



OPEN

# Evaluation of roadside air quality using deep learning models after the application of the diesel vehicle policy (Euro 6)

Hyemin Hwang<sup>1</sup>, Sung Rak Choi<sup>2</sup> & Jae Young Lee<sup>2</sup>✉

Euro 6 is the latest vehicle emission standards for pollutants such as CO, NO<sub>2</sub> and PM, that all new vehicles must comply, and it was introduced in September 2015 in South Korea. This study examined the effect of Euro 6 by comparing the measured pollutant concentrations after 2016 (Euro 6-era) to the estimated concentrations without Euro 6. The concentration without Euro 6 was estimated by first modeling the air quality using various environmental factors related to diesel vehicles, meteorological conditions, temporal information such as date and precursors in 2002–2015 (pre-Euro 6-era), and then applying the model to predict the concentration after 2016. In this study, we used both recurrent neural network (RNN) and random forest (RF) algorithms to model the air quality and showed that RNN can achieve higher R<sup>2</sup> (0.634–0.759 depending on pollutants) than RF, making it more suitable for air quality modeling. According to our results, the measured concentrations during 2016–2019 were lower than the concentrations predicted using RNN by –1.2%, –3.4%, and –4.8% for CO, NO<sub>2</sub> and PM<sub>10</sub>. Such reduction can be attributed to the result of Euro 6.

Vehicular exhaust is a major source of pollutants that affects regional and global urban air quality. The European Environment Agency reported that the road transport sector contributes majorly to nitrogen oxides (NO<sub>x</sub>) and black carbon<sup>1</sup>. Similarly, in Korea, the transport sector accounts for the largest contribution of NO<sub>x</sub> emissions<sup>2</sup>. Especially, diesel engines are well known for high NO<sub>x</sub> emissions<sup>3</sup>. In addition, diesel vehicles also emit other air pollutants, such as particulate matter (PM), volatile organic carbons, and polyaromatic hydrocarbons<sup>4</sup>. All of these pollutants are known to have adverse effects on human health<sup>5</sup>. Acute effects of exposure to diesel exhaust include irritation in the nose and eyes, lung function alterations, breathing pattern alterations, headache, fatigue, and nausea. Chronic exposure is associated with coughing, sputum production, and decreased lung functions<sup>6</sup>. Additionally, diesel exhaust directly affects ozone formation<sup>7</sup>.

The European Union has established automobile emission standards for NO<sub>x</sub>, dust, and CO. These standards, collectively referred to as Euro emission standards, have been periodically reinforced<sup>8</sup>. In Europe, Euro 1 was established in 1992, and the latest Euro 6 standards have been used since 2013. To meet these air pollutant emission regulations, various engine technologies have been developed. For example, selective catalytic reduction (SCR) devices<sup>9</sup> converted NO<sub>x</sub> from the exhaust gas into nitrogen (N<sub>2</sub>) and water (H<sub>2</sub>O) by injecting a reductant such as ammonia (NH<sub>3</sub>). Diesel particulate filters (DPFs) reduce particulate matters from the exhaust gas of a diesel engine<sup>10</sup>. Exhaust gas recirculation (EGR)<sup>11</sup> is a method that can effectively reduce NO<sub>x</sub> emitted to the atmosphere. What EGR does is recirculating a portion of the exhaust gas back to the engine and mixing it with the atmospheric air in the combustion chamber. This reduces the concentration of O<sub>2</sub> in the chamber, which in turn reduces the combustion of fuel and the peak combustion temperature, thereby reducing the generation of NO<sub>x</sub> from the process<sup>12</sup>. Generally, a combination of above technologies is used to meet the stringent emission control standards in modern diesel cars<sup>13</sup>.

To control air pollutant emissions from diesel vehicles, the Korean government has also implemented various policies. In Korea, Euro 1 was first introduced in 1994, and the Euro 6 standards were introduced in September 2015 to limit automobile exhaust gas emissions. In the context of Euro 6, all new diesel cars were mandated to install SCR to reduce NO<sub>x</sub>. In addition, Korea implemented policies such as providing subsidies for DPF attachments and early scrapping of high-emission diesel cars (since 2005)<sup>14</sup>, prohibiting old diesel vehicles from

<sup>1</sup>Environmental Engineering Department, Ajou University, Suwon 16499, Korea. <sup>2</sup>Environmental and Safety Engineering Department, Ajou University, Suwon 16499, Korea. ✉email: jaeylee@ajou.ac.kr

entering the metropolitan area from December to March (since December 2019), and implementing a vehicle emission rating system (since April 2020)<sup>15</sup>.

Previous researchers have reported that the Euro emission standards effectively reduced air pollution from automobile emissions<sup>16–22</sup>. In Spain, vehicular CO and PM emissions have considerably reduced due to Euro 4 and 5 vehicular emission standards<sup>16</sup>. In Paris and London, NO<sub>2</sub> and PM concentrations decreased owing to Euro 5<sup>17</sup>. Grigoratos et al. tested five Euro 6 heavy duty vehicles (HDVs) on-road under typical driving conditions and showed that all tested vehicles emitted less pollutants compared to HDVs with previous technologies<sup>18</sup>. Kim et al. investigated the NO<sub>x</sub> emissions of 132 certified Euro 6 light duty diesel vehicles in Korea, and they analyzed the installed after-treatment devices and the vehicles' control strategies<sup>19</sup>. Ko et al.<sup>20</sup> reported low NO<sub>x</sub> emissions of Euro 6 vehicles under actual driving conditions, regardless of the total vehicle weight. Despite many previous studies, the effects of Euro 6 policies on the air pollutant concentration in the roadside air quality in Korea have not been studied. Therefore, it is necessary to study the roadside air quality in Korea after the introduction of Euro 6 by comparing the air quality before and after the introduction.

Recently, air quality has been predicted using a combination of different artificial intelligence technologies<sup>23–26</sup>. Jo et al. compared the performance of air quality prediction systems using deep neural network and long short-term memory (LSTM) models, which are used for time-series data-based predictions<sup>23</sup>. Athira et al. predicted air quality using recurrent neural network (RNN), LSTM, and gated recurrent unit (GRU) based on pollution and meteorological time-series AirNet data<sup>24</sup>. Tao et al. predicted air pollution by presenting a convolution-based bidirectional GRU method based on 1-D convolutional neural networks and bidirectional GRU neural networks<sup>25</sup>. Feng et al. explained the causes of winter PM<sub>10</sub> fluctuations in nine Chinese cities using random forest (RF) and RNN<sup>26</sup>.

In this study, the association between air quality and the diesel vehicles was modeled using RNN and RF. RNN is a type of artificial neural network that is suitable to process sequential data<sup>27,28</sup>. It has a structure that can feedback outputs into inputs and has an internal memory that can remember important things from previous inputs. This structure makes it suitable for handling the dynamics of sequential data<sup>28</sup>. RF is an ensemble of many decision trees, and it is known as one of the best performing machine learning algorithms for various classification and regression problems<sup>29</sup>. The high model accuracy and the accurate prediction results can be achieved with RF by minimizing overfitting through various techniques<sup>30</sup>. Since RNN and RF are both very popular algorithms for great modeling performance, this study decided to use both algorithms to model the air quality.

Using RNN and RF prediction models, this study aimed to determine how the Euro 6 diesel vehicle policy influenced the roadside air quality in Korea. Specifically, we evaluated the importance of various environmental factors such as the diesel vehicle registration number and meteorological conditions in determining the concentrations of each of four pollutants (CO, NO<sub>2</sub>, O<sub>3</sub>, PM<sub>10</sub>) during 2002–2015, and then we used the information to eliminate the features with low importance. Then the air quality during 2002–2015 was modeled using the selected features based on both RF and RNN. Since the most of the modeling period is before the introduction of Euro 6 (introduced in Sep. 2015), these models didn't include the effect of Euro 6. Finally, we used these models to predict the air pollutant concentration for Euro 6–era from 2016 to 2020 and calculated any systematic deviations between the predicted and measured concentrations to assess the effect of Euro 6 in Korea. To the best of our knowledge, this is the first study model the air pollutant concentrations using RF and RNN and examined the effect of Euro 6 in Korea.

## Method

**Data collection and pre-processing.** Daily concentrations of CO, NO<sub>2</sub>, O<sub>3</sub>, PM<sub>10</sub>, and SO<sub>2</sub> measured from 49 roadside air quality monitoring stations (AQMS) in Korea during 2002–2020 were collected from the Korea Environment Corporation<sup>31</sup>. The data from 49 stations were averaged to obtain the country-wide overall roadside air quality in Korea for modeling the impact of vehicle emissions on the air pollutant concentrations.

Daily meteorological data measured from 100 monitoring stations in Korea during 2002–2020 were collected from the Korea Meteorological Administration<sup>32</sup>. The data include temperature, precipitation, relative humidity, insolation, wind direction, and wind speed, and the data from all stations were averaged to obtain the country-wide overall meteorological factors for the modeling. The behavior of the wind was initially expressed as a vector using wind direction and wind speed. For the averaging of the vectors measured from multiple stations, we converted the wind vectors into the north–south (NS) and east–west (EW) components by projecting the vectors on to those two axes. This can be expressed as the following equations.

$$\text{Wind NS} = \text{wind speed} \times \cos(\text{wind direction}) \quad (1)$$

$$\text{Wind EW} = \text{wind speed} \times \sin(\text{wind direction}) \quad (2)$$

These overall north–south and east–west winds are the factors that can be useful for modeling any possible long-range transport of air pollutants from nearby countries<sup>33</sup>.

Monthly diesel vehicle registration data during 2002–2020 were provided by the Ministry of Land, Infrastructure, and Transport<sup>2</sup>. In order to model the air quality on a daily basis using factors such as meteorological data and the number of diesel cars, the monthly registration data were linearly interpolated to obtain the daily data for the modeling.

All numerical variables were pre-processed using min–max normalization<sup>34</sup> as expressed by

$$\text{Normalized} = \frac{\text{Original} - \text{Min}}{\text{Max} - \text{Min}} \quad (3)$$

The day of the week, a categorical variable, was pre-processed using one-hot encoding<sup>35</sup>.

**Feature selection.** In this study, we used four categories of factors to model the air pollutant concentrations. The first was the number of registered diesel vehicles. The second was the meteorological factors, including temperature, relative humidity, insolation, precipitation, wind NS and wind EW. Meteorological factors were important, since they could influence the processes of diffusion, generation, and removal of air pollutants<sup>36</sup>. The third was the temporal factors. The day of the year, day of the week, and date were included in this category to model the weekly, seasonal and long-term variations in the concentrations of air pollutants. The fourth was the precursor factors. It is known that the presence of gaseous pollutants (CO, NO<sub>2</sub>, and SO<sub>2</sub>) can serve as the precursors to facilitate the generation of O<sub>3</sub> and PM<sub>10</sub>. Figure S1 presented a graphical explanation of such secondary formation<sup>37,38</sup>. Therefore, we used these factors for modeling the concentrations of O<sub>3</sub> and PM<sub>10</sub>.

Among the four categories of factors, we selected the final factors to be provided in modeling the pollutant concentrations based on the feature importance analysis<sup>30,39</sup>. The feature importance analysis evaluated how important each feature is in explaining the air pollutant concentrations. By removing the features with low importance, we can reduce the possibility of overfitting. To perform the feature importance analysis, the “GridSearchCV” function of “sklearn” package of Python (ver. 3.10.2) was used<sup>40</sup>.

**Modeling using RF and RNN in pre-Euro 6-era (2002–2015).** Air pollutant concentrations were modeled based on RF and RNN using the selected features from 2002 to 2015, when Euro 6 has not been applied yet. Due to the choice of the period, the effect of Euro 6 was not included in the modeling. RF and RNN were two of the most popular machine learning algorithms and the relative performance of those two algorithms depends on the application and the size of dataset<sup>41</sup>. Therefore, we decided that it was worth trying both RF and RNN for modeling.

We used a validation technique for robust modeling, where 70% of the modeling data were randomly selected as a training set and 30% as a validation set. The validation set were used to evaluate the performance of the training, and the iteration in training stopped when the loss function value on the validation set started to increase owing to overfitting<sup>28</sup>.

The “sklearn” package of Python was used for RF. The main parameter in the RF model was “best\_estimator\_” of the “GridSearchCV” function, and the optimal value for each variable was set automatically<sup>40</sup>. The model parameters used in the model are summarized in Table S1. The root mean square error was used as the loss function, and R<sup>2</sup> was checked to analyze its accuracy<sup>42</sup>.

The “keras<sup>43</sup>” and “tensorflow<sup>44</sup>” packages of Python were used for RNN. The main parameters of RNN are listed in Table S2. For the RNN layer, “simpleRNN,” the most basic RNN, was used<sup>45</sup>, and 64 neurons were added to the layer. An activation function, “tanh,” was used to convert the sum of the input signals into an output signal. Compared with other functions, the “tanh” function is zero-centered and can be used for optimization<sup>28</sup>. Using the Adam optimizer<sup>46</sup>, the loss function was calculated based on the mean absolute error and the learning rate was set to 0.001. Figures S2–S5 show the training and validation losses according to the RNN epochs.

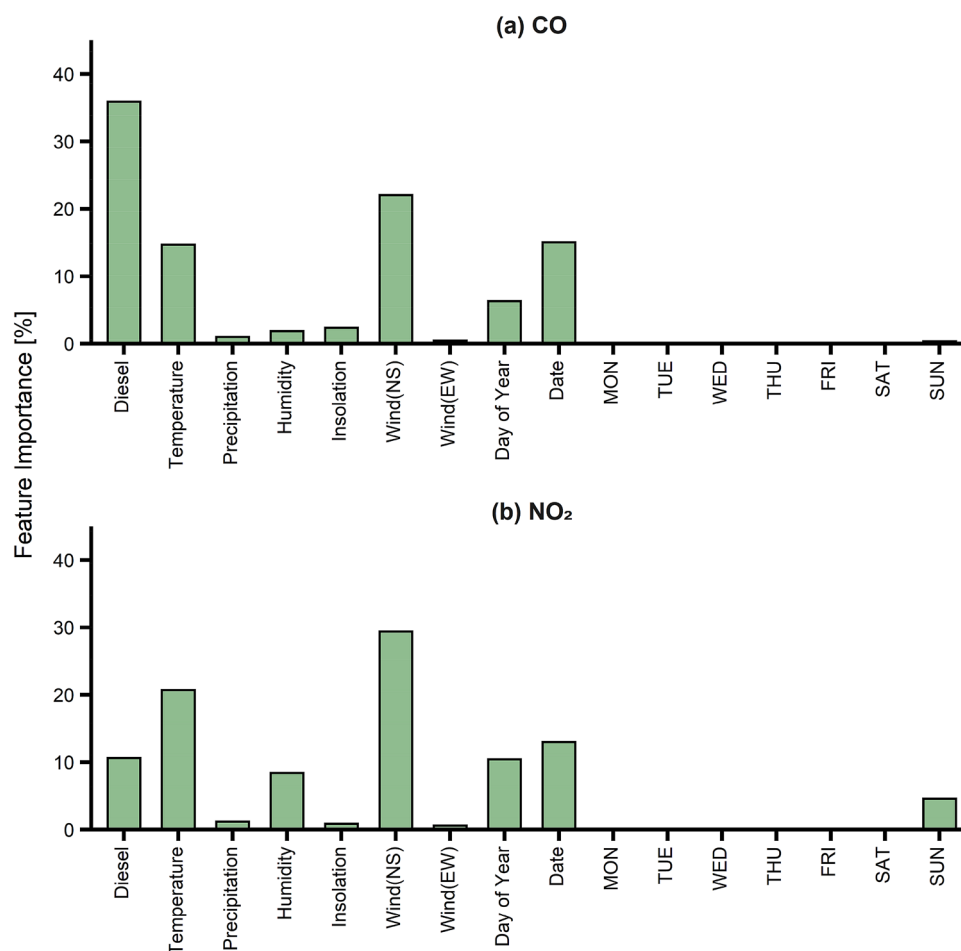
**Prediction using the models for Euro 6-era (2016–2020).** The RF and RNN models trained using the data from pre-Euro 6-era (2002–2015) were used to predict the air pollutant concentrations for Euro 6-era (2016–2020). The difference between the predicted and measured average concentrations during 2016–2020 was calculated to see any systematic deviations between the model and the reality, part of which may be due to the effect of Euro 6. The equation used to calculate the deviation is given by

$$\text{Deviation}(\%) = \frac{\text{Observed} - \text{Predicted}}{\text{Observed}} \times 100. \quad (4)$$

## Results

**Selection of features that affect pollutant concentration.** Figure S6 shows that number of diesel vehicle registrations in Korea increased steadily during 2002–2020 and CO, NO<sub>2</sub>, O<sub>3</sub>, and PM<sub>10</sub> showed periodicity due to seasonal fluctuations. CO and NO<sub>2</sub> tended to decrease, while O<sub>3</sub> tended to increase during 2002–2020. The increase in O<sub>3</sub> may be due to increased anthropogenic precursor emissions and long-term changes in meteorology. In particular, rising global temperatures may cause ozone concentrations to increase<sup>47</sup>. The strong spike in PM<sub>10</sub> could be attributed to long-range transport<sup>48</sup>. Because strong spikes caused by long-range transport increase uncertainty in the concentration prediction, PM<sub>10</sub> trends were analyzed after removing outliers. Figure S7 shows the trend of the annual number of diesel vehicles per diesel vehicle type. Except passenger cars, the number of other vehicle types did not change significantly during 2002–2020. However, the proportion of passenger cars was 20.5% in 2002 among all diesel vehicles; subsequently, this ratio steadily increased, and in 2020, passenger cars accounted for the largest percentage (58.6%) among all domestic diesel vehicles. Overall, passenger cars and trucks accounted for most number of diesel vehicle registrations.

Feature importance analysis was performed based on RF to select the final input factors for modeling. Table S3 shows the results of the feature importance analysis. Figure 1 shows the results of the feature importance analysis for CO and NO<sub>2</sub>. The number of diesel vehicles, temperature, north–south wind, and date have a major influence on CO and NO<sub>2</sub>. Figure 2 shows the results of the feature importance analysis for O<sub>3</sub> and PM<sub>10</sub>. The influence of insolation was the largest (37.4%) for O<sub>3</sub>. Insolation is involved in the photochemical reaction of O<sub>3</sub>, and is more important than the presence of gases, such as CO, NO<sub>2</sub>, and SO<sub>2</sub><sup>49</sup>. PM<sub>10</sub> was dominantly influenced by the concentrations of CO, NO<sub>2</sub>, and SO<sub>2</sub>. This is because PM<sub>10</sub> could be produced not only from primary emissions,



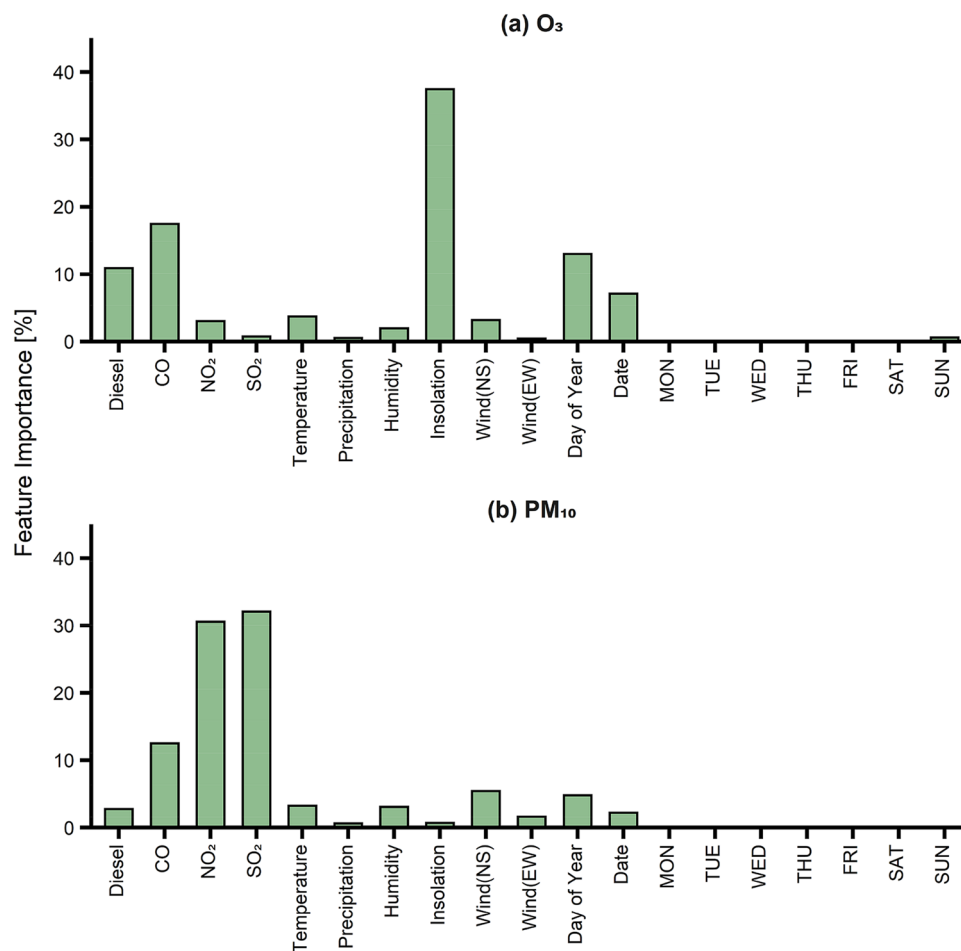
**Figure 1.** Feature importance of (a) CO and (b) NO<sub>2</sub>.

but also from gas-to-particle conversion of gaseous pollutant emitted from roadside air<sup>50</sup> (see Figure S1 for the graphical illustration of the process). Among various features, the day of the week showed only a small influence on the air pollutant concentrations. Therefore, the day of the week was not selected as the final input data in the RF and RNN models.

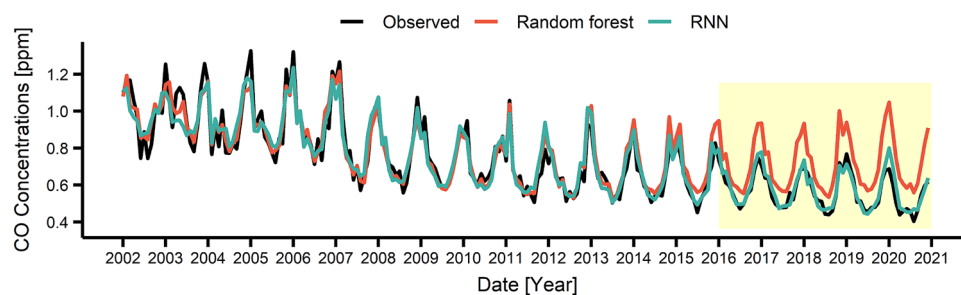
**Modeling and prediction using RF and RNN models.** The air pollutant concentrations were modeled based on RF and RNN algorithms using data from 2002 to 2015. Table S4 shows the  $R^2$  and RMSE of the models on the validation set (2002–2015) for each air pollutant. As can be seen, RF shows  $R^2$  of 0.44–0.635 and RNN shows  $R^2$  of 0.634–0.759 depending on the pollutants, suggesting that the performance of the RNN model was better than that of the RF model.

These RF and RNN models were used to predict the air pollutant concentrations during 2016–2020. Figures 3, 4, 5, 6 present the air pollutant concentrations modeled (2002–2015) and predicted (2016–2020) by the RF (in red) and RNN (in blue) models along with the observed concentrations (in black). The prediction period (2016–2020), coincided with the Euro 6 period, was depicted as yellow box in Figs. 3, 4, 5, 6. Note that the modeled curves agreed well with the observed curve during 2002–2015 and started to deviate from the observation since 2016. Table 1 lists the deviations between the modeled and measured average concentrations for 2002–2015 (before Euro 6) and between the predicted and measured average concentrations for 2016–2020 (after Euro 6). Note that the deviation during modeling period is relatively small with  $-1.3 \sim 6.7$  for RF and  $-0.3 \sim 2.4$  for RNN depending on the pollutants, while the deviation during the prediction period is larger with  $-26.6 \sim 20.0$  for RF and  $-8.8 \sim -2.1$  for RNN. This result also suggested that the performance of the RNN model was better than that of the RF model. Particularly, the RNN model succeeded in predicting the long-term decrease trend for CO, NO<sub>2</sub> and PM<sub>10</sub> as well as the long-term increase trend for O<sub>3</sub> during 2016–2020. The directions of the change were correct for all pollutants, but the magnitudes of the change were predicted to be less. However, the RF model failed to predict the long-term trends for both direction and magnitude.

For the detailed analysis, Tables S5–S8 summarized observed, modeled and predicted yearly concentrations of each pollutant. As can be seen, the observed concentrations of CO, NO<sub>2</sub>, PM<sub>10</sub> dramatically reduced in 2020, and the deviation between the observation and the RNN prediction in 2020 becomes  $-6.2\%$ ,  $-28.6\%$ , and  $-29.3\%$  for CO, NO<sub>2</sub>, PM<sub>10</sub>. This can be mostly attributed to the lockdown measures applied during COVID-19 pandemic<sup>31,52</sup>.



**Figure 2.** Feature importance of (a) O<sub>3</sub> and (b) PM<sub>10</sub>.

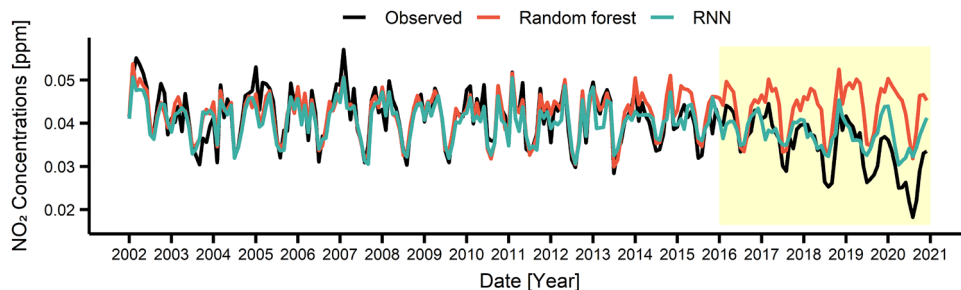


**Figure 3.** Monthly concentrations of observed and predicted values by using recurrent neural network (RNN) and random forest (RF) for CO. The area shaded in yellow showed the period when the Euro 6 was applied.

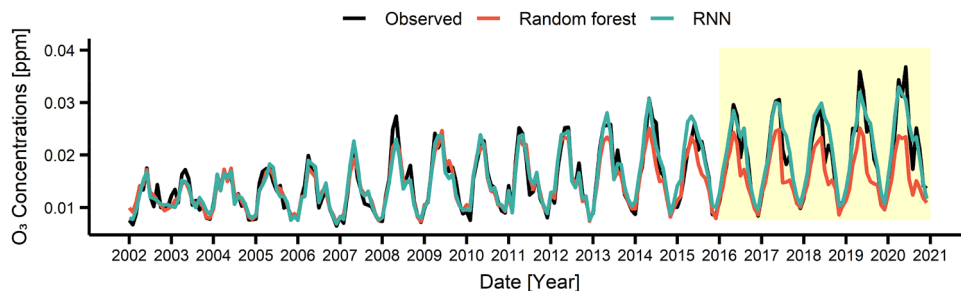
Except for 2020, the further decrease in the concentrations of air pollutants compared to the RNN prediction may be attributable to the environmental policies such as Euro 6.

## Discussion

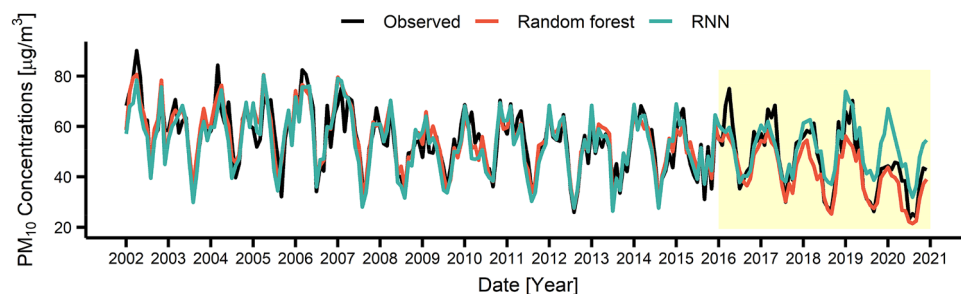
This study modeled the air pollutant concentrations based on RF and RNN models with the diesel car registration information and various environmental factors in 2002–2015. Then, we predicted the air quality in 2016–2022 using the models. Finally, we calculated any systematic deviations between the predicted and observed concentrations in order to assess the effects of various regulations applied after 2015–2016 in South Korea. The high performance of the model trained in 2002–2015 could predict the air pollutant concentration accurately. Therefore, a difference between the concentration predicted by RNN and actual concentrations during 2016–2020 suggested the result of new policies introduced after the modeling period. Any policies that have been applied during 2002–2015 (e.g. early scrapping of high-emission diesel cars applied since 2005) were assumed to be



**Figure 4.** Monthly concentrations of observed and predicted values by using RNN and RF for NO<sub>2</sub>. The area shaded in yellow showed the period when the Euro 6 was applied.



**Figure 5.** Monthly concentrations of observed and predicted values by using RNN and RF for O<sub>3</sub>. The area shaded in yellow showed the period when the Euro 6 was applied.



**Figure 6.** Monthly concentrations of observed and predicted values by using RNN and RF for PM<sub>10</sub>. The area shaded in yellow showed the period when the Euro 6 was applied.

Period	Model	CO		NO <sub>2</sub>		O <sub>3</sub>		PM <sub>10</sub>	
		Conc. (ppm)	Deviation (%)	Conc. (ppm)	Deviation (%)	Conc. (ppm)	Deviation (%)	Conc. (µg/m <sup>3</sup> )	Deviation (%)
2002–2015	Observed	0.794		0.042		0.015		54.97	
	RF	0.804	– 1.3	0.042	0	0.014	6.7	55.34	– 0.7
	RNN	0.796	– 0.3	0.041	2.4	0.015	0	53.76	2.2
2016–2020	Observed	0.564		0.035		0.020		46.73	
	RF	0.714	– 26.6	0.044	– 25.7	0.016	20.0	41.68	10.8
	RNN	0.576	– 2.1	0.038	– 8.6	0.021	– 5.0	50.84	– 8.8

**Table 1.** Difference between the predicted value of the model and the actual observed value according to air pollutants. *RF* random forest, *RNN* recurrent neural network.

captured in the modeling and not to contribute to the systematic deviations. Such policies applied after 2015–2016 include applying Euro 6 standards for new cars (since September 2015), and prohibiting old diesel vehicles from entering the metropolitan area during the cold season (since December 2019). According to our results, the measured concentrations were lower than the concentrations predicted using RNN by  $-1.2\%$ ,  $-3.4\%$ , and  $-4.8\%$  for CO, NO<sub>2</sub> and PM<sub>10</sub> during 2016–2019. Year 2020 was eliminated to rule out the effects of seasonal banning of old diesel cars and COVID-19 (both happened since December 2019). Since we eliminated the effects of other known policies, these reductions are likely to be originated from Euro 6.

This study aimed to investigate the air quality improvement associated with Euro 6 by comparing the measured concentrations with the ones predicted using a model constructed in pre-Euro 6-era. The methodology used in this study has a few limitations. First, the deviation of the observed from the predicted can stem not only from Euro 6 but also from many other differences between 2002 and 2015 and 2016–2020. In this study, we ruled out the effects of other policies and COVID-19, but there can be a yet another effect that hasn't been ruled out. Any of those effects could make us overestimate the effect of Euro 6. Second, the diesel-car emissions depended on the mileage driven by the cars not on the number of vehicles. However, due to the availability of the data, we used the number of diesel cars instead. Third, we used factors related to diesel cars, meteorology, time, and precursors to model the air quality. However, other factors such as those related to industrial and non-industrial emissions were not included in the modeling due to the lack of the data. Providing such confounding factors might have improved the accuracy of the modeling. All of these limitations were originated from the lack of proper data and can be solved by collecting and building them. This work and the corresponding analysis were left for future work.

## Conclusions

The present study is the first to model the roadside air quality in South Korea using deep learning algorithms (RNN and RF) and to investigate the improvements in air quality after the introduction of Euro 6. To model the air quality, we prepared data on various environmental factors and performed feature analysis to select the final factors for modeling. As a result, we showed that the performance of the RNN model was better than that of the RF model in terms of R<sup>2</sup>. By comparing the measured concentrations with the ones predicted using RNN, we showed that Euro 6 contributed the pollutant reduction by  $-1.2\%$ ,  $-3.4\%$ , and  $-4.8\%$  for CO, NO<sub>2</sub> and PM<sub>10</sub>. These results may contribute to future policy making related to diesel vehicles, and the methodology used in this study can be applicable in assessing the effect of other policies.

## Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 4 September 2022; Accepted: 22 November 2022

Published online: 01 December 2022

## References

1. EEA, EEA. Air quality in Europe—2019 report. *European Environment Agency* (2019).
2. Statistics Korea. Statistical Data of the Ministry of Land, Infrastructure and Transport, Korea. <https://kosis.kr/index/index.do> (2022).
3. Sowman, J. *et al.* In-use emissions testing of diesel-driven buses in Southampton: Is selective catalytic reduction as effective as fleet operators think?. *IET Intell. Transp. Syst.* **12**, 521–526 (2018).
4. Nelson, P. F., Tibbett, A. R. & Day, S. J. Effects of vehicle type and fuel quality on real world toxic emissions from diesel vehicles. *Atmos. Environ.* **42**, 5291–5303 (2008).
5. Sawyer, R. F. *et al.* Mobile sources critical review: 1998 NARSTO assessment. *Atmos. Environ.* **34**, 2161–2181 (2000).
6. Sydbom, A. *et al.* Health effects of diesel exhaust emissions. *Eur. Respir. J.* **17**, 733–746 (2001).
7. Marr, L. C. & Harley, R. A. Modeling the effect of weekday–weekend differences in motor vehicle emissions on photochemical air pollution in Central California. *Environ. Sci. Technol.* **36**, 4099–4106 (2002).
8. Suarez-Bertoa, R. & Astorga, C. Impact of cold temperature on euro 6 passenger car emissions. *Environ. Pollut.* **234**, 318–329 (2018).
9. Forzatti, P. Present status and perspectives in de-NO<sub>x</sub> SCR catalysis. *Appl. Catal. A* **222**, 221–236 (2001).
10. Kang, W., Choi, B., Jung, S. & Park, S. PM and NO<sub>x</sub> reduction characteristics of LNT/DPF+SCR/DPF hybrid system. *Energy* **143**, 439–447 (2018).
11. Pierpont, D. A., Montgomery, D. T. & Reitz, R. D. Reducing particulate and NO<sub>x</sub> using multiple injections and EGR in a DI diesel. *SAE Trans.*, 171–183. <https://www.jstor.org/stable/44615074> (1995).
12. Lopatin, O. P. Study of the influence of the degree of exhaust gas recirculation on the working process of a diesel. *J. Phys. Conf. Ser.* **1515**, 042021 (2020).
13. Konstandopoulos, A. G. *et al.* Impact of combination of EGR, SCR, and DPF technologies for the low-emission rail diesel engines. *Emiss. Control Sci. Technol.* **1**, 213–225 (2015).
14. Kim, J. & Kang, K. Evaluation and improvement of a subsidy policy on early scrapping of old diesel vehicles. *J. Environ. Policy* **14**(2), 73–99 (2015).
15. Joo, H.-S. *Comprehensive Plan on Fine Dust Management*. (2018).
16. Querol, X. *et al.* 2001–2012 trends on air quality in Spain. *Sci. Total Environ.* **490**, 957–969 (2014).
17. Font, A., Guiseppin, L., Blangiardo, M., Gherzi, V. & Fuller, G. W. A tale of two cities: Is air pollution improving in Paris and London?. *Environ. Pollut.* **249**, 1–12 (2019).
18. Grigoratos, T., Fontaras, G., Giechaskiel, B. & Zacharof, N. Real world emissions performance of heavy-duty euro VI diesel vehicles. *Atmos. Environ.* **201**, 348–359 (2019).
19. Kim, H. J., Jo, S., Kwon, S., Lee, J. T. & Park, S. NO<sub>x</sub> emission analysis according to after-treatment devices (SCR, LNT+SCR, SDPF), and control strategies in Euro-6 light-duty diesel vehicles. *Fuel* **310**, 122297 (2022).
20. Ko, S. *et al.* NO<sub>x</sub> emissions from euro 5 and euro 6 heavy-duty diesel vehicles under real driving conditions. *Energies* **13**, 218 (2020).

21. Winkler, S. L. *et al.* Vehicle criteria pollutant (PM, NO<sub>x</sub>, CO, HCs) emissions: How low should we go?. *Npj Clim. Atmos. Sci.* **1**, 1–5 (2018).
22. Tomassetti, L. *et al.* Evaluation of air quality and mobility policies in 14 large Italian cities from 2006 to 2016. *J. Environ. Sci. Health A Tox. Hazard Subst. Environ. Eng.* **55**, 886–902 (2020).
23. Jo, S. J., Kim, J., Kim, S. H., & Youn, J. A Comparative study on the performance of air quality prediction model based on DNN and LSTM. in *Proceedings of the Korea Information Processing Society Conference*. Korea Information Processing Society, 577–579 (2020).
24. Athira, V., Geetha, P., Vinayakumar, R. & Soman, K. P. Deepairnet: Applying recurrent networks for air quality prediction. *Proc. Comput. Sci.* **132**, 1394–1403 (2018).
25. Tao, Q., Liu, F., Li, Y. & Sidorov, D. Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU. *IEEE Access* **7**, 76690–76698 (2019).
26. Feng, R., Zhou, R., Shi, W., Shi, N. & Fang, X. Exploring the spatial heterogeneity and temporal homogeneity of ambient PM<sub>10</sub> in nine core cities of China. *Sci. Rep.* **11**, 8991 (2021).
27. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D Nonlinear Phenom.* **404**, 132306 (2020).
28. Tsai, Y. T., Zeng, Y. R. & Chang, Y. S. Air pollution forecasting using RNN with LSTM in 16<sup>th</sup> International Conference on Dependable, Autonomic and Secure Computing, 16<sup>th</sup> International Conference on Pervasive Intelligence and Computing, 4<sup>th</sup> International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/Cyber-SciTech) 2018 (1074–1079) (IEEE, 2018) (IEEE, 2018).
29. Yi, H. *et al.* A novel random forest and its application on classification of air quality in 8<sup>th</sup> International Congress on Advanced Applied Informatics (IIAI-AAI) 2019 (35–38) (IEEE, 2019).
30. Zhang, K., Yang, J., Sha, J. & Liu, H. Dynamic slow feature analysis and random forest for subway indoor air quality modeling. *Built. Environ.* **213**, 108876 (2022).
31. Korea Environment Corporation. AirKorea. <https://www.airkorea.or.kr/web> (2022).
32. Korea Meteorological Administration. Open MET Data Portal. <https://data.kma.go.kr/cmmn/main.do> (2022).
33. Kim, M. J. The effects of transboundary air pollution from China on ambient air quality in South Korea. *Heliyon* **5**(12), e02953 (2019).
34. Patro, S. G. K. & Sahu, K. K. Normalization: A preprocessing stage. arXiv preprint [arXiv:1503.06462](https://arxiv.org/abs/1503.06462) (2015).
35. Sethi, A. One-hot encoding versus label encoding using scikit-learn. *Analytics Vidhya-Learn Everything about Analytics* (2020).
36. Pérez, I. A., García, M. Á., Sánchez, M. L., Pardo, N. & Fernández-Duque, B. Key points in air pollution meteorology. *Int. J. Environ. Res. Public Health* **17**, 8349 (2020).
37. Zhang, J., Yongjie, W. & Zhangfu, F. Ozone pollution: a major health hazard worldwide. *Front. Immunol.* **10**, 2518 (2019).
38. Kroll, J. H. *et al.* The complex chemical effects of COVID-19 shutdowns on air quality. *Nat. Chem.* **12**(9), 777–779 (2020).
39. Font, A. *et al.* Air quality in enclosed railway stations: Quantifying the impact of diesel trains through deployment of multi-site measurement and random forest modelling. *Environ. Pollut.* **262**, 114284 (2020).
40. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
41. Roßbach, P. Neural Networks versus Random Forests – Does it always have to be Deep Learning? <https://blog.frankfurt-school.de/wp-content/uploads/2018/10/Neural-Networks-vs-Random-Forests.pdf>.
42. Alsaber, A. R., Pan, J. & Al-Hurban, A. Handling complex missing data using random forest approach for an air quality monitoring dataset: A case study of Kuwait environmental data (2012 to 2018). *Int. J. Environ. Res. Public Health* **18**, 1333 (2021).
43. Gulli, A. & Pal, S. *Deep Learning with Keras* (Packt Publishing Ltd, 2017).
44. Raschka, S. & Mirjalili, V. *Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-Learn, and TensorFlow 2* (Packt Publishing Ltd, 2019).
45. Septiawan, W. M. & Endah, S. N. Suitable recurrent neural network for air quality prediction with backpropagation through time in 2<sup>nd</sup> international conference on informatics and computational sciences (ICICoS) 2018 (1–6) (IEEE, 2018).
46. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. arXiv Preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014).
47. Seo, J. *et al.* Extensive spatiotemporal analyses of surface ozone and related meteorological variables in South Korea for the period 1999–2010. *Atmos. Chem. Phys.* **14**(12), 6395–6415 (2014).
48. Oh, H. R. *et al.* Long-range transport of air pollutants originating in China: A possible major cause of multi-day high-PM<sub>10</sub> episodes during cold season in Seoul, Korea. *Atmos. Environ.* **109**, 23–30 (2015).
49. Atkinson, R. Atmospheric chemistry of VOCs and NO<sub>x</sub>. *Atmos. Environ.* **34**, 2063–2101 (2000).
50. Röööli, M. *et al.* Temporal and spatial variation of the chemical composition of PM<sub>10</sub> at urban and rural sites in the Basel area, Switzerland. *Atmos. Environ.* **35**(21), 3701–3713 (2001).
51. Hwang, H. & Lee, J. Y. Impacts of COVID-19 on air quality through traffic reduction. *Int. J. Environ. Res. Public Health* **19**, 1718 (2022).
52. Lee, H. *et al.* The relationship between trends in COVID-19 prevalence and traffic levels in South Korea. *Int. J. Infect. Dis.* **96**, 399–407 (2020).

## Acknowledgements

This study was supported by the National Research Foundation of Korea (grant number NRF-2021R1C1C1013350) and by a grant from the National Institute of Environmental Research (NIER), funded by the Ministry of Environment (ME) of the Republic of Korea (NIER-2022-04-02-087).

## Author contributions

Conceptualization, H.H., S.R.C. and J.Y.L.; methodology, H.H., S.R.C. and J.Y.L.; software, H.H. and S.R.C.; data curation, H.H.; writing-original draft preparation, H.H. and S.R.C.; writing-review and editing, J.Y.L.; funding acquisition, J.Y.L. All authors have read and agreed to the published version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-24886-z>.

**Correspondence** and requests for materials should be addressed to J.Y.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022