Attentive Multi-Task Prediction of Atmospheric Particulate Matter: Effect of the COVID-19 Pandemic

Seona Song¹, Seongjin Bang¹, Soyoung Cho², Hyungseok Han³, and Sangmin Lee¹

¹College of Software and Convergence, Kwangwoon University, Nowon-gu, Seoul 01897, South Korea
²Graduate School of Knowledge Service Engineering, Korea Advanced Institute of Science and Technology, Yuseong-gu, Daejeon 34141, South Korea
³Department of Computer Science and Engineering, Seoul National University, Gwanak-gu, Seoul 08826, South Korea

Corresponding author: Sangmin Lee (e-mail: smlee5679@gmail.com)

This study was supported by the Translational R&D Program on Smart Rehabilitation Exercises (TRSRE-Eq01), National Rehabilitation Center, Ministry of Health and Welfare, Korea; This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2021R1G1A1012888); This research was conducted by a Research Grant of Kwangwoon University in 2021

ABSTRACT Air pollution, especially the continual increase in atmospheric particulate matter (PM), is a global environmental challenge. To reduce the PM concentration, a remarkable amount of machine learning-based research has been proposed. However, increasing the accuracy of the predictions and providing clear interpretations of the predictions are challenging. In particular, no studies have addressed models that predict and interpret PM before and after the COVID-19 pandemic. In this paper, we present a two-step predictive and explainable model to obtain insights into reducing PM. We first use attentive multi-task learning to predict the air quality of cities. To accurately predict the concentration of particles with sizes of ~10 µm or ≤2.5 µm (PM₁₀ and PM₂.₅, respectively), we demonstrate a performance difference between single-task and multi-task learning, as well as among the state-of-the-art methods. The proposed attentive model with multi-task learning outperformed the others in terms of accuracy performance. We then used Shapley additive explanations, a representative explainable artificial intelligence framework, to interpret and determine the significance of features for predicting PM₁₀ and PM₂.₅. We demonstrated the superiority of the proposed approach in predicting and explaining both PM₁₀ and PM₂.₅ concentrations, and observed a statistically significant difference in air pollution before and after the COVID-19 pandemic.

INDEX TERMS: air pollution, attentive multi-task learning, explainable artificial intelligence, particulate matter, surrogate model, Shapley value

I. INTRODUCTION

Air pollution is airborne matter that affects the climate and harms the health of living beings, including humans. Air pollution has three major constituents: gases, biological molecules, and particles. Atmospheric particulate matter (PM) comprises solid and liquid particles that are suspended in the atmosphere by human activity and natural processes. Because air pollution with fine particles has become a serious global problem, various studies have been conducted to investigate the toxic effects of PM on human health and its mechanisms. Yue et al. (2006) reported that PM can cause serious health complications in human health [1]. The World Health Organization states that PM can contribute to lung cancer, heart disease, immune system damage, stroke, Alzheimer’s disease, chronic obstructive pulmonary disease, and other diseases [2]. PM is attributed to approximately 4.2 million premature deaths worldwide annually [3]. To develop effective method for mitigating air pollution, it is important to monitor the PM level in the air. To monitor and evaluate the air quality, PM with an aerodynamic diameter of approximately 10 µm and ≤2.5 µm (PM₁₀ and PM₂.₅, respectively) is generally measured [4]. In particular, because many studies have stated that PM₂.₅ is even more harmful to living beings than PM₁₀, recent studies have
focused on identifying the direct and indirect relations between health outcomes and exposure to PM$_{2.5}$[5-8].

To determine the characteristics of PM, many studies have gathered air pollution metrics. In addition, physical and/or chemical principles have been proposed to explain the transformation and movement of PM. Mahajan et al. assessed the exposure of humans to PM$_{2.5}$ from the road network and proposed healthier alternate routes [9]. In addition, an air-quality forecast monitoring system was proposed to integrate data acquisition, data preprocessing, and quality predictions [10]. Chang and Tseng reported statistical results for the correlations between PM and factory pollution sources [11]. Further, various studies have focused on gathering, integrating, and investigating PM datasets [12-15].

Recently, many studies on air pollution have presented neural network models to forecast pollutant and particulate levels. A long short-term memory (LSTM)-based model was developed for predicting PM levels [16]. Park et al. presented a hybrid model combining convolutional neural networks (CNN) and LSTM to describe the short and long historical patterns in the time series in ozone concentration [17], while Qin et al. proposed a hybrid CNN–LSTM model to predict urban PM$_{2.5}$ [18]. Further, deep neural networks using convolutional and bidirectional gated recurrent units (GRUs) were used to predict PM$_{2.5}$ [19]. In addition, predictive models for PM$_{10}$ and PM$_{2.5}$ based on neural ensemble techniques were developed [20]. Chiang et al. proposed a hybrid time-series prediction framework including an autoencoder, dilated CNN, and GRU to predict PM$_{2.5}$ [21].

Despite the significant progress in this field, due to the diverse and complex problems of air quality predictions, it is impractical to obtain highly accurate prediction results using a single predictive model [22]. As such, various multitask learning approaches have been proposed over the last five years. For example, a deep-neural-network-based multitask learning model with GRU was developed [23]. Zhang et al. reported that multi-task learning using a support vector machine (SVM) proved to be effective in identifying heterogeneity in air pollutant sources [24]. They proposed a multi-step-ahead PM$_{2.5}$ predictor with SVM to obtain better consistency with spatiotemporal stability. Yousefi and Alvarez used a Gaussian process-based multi-task learning model to conduct joint learning of variables with different scales [25]. They confirmed that the cross-covariance between different cities, regions, and countries can be computed to build an optimized predictive model in a stochastic manner. In addition, various multi-task learning studies have been conducted to achieve both efficient and accurate models of air pollution [26-29].

To date, no universal method suitable for analyzing any type of air pollution has been presented. In particular, research on PM in Korea is further complicated by the effects of both internal and external air pollutants, including PM from neighboring countries, seasonal variations, and unknown significant sources. Furthermore, it should be noted that in the past two years, few studies of PM have been conducted, before and during the COVID-19 pandemic. We indicated that PM air pollution has reduced in South Korea in 2020. It is presumed that this is due to reduced industrial activity in Korea and its neighboring countries. To confirm this, more scientific and rational data-driven research is required.

Moreover, there are still knowledge gaps regarding accurate predictive models of air pollution. For example, it is difficult to identify which level of PM should be classed as high or low and the basic causal elements for PM prediction. A better understanding of the predictive models is expected to contribute to the development of more accurate models for various air pollutant problems. That is a need to introduce an approach to resolve the black-box problem of PM predictive models.

The contributions of this study are summarized as follows. First, we propose an attentive multi-task-learning-based predictive model. In the proposed method, the attention mechanism provides an opportunity to exploit long sequential information, whereas multi-task learning outperforms single-task learning. To confirm the accuracy of the proposed method, we compared the proposed method with the latest algorithms for ten representative cities in South Korea. Second, to identify the statistical significance of the estimated PMs before and after the COVID-19 pandemic, we conducted a rank-based nonparametric hypothesis test for change detection. The experimental results confirmed that there was a clear statistical difference before and after the COVID-19 era. In addition, we found that the predictive performance of the previous built-learning models significantly decreased after COVID-19 due to the concept drift phenomenon.

Finally, we propose using explainable artificial intelligence (XAI) techniques to interpret the prediction results in this study. In this study, we used a game-theory method, namely SHapley Additive exPlanations (SHAP), to indicate the prediction results and achieve insights into improving the PM prediction accuracy. Using SHAP, we identified the independent variables that significantly influenced the prediction of high or low PM values. Based on the results of these experiments, we conducted a study to reveal the significant factors that have a major influence on PM prediction in major cities in South Korea.

The remainder of this paper is organized as follows. Section II describes the proposed method, and Section III presents the experimental studies for the proposed method containing two main parts: prediction and interpretation. Finally, Section IV presents concluding remarks.
II. PROPOSED METHOD
This section provides an overview of the proposed system, describes the data collection and processing steps, and the verification of the prediction algorithms.

A. OVERVIEW
Fig. 1 presents an overview of the overall procedure for predicting and explaining PM<sub>10</sub> and PM<sub>2.5</sub> concentrations. The method consists of four parts: data preparation, predictions of PM<sub>10</sub> and PM<sub>2.5</sub>, air quality before and during the COVID-19 pandemic, and interpretations of these predictions. We here focused on air pollution caused by solid and liquid particles in South Korea.

First, we collected three public datasets. Weather data were measured using the Korea Meteorological Administration (KMA) and Automated Surface Observing System (ASOS) network. Air-quality data for South Korea and China were collected from the National Institute of Environmental Research (NIER) AIR KOREA network and the World Air Quality Forecast AIR CHINA network.

Second, we built models using four deep learning methods, which were used to predict PM<sub>10</sub> and PM<sub>2.5</sub> in 10 cities. The performance of the predictive model was evaluated using the root mean square error (RMSE), mean absolute error (MAE), and mean absolute square error (MASE). Furthermore, a multi-task model that can predict both PM<sub>10</sub> and PM<sub>2.5</sub> simultaneously was developed by selecting the model with the best average performance. In addition, to evaluate the performance of multitask learning, we used PM<sub>10</sub> and PM<sub>2.5</sub> data for three adjacent cities (Seoul, Incheon, and Suwon).

Third, we performed experiments to evaluate the effects of the COVID-19 pandemic on air pollution. We compared the prediction results of 2019 and 2020 in terms of PM<sub>10</sub> and PM<sub>2.5</sub> using a predictive model with data sets gathered from 2018 to 2019. The predicted values were compared using parametric and nonparametric analyses to determine whether PM<sub>10</sub> and PM<sub>2.5</sub> for 2019 were different from those of 2020 (i.e., if there has been any significant change in air quality before and during the COVID-19 pandemic).

Finally, Shapley additve explanations (SHAP) were used to interpret the significant variables affecting the predictions of PM<sub>10</sub> and PM<sub>2.5</sub>.

![FIGURE 1. Overview of the PM<sub>10</sub> and PM<sub>2.5</sub> prediction system.](image-url)

B. DATA PREPARATION
This session describes the data collection and pre-processing steps performed before PM prediction.

1) DATA SOURCES
This section describes the data sources and their features. We collected observations from the KMA ASOS, NIER AIR KOREA, and WAQI networks to predict air quality. Weather data provided by the KMA ASOS network contained the daily temperature, rainfall, wind, pressure, humidity, sunlight, snow, clouds, ground temperature, weather phenomena, and evaporation volume from a total of 102 observatories. Air quality data provided by NIER AIR KOREA contained SO<sub>2</sub>, CO, NO<sub>2</sub>, O<sub>3</sub>, PM<sub>10</sub> and PM<sub>2.5</sub> concentrations (i.e., substances recognized as the causes of air pollution) from a total of 356 observatories. Kim et al. reported that air pollutants from China generally migrate to Korea within two days [30]. As such, we collected daily air-quality datasets from China provided by the WAQI network.

Regarding gathering the air quality and weather data sets, we focused on ten major cities in South Korea (Seoul, Incheon, Daejeon, Daegu, Ulsan, Busan, Gwangju, Jeju, Suwon, and Gangneung) to identify the representative air conditions. We also collected daily air quality data from WAQI for six cities in China (Beijing, Qingdao, Shanghai, Dalian, Shenyang, and Tianjin), which were used in the source-receptor study of long-range transboundary air pollutants in the Northeast Asia project, which is an international joint
research project between South Korea, China, and Japan. Fig. 2 shows the locations of the ten cities in South Korea and six cities in China selected for this study. Data from January 2016 to December 2020 were collected.

In addition, we incorporated wind variables for multi-task learning experiments to improve the accuracy of performance, because these variables are expected to leverage the common weather characteristics among adjacent cities. Finally, the datasets used for the PM$_{10}$ and PM$_{2.5}$ predictions are described in Table I.

![Locations of air quality and weather observatories in selected cities in South Korea and China.](image)

### TABLE I

**FINAL DATASET VARIABLE DESCRIPTION**

| Feature category | Data description |
|------------------|------------------|
| Air quality      | **SO$_2$, CO, NO$_2$, O$_3$, PM$_{10}$ and PM$_{2.5}$** |
| Temperature      | Average, minimum, maximum temperature, average dew point, average ground temperature, minimum super-secondary temperature |
| Wind             | Maximum instantaneous, maximum, and average wind speed |
| Pressure         | Average, maximum, and minimum local or sea-level pressure |
| Relative humidity| Minimum and average relative humidity |
| Total duration of sunshine | Total duration of sunshine |

2) PREPROCESSING

Adequate handling of missing values is important because incomplete data leads to erroneous analysis results. We selected significant variables using two criteria: few missing values and sufficient explanatory power as a predictor variable for PM$_{10}$ and PM$_{2.5}$. Therefore, by focusing on the daily forecast, time-related variables were excluded, and datasets with more than 10% of missing values were excluded.

We compared representative imputation methods, such as isolation forest, multiple imputation by chained equations (MICE), and K-nearest neighborhood (KNN) algorithms to impute the residual missing values. The experiments confirmed that the replaced missing values with the KNN method presented the lowest error by identifying the k-nearest neighbors and averaging the nearby points.

The data from January 2016 to December 2017 were used as the training set, whereas those from January 2018 to December 2018 were used as the validation data, and finally, the data from 2019 were used as the test set. Fig. 3 shows the PM$_{10}$ and PM$_{2.5}$ values for each city. The data for the adjacent cities of Seoul, Incheon, and Suwon (which are less than 100 km apart) were similar, while those of Busan and Daegu (300–400 km apart) had different values. This indicates that there is some consistency in the amount of atmospheric dust between the geographically adjacent cities (discussed in Section II C).

Normalization was then performed between 0 and 1 to match the size of each variable:

$$x_{\text{scaled}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}},$$

where $x_{\text{scaled}}$ is the normalized value, $x$ is the observed value, and $x_{\text{max}}$ and $x_{\text{min}}$ are the maximum and minimum values, respectively, of the dataset.

Because most PM$_{10}$ and PM$_{2.5}$ concentrations are close to the average value, high values are rarely observed. Here, we used a logarithmic transformation to reduce the variability of the asymmetric PM$_{10}$ and PM$_{2.5}$ data and impose boundary constraints.

$$y_{\text{scaled}} = \ln(y + 1),$$

here, $y$ is the target PM$_{10}$ or PM$_{2.5}$ value and $y_{\text{scaled}}$ is the normalized value. To predict PM$_{10}$ or PM$_{2.5}$, we constructed a dataset to predict the PM for the next day based on the previous three days of atmospheric and meteorological data.
C. SINGLE AND MULTI-TASK PREDICTIONS

Our learning techniques were based on single-task and multi-task learning, which used recurrent neural network (RNN), LSTM, GRU, and attentive models as machine-learning algorithms. Fig. 4 schematically shows the learning techniques used in the experiments, which consisted of three methods, as described in this section.

1) SINGLE-TASK LEARNING

Single-task learning is generally considered a predictive approach based on estimating a single dependent variable. A dataset for predicting PM$_{10}$ and PM$_{2.5}$ was established considering 10 major cities in South Korea. In the single-task learning method, the independent variable $X$ consists of 59 variables (such as air quality, wind direction, and temperature), while the dependent variable $Y$ consists of PM$_{10}$ and PM$_{2.5}$. Single-task learning predicts PM$_{10}$ and PM$_{2.5}$ for a day based on weather conditions and air pollution data for the previous three days. The structure of this model is shown in Fig. 4(a).

2) MULTI-TASK LEARNING

Multi-task learning learns multiple tasks simultaneously [31], and can process related tasks in parallel and use shared information to improve learning. The variables to be predicted in this study, PM$_{10}$ and PM$_{2.5}$, only differ by diameter and are closely related to each other. Therefore, to increase the efficiency, we devised a method that includes learning the PM as a single model. In the first multi-task learning method, $X$ consists of 60 variables, such as air quality, wind direction, and temperature, and $Y$ consists of PM$_{10}$ and PM$_{2.5}$. The structure of this model is shown in Fig. 4(b).

As shown in Fig. 2, Seoul, Incheon, and Suwon are adjacent cities with no significant difference in their maximum PM values at the same time. Therefore, we expect to increase the prediction accuracy by using multi-task learning. The multi-task learning method uses $X$, which consists of 112 variables, and $Y$ contains six variables, PM$_{10}$ and PM$_{2.5}$ for three cities. The structure of this model is shown in Fig. 4(c).

Moreover, in addition to the merits of multi-task learning in the training phase, the number of predictive models can be reduced. Because maintaining high accuracy for the models over entire areas is labor intensive, using multi-task learning improves the practicability of the proposed predictive approach.
3) PREDICTIVE MODELS

RNNs have been used for processing sequential data to handle chronological order over a period. However, the information in the initial input data is lost when handling long sequence information, resulting in a vanishing gradient problem that greatly reduces the learning ability. To resolve this, LSTM was proposed by Hochreiter and Schmidhuber [32]. LSTM has a structure in which the cell state is added to the hidden state of the RNN. LSTM has three gates: forget gate, input gate, and output gate. The forget gate is used to forget past information, the input gate is used to memorize current information, and the output gate is a gate to output the final result. The GRU proposed by Cho et al. is similar to LSTM, but simpler to calculate and implement [33]. In the case of LSTM, three gates are used, but in the GRU, only two gates are required (reset and update gates). The reset gate limits the amount of information to be used from the previous state while the update gate defines to be used from the previous and current states.

In this study, sequence-to-sequence (Seq2Seq) and attention networks were used. Seq2Seq was proposed as a method for handling data in the form of a flexible sequence without any length limitations [34]. Seq2Seq is an encoder-decoder model, where the input part acts as an encoder that compresses time-series data into vectors, and the output part acts as a decoder that converts the compressed vectors back into time-series data. In the encoding process, the model learns the data distribution in the latent vector space. In this study, the RNN, GRU, and LSTM methods described above were constructed based on seq2seq, and the corresponding structure is shown in Fig. 5(a). We here compared the accuracies of various RNN-based methods.

The attention network uses a hidden representation from the source sentence to compute a fixed-length context vector as input to the decoder [35]. Fig. 5(b) represents the structure
of an attention network, in which data are input from the encoder and computed as an attention layer to predict the next point in time. The attention layer consists of the attention score, weight, and context vector. The attention score function is defined as follows:

\[ e_{ij} = a(s_{t-1}, h_t), \]  

(3)

where \( j \) is the input \( i \) is the output, \( S_{ij} \) is the hidden state of the decoder just before predicting the next time step, and \( h_t \) is the hidden state of the encoder. We applied the softmax function to the attention score to obtain the attention weight. The softmax obtains a probability distribution where the sum of all values is 1 and the attention weight is calculated as:

\[ \alpha_{ij} = \frac{\exp(e_{ij})}{\sum_k \exp(e_{ik})} \]  

(4)

where \( \alpha \) is the weight of the input time points and reflects the importance of each time point.

To obtain the attention value, the attention weight and hidden state of each encoder are mapped to the weighted sum. The context vector is expressed as:

\[ c_t = \sum_{j=1}^{T_x} \alpha_{ij} h_j, \]  

(5)

where the final output is calculated from the weighted sum of the attention weight and hidden state.

Finally, the next decoder hidden state is the output using the calculated attention output vector, the previous decoder hidden state, and the previous decoder output. The target is expressed as follows:

\[ s_t = f(s_{t-1}, c_t, y_{t-1}). \]  

(6)

This reflects the information differently every hour without encoding all the information into a fixed-length vector, whereby the analysis can focus on the important time points of the input.

Fig. 6 shows the structure of both the prediction and surrogate models with the given datasets and the steps for calculating the contributions of variables as Shapley values to interpret the predictions. The surrogate model \( g \) was built using SHAP, as described in the following section.

1) SHAP

The SHAP method was proposed by Lundberg and Lee in 2016 [36] to explain of machine learning model predictions for each dataset. The advantage of SHAP is that the description of the Shapley value is expressed in a linear combination, as follows:

\[ g(x') = \varphi_0 + \sum_{j=1}^{M} \varphi_j z'_j, \]  

(7)

where \( g \) is the explanation model; \( z' \) is the coalition vector in \( z' \in \{0,1\}^M \) and \( M \) is the number of input features; \( \varphi_j \) denotes the Shapley values, and the contribution of the \( j \)th feature in \( \varphi_j \in \mathbb{R} \). Further, \( \varphi_0 \) is a constant value when all the inputs are missing. If the value of all features is present, the model is expressed as:

\[ g(x') = \varphi_0 + \sum_{j=1}^{M} \varphi_j. \]  

(8)

The Shapley value is obtained by constructing all the possible combinations of marginal contributions to understand the importance of a corresponding feature based on cooperative game theory, by averaging changes with or without the target feature.

\[ \varphi_j(v) = \sum_{S \subseteq \{x_1, \ldots, x_p\} \backslash \{x_j\}} |S|!(p-|S|-1)! \frac{1}{p!} (v(S \cup \{x_j\}) - v(S)), \]  

(9)

where \( S \) is a subset of the features used in the model, \( x \) is the vector of the feature values of the instance to be explained, and \( p \) is the number of features. In addition, \( v(S) \) is the prediction for feature values in \( S \) that is marginalized over features not included in \( S \).

2) DEEPHAP

We here used DeepSHAP framework, which is a specialized method for explaining the predictions of deep learning models [37], such as seq2seq-based RNN, LSTM, GRU, and attention networks. DeepSHAP measures the contribution of each feature for each sample by using the DeepLIFT method [38]. We conducted an experiment using a single-task or multi-task model to construct the prediction model. We constructed an explainable model by applying DeepSHAP only to the single-task model. However, a problem occurred when using the SHAP module. We predicted the \( y \) value at \( t+1 \) using data at \( t-2 \) and \( t-1 \), and \( t \). The SHAP module generated a Shapley value for the data at \( t-2 \), \( t-1 \), and \( t \). Therefore, we generated three Shapley values for each feature for one prediction.

Furthermore, identifying the Shapley value of each feature with the existing SHAP network is still not proposed.
for the time-series model. To resolve time, we used the time-averaging method to extract the Shapley values for each feature in the time-series model. In particular, we resolve the DeepSHAP for the time-series datasets with the time step = 3. For one prediction at \( t+1 \), there are three Shapley values per feature \((t-2, t-1, \text{and } t)\), which were averaged to obtain the Shapley values, as shown in Fig. 7.

![Flowchart of the process for calculating the Shapley value of each feature. When predicting \( y_{t+1} \), the Shapley value of the instances at \( t-2, t-1, \text{and } t \) were calculated, and then these values were averaged for each feature to give a final Shapley value.](image)

**III. EXPERIMENTAL STUDIES**

This section describes the experimental setup and evaluations of single and multi-task models.

**A. EXPERIMENTAL SETTING**

The mean square error (MSE) is used as a loss function, and is defined as:

\[
\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{10}
\]

where \( \hat{y}_i \) and \( y \) are the predicted and target value, respectively. An optimization algorithm is a method for determining the path that minimizes the lost values in a loss-function graph. Among the various functions of gradient descent, we used Nesterov-accelerated adaptive moment estimation (Nadam), which combines the Nesterov accelerated gradient (NAG) and Adam functions. Nadam can find the global minimum faster and more accurately than the popular optimizer Adam [39]. We searched for the optimal hyperparameters for predictors to increase the accuracy. In this study, the model was optimized by fixing the time step to 3, setting the hyperparameters for the remaining batch sizes \{16, 20, 25, 32, 64, 72, 128\}, and cell units from 300 to 1000. The chosen parameters were trained with a fixed batch size and cell units for all regions for fairness.

The hyperparameter settings for each algorithm are listed in Table II. The batch size refers to the total number of data delivered to the model at one time from the sample data. All models used the cell units of the RNN (300 for single-task, 500 for multi-task with two dependent variables, and 1000 for multi-task with six dependent variables). The input shape refers to the time step and size of the input, and the output shape is the number of results.

Although it is known that \( \text{PM}_{10} \) and \( \text{PM}_{2.5} \) have high correlation, each was used as an independent variable to predict the other. Because the aim of this study was to both accurately predict the concentrations and identify significant factors in PM prediction, we excluded \( \text{PM}_{2.5} \) from the independent variable in the single-task prediction of \( \text{PM}_{10} \), and vice versa. To verify the performance of the comparisons, we used RMSE, MAE, MASE as defined by eq. (11), (12), and (13).

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \tag{11}
\]

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|, \tag{12}
\]

\[
\text{MASE} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{|y_i - \hat{y}_i|}{\max(1, \sum_{i=1}^{n} |y_i - y_{i-1}|)} \right) \tag{13}
\]

where \( \hat{y} \) is the predicted value, and \( y \) is the target value. These evaluations were performed after changing the normalized \( y \) value to the original state.

| TABLE II |
| --- |
| **HYPERPARAMETER SETTINGS FOR EACH ALGORITHM** |
B. PREDICTION RESULTS

In this section, we describe the prediction results of the proposed learning models. First, the single-task results for 10 major cities in South Korea are shown for each learning model. Second, we show the results of comparing the performance of single-task and multi-task methods using the learning models that showed the best performance in the first result.

1) COMPARISONS OF SINGLE-TASK MODELS

The PM$_{10}$ and PM$_{2.5}$ values in 2019 were predicted for ten major cities in South Korea using the seq2seq-based RNN, LSTM, and attention models. The predicted results for 10 cities are summarized in Table III. Vanilla-RNN was also included in the experiment, but the results were excluded because of their low accuracy.

For both PM$_{10}$ and PM$_{2.5}$ predictions, the attention approach had the best performance whereas the performances of the seq2seq-based LSTM and GRU were similar or slightly lower. As for PM$_{2.5}$ predictions, the comparison results were similar with a difference lower than 1 in terms of the MAE and MASE.

Although the attention algorithm did not always show the best performance for every city. However, it is thought to be the most appropriate model considering the overall performance for both PM$_{10}$ and PM$_{2.5}$. Therefore, we compared single-task and multi-task methods using an attentive model.

### TABLE III

PM$_{10}$ and PM$_{2.5}$ Prediction Results Using RNN, LSTM, GRU, and Attention Models. The Most Accurate Results are Shown in Bold.

| Model | RMSE | MAE | MASE | RMSE | MAE | MASE |
|-------|------|-----|------|------|-----|------|
| SEQ2SEQ(RNN) | 15.682 | 11.435 | 0.927 | 10.146 | 6.961 | 0.935 |
| SEQ2SEQ(LSTM) | 14.565 | 10.672 | 0.867 | 9.124 | 6.474 | 0.874 |
| SEQ2SEQ(GRU) | 15.149 | 11.578 | 0.911 | 8.981 | 6.476 | 0.878 |
| ATTENTION | 14.044 | 10.117 | 0.861 | 8.830 | 6.324 | 0.819 |

2) COMPARISON OF SINGLE-TASK AND MULTI-TASK LEARNING

We selected the attentive model as the optimal model through a single-task prediction experiment. We set up multi-task learning experiments in two ways. The first method used two dependent variables, PM$_{10}$ and PM$_{2.5}$ in each city, while the second method used six dependent variables, PM$_{10}$ and PM$_{2.5}$ of the three adjacent cities, Seoul, Incheon, and Suwon. The prediction accuracy of these results is compared for each city in Table IV.

Overall, multi-task performance was better than single-task performance for PM$_{2.5}$. For PM$_{10}$, the performances of the single-task and 2-variable multi-task models were the best. In particular, the results show that multi-task models had the best performance in Suwon.

### TABLE IV

PM$_{10}$ and PM$_{2.5}$ Predictions Results for Single and Multi-Task Attentive Models for Seoul, Incheon, and Suwon.

| City      | MODEL                      | RMSE  | PM$_{10}$ | MASE | RMSE  | PM$_{2.5}$ | MASE |
|-----------|----------------------------|-------|-----------|------|-------|-----------|------|
| Seoul     | Single-task                | 13.135| 9.002     | 0.748| 8.655 | 5.883     | 0.756|
| Seoul     | Multi-task (2 dep. variables) | 12.860| 9.043     | 0.751| 8.363 | 5.867     | 0.754|
| Seoul     | Multi-task (6 dep. variables) | 12.942| 9.187     | 0.763| 8.704 | 5.701     | 0.733|
| Incheon   | Single-task                | 13.713| 10.165    | 0.795| 10.347| 6.844     | 0.818|
| Incheon   | Multi-task (2 dep. variables) | 14.728| 10.796    | 0.844| 10.247| 6.837     | 0.817|
| Incheon   | Multi-task (6 dep. variables) | 14.794| 11.805    | 0.923| 10.936| 7.148     | 0.855|
| Suwon     | Single-task learning       | 14.875| 10.412    | 0.809| 9.173 | 6.402     | 0.746|
| Suwon     | Multi-task (2 dep. variables) | 13.151| 9.248     | 0.751| 9.076 | 6.368     | 0.754|
| Suwon     | Multi-task (6 dep. variables) | 14.818| 10.549    | 0.82 | 11.015| 6.96     | 0.811|
Fig. 8 compares the predictive performances between the single-task and multi-task models. In Fig. 8, the proposed multi-task showed high prediction accuracy for PM$_{10}$ and PM$_{2.5}$ in both Seoul and Suwon. Compared to the sing-task models, the multi-task models better at following the variations in the curve.

![PM$_{10}$ and PM$_{2.5}$ prediction results using single-task and multi-task attentive models](image)

**C. AIR QUALITY PREDICTIONS BEFORE AND DURING THE COVID-19 PANDEMIC**

The global spread of COVID-19 in 2020 has significantly impacted on the overall environment. Several studies have shown significant reductions in air pollution due to the restriction imposed during COVID-19 [40–42]. Similarly, in South Korea, several studies have reported a decrease in air quality after the COVID-19 outbreak related to air pollution [43–45]. Based on previous research, we analyzed whether we can claim that changes in PM$_{10}$ and PM$_{2.5}$ are statistically significant using the data and model used in this study. First, we evaluated the nonparametric changes using the Wilcoxon rank-sum test to determine the mean shift and the Brown-Forsythe test to calculate the variance shift. We used the trained attentive model from 2016 to 2018 without updating the model, and the prediction accuracies were compared for PM$_{10}$ and PM$_{2.5}$ in early 2019 (when COVID-19 only had a small global impact) with the corresponding values in early 2020 (when COVID-19 had a huge impact). We here expected that the prediction accuracies for PM$_{10}$ and PM$_{2.5}$ in 2020 were lower than earlier time periods because of concept drift.

1) NONPARAMETRIC CHANGE DETECTION

First, we used a nonparametric hypothesis to identify true changes in the distributions of both PM$_{10}$ and PM$_{2.5}$. Considering seasonal changes, South Korea shows significant increments in PM concentration during the January to April period. Thus, we compared the differences in air quality between January and March 2019 and from January to March 2020, corresponding to periods with expected small and large impact, respectively, by the COVID-19 pandemic.

To evaluate the distribution changes, the Wilcoxon rank-sum test is used as a nonparametric test of the difference in centroid positions between two populations with a small sample size, where the observations do not follow a normal distribution or the distribution shape is unknown [46]. The Z-test was determined as follows:

\[
Z = \frac{W - n(n+1)/2}{\sqrt{n(n+m+1)/12}},
\]

where $W$ is the smallest sum of the ranks of the two samples, and $n$ and $m$ are the smallest and largest samples, respectively.

The Brown-Forsythe test is a method to test whether a target group has statistically different variances [47]. The F-test is defined as:

\[
F = \frac{(N-k)\sum_{i=1}^{N} (\bar{y}_i - \bar{y})^2}{(k-1)\sum_{i=1}^{N} (\bar{y}_i - \bar{y})^2}
\]

where $k$ is the group, $N$ is the total sample size for all groups, $\bar{y}$ and $\bar{y}_i$ are the overall sample mean and the group mean, respectively.

Fig. 9 shows the results of the experiment conducted to evaluate the differences in PM before and after the COVID-19. The brown line indicates the point where the p-value given by the Brown-Forsythe test was the smallest, and the cyan line indicates the equivalent point from the Wilcoxon rank-sum test. The purple line indicates the point with the largest difference in the 1st quartile, and the red line indicates the point with the largest difference between the maximum and minimum.

![Results of nonparametric hypothesis tests for change-point detection using the PM data](image)
For PM$_{2.5}$, the Brown-Forsythe test showed the greatest difference at the end of February 2019, and for PM$_{10}$, the Brown-Forsythe and Wilcoxon tests showed the greatest difference at the end of February 2019. Although the results from the end of March 2019 or early January 2020 were not precisely divided, the proposed experiment determined that there was a clear difference in PM before and after the COVID-19.

2) COMPARISONS OF PM PREDICTION ACCURACY BETWEEN 2019 AND 2020

We predicted PM data for 2020 using the attentive model with either a single-task or multi-task model with two dependent variables. The attentive model was trained from 2016 to 2018. The PM$_{10}$ and PM$_{2.5}$ results for Seoul are presented in Tables V and VI, respectively. The 2019 PM prediction results are compared here to those for 2020. Comparing 2020 and 2019 PM$_{10}$ results, the MASE of the single-task results increased by 49%, while that of the multi-task results increased by 40%. In the case of PM$_{2.5}$, the single-task results for 2020 increased by 22% compared to the previous year, while the corresponding multi-task results increased by 44%. The accuracy of PM$_{10}$ and PM$_{2.5}$ predictions in 2020 was significantly lower than those in 2019.

Fig. 10 shows a graph of the actual and predicted PM$_{10}$ and PM$_{2.5}$ values for 2019 and 2020. The forecasts for 2019 followed the actual values fairly well, while the forecasts for 2020 were significantly lower.

Based on these experiments, we concluded that there was a significant difference in the air quality of South Korea before and after COVID-19. The PM$_{10}$ and PM$_{2.5}$ values were significantly lower in 2020 compared to the same period in 2019, which means that the COVID-19 pandemic decreased the PM in South Korea. Air quality improvements related to the reduction in PM after the COVID-19 pandemic could be related to several factors. In Daegu, the number of confirmed COVID-19 cases in has increased rapidly since February 2020 because of the religious activities. Further, the remarkable decline in industrial activities had a major impact on air quality improvement in South Korea from early February 2020, when the COVID-19 virus spread rapidly.

### TABLE V

| Model         | 2019 RMSE | 2020 RMSE | 2019 MAE | 2020 MAE | 2019 MASE | 2020 MASE |
|---------------|-----------|-----------|----------|----------|-----------|-----------|
| Single-task   | 13.135    | 20.239    | 9.002    | 16.040   | 0.748     | 1.481     |
| Multi-task    | 12.860    | 18.304    | 9.043    | 13.638   | 0.751     | 1.259     |

At this point, there have been significant improvements in air pollution in most cities in China [48]. Furthermore, a study showed that air pollutants from China travel over long distances depending on the wind direction and affect the air quality in South Korea [49]. As such, the reason that air quality in major cities of South Korea improved before and during the COVID-19 pandemic is directly and indirectly related to the decline in industrial activities of China and fewer pollutants within South Korea. This will be discussed in more detail in Section III D, which analyzes independent variables that have an important influence on the PM prediction.

### TABLE VI

| Model         | 2019 RMSE | 2020 RMSE | 2019 MAE | 2020 MAE | 2019 MASE | 2020 MASE |
|---------------|-----------|-----------|----------|----------|-----------|-----------|
| Single-task   | 8.655     | 9.184     | 5.883    | 6.683    | 0.756     | 0.981     |
| Multi-task    | 8.363     | 12.097    | 5.867    | 9.251    | 0.754     | 1.358     |
D. EXPERIMENTAL RESULTS FOR SHAP METHODOLOGY

We here describe the results of applying the explainable model SHAP to the single task learning model. Fig. 11 is a force plot that showing the contribution of each feature of an instance in Seoul on January 7, 2020, when the PM$_{2.5}$ predicted value was 0.21. The features in red are those that contributed to raising PM$_{2.5}$ prediction to 0.21, for example, air features in China such as Shenyang_CO, Shenyang_PM$_{2.5}$, and Dalian_PM$_{10}$. Meanwhile, features such as Qingdao_SO$_2$, Shanghai_CO, and O$_3$, indicated in blue, contributed to lowering PM$_{2.5}$ predicted value to 0.21. Additionally, in the plot, a larger area occupied denotes the higher the contribution of a feature.

Fig. 12 shows results for when the PM$_{2.5}$ predicted value is lower than the average base value of 0.1366, unlike the SHAP force plot in Fig. 11. It shows the contribution of each feature of an instance in Seoul on January 25, 2020, when the PM$_{2.5}$ prediction value was 0.11. In this case, as opposed to SHAP.

The graphs in Fig. 13 are the summary plots of the SHAP values in Seoul; they visualize how each feature affects the Shapley value distribution. Each point in the summary plot is the Shapley value of each feature in one instance. If more red points distributed to the right based on an axis with a SHAP value of 0, then the feature and Shapley value are proportional to each other. If, on the contrary, more blue points are distributed on the right with the same axis, then the feature and Shapley value are inversely related.

The summary plots in Fig. 13 show 20 features that contributed the most to predicting PM$_{10}$ and PM$_{2.5}$ in Seoul. The features that contributed the most to PM$_{10}$ predictions are: PM$_{10}$, Shenyang_PM$_{10}$, Shanghai_CO, Dalian_SO$_2$, O$_3$ etc. In this case, Shanghai_CO is inversely proportional to SHAP value and the other three features are proportional to SHAP value.

The features that contributed the most to PM$_{2.5}$ predictions are: PM$_{2.5}$, Shanghai_NO$_2$, O$_3$, Qingdao_SO$_2$, Shenyang_PM$_{2.5}$, etc. However, there is a difference between Qingdao_SO$_2$ and the other four variables. The SHAP value increase when the value of Qingdao_SO$_2$ decreases. That is, Qingdao_SO$_2$ is inversely proportional to the SHAP value. However, the other four features are directly proportional to SHAP value.

We also drew visualization plots such as Fig. 14 for Busan, Incheon, and Suwon areas besides Seoul. After checking the SHAP value-feature summary plots for four regions, we can compare the features of the top distributions by region in the following tables.
Table VII shows 10 features with high contribution based on SHAP summary plots for PM$_{10}$ and PM$_{2.5}$ in all four regions. When predicting PM$_{2.5}$ in all four regions, PM$_{2.5}$ and O$_3$ generally contribute to increasing the prediction value. Another point of interest is that the features in China contribute significantly to predicting PM$_{2.5}$ values in the four regions of Korea. Shanghai_NO$_2$ and Shenyang_PM$_{2.5}$ contribute significantly to raising PM$_{2.5}$ prediction values in Seoul, Suwon and Incheon, but not in Busan. We estimate that the lower impact on the Busan region may be because it is relatively far from Shenyang and Shanghai in China, unlike the other three regions. However, features in the Qingdao region, such as Qingdao_SO$_2$ and Qingdao_NO$_2$ are variables that contribute to lowering PM$_{2.5}$ values in four regions.

The features that contribute to predicting PM$_{10}$ also display similarities with those contributing to predicting PM$_{2.5}$. Although O$_3$ does not contribute much in Busan, when predicting PM$_{10}$ in all four regions, PM$_{10}$ and O$_3$ generally contribute greatly to increasing the PM$_{10}$ prediction. In addition, the Shenyang_PM$_{10}$ feature contributes greatly to increasing PM$_{10}$ values in all four regions while features such as Dalian_SO$_2$, Dalian_NO$_2$, and Dalian_O$_3$ in Dalian region contribute greatly to increasing PM$_{10}$ values in these regions. However, when predicting PM$_{10}$, features such as Shanghai_CO and
**TABLE VII**

**Top 10 contributing features for PM$_{10}$ and PM$_{2.5}$ predictions in Seoul, Suwon, Busan, and Incheon.**

| No. | Seoul | Suwon | Busan | Incheon | Seoul | Suwon | Busan | Incheon |
|-----|-------|-------|-------|---------|-------|-------|-------|---------|
| 1   | PM$_{10}$ | PM$_{10}$ | PM$_{10}$ | PM$_{10}$ | PM$_{2.5}$ | PM$_{2.5}$ | PM$_{2.5}$ | PM$_{2.5}$ |
| 2   | Shenyang | O$_3$ | MIWS (16bearings) | O$_3$ | Shanghai | O$_3$ | Shenyang | O$_3$ | PM$_{2.5}$ |
| 3   | Shanghai | Qingdao | Dalian | Dalian | O$_3$ | Shanghai | O$_3$ | Shenyang | O$_3$ | Shanghai |
| 4   | Dalian | Shanghai | Shenyang | Qingdao | Shenyang | Shenyang | MIWS (16bearings) | Qingdao |
| 5   | O$_3$ | Shenyang | PM$_{10}$ | CO | CO | PM$_{2.5}$ | PM$_{2.5}$ | CO | Shanghai | PM$_{10}$ |
| 6   | Dalian | Shanghai | Shenyang | Dalian | NO$_2$ | PM$_{10}$ | NO$_2$ | Shenyang | NO$_2$ | Dalian |
| 7   | Shanghai | Dalian | Shenyang | Dalian | NO$_2$ | Shenyang | PM$_{2.5}$ | Dalian | NO$_2$ | Beijing | O$_3$ | SO$_2$ |
| 8   | Shenyang | Qingdao | Tianjin | Qingdao | Shenyang | Shenyang | PM$_{2.5}$ | Qingdao | Shenyang | PM$_{2.5}$ | Shenyang |
| 9   | Dalian | O$_3$ | Shenyang | Dalian | NO$_2$ | AVP(hPa) | MWS (m/s) | Shenyang | PM$_{2.5}$ | Dalian | O$_3$ |
| 10  | NO$_2$ | Beijing | Shenyang | Dalian | O$_3$ | AGT(°C) | MRH(%) | O$_3$ | MWS (16bearings) |

Qingdao_NO$_2$ in Shanghai and Qingdao contribute significantly to lowering PM$_{10}$ value.

When predicting PM$_{10}$ and PM$_{2.5}$, features in some regions of China contribute greatly to raising the prediction values. The correlation of each feature needs to be meteorologically studied.

AVP (hPa) : avg_vapor_press, AGT (°C) : avg_ground_temp, MWS (m/s) : max_wind_speed, MRH (°C) : min_relative_humidity, MIWS (16bearings) : max_instant_wind_speed_direction, MIWS (m/s) : max_instant_wind_speed

**E. Limitations of the Study**

There are several limitations to our study. We only addressed the proposed method with batch learning, but the features used in the model for air pollution inevitably changes over time. Thus, to ensure robust and accurate performance for reliable air pollution prediction, we should investigate the phenomenon of concept drift in time series manner [50]. Thus, the explicit self-updating framework should be handled to reflect these nonstationary air pollution conditions. Second, we used the air quality features of six cites in China and ten cities in South Korea to design a predictive model, but we did not consider exogenous features such as seasonality factors, exhaust engine particulates which generated from the cities, or of fossil fuels consumption. Finally, regarding the interpretation for PM prediction, the proposed approach did not provide the causalities, but significant relations between the independent and dependent variables. To identify the causal relations for PM, the predefined causal ordering or direction among variables should be discussed.

**IV. CONCLUSION**

This study aimed to determine the effectiveness of attentive multi-task learning for accurate performance and the interpretability of XAI for PM$_{10}$ and PM$_{2.5}$ predictions. There are few cities in the world that are free from PM air pollution. All ten representative cities in South Korea suffer from severe PM problems. The proposed framework accurately predicted the time-series patterns of both PM$_{10}$ and PM$_{2.5}$ and described the significant factors of PM for a city. The experimental results demonstrated that the proposed attentive multi-task learning outperforms state-of-the-art alternatives in terms of accurate performance.

The implication of this study is that the proposed method can improve the accurate performance of air pollution.
We can summarize the merit of the proposed method as follows. First, the proposed attentive multi-task prediction approach was validated with experiments, and we presented its outperformance in terms of MAE, RMSE, and MASE. Second, we showed that the COVID-19 pandemic significantly affected on air pollution by conducting the nonparametric hypothesis test for change detection. Third, we illustrated how to utilize XAI techniques for the PM$_{10}$ and PM$_{2.5}$ predictions.

In future work, we have several research directions for extending the present study. We first plan to identify deep causal relations by adding more significant variables to increase the prediction accuracy. We will then explore the statistically significant relationships between independent variables through the structural learning of Bayesian networks. Furthermore, we plan to use an incremental learning method to increase the accuracy of the proposed model for more complex prediction environments with high concept drift problems.

REFERENCES

[1] M. Tang et al., “Integration of diverse data sources for spatial PM$_2.5$ Data interpolation,” *IEEE Trans. Multimedia*, vol. 19, no. 2, pp. 408-417, 2016. doi:10.1109/TMM.2016.2613639

[2] World Health Organization, “Ambient air pollution: Health impacts”, Available at: https://www.who.int/appr/diagnosis/ambient/health-impacts/en/.

[3] W. Yue et al., “Characterization of PM$_2.5$ in the ambient air of Shanghai city by analyzing individual particles,” *Sci. Total Environ.*, vol. 368, no. 2-3, pp. 916-925, 2006. doi:10.1016/j.scitotenv.2006.03.043.

[4] M. Krzyzanowski, “WHO air quality guidelines for Europe,” *J. Toxicol. Environ. Health A.*, vol. 71, no. 1, pp. 47-50, 2008. doi:10.1080/15271350701557834.

[5] M. Dicciaiu et al., “The two PM$_2.5$,” *Environ. Res.*, vol. 86, no. 3, pp. 254-262, 2001. doi:10.1006/ensr.2001.4275.

[6] Y. F. Xing et al., “The impact of PM$_2.5$ on the human respiratory system,” *J. Thorac. Dis.*, vol. 8, no. 1, p. E69-E74, 2016. doi:10.3978/j.issn.2072-1439.2016.01.19.

[7] D. J. Nowak et al., “Modelled PM$_2.5$ Removal by trees in ten US cities and associated health effects,” *Environ. Pollut.*, vol. 178, pp. 395-402, 2013. doi:10.1016/j.envpol.2013.03.050.

[8] L. Miller and X. Xu, “Ambient PM$_2.5$ Human health effects-Findings in China and research directions,” *Atmosphere*, vol. 9, no. 11, p. 424, 2018. doi:10.3390/atmos9110424.

[9] S. Mahajan et al., “CAR: The clean air routing algorithm for path navigation with minimal PM$_2.5$ Exposure on the move,” *IEEE Access*, vol. 7, pp. 147382-147392, 2019. doi:10.1109/ACCESS.2019.2946419, PMID:147373.

[10] Y. Huang et al., “Air quality forecast monitoring and its impact on brain health based on big data and the internet of things,” *IEEE Access*, vol. 6, pp. 78678-78688, 2018. doi:10.1109/ACCESS.2018.2855142.

[11] J. H. Chang and C. Y. Tseng, “Analysis of correlation between secondary PM$_2.5$ and factory pollution sources by using ANR and the correlation coefficient,” *IEEE Access*, vol. 5, pp. 22812-22822, 2017. doi:10.1109/ACCESS.2017.2765337.

[12] M. Tang et al., “Integration of diverse data sources for spatial PM$_2.5$ Data interpolation,” *IEEE Trans. Multimedia*, vol. 19, no. 2, pp. 408-417, 2016. doi:10.1109/TMM.2016.2613639.

[13] J. L. Chen et al., “An open framework for participatory PM$_2.5$ monitoring in smart cities,” *IEEE Access*, vol. 5, pp. 14441-14454, 2017. doi:10.1109/ACCESS.2017.2723919.

[14] D. Qin et al., “A novel combined prediction scheme based on CNN and LSTM for urban PM$_2.5$ concentration,” *IEEE Access*, vol. 7, pp. 20050-20059, 2019. doi:10.1109/ACCESS.2019.2897028.

[15] K. Kim et al., “Stable forecasting of environmental time series via long short term memory recurrent neural network,” *IEEE Access*, vol. 6, pp. 75216-75228, 2018. doi:10.1109/ACCESS.2018.2884827.

[16] X. Chen et al., “High spatial resolution PM$_2.5$ retrieval using MODIS and ground observation station data based on ensemble random forest,” *IEEE Access*, vol. 7, pp. 44416-44430, 2019. doi:10.1109/ACCESS.2019.2908975.

[17] U. Pak et al., “A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction,” *Air Qual. Atmos. Health*, vol. 11, no. 8, pp. 883-895, 2018. doi:10.1007/s11869-018-0585-1.

[18] D. Qin et al., “A novel combined prediction scheme based on CNN and LSTM for urban PM$_2.5$ concentration,” *IEEE Access*, vol. 7, pp. 20050-20059, 2019. doi:10.1109/ACCESS.2019.2897028.

[19] Q. Tao et al., “Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU,” *IEEE Access*, vol. 7, pp. 76690-76699, 2019. doi:10.1109/ACCESS.2019.2921578.

[20] P. S. G. De Mattos Neto et al., “Neural-based ensembles for particulate matter forecasting,” *IEEE Access*, vol. 9, pp. 14470-14490, 2021. doi:10.1109/ACCESS.2021.3050427.

[21] P. W. Chang and S. J. Hong, “Hybrid time-series framework for daily-based PM$_2.5$ forecasting,” *IEEE Access*, vol. 9, pp. 104162-104176, 2021. doi:10.1109/ACCESS.2021.3099111.

[22] S. Lee, D.-E. Lim, Y. Kang, and H. J. Kim, “Clustered multi-task sequence-to-sequence learning for Autonomous Vehicle Repositioning,” *IEEE Access*, vol. 9, pp. 14504-14515, 2021. doi:10.1109/ACCESS.2021.3051763.

[23] B. Wang et al., “Deep multi-task learning for air quality prediction” in, *Lecture Notes in Computer Science* Intl. Conf. on Neural Inf. Process. Cham: Springer, pp. 93-103, 2018, Dec. doi:10.1007/978-3-030-42721-9_9.

[24] Q. Zhang et al., “A PM$_2.5$ Concentration prediction model based on multitask deep learning for intensive air quality monitoring stations,” *J. Cleaner Prod.*, vol. 275, 2020. doi:10.1016/j.jclepro.2020.122722, PMID:122722.
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3144588, IEEE Access