applied sciences

MDPI

Article

Application of Deep Learning Models and Network Method for Comprehensive Air-Quality Index Prediction

Donghyun Kim [1], Heechan Han [2], Wonjoon Wang [1], Yujin Kang [1], Hoyong Lee [1] and Hung Soo Kim [1,*]

[1] Department of Civil Engineering, Inha University, Incheon 22212, Korea; yesdktpdi@naver.com (D.K.); makelest@naver.com (W.W.); rkddb1215@naver.com (Y.K.); hy5890@nate.com (H.L.)
[2] Blackland Research and Extension Center, Texas A&M AgriLife, Temple, TX 76502, USA; heechan.han@ag.tamu.edu
* Correspondence: sookim@inha.ac.kr; Tel.: +82-10-3441-1038

Abstract: Accurate pollutant prediction is essential in fields such as meteorology, meteorological disasters, and climate change studies. In this study, long short-term memory (LSTM) and deep neural network (DNN) models were applied to six pollutants and comprehensive air-quality index (CAI) predictions from 2015 to 2020 in Korea. In addition, we used the network method to find the best data sources that provide factors affecting comprehensive air-quality index behaviors. This study had two steps: (1) predicting the six pollutants, including fine dust (PM_{10}), fine particulate matter ($PM_{2.5}$), ozone (O_3), sulfurous acid gas (SO_2), nitrogen dioxide (NO_2), and carbon monoxide (CO) using the LSTM model; (2) forecasting the CAI using the six predicted pollutants in the first step as predictors of DNNs. The predictive ability of each model for the six pollutants and CAI prediction was evaluated by comparing it with the observed air-quality data. This study showed that combining a DNN model with the network method provided a high predictive power, and this combination could be a remarkable strength in CAI prediction. As the need for disaster management increases, it is anticipated that the LSTM and DNN models with the network method have ample potential to track the dynamics of air pollution behaviors.

Keywords: comprehensive air-quality index; deep learning model; kriging; network method

check for updates

Citation: Kim, D.; Han, H.; Wang, W.; Kang, Y.; Lee, H.; Kim, H.S. Application of Deep Learning Models and Network Method for Comprehensive Air-Quality Index Prediction. *Appl. Sci.* **2022**, *12*, 6699. https://doi.org/10.3390/app12136699

Academic Editor: Prashant Kumar

Received: 6 May 2022
Accepted: 29 June 2022
Published: 1 July 2022

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

Copyright: © 2022 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The aggravation of air pollution significantly impacts various fields worldwide, including air-quality-related disasters. According to the World Health Organization, air pollution exposure causes over 6 million premature deaths [1]. As the extent of air-quality-related damage in South Korea increases, studies on the damage caused by air pollution are essential [2,3]. As the damage caused by highly concentrated air pollution in South Korea has increased, research related to air pollution has also increased. Therefore, an accurate air pollution forecasting system is needed to reduce indirect damage.

The forecast has been providing air pollution information since 2014. This information is categorized into six types, including fine dust (PM_{10}), fine particulate matter ($PM_{2.5}$), ozone (O_3), sulfurous acid gas (SO_2), nitrogen dioxide (NO_2), and carbon monoxide (CO), which can be used to predict the value of the comprehensive air-quality index (CAI). However, the accuracy of these forecasts is limited [4].

Recently, with the progress of computing algorithm technology, deep learning models have begun to be used to analyze and forecast nonlinear relationships between data variables. In a similar way, deep learning models have significantly improved data analysis performance because they provide highly reliable results [5]. In addition, deep learning techniques such as deep neural networks (DNNs) and long short-term memory (LSTM) models have been applied for various purposes in the fields of meteorology, hydrology, precipitation, and drought analysis, including for forecasting [6–11]. Moreover, over the past decades, many researchers have analyzed and predicted the concentrations of air pollutants

to reduce the damage caused by air pollution and enhance prediction accuracy [12–15]. In most previous studies, all data from all stations around the study area were considered and used as independent variables. However, in using all of the data from the surrounding stations, distorted results are obtained, since data that are unnecessary for model learning are used.

Network theory studies graphs that represent either symmetric or asymmetric relations between discrete objects. Network theory is a part of graph theory in computer science and network science: a network can be defined as a graph in which nodes and links have attributes. Network theory has applications in many disciplines, including statistics, computer science, and climatology [16–20].

In recent years, kriging has been applied in various research fields owing to its excellent predictive performance, and many studies using kriging have been carried out internationally in areas such as hydrology, meteorology, and environmental science [21–25]. In particular, kriging has been used to understand the spatial distribution of data and predict the spatial distribution of target factors by integrating various remote sensing data in the fields of hydrology and meteorology [26–35].

The number of national air-quality warnings issued in South Korea has gradually increased since 2015. Missing data occur for various reasons at each measuring station when they provide incorrect information. In addition, the prediction forecast is only for each administrative unit and significant region. It is necessary to verify correct information and increase the predictive power. Therefore, a need for the spatial distribution of predictive data has emerged. To overcome these limitations, this study aimed to develop CAI prediction models using deep learning models and the network method. We applied network theory, kriging, LSTM, and DNN models in order to predict CAI changes in South Korea.

2. Materials and Methods

2.1. Study Area

South Korea was selected as the study area (Figure 1). According to local government ordinances, fine dust prediction and warning systems were operated mainly by local governments until 2014; however, since 2015, a nationwide air-quality warning system has been implemented. In South Korea, over 300 air-quality stations are used for air pollution warnings. Recently, with an increase in air pollution, the importance of CAI and the concentrations of fine dust and fine particulate matter has increased. In addition, there is a negative opinion that the forecasts do not reflect the six pollutants measured. Currently, there are national industrial complexes in Jeong-wang. Numerous development works are being carried out in neighboring cities, such as subway construction and housing site construction in various parts of the city. As a result, the health of citizens is threatened by high concentrations of fine dust compared with neighboring areas. This study focused on Jeong-wang Station, one of the most densely populated areas in South Korea, with 18 nearby stations. This study attempted to strengthen South Korea's disaster management capabilities by presenting a CAI prediction model using deep learning models with a network method.

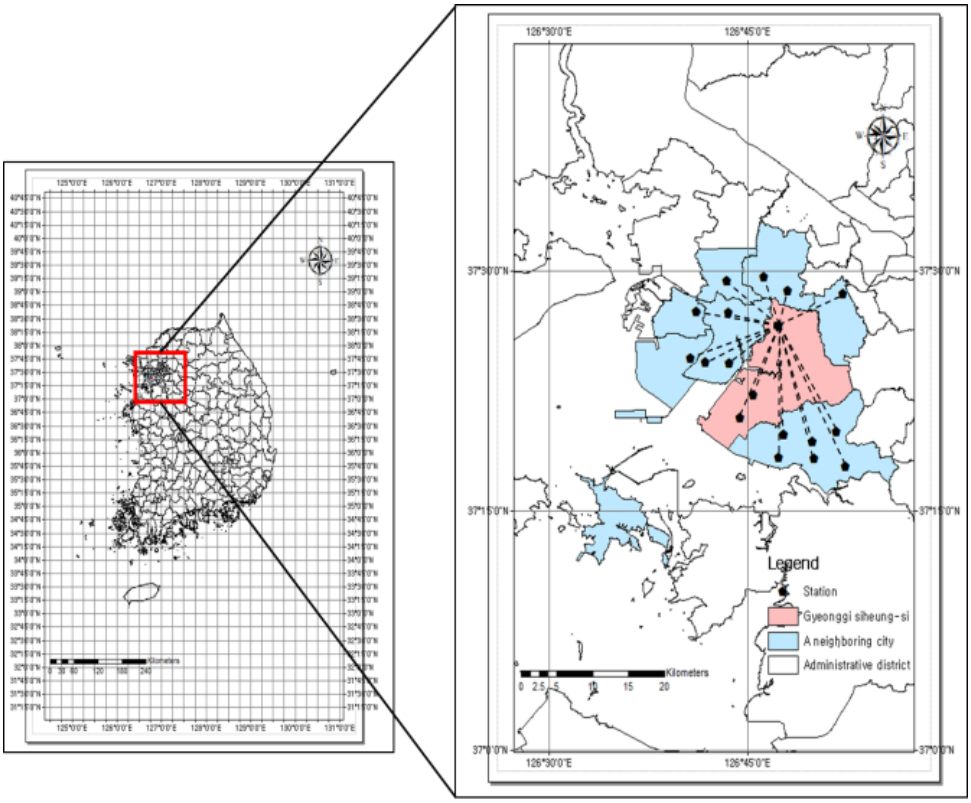


Figure 1. Study area. Black pentagonal (right panel) points are air pollutant monitoring stations in Korea (data source: Korea Environment Corporation).

2.2. Flowchart

This study attempted to develop a model to address the problems mentioned above and to predict the CAI. Figure 2 shows a flowchart for developing a CAI prediction model. The calculation process for predicting the CAI, which was the purpose of this study, was as follows. (1) Network methods were applied to air-quality stations located across the country to identify clusters with similar characteristics. (2) As for the dependent and independent variables, Jeong-wang's six-pollutant hourly data were collected from 1 January 2015 to 31 December 2020. In addition, a CAI was calculated using the six pollutants. (3) In order to develop a model for predicting the six pollutants, the model was divided into learning and evaluation sections. Data from 2015 to 2018 were used for the learning part, and data from 2019 to 2020 were used for the evaluation part. Six pollutants were predicted using LSTM. (4) In order to develop a CAI prediction model, the model was divided into learning and evaluation sections. Data from 2015 to 2018 were used for the learning part, and data from 2019 to 2020 were used for the evaluation part. A comprehensive air-quality index was predicted using the DNN model. (5) The prediction accuracy (predictive power) of each model was evaluated using a normalized root mean squared error (NRMSE), a Nash–Sutcliffe efficiency coefficient (NSE), and a correlation coefficient (CC). (6) If the prediction were made using point data, there would be the disadvantage that only the value located at the point could be predicted. To solve this problem, the predicted values were spatially distributed by applying ordinary kriging.

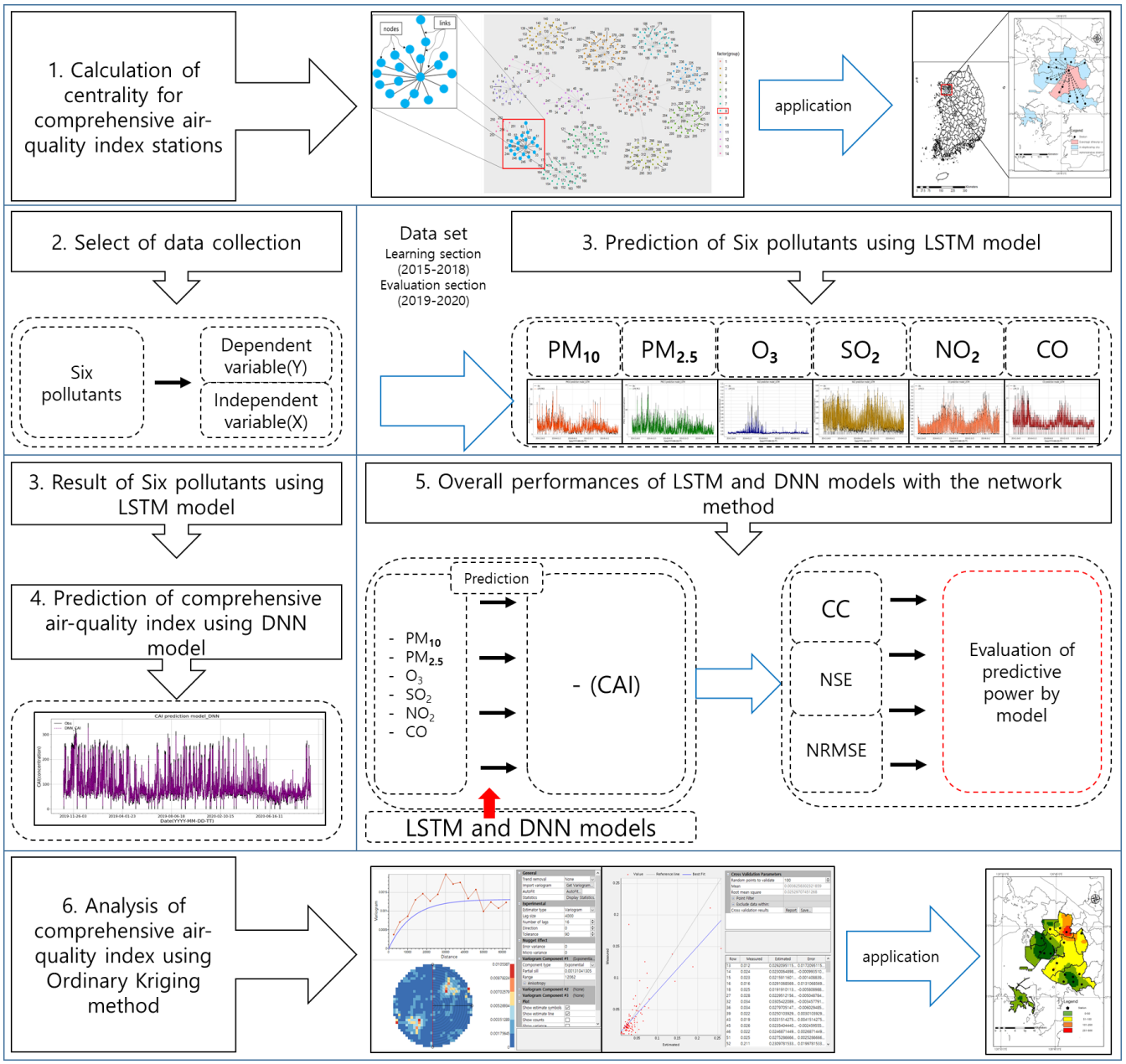


Figure 2. Application of deep learning models and network methods for CAI prediction.

2.3. Data Description

The CAI has six types of air pollution: fine dust (PM_{10}), fine particulate matter ($PM_{2.5}$), ozone (O_3), sulfurous acid gas (SO_2), nitrogen dioxide (NO_2), and carbon monoxide (CO). These six pollutants were used to predict the concentration. The CAI has been forecasted since 2014 to inform about and prepare for air pollution in advance, and air-quality information has been provided nationwide since 2015. Therefore, in this study, six pollutants observed on an hourly basis across the country were collected from 2015 to 2020 to predict the value of the CAI. Table 1 shows the basic statistics of the pollutants.

Table 1. Basic statistics for the variables.

Classification	Max.	Min.	Average	Standard Deviation	Coefficient of Variation
Fine dust (ug/m^3)	1484.00	0.00	357.51	176.49	31,150.37
Fine particulate matter (ug/m^3)	843.00	0.00	101.71	65.40	4277.58
Ozone (ppm)	0.23	0.00	0.13	0.03	0.01
Sulfurous acid gas (ppm)	0.50	0.00	0.03	0.04	0.01
Nitrogen dioxide (ppm)	0.36	0.00	0.09	0.03	0.01
Carbon monoxide (ppm)	11.00	0.00	2.05	0.99	0.99

2.4. Network Method

Network analysis has its theoretical roots in early sociologists, such as Georg Simmel and Émile Durkheim, who wrote about the importance of studying patterns of relationships that connect social actors [36,37]. A network is a very effective method for representing complex and fluid systems [38,39].

The network method consists of links and nodes, as shown in Figure 3. When building a network, the most important factor is the presence of links connected to nodes, which are used to calculate various analysis indicators for the network, such as centrality and the clustering coefficient [40,41].

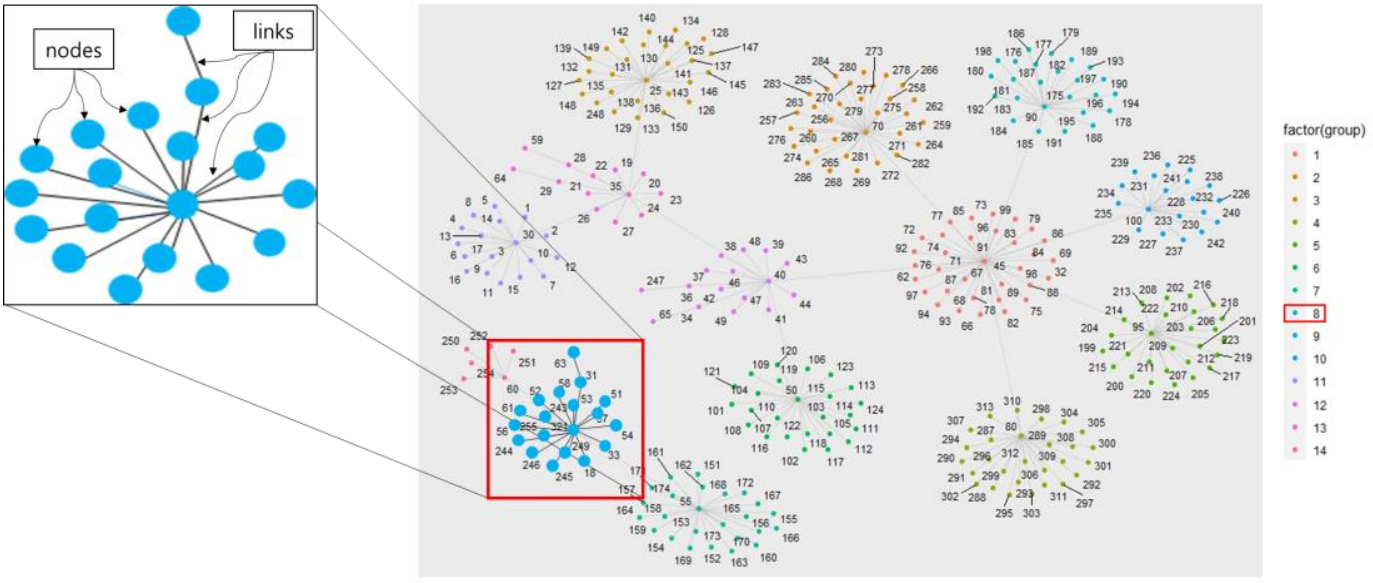


Figure 3. Structure of the network.

This study evaluated the importance of each of the six pollutant stations using degree centrality (DC) among the centrality calculation methodologies. DC is a method for assessing the significance of nodes based on the number of links connected to each node. However, when comparing the DC values for each node with those of other networks, a fair comparison is difficult if the network size differs. Therefore, in this study, the importance

of DC normalization by node was evaluated by dividing it by N-1, the maximum DC value of the network.

$$\text{Degree Centrality (DC)} = \frac{N_c}{N-1}. \quad (1)$$

In Equation (1), N expresses the total number of nodes, and N_C expresses the number of links connected to individual nodes.

2.5. Long Short-Term Memory Model

The long short-term memory model (LSTM) is a type of recurrent neural network that directly learns from time-series data [42,43]. However, although LSTM is a type of recurrent neural network (RNN), it adds a cell state structure to the hidden layer and stores information about the input data for a more extended period, thereby resolving the limitations of the RNN (Figure 4). The compact forms of equations calculate the LSTM cell with a forget gate [44].

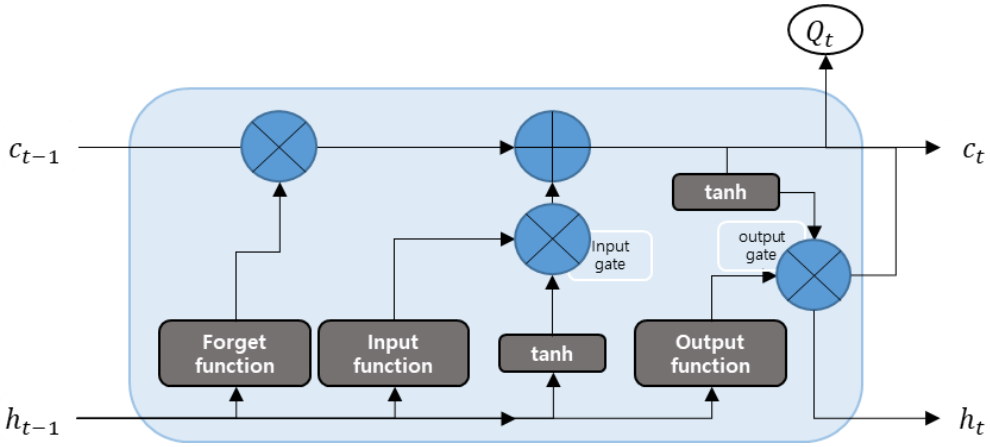


Figure 4. Conceptual diagram of the LSTM model.

The LSTM cell contains three gates that regulate the data flow: a forget gate (f_t), an input gate (i_t), and an output gate (o_t). The forget gate (f_t) performs a calculation to determine which information to discard and applies the h_{t-1} of the previous step and x_t of the current step to the sigmoid function to obtain a value between 0 and 1. Moreover, the gate maintains and adjusts the cell state ($\widetilde{C}_t$) and the hidden state (h_t). The forget gate determines the information to be kept in the cell state (C_{t-1}) from the previous LSTM cell. The LSTM computes the algorithms from an input sequence x_t to output o_t by looping through Equations (2)–(7) with initial values of $C_o = 0$ and $h_o = 0$.

$$f_t = \sigma\left(W_f \times [h_{t-1,}x_t] + b_f\right), \quad (2)$$

$$i_t = \sigma(W_i \times [h_{t-1,}x_t] + b_i, \quad (3)$$

$$\widetilde{C}_t = \tan h(W_C \times [h_{t-1,}x_t] + b_C, \quad (4)$$

$$C_t = f_t \times C_{t-1} + i_t \times \widetilde{C_t}, \quad (5)$$

$$o_t = \sigma(W_o \times [h_{t-1,}x_t] + b_o, \quad (6)$$

$$h_t = o_t \times \tanh(C_t). \quad (7)$$

2.6. Deep Neural Network Model

Deep learning (also known as deep structured learning) is a part of a broader family of machine learning methods based on artificial neural networks (ANNs) with representation learning. A DNN is an ANN with multiple layers between the input and output layers [45]. The structure of the DNN is similar to that of the ANN, but there are differences in the

number of layers in the hidden layer (Figure 5). Other neural networks consist of the same components, such as neurons, synapses, weights, biases, and functions [46]. These components behave similarly to the human brain and can be trained using any machine learning algorithm. DNN architectures generate compositional models in which an object is expressed as a layered composition of primitives [47,48]. The extra layers enable the design of lower layers, potentially modeling complex data with fewer units than a similarly performing shallow network.

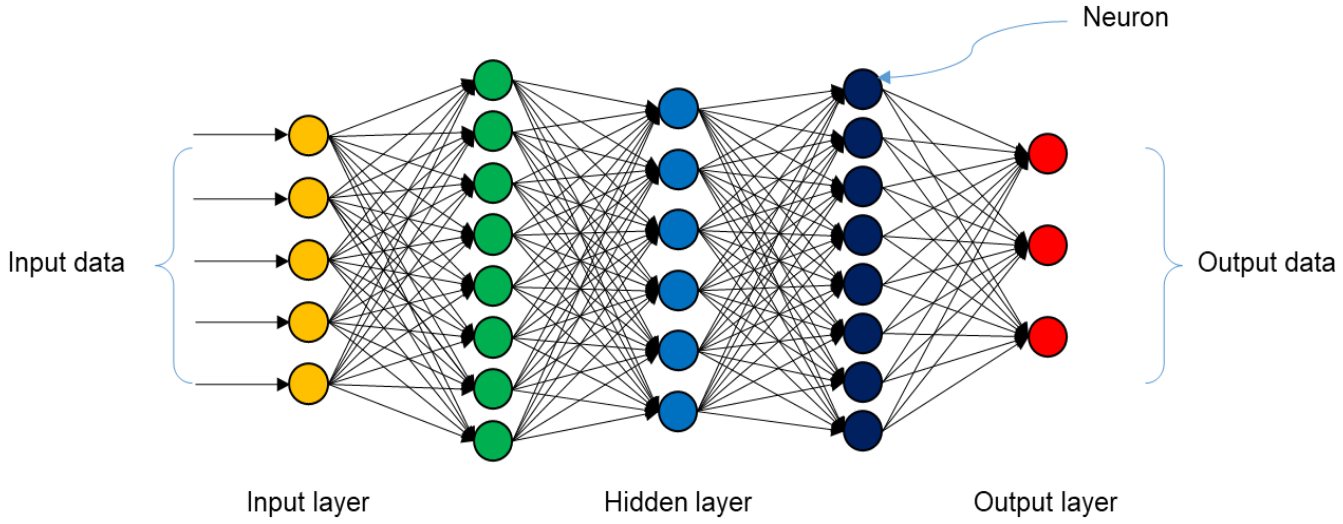


Figure 5. Conceptual diagram of the deep neural network.

2.7. Evaluation Strategy

In this study, the correlation coefficient, the Nash–Sutcliffe efficiency coefficient, and the normalized root mean squared error were used as indicators of predictive power. Correlation coefficients are methods used to analyze the relationship between two variables.

$$\text{correlation coefficient} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}. \quad (8)$$

In Equation (8), x_i is the observed value, $\overline{x}$ is its mean, y_i is the predicted value, and $\overline{y}$ is its mean. That is, x denotes the data from the station, and y denotes the data from the prediction model.

The Nash–Sutcliffe efficiency coefficient means that the predicted result is poor or inconsistent if the value is negative and can be computed as:

$$\text{Nash–Sutcliffe efficiency coefficient} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^2 - \sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}, \quad (9)$$

where y_i is the observed value, $\overline{y}$ is the mean, and $\hat{y}_l$ is the predicted value. A positive value means that using the predicted result will provide better results than using the average of the observations, and a value closer to 1 means an ideal result.

Furthermore, the normalized root mean squared error is the value obtained by dividing the numerator root mean squared error by the range (maximum–minimum) of the denominator, the actual value. The closer it is to 0, the smaller the degree of error.

$$\text{Normalized root mean squared error } (\%) = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{Max(y_i) - Min(y_i)} \times 100. \quad (10)$$

In Equation (10), y_i is the observed value, and $\hat{y}_l$ is the predicted value.

2.8. Ordinary Kriging

Generally, two assumptions are made when using the kriging method. First, the data must have a normal distribution; if they do not, then a variable transformation is performed such that the predictor and dependent variables are linear. In this case, the value estimated by kriging can be easily inversely transformed; however, the variance estimate is not accurate [49]. The following assumption is made: the data should be continuous and stationary. Most spatial data exhibit a specific trend or have a problem in that the mean changes with the location. When kriging is applied to data with strong trends, misspecification of spatial dependence may occur, and the predicted values may be biased [50]. Therefore, the predicted value reflecting the spatial characteristics of the target data can be calculated only when kriging is applied by removing the tendency of the data.

The unknown value $Z(x_0)$ is interpreted as a random variable located in x_0, as well as the values of the neighboring samples $Z(x_i)$, i = 1, ... ,N. The estimator $\hat{Z}(x_0)$ is also interpreted as a random variable located in x_0, which is a result of the linear combination of variables. To deduce the kriging system from the assumptions of the model, the following error committed while estimating Z(x) in x_0 is declared: the two quality criteria previously referred to can now be expressed in terms of the mean and variance of the new random variable, $\epsilon(x_0)$.

$$\epsilon(x_o) = \hat{Z}(x_0) - Z(x_0) = \sum_{i=1}^{N} w_i(x_0) \times Z(x_i) - Z(x_0). \quad (11)$$

3. Results

3.1. Calculation of Centrality for CAI Stations

In most previous studies, all data from the surrounding stations around the study area were taken into account and used as independent variables. If all data from neighboring stations are used without classification, unnecessary data for model learning can lead to the distortion of predictions. [51].

Network analysis was applied to CAI stations in Korea. Each station is marked as a node in the network. The station–station correlation coefficient was expressed as an inter-node link and calculated using the correlation coefficient. Centrality was used to analyze the network. Fourteen groups were formed with a high correlation. It can be said that Jeong-wang Station formed 18 stations. Jeong-wang Station was networked with 18 stations (Figure 6). Therefore, to predict the CAI in South Korea, in this study, a CAI prediction model was developed using Jeong-wang Station and 18 stations.

3.2. Overall Performance of LSTM and DNN Models with the Network Method

The performance of the LSTM model depends on the input data, and it is determined by the parameters constituting the model. The weights and biases for model learning are called parameters. The parameters included activation, time period, optimizer, learning rate, and loss. The modeling performance of DNN models depends on the input data, and optimal values are determined by the hyperparameters constituting the model. The weights and biases that the models learn are called hyperparameters. Hyperparameters include the learning rate, hidden layer, hidden nodes, drop-out, and time period.

Depending on the user, parameters and hyperparameters must be manually set for the best combination for the model. In addition, since the hyperparameter value for each model differs depending on the input and output variables, a specific value cannot be defined as an optimal value. Therefore, it is necessary to derive the optimal hyperparameter value using the random search method. The optimal parameter settings used in this study are listed in Table 2. The rectified linear unit (ReLU) is one of the most frequently used activation functions. It is a function to solve the gradient vanishing problem of sigmoid and tanh. Adaptive moment estimation (Adam) is a method for adaptively adjusting the updated intensity of learning by reducing the learning rate and calculating the rate. The mean squared error (MSE) is the average value after squaring the difference between the predicted value and the actual value.

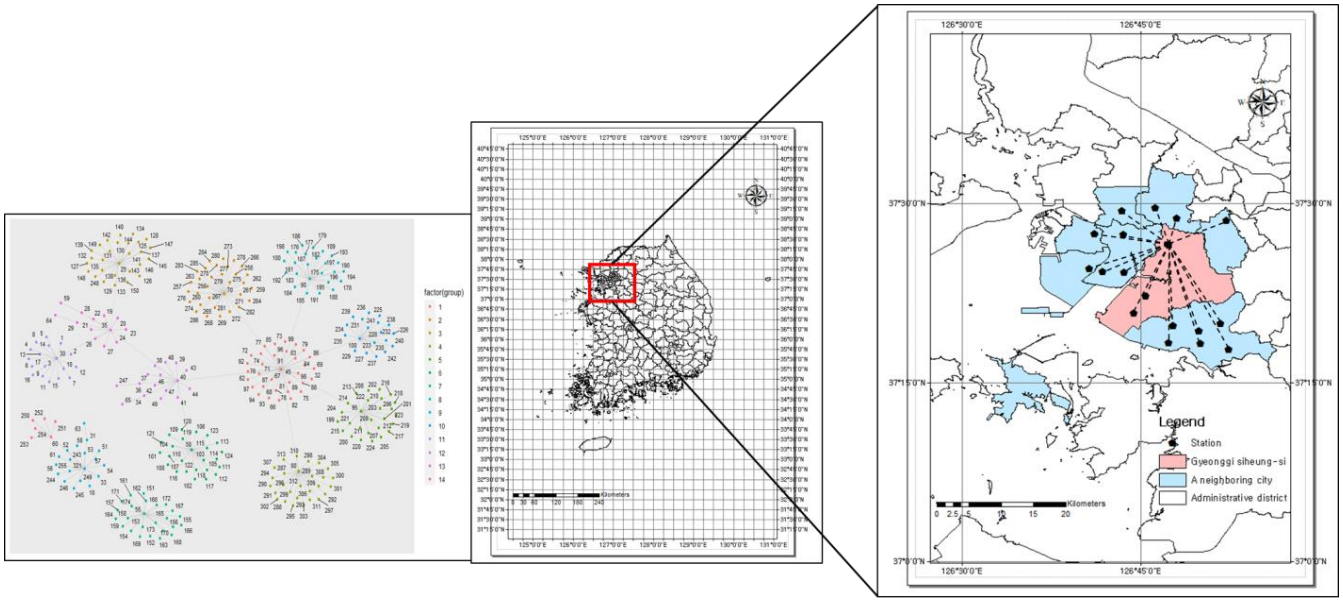


Figure 6. Construction of links in station network. Black pentagonal points represent six pollutant stations located in a neighboring city, and black lines represent six pollutant stations connected in neighboring stations.

Table 2. Setting of parameters in LSTM. Each of the six pollutants from the station is predicted using each of the six pollutants' data as an independent variable (pollutant prediction models). The six pollutants are fine dust (PM_{10}), fine particulate matter ($PM_{2.5}$), ozone (O_3), sulfurous acid gas (SO_2), nitrogen dioxide (NO_2), and carbon monoxide (CO).

Parameter	**Values (PM_{10} Prediction Model (1))**	**Values ($PM_{2.5}$ Prediction Model (2))**	**Values (O_3 Prediction Model (3))**
Activation	ReLU	ReLU	ReLU
Time period	14	11	27
Optimizer	Adam	Adam	Adam
Learning-rate	0.01	0.01	0.01
Loss	MSE	MSE	MSE
Parameter	**Values (SO_2 Prediction Model (4))**	**Values (NO_2 Prediction Model (5))**	**Values (CO Prediction Model (6))**
Activation	ReLU	ReLU	ReLU
Epoch	27	12	13
Optimizer	Adam	Adam	Adam
Learning rate	0.01	0.01	0.01
Loss	MSE	MSE	MSE

The DNN model also needs to derive the optimal hyperparameter values using a random search method. The optimal hyperparameter settings used in this study are listed in Table 3.

Table 3. Setting of hyperparameters in DNN. The CAI of the station is predicted using six pollutants' data as an independent variable (CAI prediction models).

Hyperparameters	Values (CAI Prediction Model (7))	Hyperparameters	Values (CAI Prediction Model (7))
Learning Rate	0.1	Epoch	47
Hidden layer	3	Batch Size	10
Hidden nodes	4	Optimizer	Adam
Drop-out	0.5	Activation	ReLU

The LSTM model provided a high performance for predicting the variability in six pollutants and peak values. The CC values were 0.85–0.90 and the NRMSE values were 0.12–0.19 for the LSTM models. The evaluation metrics indicated that all six models were suitable for pollutant prediction (Table 4 and Figures 7–13).

Table 4. Evaluation of predictive power by the LSTM and DNN models.

Models	Correlation Coefficient	Normalized Root Mean Squared Error (%)	Nash–Sutcliffe Efficiency Coefficient
PM_{10} prediction model (1)	0.88	0.16	0.93
$PM_{2.5}$ prediction model (2)	0.90	0.12	0.95
O_3 prediction model (3)	0.85	0.19	0.91
SO_2 prediction model (4)	0.86	0.18	0.92
NO_2 prediction model (5)	0.88	0.16	0.93
CO prediction model (6)	0.90	0.12	0.95
CAI prediction model (7)	0.91	0.10	0.96

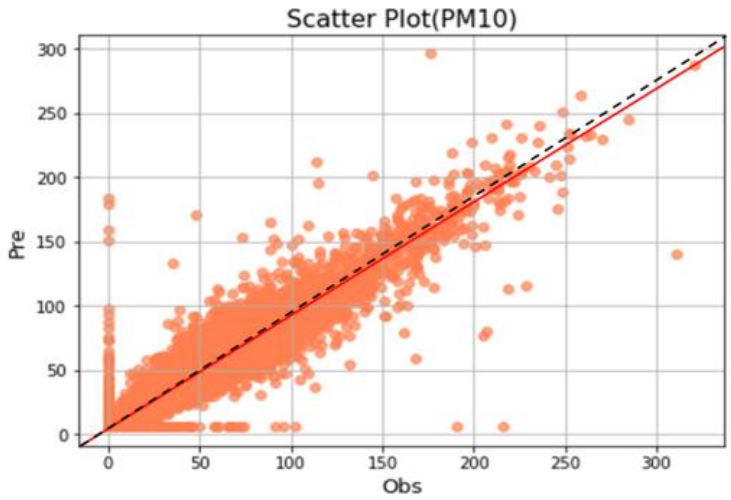


Figure 7. Comparison results of predicted fine dust with observations. The regression line of observed and predicted data is indicated in red.

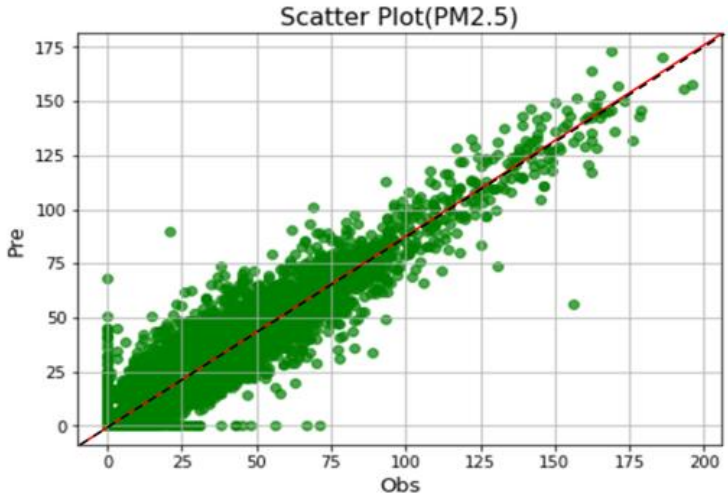


Figure 8. Comparison results of predicted fine particulate matter with observations. The regression line of observed and predicted data is indicated in red.

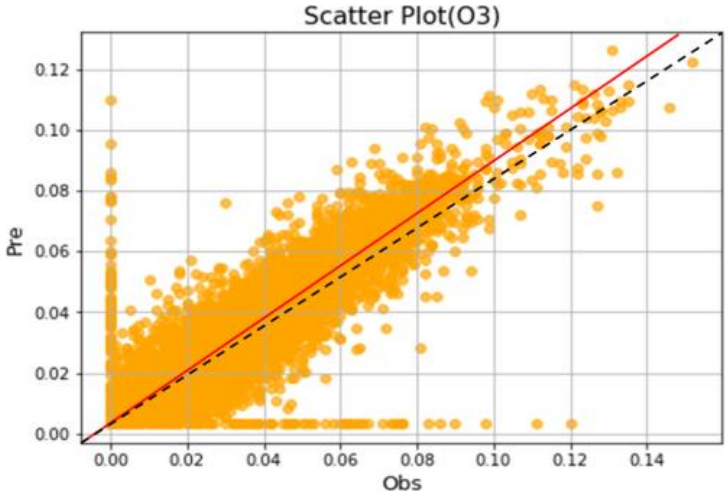


Figure 9. Comparison results of predicted ozone with observations. The regression line of observed and predicted data is indicated in red.

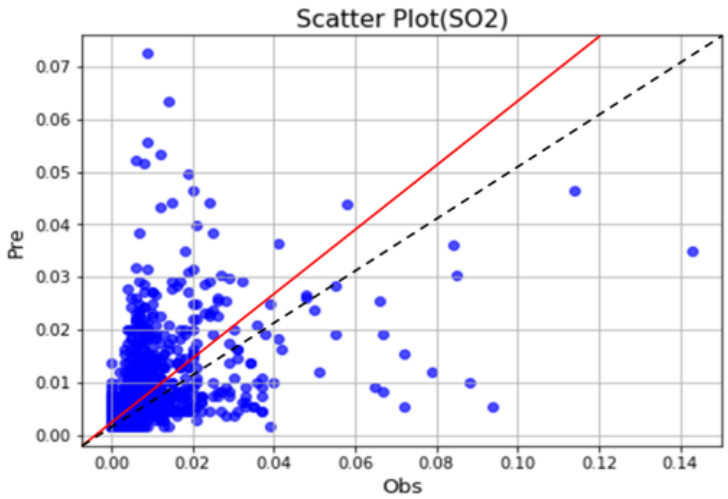


Figure 10. Comparison results of predicted sulfurous acid gas with observations. The regression line of observed and predicted data is indicated in red.

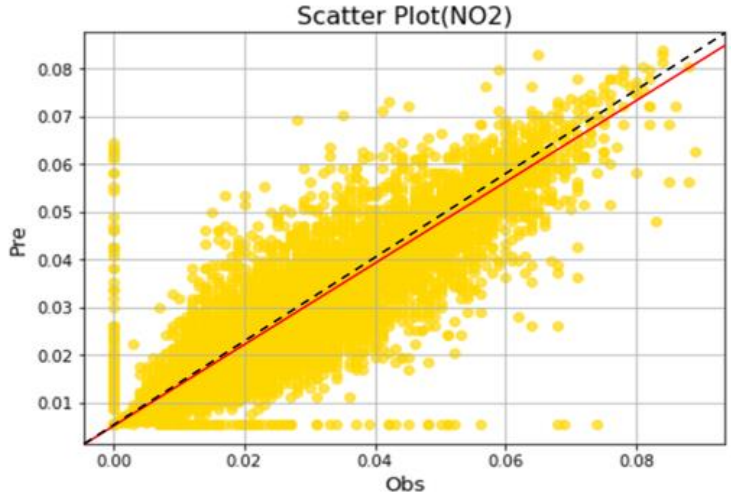


Figure 11. Comparison results of predicted nitrogen dioxide with observations. The regression line of observed and predicted data is indicated in red.

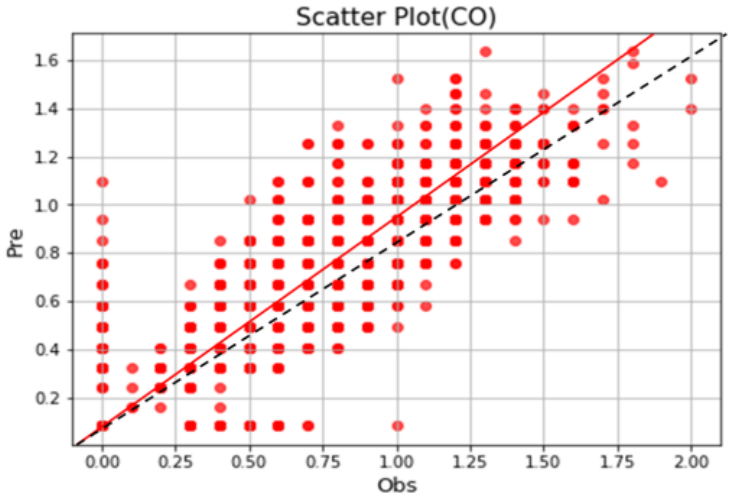


Figure 12. Comparison results of predicted carbon monoxide with observations. The regression line of observed and predicted data is indicated in red.

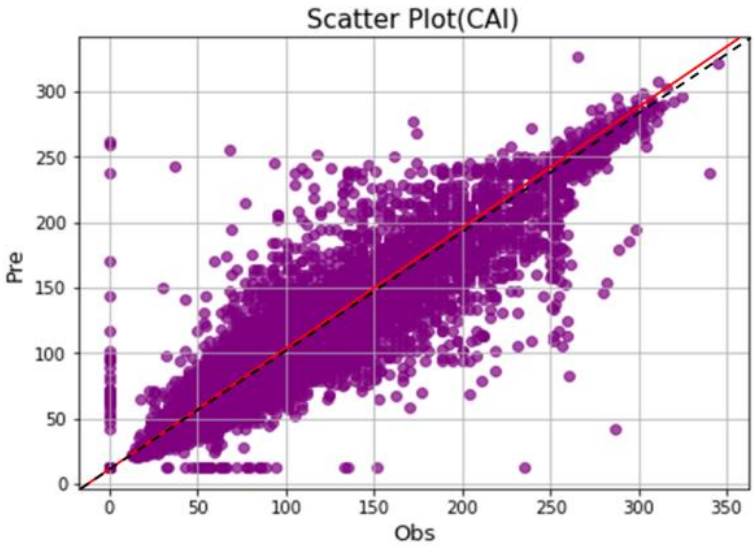


Figure 13. Comparison results of predicted comprehensive air-quality index with observations. The regression line of observed and predicted data is indicated in red.

The DNN model can predict the dependent variable by using several independent variables. Therefore, we predicted the CAI using the six predicted pollutants. In other words, Jeong-wang Station was used as a dependent variable, and data from 18 neighboring stations were used. The DNN model provided high performance in predicting hourly CAI variability and the peak value of the observed comprehensive air-quality index. The CC value was greater than 0.91, and the NRMSE value was 0.10 for the DNN model. The evaluation strategy metrics showed that a CAI model was suitable for prediction.

Scatter plots were used to observe the relationships between the variables. A comparison was performed to investigate the agreement between the observed and the re-analysis indices. Figure 7 shows the scatter plots of the observed and predicted values. The results predicted by the models developed with LSTM and DNN and the results of the comparison with observation were generally consistent.

3.3. Analysis of CAI Using Ordinary Kriging Method

In this study, the ordinary kriging method was used to determine the spatial distribution of the predicted CAI over South Korea. A variogram is expected to square the difference between two points separated by a lag. The variogram measures the similarity of data at a certain distance and decreases as the distance between two points decreases. As the distance increases, the variogram increases or the correlation is lost and, thus, has a meaningless value. Therefore, the isolation distance at which the correlation and tendency are lost in the variogram is defined as a range. The variogram value at that distance is defined as a sill. There are empirical variograms and theoretical variograms, and a theoretical variogram that minimizes prediction error was selected for considering spatial interrelationships. In the case of ordinary kriging, the sill converges to the variance of the entire dataset. The sill was set as the variance of all the data, and kriging was applied using an exponential variogram.

The work configuration for using the Surfer software is as follows: The coordinate system was set as Korea 2000/Unified CS-EPSG:5179. The cell size was set to 500 m × 500 m; the raster size was set to the X-axis (from 746,000 m to 1,394,500 m) and Y-axis (from 1,458,500 m to 2,068,500 m); the lag size was 15,500 m; and the number of lags was set to 16 (Figure 14). The Surfer software was used to calculate the range of the variogram model. The variogram is the expected value of the square of the difference in data values between two points separated by a separation distance (h). The variogram appears smaller as the distance between two points increases, and as the distance increases, the variogram increases or the correlation is lost and thus has a meaningless value. Therefore, the separation distance at which the correlation and tendency are lost in the variogram is called the range, and the variogram value at that distance is defined as the sill. In ordinary kriging, the threshold theoretically converges to the variance of the entire dataset. Based on the above theory, the sill was set as the variance of the entire dataset in Surfer, and kriging was applied to the CAI using an exponential variogram.

If a prediction is made using data from a point station, there is a disadvantage in that only a value located at one point can be predicted. However, if ordinary kriging is used, the ground station data are spatially smoothed, and the local area can be predicted. Therefore, a comprehensive air-quality index was predicted using the six predicted pollution data points in this study. We propose a methodology that enables the spatial distribution of predicted data using ordinary kriging to identify predicted values in the spatial distribution of predicted CAI in South Korea (Figure 15).

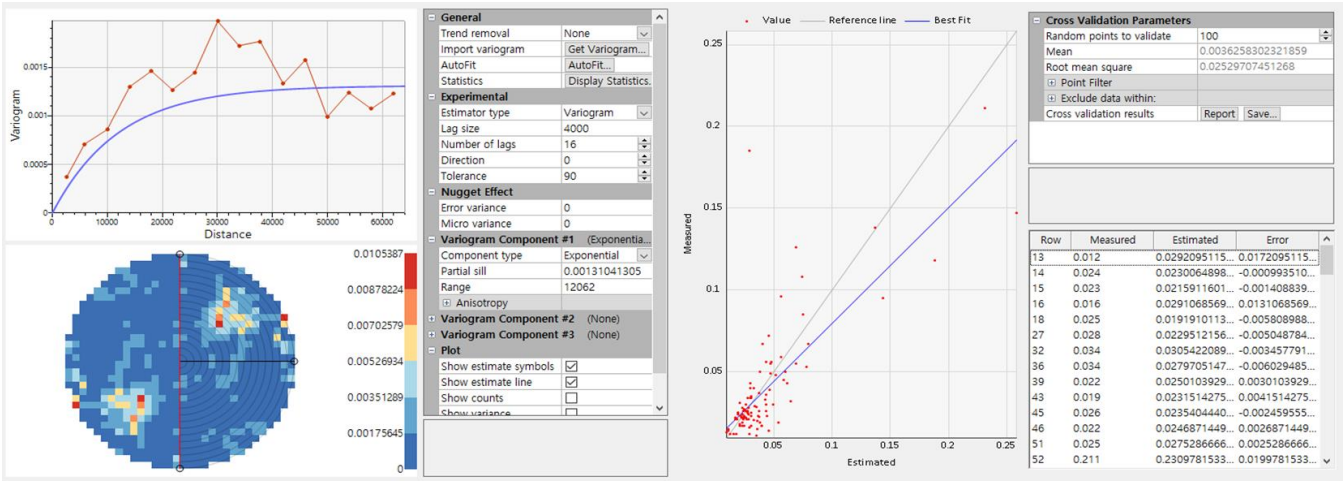


Figure 14. Parameters that constitute the exponential variogram in the Surfer software. Empirical variogram: a set of variogram values for separation distances at regular intervals (actual variogram) → (red dot). Theoretical variogram: an expression of the empirical variogram (modeled variogram) → (blue line).

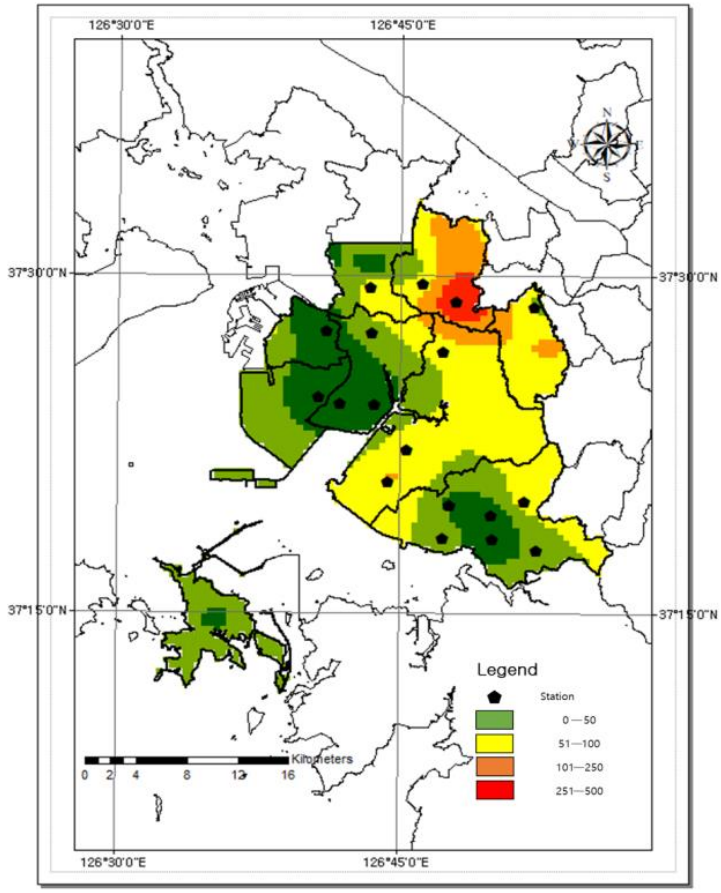


Figure 15. Spatial distribution of comprehensive air-quality index provided from ordinary kriging. Black points represent the location of air-quality stations, and green, yellow, and orange represent negligible, minor, and catastrophic air quality, respectively.

4. Discussion

The most challenging aspect of developing a CAI prediction model is reliable data collection. One of the limitations of this study is that only six years of comprehensive

air-quality indices and six pollutants' data were used as dependent and independent variables. More diverse observational datasets allow for systematically developing more predictive functions.

It is necessary to establish air management reduction measures by recognizing the importance of a CAI. It is necessary to establish air pollutant standards and reduction measures while considering the interaction of each air pollutant for the management of not only single substances but also pollutant composition and secondary substances generated in the atmosphere. More diverse studies and trials are needed for air environment standards, and accurate results should be obtained through a comparative analysis of each single pollutant and new integrated air environment standards.

The purpose of this study is to perform disaster management in advance with regard to damage to Korea's large-scale air-quality index. As the need for disaster management gradually increases, it is believed that it can be expanded even to countries without a CAI disaster response system.

5. Conclusions

This study proposes network deep learning models for a CAI prediction approach in South Korea to evaluate model performance. Based on the predicted results, the spatial distribution was determined using ordinary kriging. The results are as follows.

Unnecessary data for model learning using all data from neighboring stations can cause distortions in predictions. Therefore, in this study, the importance of stations was evaluated based on the network method. Fourteen groups were formed, and each group was identified as having a high correlation. Jeong-wang Station formed a network based on a total of 18 stations.

The LSTM models provided good performance in predicting the time variability and peak values of the six observed pollutants. The CC, NSE, and NRMSE values were 0.85–0.90, 0.91–0.95, and 0.12–0.19 for the LSTM models. The evaluation strategy indicators showed that the LSTM model was good at predicting the six pollutants. A DNN can predict the dependent variable using several independent variables. In this manner, the station can correct the missing values and outliers. The DNN model provided high performance in predicting hourly CAI variability and the peak value of the observed CAI. The CC value was greater than 0.91, the NSE value was 0.96, and the NRMSE value was 0.10 for the DNN model.

The ordinary kriging method is used to predict an observation at an unobserved location based on spatially related observed points. A CAI was spatially distributed in the study area through kriging. In other words, spatial distribution can compensate for the shortcomings of weather forecasts issued in administrative district units.

Author Contributions: Conceptualization, D.K. and H.S.K.; formal analysis, D.K.; methodology, H.H., Y.K., H.L., and W.W.; supervision, H.S.K. and H.H.; writing—original draft, D.K.; writing—review and editing, D.K. and H.H. All authors have read and agreed to the published version of the manuscript.

Funding: Ministry of Interior and Safety, Korea: 2021-MOIS36-002.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are available upon request from the corresponding author.

Acknowledgments: This research was supported by a grant (2021-MOIS36-002) from the Technology Development Program on Disaster Restoration Capacity Building and Strengthening funded by the Ministry of Interior and Safety (MOIS, Korea).

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

References

1. World Health Organization. *Evolution of WHO Air Quality Guidelines: Past, Present and Future*; World Health Organization: Geneva, Switzerlands, 2017.
2. Bollen, J.; Guay, B.; Jamet, S.; Corfee-Morlot, J. *Co-Benefits of Climate Change Mitigation Policies: Literature Review and New Results*; OCED: Paris, France, 2009. [CrossRef]
3. Lanzi, E. The Economic Consequences of Outdoor Air Pollution. Organization for Economic Cooperation and Development. 2016. Available online: https://www.oecd.org/environment/indicators-modelling-outlooks/Policy-Highlights-Economic-consequences-of-outdoor-air-pollution-web.pdf.
4. Shin, J. New Comprehensive Air-Quality Index (NCAI) and Its Effects on Respiratory and Cardiovascular Diseases. *J. Environ. Policy Adm.* **2020**, *28*, 113–157.
5. Mosavi, A.; Ozturk, P.; Chau, K.W. Flood prediction using machine learning models: Literature review. *Water* **2018**, *10*, 1536. [CrossRef]
6. Xingjian, S.H.I.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Proceedings of the Advances in Neural Information Processing Systems 28 (NIPS 2015), Montreal, QC, Canada, 7–12 December 2015; pp. 802–810.
7. Kratzert, F.; Klotz, D.; Brenner, C.; Schulz, K.; Herrnegger, M. Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 6005–6022. [CrossRef]
8. Poornima, S.; Pushpalatha, M. Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network. *Soft Comput.* **2019**, *23*, 8399–8412. [CrossRef]
9. Xiang, Z.; Yan, J.; Demir, I. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Resour. Res.* **2020**, *56*, e2019WR025326. [CrossRef]
10. Han, H.; Choi, C.; Jung, J.; Kim, H.S. Deep Learning with Long Short Term Memory Based Sequence-to-Sequence Model for Rainfall-Runoff Simulation. *Water* **2021**, *13*, 437. [CrossRef]
11. Wu, X.; Zhou, J.; Yu, H.; Liu, D.; Xie, K.; Chen, Y.; Hu, J.B.; Sun, H.Y.; Xing, F. The Development of a Hybrid Wavelet-ARIMA-LSTM Model for Precipitation Amounts and Drought Analysis. *Atmosphere* **2021**, *12*, 74. [CrossRef]
12. Russo, A.; Raischel, F.; Lind, P.G. Air quality prediction using optimal neural networks with stochastic variables. *Atmos. Environ.* **2013**, *79*, 822–830. [CrossRef]
13. Singh, K.P.; Gupta, S.; Rai, P. Identifying pollution sources and predicting urban air quality using ensemble learning methods. *Atmos. Environ.* **2013**, *80*, 426–437. [CrossRef]
14. Qi, Y.; Li, Q.; Karimian, H.; Liu, D. A hybrid model for spatiotemporal forecasting of PM2. 5 based on graph convolutional neural network and long short-term memory. *Sci. Total Environ.* **2019**, *664*, 1–10. [CrossRef]
15. Torrisi, M.; Pollastri, G.; Le, Q. Deep learning methods in protein structure prediction. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 1301–1310. [CrossRef] [PubMed]
16. Bouchaud, J.P.; Mézard, M. Wealth condensation in a simple model of economy. *Phys. A: Stat. Mech. Its Appl.* **2000**, *282*, 536–545. [CrossRef]
17. Newman, M.E. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 404–409. [CrossRef] [PubMed]
18. Liljeros, F.; Edling, C.R.; Amaral LA, N.; Stanley, H.E.; Åberg, Y. The web of human sexual contacts. *Nature* **2001**, *411*, 907–908. [CrossRef]
19. Tsonis, A.A.; Roebber, P.J. The architecture of the climate network. *Phys. A Stat. Mech. Its Appl.* **2004**, *333*, 497–504. [CrossRef]
20. Davis, K.F.; D'Odorico, P.; Laio, F.; Ridolfi, L. Global spatio-temporal patterns in human migration: A complex network perspective. *PLoS ONE* **2013**, *8*, e53723. [CrossRef]
21. Simpson, T.W.; Mauery, T.M.; Korte, J.J.; Mistree, F. Kriging models for global approximation in simulation-based multidisciplinary design optimization. *AIAA J.* **2001**, *39*, 2233–2241. [CrossRef]
22. Webster, R.; Oliver, M.A. *Geostatistics for Environmental Scientists*; John Wiley & Sons: Hoboken, NJ, USA, 2007.
23. Lim, W.S.; Lee, K.H.; Kyung, M.S.; Kim, H.S. Potential Risk of Flood Damage and Estimation of Design Frequency in Small River Basins. *J. Korean Soc. Civ. Eng.* **2007**, *27*, 631–640.
24. Park, N.W.; Jang, D.H. Mapping of temperature and rainfall using DEM and multivariate kriging. *J. Korean Geogr. Soc.* **2008**, *43*, 1002–1015.
25. Shin, H.; Chang, E.; Hong, S. Estimation of near surface air temperature using MODIS land surface temperature data and geostatistics. *Spat. Inf. Res.* **2014**, *22*, 55–63. [CrossRef]
26. Hevesi, J.A.; Istok, J.D.; Flint, A.L. Precipitation estimation in mountainous terrain using multivariate geostatistics. Part I: Structural analysis. *J. Appl. Meteorol. Climatol.* **1992**, *31*, 661–676. [CrossRef]
27. Goovaerts, P. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.* **2000**, *228*, 113–129. [CrossRef]
28. Yu, C.S.; Jeong, G.S. Estimation of area average rainfall amount and its error. *J. Korea Water Resour. Assoc.* **2001**, *34*, 317–326.
29. Lee, J.H.; Yu, Y.G. Optimal network design for the estimation of areal rainfall. *J. Korea Water Resour. Assoc.* **2002**, *35*, 187–194. [CrossRef]

30. Yoon, K.H.; Seo, B.C.; Shin, H.S. Spatial analysis of flood rainfall based on kriging technique in Nakdong river basin. *J. Korea Water Resour. Assoc.* **2004**, *37*, 233–240. [CrossRef]
31. Cho, H.L.; Jeong, J.C. Application of spatial interpolation to rainfall data. *Spat. Inf. Res.* **2006**, *14*, 29–41.
32. Jung, J.Y.; Jin, S.H.; Park, M.S. Precipitation analysis based on spatial linear regression model. *Korean J. Appl. Stat.* **2008**, *21*, 1093–1107. [CrossRef]
33. Heo, T.Y.; Park, M.S. Bayesian spatial modeling of precipitation data. *Korean J. Appl. Stat.* **2009**, *22*, 425–433. [CrossRef]
34. Park, M.; Park, C.; Shin, K.I.; Yoo, C. On proper variograms of daily rainfall data. *J. Korean Soc. Civ. Eng.* **2010**, *30*, 525–532.
35. Kim, K.H.; Kim, M.S.; Lee, G.W.; Kang, D.H.; Kwon, B.H. The adjustment of radar precipitation estimation based on the kriging method. *J. Korean Earth Sci. Soc.* **2013**, *34*, 13–27. [CrossRef]
36. Freeman, L. The development of social network analysis. *A Study Sociol. Sci.* **2004**, *1*, 159–167.
37. Paradowski, M.B.; Jarynowski, A.; Jelińska, M.; Czopek, K. Selected poster presentations from the American Association of Applied Linguistics conference, Denver, USA, March 2020: Out-of-class peer interactions matter for second language acquisition during short-term overseas sojourns: The contributions of Social Network Analysis. *Lang. Teach.* **2012**, *54*, 139–143. [CrossRef]
38. Kim, K.; Joo, H.; Han, D.; Kim, S.; Lee, T.; Kim, H.S. On complex network construction of rain gauge stations considering nonlinearity of observed daily rainfall data. *Water* **2019**, *11*, 1578. [CrossRef]
39. Joo, H.; Kim, H.S.; Kim, S.; Sivakumar, B. Complex networks and integrated centrality measure to assess the importance of streamflow stations in a River basin. *J. Hydrol.* **2021**, *598*, 126280. [CrossRef]
40. Estrada, E. *The Structure of Complex Networks: Theory and Applications*; Oxford University Press: Oxford, UK, 2012.
41. Newman, M. *Networks*; Oxford University Press: Oxford, UK, 2018.
42. Hu, C.; Wu, Q.; Li, H.; Jian, S.; Li, N.; Lou, Z. Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. *Water* **2018**, *10*, 1543. [CrossRef]
43. Fan, H.; Jiang, M.; Xu, L.; Zhu, H.; Cheng, J.; Jiang, J. Comparison of long short term memory networks and the hydrological model in runoff simulation. *Water* **2020**, *12*, 175. [CrossRef]
44. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. *Neural Comput.* **2000**, *12*, 2451–2471. [CrossRef]
45. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef]
46. Bengio, Y. *Learning Deep Architectures for AI*; Foundations and Trends®in Machine Learning; Now Publishers Inc.: Lange Geer, The Netherlands, 2009; Volume 2, pp. 1–127.
47. Fernandez, A.G.M.L.S.; Bunke, R.B.H.; Schmiduber, J. A novel connectionist system for improved unconstrained handwriting recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 855–868.
48. Szegedy, C.; Toshev, A.; Erhan, D. Deep Neural Networks for Object Detection. In Proceedings of the Advances in Neural Information Processing Systems 26 (NIPS 2013), Lake Tahoe, NV, USA, 5–10 December 2013.
49. Hengl, T.; Heuvelink, G.B.; Stein, A. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* **2004**, *120*, 75–93. [CrossRef]
50. Karl, J.W. Spatial predictions of cover attributes of rangeland ecosystems using regression kriging and remote sensing. *Rangel. Ecol. Manag.* **2010**, *63*, 335–349. [CrossRef]
51. Kim, D.; Han, H.; Wang, W.; Kim, H.S. Improvement of Deep Learning Models for River Water Level Prediction Using Complex Network Method. *Water* **2022**, *14*, 466. [CrossRef]