Concentration Separation Prediction Model to Enhance Prediction Accuracy of Particulate Matter

[1]Yong-jin Jung & [1]Chang-heon Oh
[1]Department of Electrical, Electronics, and Communication Engineering, Korea University of Technology and Education(KOREATECH), Korea

jungyj0211@koreatech.ac.kr; *choh@koreatech.ac.kr

ABSTRACT

Demand for more accurate particulate matter forecasts is accumulating owing to the increased interest and issues regarding particulate matter. Incredibly low concentration particulate matter, which accounts for most of the overall particulate matter, is often underestimated when a particulate matter prediction model based on machine learning is used. This study proposed a concentration-specific separation prediction model to overcome this shortcoming. Three prediction models based on Deep Neural Network (DNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM), commonly used for performance evaluation of the proposed prediction model, were used as comparative models. Root mean squared error (RMSE), mean absolute percentage error (MAPE), and accuracy were utilized for performance evaluation. The results showed that the prediction accuracy for all Air Quality Index (AQI) segments was more than 80 percent in the entire concentration spectrum. In addition, the

study confirmed that the over-prediction phenomenon of single neural network models concentrated in the 'normal' AQI region was alleviated.

Keywords: DNN, RNN, LSTM, particulate matter.

INTRODUCTION

Particulate matter comprises particles of various sizes, shapes, and components. Tiny particulate matter can penetrate the respiratory system and cause severe effects, especially if they contain heavy metals. Pope III (2006), Valavanidis et al. (2008), and Anderson et al. (2012) investigated the drastic effects of particulate matter on the human body. The International Agency for Research on Cancer (IARC) has designated particulate matter as having the same carcinogenic level as asbestos, while the World Health Organization (WHO) (2013) has classified particulate matter as a Group 1 carcinogen. In addition, the Organization for Economic Co-operation and Development (OECD) (2016) has reported that premature death from outdoor particulate matter and ozone is the highest in OECD countries, at 1,109 per million people (Jo et al., 2018). The Korean government has designated particulate matter as a social disaster, and a special law on reducing and managing particulate matter has been in effect since February 2019.

Many people know the dangers of particulate matter and check relevant forecasts to decide on leaving their houses or wearing a mask before leaving. Therefore, the need for highly accurate particulate matter forecasts. Particulate matter forecasts rely on the Air Quality Index (AQI), which is classified into four levels depending on particulate matter concentration: 'good', 'moderate', 'bad', and 'very bad'. A report issued by the Board of Audit and Inspection of Korea (BAI) (2017) on operation conditions of the weather forecast and earthquake alert systems showed that forecast accuracy of particulate matter above the 'bad' level is approximately 60 percent, which does not satisfy the public expectations. Incredibly low concentration particulate matter, which accounts for most of the overall particulate matter, is often underestimated when a particulate matter prediction model based on machine learning is used. Thus, research is underway to improve particulate matter forecast accuracy based on weather and air pollution data via machine learning and deep learning.

The current research team sought to enhance the accuracy of particulate matter prediction based on a study conducted in 2021 (Jung & Oh, 2021). Particulate matter concentration has the characteristic of a non-uniform distribution. The prevalence of the concentrations corresponding to 'bad' and 'very bad' in AQI is significantly less than that corresponding to 'good' or 'moderate'. If such prevalence frequency characteristics are learned using deep learning, problems can arise due to the imbalance of learning volume. Therefore, the prediction model is designed by implementing a section containing the AQI 'good' and 'moderate' particulate matter concentrations (low concentration: below 81 μg/m³) and another with the AQI 'bad' and 'very bad' concentrations (high concentration: above 81 μg/m³). The prediction model includes a classification function that distinguishes between low and high concentrations. Deep Neural Network (DNN) was used for the proposed prediction model. Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM), which underwent learning without distinguishing low and high concentrations, were utilized to compare the performance of the particulate matter concentration prediction. Performance comparison and analysis were conducted through the prediction results of the three prediction models. The performance evaluation proceeded through the prediction accuracy corresponding to the entire spectrum and the low and high concentration areas.

RELATED STUDIES

Particulate matter is either naturally occurring or artificial. Natural occurrences include dust and pollen, while artificial occurrences include fumes from burning fossil fuels (e.g., coal and oil), exhaust gas, and dust from industrial sites. Artificial particulate matter can be further classified into primary and secondary. Primary particulate matter is directly discharged from smoke in incineration plants, exhaust gas, and industrial sites. Secondary particulate matter is released as gas from a source and then converted to particulate matter through chemical reactions with other substances in the air. Particulate matter is defined by diameter for air quality regulatory purposes. PM_{10} is a particulate matter of 10 micrometers or less in diameter, whereas $PM_{2.5}$ is a particulate matter of 2.5 micrometers or less in diameter. Bae (2016) and Han et al. (2017) analyzed sulfur dioxide (SO_2), nitrogen oxide (NO_x), and ammonia (NH_3) as air pollutants that

affect the chemical reactions involved in the generation of secondary particulate matter. Furthermore, Jeon et al. (2016) found that SO_2 and nitrogen dioxide (NO_2) produced the most particulate matter among the substances that caused secondary particulate matter, and ozone (O_3) also needed to be managed as such a substance.

Related studies have confirmed that air pollutants and meteorological elements are the primary factors affecting changes in particulate matter concentrations. Analyses of changes in the particulate matter concentrations due to meteorological elements have been conducted. Shin et al. (2007) analyzed how particulate matter concentrations gradually decreased as the wind speed increased and how the concentrations increased along with humidity. Zhou (2014) scrutinized changes in particulate matter concentrations with humidity, wind speed, temperature, and rainfall, and confirmed that meteorological elements affected changes in particulate matter concentrations. Based on this information, data on air pollutants and meteorological elements were used as learning data for the proposed particulate matter concentration prediction model.

While studying the particulate matter prediction model using deep learning, Cha and Kim (2018) designed a model for predicting particulate matter concentrations using air pollutant data collected for four years. The DNN and K-Nearest Neighbor (K-NN) algorithms were applied, and the improvement in prediction performance was confirmed through comparisons with general DNN models. The study conducted by Jeon and Sook (2018) proceeded with the prediction by classifying the particulate matter concentrations into four categories based on AQI standards. Thus, data on air pollutants, meteorological elements, particulate matter concentrations from China, and seasonal variables were used. The DNN model with 200 nodes and three hidden layers was confirmed to entail higher performance in predicting high concentration particulate matter than other comparative models. Choi et al. (2022) conducted a predictive study of $PM_{2.5}$ using DNN. Data collected from 2016 to 2020 (SO_2, CO, NO_2, and PM_{10}) were used. Monthly modeling was proposed to improve the prediction performance, and it was used for performance analysis using root mean squared error (RMSE). The proposed model confirmed that the error was reduced by about 46 percent compared to the comparison model. Bihter et al. (2022) studied the predictions of PM_{10} and SO_2 in 2022. The algorithms of the prediction model performed performance

comparisons using DNN, RNN, and LSTM. Mean squared error (MSE), mean absolute error (MAE), RMSE, and R-squared regression analysis were employed to compare the predictive performance of the model. The LSTM model showed higher performance than the other models, and it was confirmed that the prediction results were the most similar to the actual value.

Lim et al. (2019) conducted their research using the RNN algorithm. The length of the input data, optimization function, and the number of layers and nodes of the RNN-based model were altered to design a more enhanced prediction model through optimal parameter settings. Zhao et al. (2018) utilized the six Individual Air Quality Indexes (IAQI) provided by the Environmental Protecting Agency (EPA) to design an RNN-based particulate matter prediction model. This model utilized SO_2, NO_2, carbon monoxide (CO), and O_3 as the learning data and was confirmed to have enhanced performances compared to a typical prediction model. Ong et al. (2014) investigated the dynamic pre-training of RNN to predict particulate matter data for monitoring. Time series training data were dynamically learned during learning in the RNN models, and wind speed, wind direction, temperature, light intensity, and humidity were used as parameters. A higher prediction capability of particulate matter was confirmed compared to existing auto-encoder methods, which involved learning through backpropagation without training data. An LSTM algorithm can be effectively used to solve the vanishing gradient issue of RNN. This advantage initiated various studies considering the time series characteristics of particulate matter concentration.

Ma et al. (2019) proposed an LSTM model to predict particulate matter data in areas where no monitoring stations that measure particulate matter are present. Bi-directional Long Short-Term Memory (BiLSTM), which uses past, present, and future data, was applied to enhance the performance of particulate matter data. Furthermore, the correlation between air pollution and space was considered, and the spatial distribution of particulate matter data was interpolated through the inverse distance weighted (IDW). The subsequent comparison of prediction performance with Support Vector Regression (SVR), Gradient Boosting Decision Tree (GBDT), Artificial Neural Network (ANN), RNN, and the existing LSTM confirmed that the accuracy of particulate matter prediction was enhanced in the proposed model. Jiao et al. (2019) designed an environmental quality model based on

LSTM to predict particulate matter concentrations. The said model was structured with ten hidden layers. the study confirmed the improvement in prediction efficiency using nine parameters (SO_2, NO_2, CO, O_3, maximum temperature, minimum temperature, wind direction, PM_{10}, and $PM_{2.5}$) in the LSTM model.

Beck et al. (2022) used Kalman, LPF, Savitzky-Golay, and Moving Average Filters to correct noise in time series data. The study established an excellent correction rate of the Savitzky-Golay Filter and Moving Average Filter. Ipek et al. (2022) conducted a study to predict the concentration of PM_{10} per hour using LSTM via the data pre-processing and feature selection (DPFS) process to improve its prediction accuracy. The LSTM model learned six data (SO_2, O_3, CO, relative humidity, wind direction, and wind speed). RMSE and MAE were used for performance evaluation, and overall, it was confirmed that the performance of the proposed model significantly improved. Kim et al. (2022) used the LSTM and DNN models to predict the Comprehensive Air-Quality Index (CAI). Moreover, network techniques were applied to improve the performance of the model. Network techniques are used in various fields, such as computer science and climatology. Data collected from 2016 to 2020 (PM_{10}, $PM_{2.5}$, O_3, SO_2, NO_2, and CO) were used.

The correlation coefficient, Nash–Sutcliffe efficiency coefficient, and RMSE were utilized to evaluate the performance of the model. The results of comparing the performance of the models confirmed that the predictive performance of the DNN model applied with the network technique was higher. Zihan and Zhe (2022) proposed a Bayesian Optimized CNN-RNN (BO-CNN-RNN) hybrid model for the accurate prediction of air pollution. Data collected for five years from 2013 ($PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, and O_3) were used. The study utilized RMSE and MAPE to evaluate the model. It showed RMSE of 10.3 and MAPE of 8.39 percent, and performed better than the LSTM and CNN models used for comparison. Kristiani et al. (2021) conducted a study of the $PM_{2.5}$ prediction model using LSTM combined with the statistical method. Air pollution data collected from 2014 to 2018 were used. Five prediction models were designed based on the data used for learning the LSTM prediction model. RMSE was applied to evaluate the performance of the prediction model. Among the prediction models, the models using PM_{10}, SO_2, and NO_2 for learning showed the lowest RMSE value. It was confirmed that PM_{10}, SO_2, and NO_2 were the main variables in predicting $PM_{2.5}$.

DATA COLLECTION AND COMPOSITION

Air pollutants and meteorological elements data collected from three measurement stations in Cheonan, South Korea, were used as the prediction model's learning and testing data. The data for each attribute were taken at hourly intervals over ten years from 2009 (Table 1). However, data were not collected for some hours due to maintenance of the measuring stations and other external factors.

If data were unmeasured at all three stations, data corresponding to the unmeasured duration were excluded from the learning data. Air pollutant data were reconstructed using the average values of the variables measured simultaneously to decrease the effect of unmeasured data. In addition, the data were rearranged so that the particulate matter concentration of the next hour could be predicted from the previous hour. Table 2 shows the data structure configured to apply collected data to the prediction model.

Table 1

Collected Data

Category	Variable	Number of data	Number of missing data
	O_3	255,464	8,480
	NO_2	254,244	8,700
Air pollutants	CO	252,924	10,020
	SO_2	252,623	10,321
	PM_{10}	249,268	13,676
	Temperature	87,622	26
Meteorological elements	Wind speed	87,612	36
	Wind direction	87,601	47

The configuration of the training set for the prediction model to learn and the testing set to be used in the evaluation of the model needed to be established using the data in Table 2, and the composition was set to 75 percent and 25 percent, respectively, as shown in Figure 1. 20 percent of the training datasets were separated and constructed into a validation set as data needing verification during the learning process may exist.

Each variable of the configured data had a different range of measured values and data characteristics, indicating that problems can arise in the learning performance due to specific algorithms. The data were pre-processed so that each variable was suitable for learning. The most frequent wind direction was not a continuous variable, but a categorical variable expressed in 16 directions. Each variable was converted to 16 vector values represented by 0 and 1 through one-hot encoding.
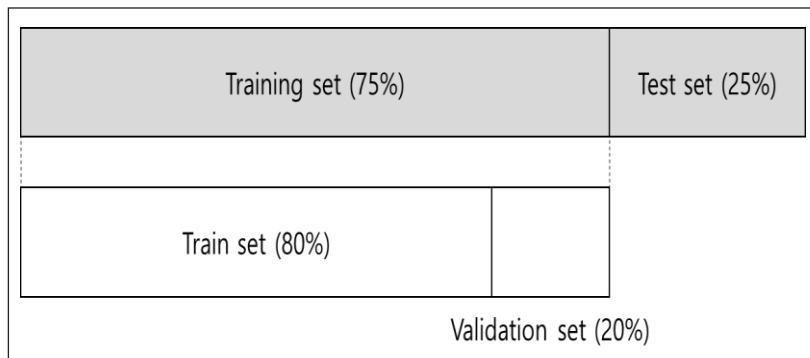
Table 2

Data Structure

Category	Variable	Number of samples
Input	O_3 (t-1)	87,215
	NO_2 (t-1)	
	CO (t-1)	
	SO_2 (t-1)	
	PM_{10} (t-1)	
	Temperature (t-1)	
	Wind speed (t-1)	
	Wind direction (t-1)	
Output	PM_{10} (t)	87,215

Figure 1

Dataset Composition

Furthermore, the scales of all numerical variables in different ranges were converted to a value between 0 and 1 using min-max scaling.

Subsequently, relevant datasets were employed directly in learning and evaluation for the RNN and LSTM prediction models, being the comparison groups. For the proposed model that conducted prediction after distinguishing low and high concentrations, the data were used separately based on 81 µg/m³ to differentiate between low and high concentrations in the training, validation, and testing datasets so that they can be used in the learning and evaluation of DNN-based classification and prediction models.

MODEL DESIGN

DNN-Based Concentration Classification Prediction Model

The proposed model consisted of a classification model that divided particulate matter concentrations into low and high concentrations and a model that predicted the separated low and high concentrations. The two models were designed based on a DNN algorithm. Figure 2 shows the structure of the proposed model. After pre-processing the data collected for learning the model, they were classified into low and high concentrations through the classification model. If zero, which meant low concentration, was the output according to the prediction result of the classification model, the data used in the classification model were transferred to the low concentration prediction model. When 1 was the output, the data used in the classification model were delivered as a high concentration prediction model. Among the prediction results of the low concentration prediction model, a value corresponding to the high concentration might be predicted.

In addition, a value corresponding to the low concentration could be predicted among the results of the high concentration prediction model. Incorrect predicted values needed to be corrected. If the high concentration value was predicted in the low concentration prediction model, it is modified to the maximum value of the low concentration range. If the low concentration value was predicted in the high concentration prediction model, it was modified to the minimum value of the high concentration range. After correcting the incorrectly predicted values, all predicted values were integrated to determine the final predicted results.

Figure 2

Structure of the Proposed Model
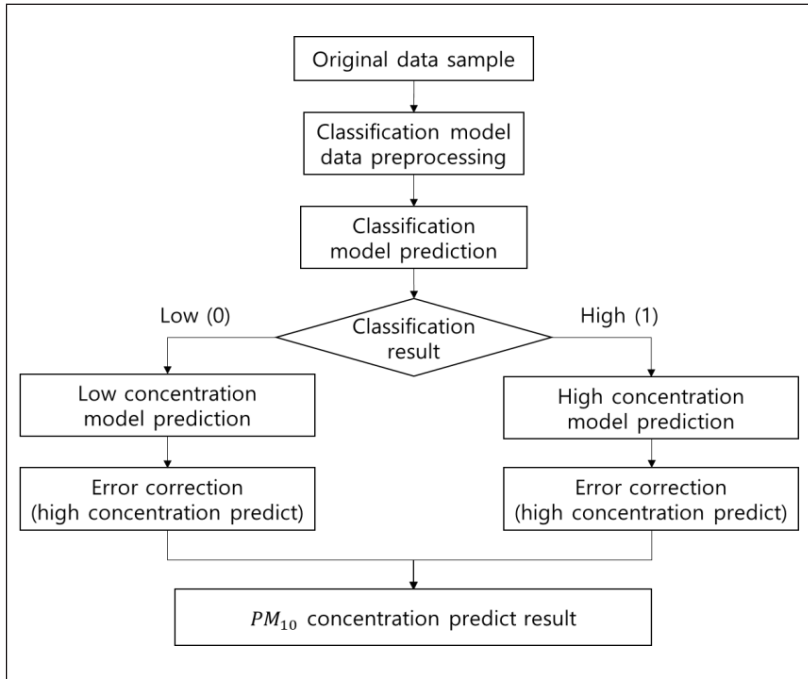
The individual settings of the model used in the proposed method are as follows. Sigmoid was applied as the activation function, and root mean squared propagation (RMSprop) was used as the optimization function of the model, as classification models must conduct a binary classification of low and high concentrations. Binary cross-entropy was employed as the cost function. Prediction models by concentration comprise a regression model that directly predicts the concentration of particulate matter. Therefore, ReLU, Adam, and MSE were used as the activation, optimization, and cost functions, respectively.

Subsequently, the optimal parameter values were derived from the common number of layers, hidden nodes, batch size, L2, and dropout to minimize overfitting through searching for hyperparameters. The epoch was designed to be 100. Table 4 shows the hyperparameter search results.

Table 3

Functions Used for Each Model

Category	Activation function	Optimization function	Cost function
Low and high concentration classification models	Sigmoid	RMSprop	Binary cross-entropy
Low concentration prediction model	ReLU	Adam	Mean square error
High concentration prediction model	ReLU	Adam	Mean square error

Table 4

Hyperparameters for Each Model

Category	Layer	Hidden node	L2	Dropout rate	Batch size	Epoch
Classification model	2	20	0.001	0.3	80	
Low concentration prediction model	4	20	0	0	40	
High concentration prediction model	2	140	0.001	0.3	40	100

Designing Comparative Models

The prediction models for performance comparison were based on DNN, RNN, and LSTN algorithms. The three models commonly set ReLU and Adam as their activation and optimization functions, respectively, and their epoch was set to 100. The learning in RNN and LSTM models must proceed through regression, and the amount of regression learning was determined by the timesteps parameter, which was set to 24. In addition, optimal parameter values were set for each model through hyperparameter searches for the layer, hidden node, L2, dropout rate, and batch size commonly applied to the three models. Three comparative prediction models were designed based

on these results. Table 5 shows the hyperparameter search results for the comparative models.

Table 5

Hyperparameters of the Comparative Models

Category	Layer	Hidden node	L2	Dropout rate	Batch size	Epoch
DNN predicion model	2	100	0.01	0.1	100	
RNN prediction model	1	80	0	0	20	100
LSTM prediction model	2	80	0	0.3	60	

PERFORMANCE EVALUATION AND ANALYSIS

As an evaluation criterion for prediction performance, RMSE, as indicated in Equation 1, was used to compare the average error between the actual value and the predicted value, and MAPE, as indicated in Equation 2, was employed to confirm the error ratio of the predicted value.

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

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{A_t - P_t}{P_t}\right| \tag{2}$$

where, n is the total number of predicted targets, A_t is the actual value, and P_t is the predicted value.
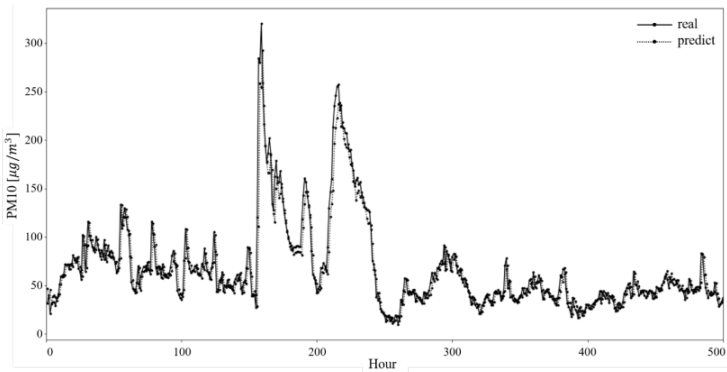
Subsequently, the accuracy of the entire spectrum of the particulate matter concentration and the AQI levels of 'good', 'moderate', 'bad', and 'very bad' were compared. Figure 3 shows the prediction results of the proposed model and the comparison model (i.e., DNN, RNN, and LSTM).
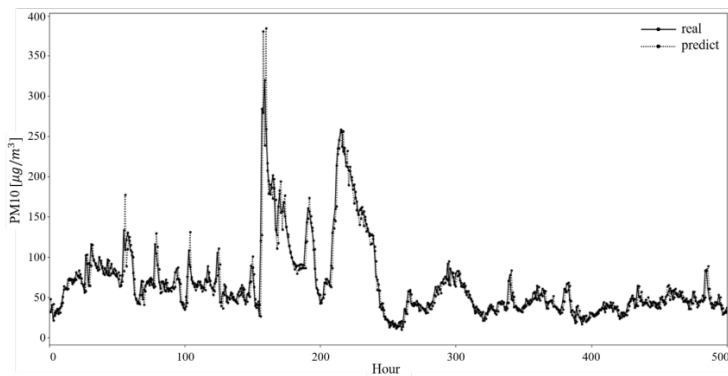
Figure 3

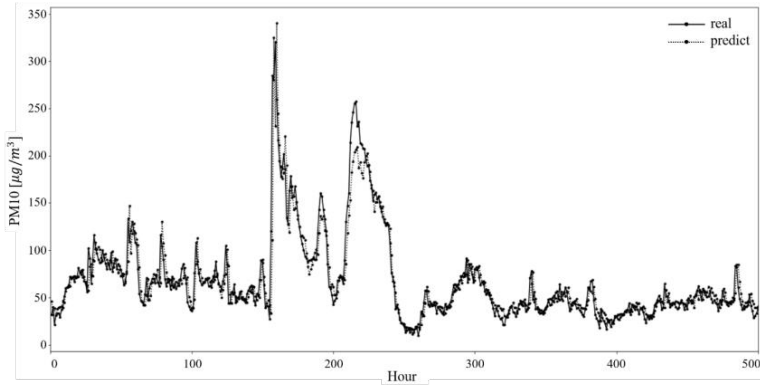*Prediction Result of Each Model*
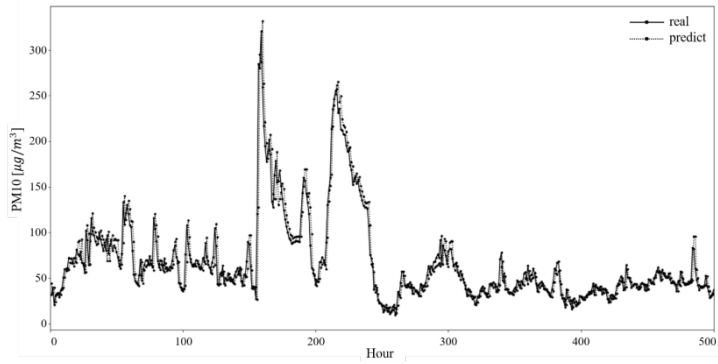
(a) Proposed Model

(b) DNN Model

(c) RNN Model

(d) LSTM Model

In the case of DNN, over-prediction and under-prediction were shown in the 150–200 concentration section. In addition, in the case of a concentration section of 100 or more, it can be seen that the range of the prediction error is relatively larger than in other sections. In the case of RNN and LSTM, over-prediction can be mainly confirmed in the concentration section of 100 or more. However, it was confirmed that the error between the actual and predicted values was large in the section with a rapid change as in the 150–170 concentration section. This can be seen as a result of the characteristics of RNN that used past information for learning. In the case of the proposed model, a predicted value close to the actual value could be confirmed in the section showing a rapid change in concentration. As a result of checking the graph of each model, the actual and predicted values were similar. Therefore, achieving accurate comparison and analysis was difficult.

Table 6 shows the prediction performance of the comparative and proposed models. The results of comparing the prediction performance of the testing dataset for each model showed that the single DNN model had the lowest error with 8.3459 µg/m³ in RMSE, which signified the error in the predicted concentration. Meanwhile, the single LSTM model had the best performance of 14.1329 percent in MAPE.

Table 6

Comparison of Prediction Performance

Category	Comparative model			Proposed model
	DNN	RNN	LSTM	
RMSE	8.3459	8.3653	8.3964	8.853
MAPE	14.2333	14.2131	14.1329	14.3944
Overall accuracy	87.38%	87.58	87.1	87.35
'Good' accuracy	79.9	87.28	79.2	84.76
'Moderate' accuracy	93.14	91.22	92.8	89.53
'Bad' accuracy	75.5	75.24	76.83	81.99
'Very bad' accuracy	65.44	72.79	66.18	80.15

Comparing the RMSE and MAPE values for each model showed that the performance of the proposed separation prediction model by concentration was lower than the single neural network models. However, both the RMSE and MAPE values showed slight differences in the decimal range. It was deemed that the performance was similar to the single model without any marginal errors, considering the structure of the proposed model where low and high concentrations of the particulate matter classification model and the prediction model by concentration were combined.

The single RNN model showed the highest performance at 87.58 percent for the overall accuracy, and the proposed model demonstrated the highest performance in the accuracy of 'good', 'bad', and 'very bad' levels at 84.76 percent, 81.99 percent, and 80.15 percent, respectively. The single DNN model indicated the highest accuracy for the 'moderate' level at 93.14 percent. Compared to the RMSE and MAPE values, which showed performance differences below the decimal range, the single neural network models and the proposed model showed large deviations in the accuracy of different AQI concentration categories. Considering the accuracy of the relevant AQI concentrations, the DNN, RNN, and LSTM models all revealed an accuracy higher than 90 percent in the 'moderate' category. However, the accuracy was relatively lower for other concentration levels.

The accuracy of each concentration based on the particulate matter concentration of 81 μg/m³ was also high in low concentrations at

88 percent but was relatively lower in high concentrations at 74~76 percent. Meanwhile, the overall AQI accuracy and the accuracy in the 'moderate' level decreased by 0.23 percent and 3.61 percent, respectively, in the proposed model, compared to the single neural network model. On the other hand, a stable accuracy of more than 80 percent resulted in the entire concentration spectrum, indicating that the over-prediction issue of low concentrations in the single neural network model had been alleviated.

CONCLUSION

Low concentration particulate matter data, which account for most of the total particulate matter generation, can cause low concentration over-prediction problems when creating a particulate matter prediction model based on machine learning. In this study, a prediction model by concentration based on deep learning conducted predictions by combining a classification model that distinguished low and high concentrations and models that predicted low and high concentrations of particulate matter. Air pollutants and meteorological elements data collected from three measurement stations in Cheonan, South Korea were used as the prediction model's learning and testing data.

The evaluation of the prediction model was configured through pre-processing. The classification and prediction models were designed based on a DNN algorithm and were used in models that predicted low or high concentrations after the particulate matter concentrations were distinguished into the relevant category in the classification model. The low concentration prediction model was designed to predict low concentrations, whereas the high concentration prediction model was constructed to predict high concentrations. Three comparative prediction models were also designed based on DNN, RNN, and LSTM algorithms, respectively. Hyperparameter searching was used to optimize these models.

The performance evaluation results and the subsequent comparison of prediction performance showed that the RMSE and MAPE values of the proposed separation prediction model by concentration were slightly lower than those of the single neural network models by figures in the decimal range. For AQI accuracy, the DNN model showed the highest accuracy of 93.14 percent in the 'moderate' category among

all the single neural network models, and the accuracy of the proposed model was less than 3.61 percent. However, the proposed model exhibited stable accuracies in the 80 percent range throughout the entire AQI spectrum, and the accuracy of the 'very bad' category was 80.15 percent, which was 7.36 percent higher than the accuracy of the RNN model at 72.79 percent. Considering the characteristics of particulate matter prediction problems where the prediction of high concentrations is crucial, it was confirmed that the separation prediction model by concentration entailed similar errors in concentration with single neural networks, and the prediction performance of high concentrations of particulate matter was enhanced.

In addition, when checking the prediction accuracy of the AQI segments, the accuracy was at a stable level of over 80 percent throughout the entire concentration spectrum. The over-prediction phenomenon at low concentrations in single neural network models concentrated in the 'moderate' region of AQI was confirmed to have been alleviated. A classification model was designed and proceeded to classify low and high concentrations. In this process, the problem of not being accurately classified was identified. This study plans to improve prediction performance by solving the problem in the future. It will also conduct a performance analysis of the prediction model by converging various types of algorithms. Furthermore, this study intends to design an improved prediction algorithm.

ACKNOWLEDGMENT

REFERENCES

Anderson, J. O., Thundiyil, J. G., & Stolbach, A. (2012). Clearing the air: A review of the effects of particulate matter air pollution on human health. *Journal of Medical Toxicology*, *8*(2), 166–175. https://doi.org/10.1007/s13181-011-0203-1

Bea, G. N. (2016). *Scientific understanding and rational judgment of fine dust pollution*. The Board of Audit and Inspection of Korea.

Beak, J. W., Choi, J. Y., & Jung, H. K. (2022). Filtering correction method and performance comparison for time series data. *Journal of Information and Communication Convergence Engineering*, *20*(2), 125–130. https://doi.org/10.6109/jicce.2022.20.2.125

Bihter, D., Omer, O. D., & Suat, T. (2022). Prediction of air pollutants for air quality using deep learning methods in a metropolitan city. *Urban Climate*, 46, https://doi.org/10.1016/j.uclim.2022.101291

Cha, J. W., & Kim, J. Y. (2018). Development of data mining algorithm for implementation of fine dust numerical prediction model. *Journal of the Korea Institute of Information and Communication Engineering*, *22*(4), 595–601, https://doi.org/10.6109/jkiice.2018.22.4.595

Choi, I. H., Lee, W. Y., Eun, B. J., Heo, J. S., Chang, K. H., & Oh, J. M. (2022). A study on prediction of PM2.5 concentration using DNN. *Journal of Environmental Impact Assessment*, *31*(2), 83–94. https://doi.org/10.14249/eia.2022.31.2.83

Han, J. M., Kim, J. G., & Cho, K. H. (2017). Verify a causal relationship between fine dust and air condition-weather data in selected area by contamination factors. *The Korea Journal of BigData*, *2*(1), 17–26.

Ipek, A., Beytullah, E., & Caner, Erden. (2022). Evaluation of data pre-processing and feature selection process for prediction of hourly PM10 concentration using long short-term memory models. *Environmental Pollution*, *311*, 119973. https://doi.org/10.1016/j.envpol.2022.119973

Jo, H. J., Park, S. W., Lee, H. I., & Lee, S. W. (2018). *Health effects caused by particulate matter and guidelines for health care*. Korea Disease Control and Prevention Agency. *Public Health Weekly Report*, *11*(11), 458–462.

Jung, I. S., & Kwon, J. A. (2017). *Current state of operation of weather forecasting and earthquake reporting system*. The Board of Audit and Inspection of Korea.

Jeon, K. H., Lee, J. H., Park, J. H., Park, H. J., Lee, Y. H., Jung, M. S., Lee, H. S., Nam, K. P., Myoung, J. S., Choi, K. C., & Kim, T. H. (2016). A study of data accuracy improvement for national air quality forecasting (III). *National Institute of Environmental Research*. https://doi.org/10.23000/TRKO201700008074

Jeon, S. H., & Son, Y. S. (2018). Prediction of fine dust PM10 using a deep neural network model. *The Korean Journal of Applied Statistics*, *31*(2), 265–285. https://doi.org/10.5351/KJAS.2018.31.2.265

Jiao, Y., Wang, Z., & Zhang, Y. (2019, May). Prediction of air quality index based on LSTM. In *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)* (pp.17–20). https://doi.org/10.1109/ITAIC.2019.8785602

Jung, Y. J, & Oh, C. H. (2021). Comparative analysis of PM10 prediction performance between neural network models. *Journal of Information and Communication Convergence Engineering*, *19*(4), 241–247.https://doi.org/10.6109/jicce.2021.19.4.241

Kim, D. H., Han, H. C., Wang, W. J., Kang, Y. J., Lee, H. Y., & Kim, H. S. (2022). Application of deep learning models and network method for comprehensive air-quality index prediction. *Applied Sciences*, *12*(13), 6699. https://doi.org/10.3390/app12136699

Kristiani, E., Kuo, T. Y., Yang, C. T., Pai, K. C., Huang, C. Y., & Nguyen, K. L. P. (2021). PM2.5 forecasting model using a combination of deep learning and statistical feature selection. *IEEE Access*, *9*, 68573–68582. https://doi.org/10.1109/ACCESS.2021.3077574

Lim, Y. B., Aliyu, I., & Lim, C. G. (2019, March). Air pollution matter prediction using recurrent neural networks with sequential data. In *Proceedings of the 2019 3rd International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence (ISMSI 2019)* (pp. 40–44). https://doi.org/10.1145/3325773.3325788

Ma, J., Ding, Y., Gan, V. J. L., Lin, C., & Wan, Z. (2019). Spatiotemporal prediction of PM2.5 concentrations at different time granularities using IDW-BLSTM. *IEEE Access*, *7*, 107897–107907. https://doi.org/10.1109/ACCESS.2019.2932445

Ong, B. T., Sugiura, K., & Zettsu, K. (2014, October). Dynamic pre-training of deep recurrent neural networks for predicting environmental monitoring data. In *2014 IEEE International Conference on Big Data (Big Data)* (pp. 760–765). IEEE. https://doi.org/10.1109/BigData.2014.7004302

Organization for Economic Co-operation and Development (OECD). (2016). *The economic consequences of outdoor air pollution. Policy highlights*. Organization for Economic Co-operation and Development. https://www.oecd.org/environment/indicators-modelling-outlooks/Policy-Highlights-Economic-consequences-of-outdoor-air-pollution-web.pdf

Pope III, C. A., & Dockery, D. W. (2006). Health effects of fine particulate air pollution: Line that connect. *Journal of the Air & Waste Management Association*, *56*(6), 709–742. https://doi.org/10.1080/10473289.2006.10464485

Shin, M. K., Lee, C. D., Ha, H. S., Choe, C. S., & Kim, Y. H. (2007). The influence of meteorological factors on PM10 concentration in Incheon. *Journal of Korean Society for Atmospheric Environment*, *23*(3). 322–331. https://doi.org/10.5572/KOSAE.2007.23.3.322

Valavanidis, A., Fiotakis, K., & Vlachogianni, T. (2008). Airborne particulate matter and human health: Toxicological assessment and importance of size and composition of particles for oxidative damage and carcinogenic mechanisms. *Journal of Environmental Science and Health, Part C*, *26*(4), 339–362. https://doi.org/10.1080/10590500802494538

World Health Organization (WHO). (2013). *Health effects of particulate matter: Policy implications for countries in eastern Europe, Caucasus and central Asia*. World Health Organization Regional Office for Europe. https://apps.who.int/iris/handle/10665/344854

Zhao, X., Zhang, R., Wu, J. L., & Chang, P. C. (2018). A deep recurrent neural network for air quality classification. *Journal of Information Hiding and Multimedia Signal Processing*, *9*(2), 346–354.

Zhou, X. (2014). Statistical characteristics of meteorological factors and particulate matters (PM10, PM2.5) in Seoul [Master's dissertation, Seoul National University of Science and Technology, Seoul, Korea].

Zihan, T., & Zhe, W. (2022, March). Predicting Beijing air quality using Bayesian optimized CNN-RNN hybrid model. In *2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML)* (pp. 581–587). IEEE. https://doi.org/10.1109/CACML55074.2022.00104