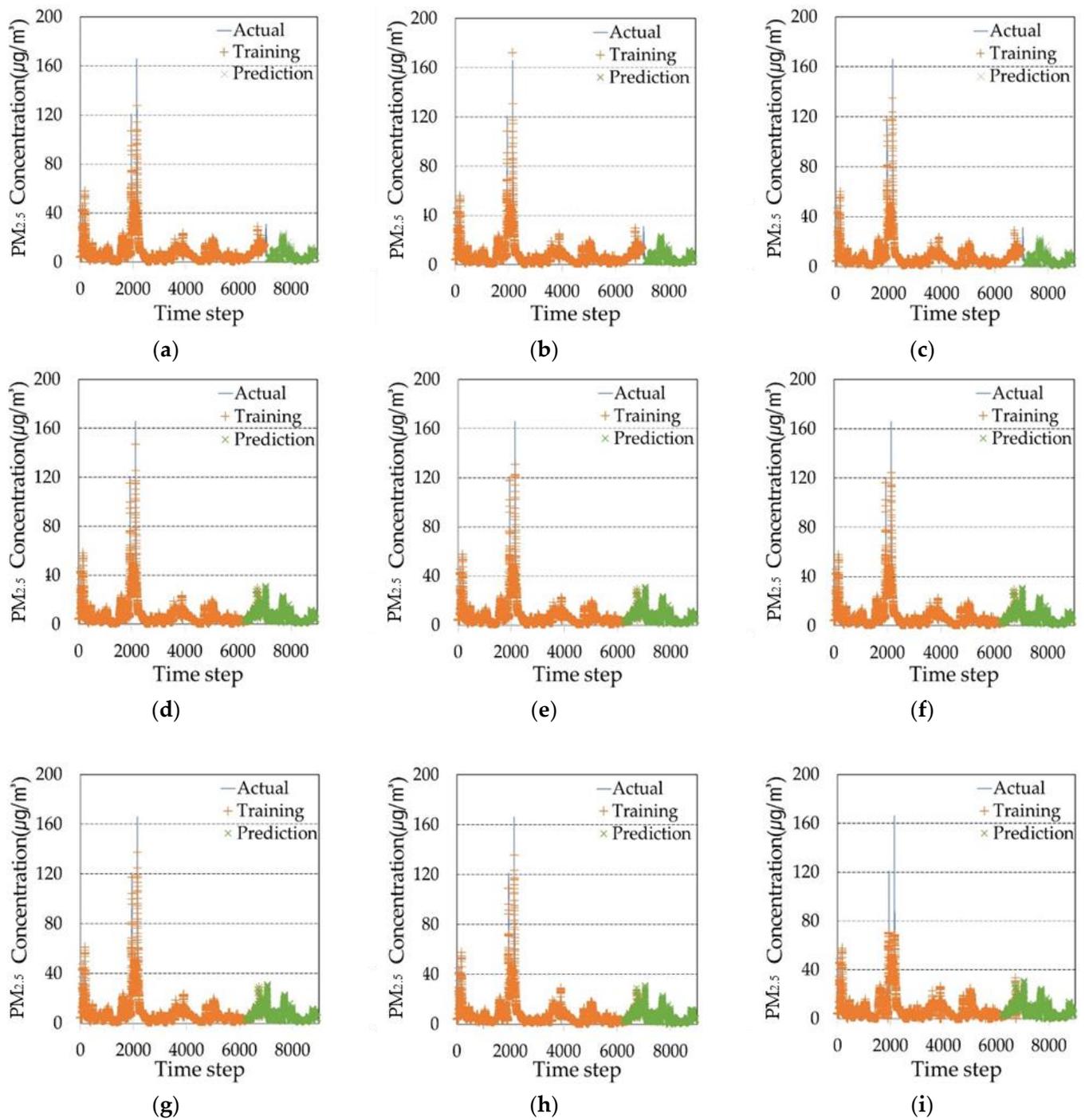
Figure 5. Prediction results of PM$_{2.5}$ concentration. (**a**) LM-Hidden layer-1; (**b**) LM-Hidden layer-3; (**c**) LM-Hidden layer-5; (**d**) BR-Hidden layer-1; (**e**) BR-Hidden layer-3; (**f**) BR-Hidden layer-5; (**g**) BFG-Hidden layer-1; (**h**) BFG-Hidden layer-3; (**i**) BFG-Hidden layer-5.

Table 4. CV(RMSE) and MBE of PM$_{2.5}$ concentration prediction result.

Training Algorithm	Hidden Layers-1		Hidden Layers-3		Hidden Layers-5	
	CV(RMSE) (%)	MBE (%)	CV(RMSE) (%)	MBE (%)	CV(RMSE) (%)	MBE (%)
LM	13.24	4.44	13.76	5.17	13.73	5.49
BR	13.17	3.62	13.27	3.91	13.31	4.25
BFG	13.79	5.66	18.60	6.52	27.90	6.69

For the R^2 of the PM$_{2.5}$ concentration prediction, high suitability for both LM and BR models is achieved (Figure 6). When the number of the hidden layers is one, the R^2 for the LM model and BR model are 0.9994 and 0.9998, respectively. In the case of the BFG model, the R^2 is 0.9984 when the number of the hidden layer is one, which shows high suitability. However, the R^2 is decreased to 0.9734 when the number of the hidden layers is increased.

4.3. VOCs

As shown in Figure 7, the indoor VOCs concentration prediction results are compared with the measurement data. In the training phase, the VOCs concentration is significantly increased to about 16,000 μg/m^3 and all the models show the difference in the VOCs concentration. The VOCs concentration for all the models is close to the measurement data in the prediction phase.

Table 5 summarizes the CV(RMSE) and MBE of the prediction results. The LM model shows the lowest CV(RMSE) among the models and the MBE of the LM model is only within the acceptable range of the ASHRAE guidelines 14. While the CV(RMSE) and MBE for the BR model can meet the acceptable criteria of the ASHRAE guidelines 14 when the number of the hidden layer was one, the CV(RMSE) and MBE cannot satisfy them when the number of the hidden layers is increased. A similar trend was observed for the BFG model. Specifically, the CV(RMSE) is increased to 26.52% when the number of the hidden layers is increased to 5.
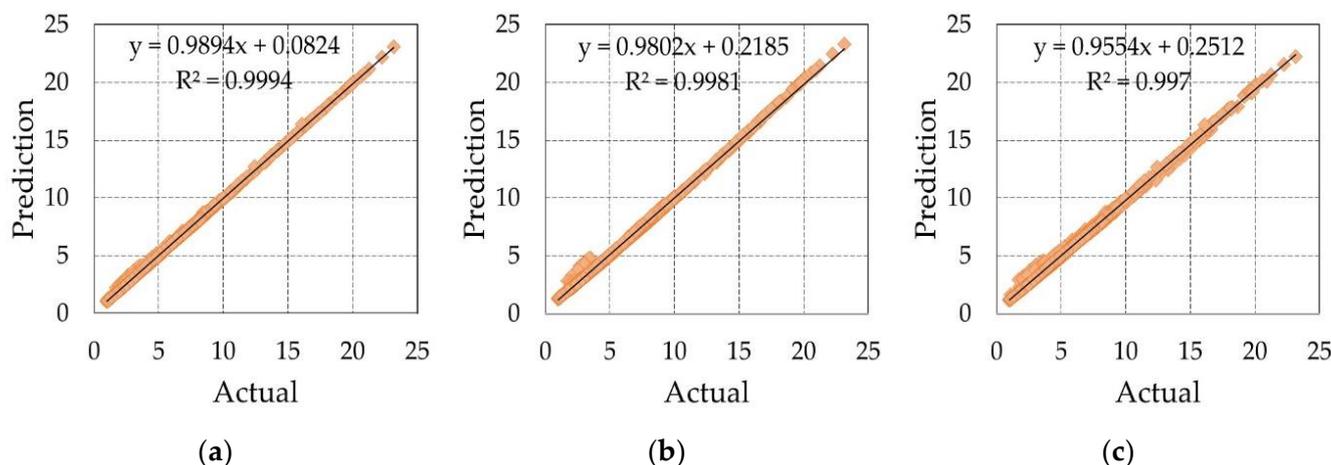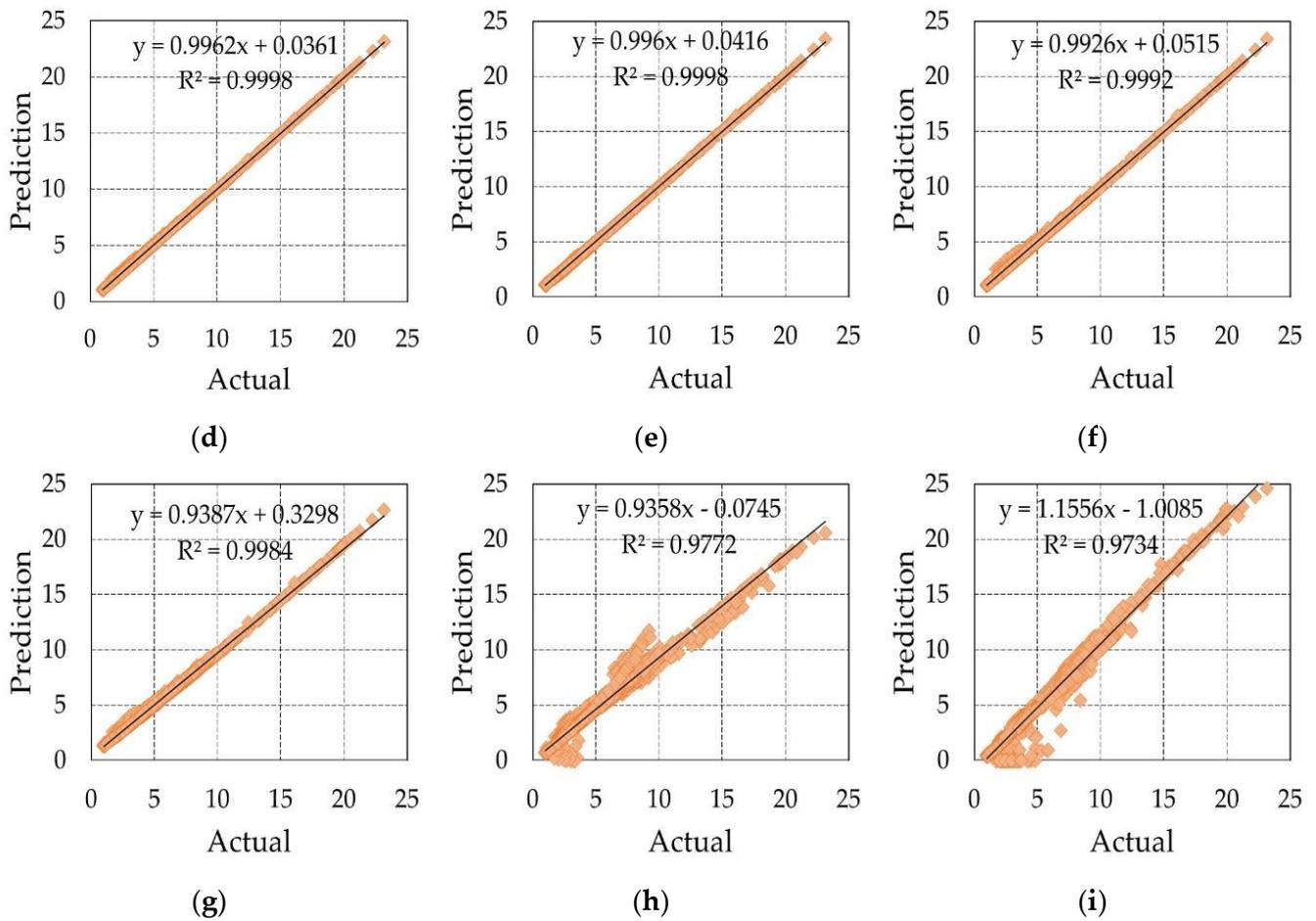

(a)   (b)   (c)

Figure 6. *Cont.*

Figure 6. R^2 of prediction results of PM$_{2.5}$ concentration. (**a**) LM-Hidden layer-1; (**b**) LM-Hidden layer-3; (**c**) LM-Hidden layer-5; (**d**) BR-Hidden layer-1; (**e**) BR-Hidden layer-3; (**f**) BR-Hidden layer-5; (**g**) BFG-Hidden layer-1; (**h**) BFG-Hidden layer-3; (**i**) BFG-Hidden layer-5.
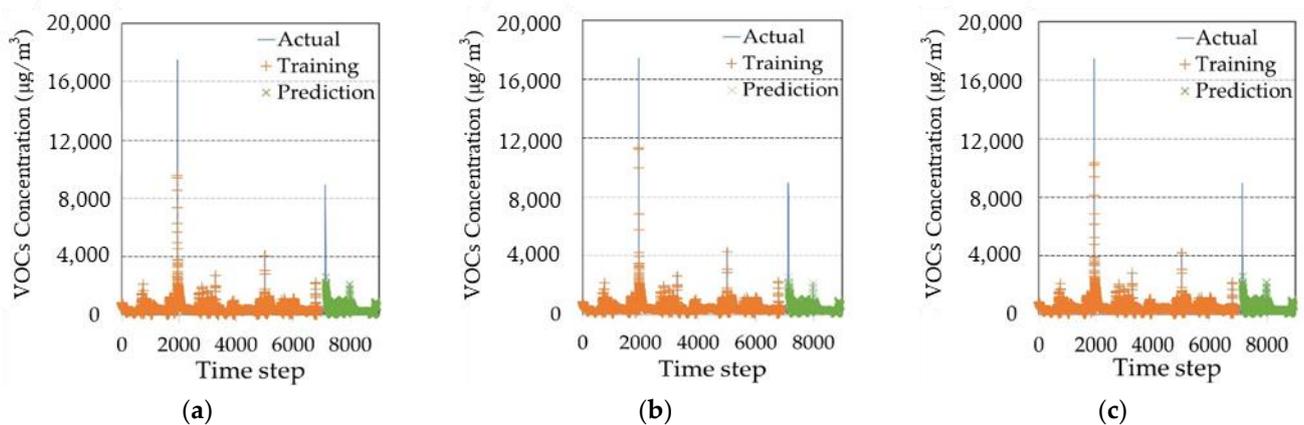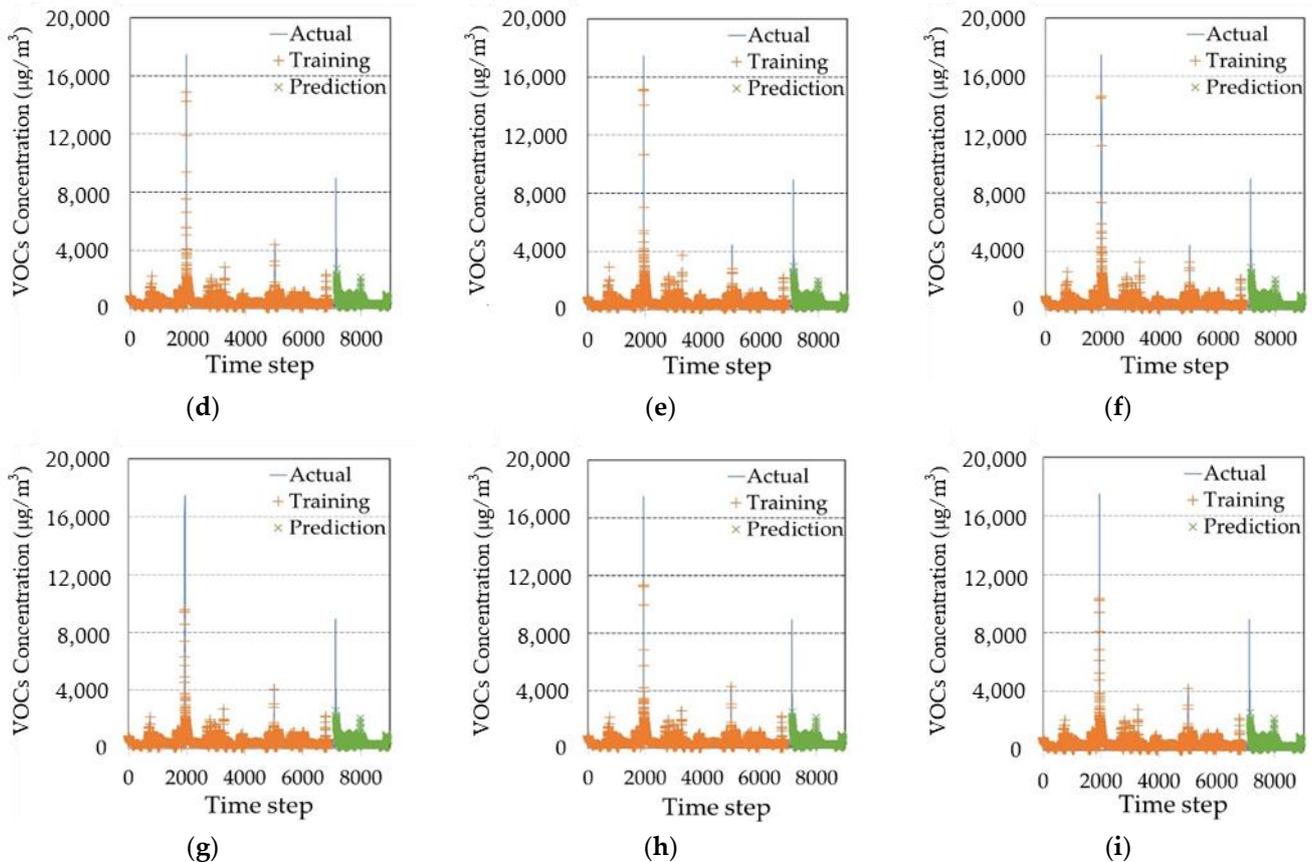

Figure 7. *Cont.*

Figure 7. Prediction results of VOCs concentration. (**a**) LM-Hidden layer-1; (**b**) LM-Hidden layer-3; (**c**) LM-Hidden layer-5; (**d**) BR-Hidden layer-1; (**e**) BR-Hidden layer-3; (**f**) BR-Hidden layer-5; (**g**) BFG-Hidden layer-1; (**h**) BFG-Hidden layer-3; (**i**) BFG-Hidden layer-5.

Table 5. CV(RMSE) and MBE of VOCs concentration prediction result.

Training Algorithm	Hidden Layers-1		Hidden Layers-3		Hidden Layers-5	
	CV(RMSE) (%)	MBE (%)	CV(RMSE) (%)	MBE (%)	CV(RMSE) (%)	MBE (%)
LM	10.81	8.89	10.98	9.35	11.12	8.96
BR	11.06	9.10	14.11	10.19	15.65	10.37
BFG	14.57	10.68	14.58	10.79	26.52	11.25

Figure 8 presents the R^2 of the VOCs prediction results. When the number of the hidden layer is one, the LM model shows the highest suitability among the models. When the number of the hidden layer is increased to 5, the R^2 for the LM model is 0.9984. While the BR model shows a relatively lower R^2 (0.9998) than that of the LM model, it is highly suitable. When the number of the hidden layers is increased, the R^2 for the BR model is ranging from 0.9946 to 0.9966. In the case of the BFG model, the R^2 is 0.9974 when the number of the hidden layer is one. However, it is decreased to 0.9445 with the increase in the number of the hidden layers.
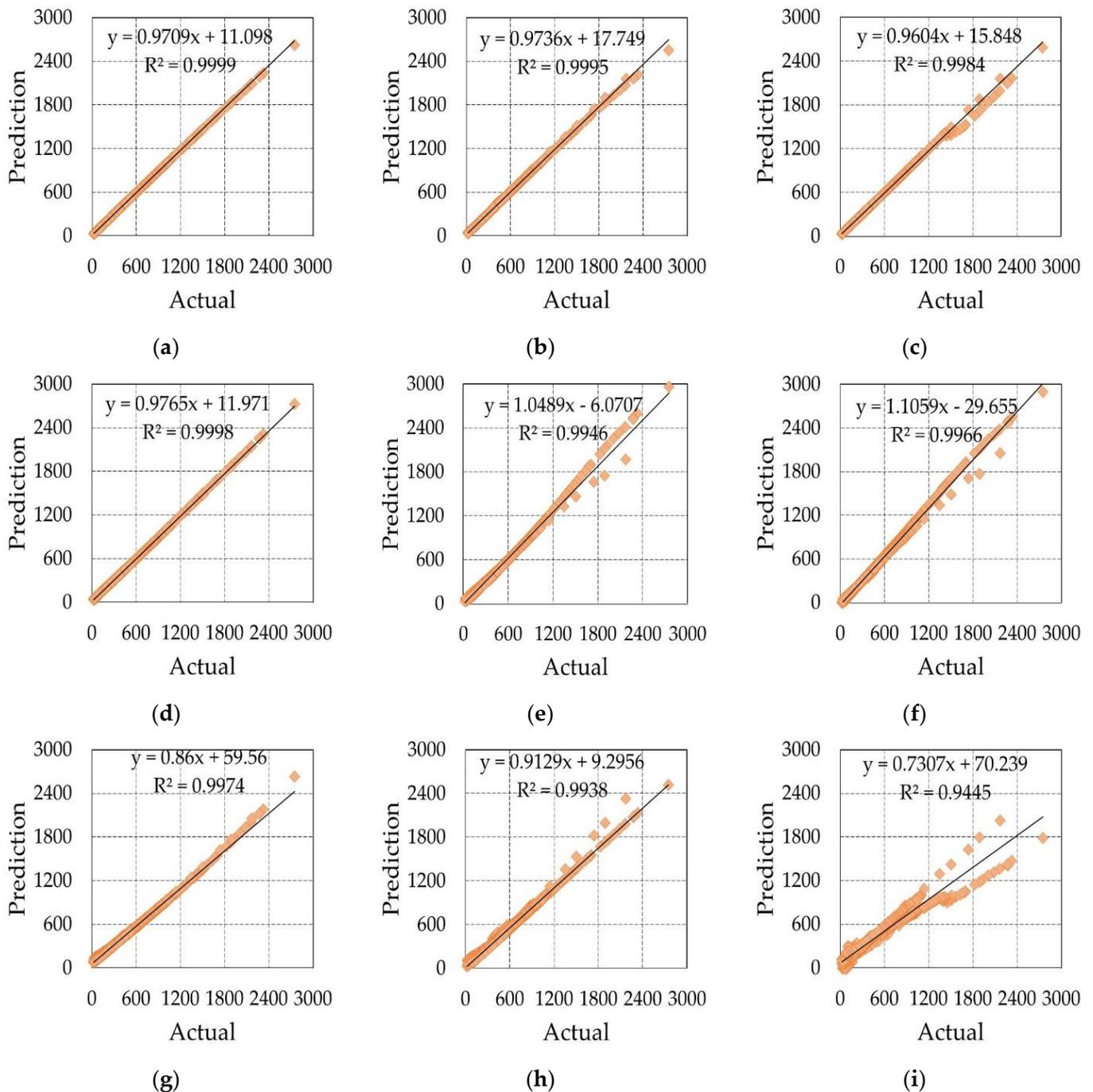
Figure 8. R^2 of prediction results of VOCs concentration. (**a**) LM-Hidden layer-1; (**b**) LM-Hidden layer-3; (**c**) LM-Hidden layer-5; (**d**) BR-Hidden layer-1; (**e**) BR-Hidden layer-3; (**f**) BR-Hidden layer-5; (**g**) BFG-Hidden layer-1; (**h**) BFG-Hidden layer-3; (**i**) BFG-Hidden layer-5.

5. Discussion

For the present study, the concentration of CO_2, $PM_{2.5}$, and VOCs was predicted using the feed-forward neural network model with three different back-propagation training algorithms, which were LM, BR, and BFG. Among the models, the LM model showed the best performance for the prediction of those indoor pollutants and the obtained results can meet the acceptable criteria of the ASHRAE guidelines 14. While the prediction for CO_2 and $PM_{2.5}$ by the BR model showed a good performance, the performance was decreased as the number of the hidden layers was increased in the case of VOCs concentration prediction.

The BFG model showed the lowest performance for the prediction of indoor pollutants and it was decreased largely as with the increase in the number of the hidden layers.

It was commonly observed that the prediction performance decreased as the number of the hidden layers increased. As shown, the number of the hidden layers in the ANN model can highly affect the prediction results. In addition, over-fitting data can be produced as the number of the hidden layers is increased [51]. Thus, the choice of the number of the hidden layers is one of the most important variables in the ANN model [52]. According to the study of Yeon et al., the number of hidden layers can be determined, when the CV(RMSE) is lowest [53]. In their study, the number of hidden layers was tested from one to three layers and one layer was selected based on the lowest CV(RMSE) value. This was similarly observed in the present study. Bui et al. also chose one hidden layer for identifying the impact of each building parameter on the energy performance [51]. In the case of the ANN model performed by Cho and Moon, they proved that the structured model with two hidden layers was suitable to predict the concentrations of CO_2, PM_{10}, and $PM_{2.5}$ in a school building [54]. As can be shown, the optimized number of hidden layers can be different based on the structured ANN model and it is carefully chosen by assessment of the prediction performance.

Moreover, the prediction performance differed based on the type of indoor pollutant. Based on the CV(RMSE), all models showed the best performance for the prediction of CO_2 concentration, while the poorest performance was shown for the prediction of $PM_{2.5}$ concentration. This can be caused by the difference in data patterns. In the case of the CO_2, the measured CO_2 concentration showed a regular pattern. For the measured concentration of $PM_{2.5}$ and VOCs, the data showed irregular patterns due to the outdoor pollutants through building openings. This can affect the prediction results in the training phase causing a low degree of dispersion. While these irregular data patterns caused low suitability, the high fitness of the R^2 (0.99) was achieved for all the models when the number of the hidden layer was one.

6. Conclusions

As the amount of time spent by people indoors is significantly increasing, much attention has been paid to the indoor air quality in buildings. Many studies have investigated indoor air pollutants to improve the IAQ. While most investigations were performed in residential or commercial buildings, a few studies have focused on the IAQ in child daycare centers in which, a high concentration of indoor air pollutants was reported. Moreover, investigations of indoor air pollutants, in general, were conducted through field measurements which are cost-intensive and time-consuming.

For the present study, the concentrations of indoor air pollutants such as CO_2, $PM_{2.5}$, and VOCs in child daycare centers were predicted by using a feed-forward neural network model with three different back-propagation training algorithms such as LM, BR, and BFG. For the training and validation, the data of the indoor pollutants measured in child daycare facilities for a month were used. The number of hidden layers in each model was set at one, three, and five, and other training parameters were the same for all models.

The results showed that all the models produced a good performance for the prediction of indoor pollutants compared with the measured data. Among the models, the prediction by the LM model met the acceptable criteria of the ASHRAE guideline 14 under all conditions. While the CO_2 and $PM_{2.5}$ concentrations predicted by the BR model satisfied the acceptance criteria of the ASHRAE guideline 14, the predictive performance decreased, when the number of the hidden layers increased. The BFG model showed the poorest performance for the prediction of indoor pollutants among the models under all conditions. Moreover, a large difference between the prediction and the measured data was observed by the BFG model when the number of the hidden layers increased.

While the predictions by using machine learning techniques have been widely used in various fields, there were few studies available for indoor air pollutant prediction. In addition, the predictive performance by different training algorithms was rarely investigated.

This study presented the difference in the predictive performance of indoor air pollutants by applying different training algorithms in the ANN model. Considering the outcomes of the study, better prediction results can be obtained through the proper selection of training algorithms for time series data. Furthermore, the outcome of the present study can be used as information for more sophisticated machine learning applications for improving indoor air quality and related predictions in child daycare centers. For further studies, other indoor pollutants such as CO and PM_{10}, etc. will be investigated for improving indoor air quality in child daycare centers.

Author Contributions: J.K. contributed to the project idea development and wrote a draft version; Y.H. prepared the data and contributed to the results; N.S. performed the data analysis; D.D.K. reviewed the final manuscript and contributed to the discussion. 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: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest regarding the publication of this article.

References

1. Awada, M.; Becerik-Gerber, B.; Hoque, S.; O'Neill, Z.; Pedrielli, G.; Wen, J.; Wu, T. Ten questions concerning occupant health in buildings during normal operations and extreme events including the COVID-19 pandemic. *Build. Environ.* **2021**, *188*, 107480. [CrossRef]
2. Adam, M.G.; Tran, P.T.M.; Balasubramanian, R. Air quality changes in cities during the COVID-19 lockdown: A critical review. *Atmos. Res.* **2021**, *264*, 105823. [CrossRef] [PubMed]
3. Schibuola, L.; Tambani, C. High energy efficiency ventilation to limit COVID-19 contagion in school environments. *Energy Build.* **2021**, *240*, 110882. [CrossRef] [PubMed]
4. Sha, H.; Zhang, X.; Qi, D. Optimal control of high-rise building mechanical ventilation system for achieving low risk of COVID-19 transmission and ventilative cooling. *Sustain. Cities Soc.* **2021**, *74*, 103256. [CrossRef]
5. Gil-Baez, M.; Lizana, J.; Becerra Villanueva, J.A.; Molina-Huelva, M.; Serrano-Jimenez, A.; Chacartegui, R. Natural ventilation in classrooms for healthy schools in the COVID era in mediterranean climate. *Build. Environ.* **2021**, *206*, 108345. [CrossRef]
6. Szczepanik-Ścisło, N.; Ścisło, Ł. Air leakage modelling and its influence on the air quality inside a garage. *E3S Web Conf.* **2018**, *44*, 00172. [CrossRef]
7. Ye, W.; Won, D.; Zhang, X. A preliminary ventilation rate determination methods study for residential buildings and offices based on voc emission database. *Build. Environ.* **2014**, *79*, 168–180. [CrossRef]
8. Földváry, V.; Bekö, G.; Langer, S.; Arrhenius, K.; Petráš, D. Effect of energy renovation on indoor air quality in multifamily residential buildings in slovakia. *Build. Environ.* **2017**, *122*, 363–372. [CrossRef]
9. Kaunelienė, V.; Prasauskas, T.; Krugly, E.; Stasiulaitienė, I.; Čiužas, D.; Šeduikytė, L.; Martuzevičius, D. Indoor air quality in low energy residential buildings in lithuania. *Build. Environ.* **2016**, *108*, 63–72. [CrossRef]
10. Yang, S.; Yuk, H.; Yun, B.Y.; Kim, Y.U.; Wi, S.; Kim, S. Passive pm2.5 control plan of educational buildings by using airtight improvement technologies in south korea. *J. Hazard. Mater.* **2022**, *423*, 126990. [CrossRef]
11. Bralewska, K.; Rogula-Kozłowska, W.; Bralewski, A. Indoor air quality in sports center: Assessment of gaseous pollutants. *Build. Environ.* **2022**, *208*, 108589. [CrossRef]
12. Connolly, R.E.; Yu, Q.; Wang, Z.; Chen, Y.-H.; Liu, J.Z.; Collier-Oxandale, A.; Papapostolou, V.; Polidori, A.; Zhu, Y. Long-term evaluation of a low-cost air sensor network for monitoring indoor and outdoor air quality at the community scale. *Sci. Total Environ.* **2022**, *807*, 150797. [CrossRef]
13. Mikola, A.; Hamburg, A.; Kuusk, K.; Kalamees, T.; Voll, H.; Kurnitski, J. The impact of the technical requirements of the renovation grant on the ventilation and indoor air quality in apartment buildings. *Build. Environ.* **2022**, *210*, 108698. [CrossRef]
14. Kang, I.; McCreery, A.; Azimi, P.; Gramigna, A.; Baca, G.; Abromitis, K.; Wang, M.; Zeng, Y.; Scheu, R.; Crowder, T.; et al. Indoor air quality impacts of residential mechanical ventilation system retrofits in existing homes in chicago, il. *Sci. Total Environ.* **2022**, *804*, 150129. [CrossRef]
15. Al-Rawi, M.; Ikutegbe, C.A.; Auckaili, A.; Farid, M.M. Sustainable technologies to improve indoor air quality in a residential house—A case study in waikato, new zealand. *Energy Build.* **2021**, *250*, 111283. [CrossRef]
16. Oh, H.-J.; Nam, I.-S.; Yun, H.; Kim, J.; Yang, J.; Sohn, J.-R. Characterization of indoor air quality and efficiency of air purifier in childcare centers, korea. *Build. Environ.* **2014**, *82*, 203–214. [CrossRef]

17. Roda, C.; Barral, S.; Ravelomanantsoa, H.; Dusséaux, M.; Tribout, M.; Le Moullec, Y.; Momas, I. Assessment of indoor environment in paris child day care centers. *Environ. Res.* **2011**, *111*, 1010–1017. [CrossRef]

18. Hwang, S.H.; Seo, S.; Yoo, Y.; Kim, K.Y.; Choung, J.T.; Park, W.M. Indoor air quality of daycare centers in seoul, korea. *Build. Environ.* **2017**, *124*, 186–193. [CrossRef]

19. Madureira, J.; Paciência, I.; Rufo, J.C.; Pereira, C.; Teixeira, J.P.; de Oliveira Fernandes, E. Assessment and determinants of airborne bacterial and fungal concentrations in different indoor environments: Homes, child day-care centres, primary schools and elderly care centres. *Atmos. Environ.* **2015**, *109*, 139–146. [CrossRef]

20. St-Jean, M.; St-Amand, A.; Gilbert, N.L.; Soto, J.C.; Guay, M.; Davis, K.; Gyorkos, T.W. Indoor air quality in montréal area day-care centres, canada. *Environ. Res.* **2012**, *118*, 1–7. [CrossRef]

21. Harbizadeh, A.; Mirzaee, S.A.; Khosravi, A.D.; Shoushtari, F.S.; Goodarzi, H.; Alavi, N.; Ankali, K.A.; Rad, H.D.; Maleki, H.; Goudarzi, G. Indoor and outdoor airborne bacterial air quality in day-care centers (dccs) in greater ahvaz, iran. *Atmos. Environ.* **2019**, *216*, 116927. [CrossRef]

22. Zuraimi, M.S.; Tham, K.W. Indoor air quality and its determinants in tropical child care centers. *Atmos. Environ.* **2008**, *42*, 2225–2239. [CrossRef]

23. Araújo-Martins, J.; Carreiro Martins, P.; Viegas, J.; Aelenei, D.; Cano, M.M.; Teixeira, J.P.; Paixão, P.; Papoila, A.L.; Leiria-Pinto, P.; Pedro, C.; et al. Environment and health in children day care centres (envirh)—Study rationale and protocol. *Rev. Port. De Pneumol.* **2014**, *20*, 311–323. [CrossRef] [PubMed]

24. Heibati, S.; Maref, W.; Saber, H.H. Assessing the energy, indoor air quality, and moisture performance for a three-story building using an integrated model, part two: Integrating the indoor air quality, moisture, and thermal comfort. *Energies* **2021**, *14*, 4915. [CrossRef]

25. Olu-Ajayi, R.; Alaka, H.; Sulaimon, I.; Sunmola, F.; Ajayi, S. Machine learning for energy performance prediction at the design stage of buildings. *Energy Sustain. Dev.* **2022**, *66*, 12–25. [CrossRef]

26. Elbeltagi, E.; Wefki, H. Predicting energy consumption for residential buildings using ann through parametric modeling. *Energy Rep.* **2021**, *7*, 2534–2545. [CrossRef]

27. Olu-Ajayi, R.; Alaka, H.; Sulaimon, I.; Sunmola, F.; Ajayi, S. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. *J. Build. Eng.* **2022**, *45*, 103406. [CrossRef]

28. Zhu, Y.F.; Yao, Y.; Huang, Y.; Chen, C.H.; Zhang, H.Y.; Huang, Z. Machine learning applications for assessment of dynamic progressive collapse of steel moment frames. *Structures* **2022**, *36*, 927–934. [CrossRef]

29. Chalapathy, R.; Khoa, N.L.D.; Sethuvenkatraman, S. Comparing multi-step ahead building cooling load prediction using shallow machine learning and deep learning models. *Sustain. Energy Grids Netw.* **2021**, *28*, 100543. [CrossRef]

30. Ding, Y.; Fan, L.; Liu, X. Analysis of feature matrix in machine learning algorithms to predict energy consumption of public buildings. *Energy Build.* **2021**, *249*, 111208. [CrossRef]

31. Zhang, H.; Srinivasan, R.; Yang, X. Simulation and analysis of indoor air quality in florida using time series regression (tsr) and artificial neural networks (ann) models. *Symmetry* **2021**, *13*, 952. [CrossRef]

32. Amuthadevi, C.; Ds, V.; Ramachandran, V. Development of air quality monitoring (aqm) models using different machine learning approaches. *J. Ambient Intell. Humaniz. Comput.* **2021**, 1–13. [CrossRef]

33. Tagliabue, L.C.; Re Cecconi, F.; Rinaldi, S.; Ciribini, A.L.C. Data driven indoor air quality prediction in educational facilities based on iot network. *Energy Build.* **2021**, *236*, 110782. [CrossRef]

34. Jang, H.-S.; Xing, S. A model to predict ammonia emission using a modified genetic artificial neural network: Analyzing cement mixed with fly ash from a coal-fired power plant. *Constr. Build. Mater.* **2020**, *230*, 117025. [CrossRef]

35. Jeong, J.-H.; Jo, J.-H.; Kim, S.-H.; Bang, S.-H.; Lee, S.-S. A study on machine learning model for predicting uncollected parameters in indoor environment evaluation. *J. Korea Inst. Inf. Electron. Commun. Technol.* **2021**, *14*, 413–420.

36. Scislo, L.; Szczepanik-Scislo, N. Air quality sensor data collection and analytics with iot for an apartment with mechanical ventilation. In Proceedings of the 2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Cracow, Poland, 22–25 September 2021; pp. 932–936.

37. Li, Z.; Tong, X.; Ho, J.M.W.; Kwok, T.C.Y.; Dong, G.; Ho, K.-F.; Yim, S.H.L. A practical framework for predicting residential indoor pm2.5 concentration using land-use regression and machine learning methods. *Chemosphere* **2021**, *265*, 129140. [CrossRef] [PubMed]

38. Kallio, J.; Tervonen, J.; Räsänen, P.; Mäkynen, R.; Koivusaari, J.; Peltola, J. Forecasting office indoor co2 concentration using machine learning with a one-year dataset. *Build. Environ.* **2021**, *187*, 107409. [CrossRef]

39. Taheri, S.; Razban, A. Learning-based co2 concentration prediction: Application to indoor air quality control using demand-controlled ventilation. *Build. Environ.* **2021**, *205*, 108164. [CrossRef]

40. Sassi, M.S.H.; Fourati, L.C. In Deep learning and augmented reality for iot-based air quality monitoring and prediction system. In Proceedings of the 2021 International Symposium on Networks, Computers and Communications (ISNCC), Dubai, United Arab Emirates, 31 October–2 November 2021; pp. 1–6.

41. Saad, S.M.; Andrew, A.M.; Shakaff, A.Y.M.; Saad, A.R.M.; Kamarudin, A.M.Y.; Zakaria, A. Classifying sources influencing indoor air quality (iaq) using artificial neural network (ann). *Sensors* **2015**, *15*, 11665–11684. [CrossRef]

42. Putra, J.C.P.; Safrilah; Ihsan, M. The prediction of indoor air quality in office room using artificial neural network. *AIP Conf. Proc.* **2018**, *1977*, 020040.

43. Dai, X.; Liu, J.; Zhang, X.; Chen, W. An artificial neural network model using outdoor environmental parameters and residential building characteristics for predicting the nighttime natural ventilation effect. *Build. Environ.* **2019**, *159*, 106139. [CrossRef]

44. Segala, G.; Doriguzzi Corin, R.; Peroni, C.; Gazzini, T.; Siracusa, D. A practical and adaptive approach to predicting indoor co2. *Appl. Sci.* **2021**, *11*, 10771. [CrossRef]

45. Big Data Environment Platform. Available online: https://www.Bigdata-environment.Kr/user/main.Do (accessed on 1 November 2021).

46. Matlab, The Mathworks, Inc. Available online: https://www.Mathworks.Com/?S_tid=gn_logo (accessed on 1 December 2021).

47. Golafshani, E.M.; Behnood, A. Application of soft computing methods for predicting the elastic modulus of recycled aggregate concrete. *J. Clean. Prod.* **2018**, *176*, 1163–1176. [CrossRef]

48. Koçer, M.; Öztürk, M.; Hakan Arslan, M. Determination of moment, shear and ductility capacities of spiral columns using an artificial neural network. *J. Build. Eng.* **2019**, *26*, 100878. [CrossRef]

49. Wang, R.; Lu, S.; Li, Q. Multi-criteria comprehensive study on predictive algorithm of hourly heating energy consumption for residential buildings. *Sustain. Cities Soc.* **2019**, *49*, 101623. [CrossRef]

50. American Society of Heating, Refrigerating and Air Conditioning Engineers. Ashrae Guideline 14-2002, Measurement of Energy and Demand Savings—Measurement of Energy, Demand and Water Savings. 2002. Available online: http://www.eeperformance.org/uploads/8/6/5/0/8650231/ashrae_guideline_14-2002_measurement_of_energy_and_demand_saving.pdf (accessed on 1 February 2021).

51. Bui, D.-K.; Nguyen, T.N.; Ngo, T.D.; Nguyen-Xuan, H. An artificial neural network (ann) expert system enhanced with the electromagnetism-based firefly algorithm (efa) for predicting the energy consumption in buildings. *Energy* **2020**, *190*, 116370. [CrossRef]

52. Lu, C.; Li, S.; Lu, Z. Building energy prediction using artificial neural networks: A literature survey. *Energy Build.* **2022**, *262*, 111718. [CrossRef]

53. Yeon, S.; Yu, B.; Seo, B.; Yoon, Y.; Lee, K.H. Ann based automatic slat angle control of venetian blind for minimized total load in an office building. *Sol. Energy* **2019**, *180*, 133–145. [CrossRef]

54. Cho, J.H.; Moon, J.W. Integrated artificial neural network prediction model of indoor environmental quality in a school building. *J. Clean. Prod.* **2022**, *344*, 131083. [CrossRef]