Forecasting the Effects of Real-Time Indoor PM2.5 on Peak Expiratory Flow Rates (PEFR) of Asthmatic Children in Korea: A Deep Learning Approach

JIYOUNG WOO [1], JI-HYUN LEE[2], YEONJIN KIM[3], GUILLAUME RUDASINGWA [4], DAE HYUN LIM[5], AND SUNGROUL KIM[3,4]

[1]Department of Bigdata Engineering, Soonchunhyang University, Asan-si 31538, South Korea
[2]Department of Radiology, SMG-SNU Boramae Medical Center, Seoul 07061, South Korea
[3]Department of ICT Environmental Health System, Graduate School, Soonchunhyang University, Asan-si 31538, South Korea
[4]Integrated Research Center for Risk Assessment, Soonchunhyang University, Asan-si 31538, South Korea
[5]Department of Pediatrics, School of Medicine, Inha University, Incheon 22332, South Korea

Corresponding author: Sungroul Kim (sungroul.kim@gmail.com)

ABSTRACT We built a deep learning algorithm to predict the deterioration of health symptoms among asthmatic children between 8-12 years of age. It is based on Peak Expiratory Flow Rates (PEFR) and indoor air pollution data, as well as meteorological data collected at their indoor residences every 2 minutes using portable monitoring devices with a low-cost sensor between November 2018 and March 2019. The PEFR results collected twice a day were matched with daily PM2.5. A personalized model has been developed to predict the peak expiratory flow rate of the next day, considering indoor air quality data including PM2.5, humidity, temperature, and CO_2 level in previous days. Two models were developed incorporating Indoor Air Quality (IAQ) with the PEFR-only model. The IAQ uses the daily IAQ, and 10-minute basis IAQ in predicting the future PEFR. Recurrent Neural Networks (RNN) and Deep Neural Networks (DNN) models were trained using 4 months of linked data to predict PEFR for the next days during the study period. The 10-minute RNN model was found to predict better PEFR with a Root Mean Square Error (RMSE) of 42.5 and a Mean Absolute Percentage Error (MAPE) of 14.0, as it consolidates the cumulative effects of PM2.5 concentrations over time. The highly accurate estimation showed that indoor air quality significantly affects PEFR.

INDEX TERMS Asthma, big data, machine learning, recurrent neural network, peak expiratory flow rates (PEFR).

I. INTRODUCTION

Asthma is a chronic inflammatory disorder of the airways that affects the quality of life in all age groups. According to global Disease Adjusted Life Years (DALY), asthma is one of the leading global chronic disease burden [1], [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Wei-Yen Hsu.

Exacerbation of asthma may require critical medical care and hospitalization. According to the Center for Disease Control (CDC), 8.3% of American children were asthmatic in 2016 [3]. In Korea, pediatric asthma accounts for the largest disease burden among children, and the prevalence of asthma has steadily increased from 1998 to the present [4]. The prevalence of childhood asthma in the Seoul metropolitan city was estimated between 4.0-6.7%

IEEE Access·

J. Woo *et al.*: Forecasting Effects of Real-Time Indoor PM2.5 on PEFR of Asthmatic Children in Korea: Deep Learning Approach

between 2011 and 2017 [5]. The hospital admission rate for asthma among Koreans was twice higher than the overall average for the Organization for Economic Cooperation and Development (OECD) [6]. The relative risk of hospital admissions due to asthma among children in Seoul, Korea, was shown to be related to particle matter in the air [7], and there is immense evidence that the increase in the concentration of air pollutants is related to excessive emergency room visits, as well as hospital admissions due to asthma [6].

Environmental health forecasting is useful for measuring environmental risks and environmental factors that affect the health status of populations, such as air quality and weather conditions, which are of great importance in forecasting asthma [8]. Currently, rules to predict asthma exacerbations use various types of factors, including inflammatory biomarkers [9], [10], genes [11], airway function [12], [13], symptom scores [13], [14], use of asthma medications [15] and environmental factors [16]. Although most of these rules can only identify an individual who is more likely to develop an asthma attack, rather than the time of experiencing an asthma attack, identifying individuals exposed to a high risk of lung deterioration leading to developing asthma exacerbations is of massive importance, as it allows healthcare providers and decision-makers to focus on individuals at higher risk.

As children tend to spend most of their time indoors [17], modeling indoor air quality is important to predict asthma among children [18]. A detailed and accurate model is required that can be used to predict when a child may suffer asthma attacks based on clinical knowledge. Predictive clinical data modeling is characterized by the comprehensive use of machine learning algorithms to develop forecasting or classification models [19]. Accurate predictive models can provide a useful advanced warning to asthmatic children to seek or take adequate medication on time, as well as help asthmatics plan their mobility accordingly. In addition, the ability to predict health risk based on indoor monitoring and personal data is of great importance as it provides real-time risk assessment to asthma exacerbation based on general health conditions and dynamic patterns of indoor air pollutants.

Despite substantial studies demonstrating the association between indoor air pollution and asthma exacerbation [18], [20]–[24], research that predicts personalized health risks based on real-time indoor air quality monitoring data is still limited. Some studies have used one-time data or monthly measurements rather than real-time monitoring [25]. Such approaches are useful in identifying a correlation between indoor air pollution and asthma morbidity but are less practical in preventive intervention, such as early warning, to susceptible individuals that require timely monitoring.

Recently, the development of artificial health intelligence started from deep learning, which is mainly useful in healthcare-related predictions due to its power in learning capabilities, flexibility in dealing with time series and longitudinal data, and the ability to alleviate irregularities in data [26], [27]. Deep learning, mainly using RNNs has successfully predicted various incidences and morbidity of diseases [28]. Especially, regarding the air quality, RNN and its variants have also been adopted to predict PM2.5 [29]–[31]. These new models advanced the previous models that had mainly adopted regression models [32], [33]. RNN models are good at characterizing time-series data such as PM2.5 by applying different weights to previous values in time-series, transforming the weighted sum into abstract values, and repeating this transformation. The differences of the deep learning from the regression, and time-series models, such as Autoregressive Integrated Moving Average (ARIMA) exist in that the higher-level aggregation toward time-series by adding up layers is possible in the deep learning model.

Artificial Intelligence (AI) based on machine learning, particularly deep learning, has been applied in various fields [35]. However, the predictability level provided by a deep learning model is based on real-time risk monitoring data through personal and environmental sensors using the Internet of Things (IoT) [36]. Deep learning has been less utilized for asthma or other respiratory diseases compared to other health applications, mainly due to the arduousness of obtaining real-time measurements of lung function data [37] since most existing machine learning algorithms developed to predict asthmatic risks used historical patient records in the clinical setting [38]. Deep learning algorithms can forecast the personal risk of developing asthma by linking with air pollution, weather, and real-time monitoring data. PEFR is a popular estimate used to measure the extent of airway obstruction among asthmatic patients due to its convenience of using a small, portable and inexpensive device even at home [39].

Previously, some studies used weekly PEFR measurements to predict asthma deterioration among children using machine learning models [40]. Recently, very limited research attempted to collect daily PEFR measurements over a period and match them with temporal variations in indoor air quality and weather conditions to predict the deterioration of asthma.

In this study, we develop a prediction model to evaluate lung function deterioration among asthmatic children reflected by PEFR using a deep learning technique to real-time indoor particulate matter and meteorological data among asthmatic children aged 8-12 years who live in the metropolitan city of Incheon, Cheonan, and Asan in South Korea. As far as we know, this is the first work to predict PEFR based on previous patterns of it and the air quality data.

The rest of this paper is organized as follows. A review of the literature on current and previous research related to machine learning including deep learning. Deep learning applications in respiratory diseases are mentioned in Section I. A detailed description of the proposed methods, deep learning model description, and model training are presented in Section II. The experimental results are shown

J. Woo *et al.*: Forecasting Effects of Real-Time Indoor PM2.5 on PEFR of Asthmatic Children in Korea: Deep Learning Approach

IEEE *Access*

in Section III, and conclusions with the contribution of this study are drawn in Section IV.

II. MATERIALS AND METHODS

A. STUDY DESIGN

This study was carried out with a randomized, double-blind crossover intervention design that included 2 separate intervention periods, each lasting 2 weeks and separated by a washout period of 1 week. Air filtration units (Tower_XQ600, ATXH663-HWK, Winix, South Korea) were placed in bedrooms or main living rooms at each participant's residence, and each participant served as his/her own control. IAQ and lung function measurements were performed in the same season (Fall of 2018 from September to October) to adjust for significant seasonal variation.

Each participant was exposed to 2 different scenarios of each intervention period, such as unfiltered indoor air, that is, no filter installed in the filtration unit (control) and filtered indoor air, that is, filter installed in the filtration unit (experiment). A brand new HEPA filter (CAF-E0S4, Winix, South Korea) was used for the experimental group in phase 1 (first 2 weeks) and phase 2 (second 2 weeks), despite the possibility of using High-Efficiency Particle Absorbing filters (HEPA) for at least 6 months in the home environment. As a result, all air purifiers exhibited an identical outward appearance regardless of filtration status. This filter system has a certified Clean Air Delivery Rate of 98 ft^3/min and covers approximately 628 ft^2.

As this study involved human participants, the Inha University Hospital research ethics committee (IRB No. 2018-07-007) approved this study and written informed consent was obtained from the legal guardians of all participants.

B. STUDY POPULATION

This study collected and used data from 26 asthmatic children registered with the Center for Environmental Health Research for Allergy and Respiratory Diseases of Inha University Hospital, South Korea. Following the Global Initiative for Asthma [41], all study participants were diagnosed as mild asthmatics. Therefore, this study included only asthmatic children with a PEFR range between $100 \sim 500$ L/min residing in the Incheon metropolitan city. Furthermore, according to the international guideline, study participants underwent the baseline clinical test before the intervention design at their visit [42]. Therefore, this study excluded children who were born prematurely and had immune disorders for validation purposes. Furthermore, data on children's demography and social and economic status were collected before the intervention process.

C. MEASUREMENT OF PEFR

Prior to PEFR measurements, our field manager demonstrated the SmartOne user guide to children and guardians prior to the actual recording. Each child was asked to take a deep breath and then blow into the maximum flow meter as hard and fast as possible. Each child was given 2 trials and encouraged to blow harder each time. According to the manufacturer's guide, a tight seal was maintained between the lips and the mouthpiece [43]. Self-reported PEFR records were obtained using a SmartOne Spirometer (Medical International Research, Roma, Italy). The daily maximum PEFRs were selected from the morning and afternoon measurements. SmartOne had an accuracy test result that complied with the American Thoracic Society (ATS) and European Respiratory Society (ERS) 2005 standards; the International Organization for Standardization (ISO) 26782 standard (for spirometry parameters), and ISO 23747 standard (only for the peak flow parameter). The ISO 26782 standard for 2009 and the 2005 ATS/ERS Spirometry Statement agree that the accuracy of both volume and flow-type spirometers should be checked [22].

D. INDOOR AIR POLLUTION DATA

To measure IAQ data, including PM2.5, temperature, and Relative Humidity (RH), continuous measurement devices with an onboard laser light scattering sensor were installed in each study participant. Data were obtained from PurpleAir (PA) (purchased through ESCORT&CARE.com, manufactured by PurpleAir, LLC, Draper, Utah, USA) with a 2-minute interval. CO2 was obtained from ESCORTAIR. The PM2.5 data was finally matched with PEFR records for all asthmatic children. Before IAQ data collection, real-time monitors were compared with FEM from the US EPA, and the results were published in a separate article [44].

PM2.5 levels were continuously measured for 7 days at the home of each participant. Before sending IAQ devices to study participants for data collection, the devices were calibrated in the Soonchunhyang University test laboratory with a reference device, GRIMM (MODEL 11-D, GRIMM, AEROSOL TECHNIK). For the precision calculation, the relative standard deviation was $< 15\%$.

The devices were further calibrated for short-term performance by passing a national certification guideline prepared by the special act on reducing and managing fine dust, South Korea, amended by Act No. 16303 since March 26, 2019 [45]. As stated in the guidelines, the measured ambient PM2.5 records using PAs were compared with reference filter samplers developed by the nationally approved test site. During the 4 weeks of the outdoor certification test period, the conditions were characterized by cold weather ranging from -7 to $8\,°$ C as a daily mean value and 25% of the test period was predominated by rain. The concentration of time-averaged hourly PM2.5 ranges from 5 to 40 $\mu g/m^3$. A linear model related to the national site reported by a tapered oscillating microbalance element and PA monitors had a coefficient of determination (R^2) of 99%. Our final hourly results were within 14% of the error rate, compared to the national monitoring site values for 100% of the samples. Since the variance of daily outdoor conditions was much greater, we used PA in the indoor environment.

IEEE *Access*

J. Woo *et al.*: Forecasting Effects of Real-Time Indoor PM2.5 on PEFR of Asthmatic Children in Korea: Deep Learning Approach

E. RESEARCH FRAMEWORK

In this work, we examined how much the indoor air quality of the previous day affects the breathing condition of the very next day, represented as PEFR. We developed a personalized model to predict the peak expiratory flow rate of the next day after seeing the indoor air quality data including PM2.5, humidity, temperature and CO2 level in previous days. The basic assumption was that the past experience under certain air conditions influenced the next days' PEFR. We formulated this assumption as the PEFR forecasting model using past PEFR and indoor air quality records.

For that, we compared 2 models incorporating IAQ with the PEFR-only model. The PEFR-only model just predicted the future value based on the history of PEFR. The IAQ model was specified into 2 models, one reflected the daily IAQ, and the other reflected the 10-minute basis IAQ in predicting the future PEFR. We named these models as the daily IAQ model and the granular IAQ model, respectively. The general research framework is formulated as shown in Figure 1.
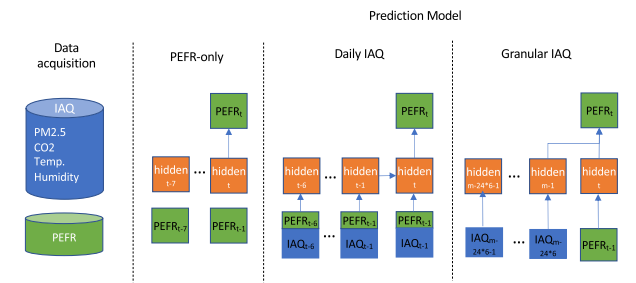

FIGURE 2. Data framing to make the input data fit the RNN models of the daily IAQ model.

The number of data points was calculated 24h*6, and the target PEFR was set to the value after 6 hours. Therefore, the delay was set to 6 hours*6 points per hour. The data framing process for the second model is described in Figure 3.

FIGURE 3. Data framing to make input data fit the RNN models of the granular IAQ model on a 10-minute basis.

FIGURE 1. Research framework.

Two models had different look-backs to observe the past data and derive patterns from them. The framework for the 2 models was elaborated, as shown in Figures 2 and 3. These models were developed to investigate whether IAQ affects future PEFR. The comparison between the 2 models tells the effect of fine granular examination of IAQ on PEFR.

To feed the input data to the model, a data framing process was required. For the first model, the data sample was formed with a duration of 7 days. Next, the input vector was thrown into samples. Each sample was cut to a length of 7; we looked back over the past 7 days to generate a forecast value. We predicted a PEFR the next day, so the delay was set to 0. However, the interval between the 2 observations was 24 hours on average. The data framing process is described in Figure 2. In this case, the IAQ was aggregated into daily average to be consistent with the PEFR.

Regarding the second model, the data sample was formed on a 10-minute basis to integrate CO2, temperature, humidity, and PM2.5 recorded within their own periods. We took a 10-minute average value for IAQ data to reduce the data volume and the effect of fluctuation. The PEFR was recorded daily, but IAQ data were recorded every 10 minutes, so we adopted the sequence-to-vector model.
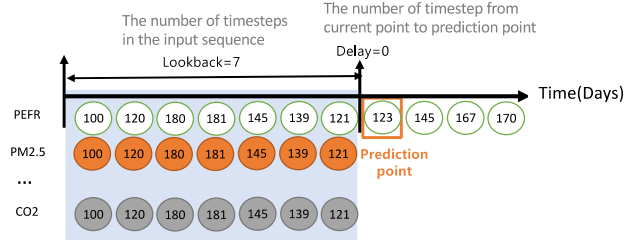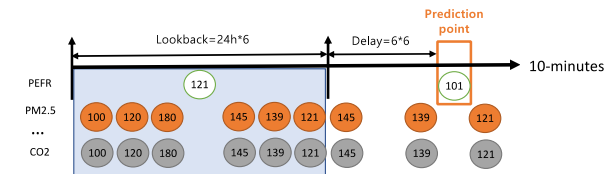
F. ALGORITHMS

As a forecasting model on time-series of indoor air quality and PEFR itself, we adopted the state-of-the-art model, deep learning, especially recurrent neural network that can consider the sequence of data. We considered RNN, which reflects the sequence of input data in different ways. As a comparative model to highlight the effect of consideration on the sequence of PEFR and indoor air quality records, we also adopted the DNN model. The architecture of deep learning models was specified with the number of layers and the number of nodes in each layer. The hyperparameters of the deep learning model, such as optimization algorithm, activation function, and learning rate, were explored to derive the best model.

RNN was adopted as a deep learning algorithm to predict the maximum PEFR. The algorithms adopted are briefly introduced as follows. In feed-forward neural networks, when data was fed to the network, the operation proceeded from the input layer through the hidden layer to the output step by step. Unlike feed-forward neural networks, RNNs were neural networks with a recurrent structure [35]. As shown in Figure 4, the current input is remembered as the state and feeds to the RNN node by combining with the next input. The basic structure of RNN is shown in Figure 4. h_t stands for hidden state and x_t stands for input data. The hidden state h_t of the current time is updated by receiving the hidden

J. Woo *et al.*: Forecasting Effects of Real-Time Indoor PM2.5 on PEFR of Asthmatic Children in Korea: Deep Learning Approach

IEEE *Access*

state h_{t-1} of the previous time point. The current output y_t is a structure that is updated by receiving h_t. As can be seen from the formula in Figure 4, the hidden state activation function is set to the hyperbolic tangent (tanh), which is a non-linear function. A recurrent neural network is an artificial neural network specialized in sequential data learning and is characterized by an internal cyclic structure. By allowing the state to be stored inside the neural network, the input in a sequence is processed using the internal memory to predict future data. In addition, because the RNNs have a memory to store the state of the past, it is suitable for time-series data processing or when it is necessary to capture sequential context.
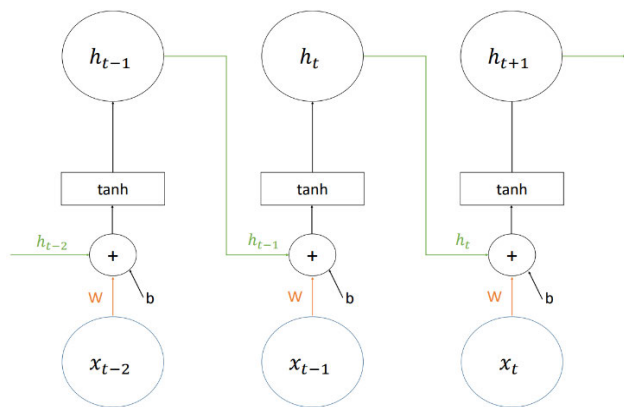

FIGURE 4. Recurrent neural networks (RNN).

GRU (Gated Recurrent Unit) upgrades RNN model incorporating the gated unit. GRU has variations of the update gate and the reset gate denoted as z and r in Figure 5. The update gate aims to determine how much of the past information needs to be traced to the future. The reset gate aims to decide how much of the past information should be forgotten. The tanh part decides what to forget from previous states.
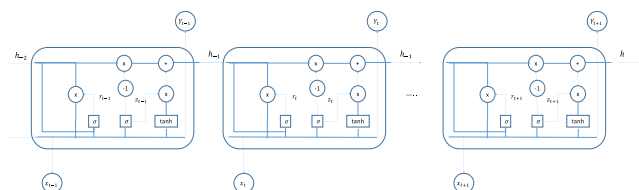

FIGURE 5. Gated recurrent unit (GRU).

DNN is a feed-forward neural network with two or more hidden layers, an input layer and an output layer. DNNs can model complex nonlinear relationships like general artificial neural networks. The DNN model takes sequential input data independently without considering the sequence and transforms the input data to abstract features through a full connection. For example, the PEFR and IAQ data during the lookback period are combined with being transformed to the next day PEFR.

Figure 6 shows the structure of a DNN. In this figure, the target of the DNN model is denoted as y, which indicates the PEFR, and the input of DNN, including previous IAQ variables and the PEFR is denoted as x. The hidden nodes denoted as h take all input from the previous layer with some weights. Therefore, PEFR and IAQ during the previous days are transformed into PEFR of the next day through a fully connected network. n represents the id of a hidden node, and m represents the id of a hidden layer.
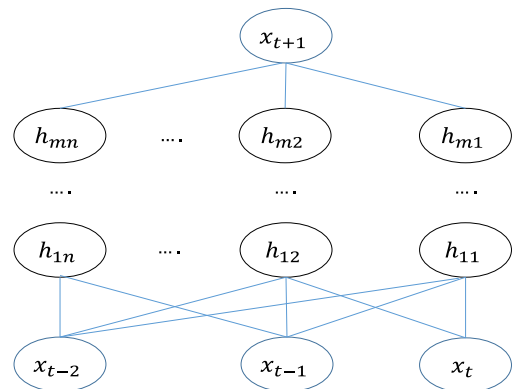

FIGURE 6. Deep neural network (DNN).

G. MODEL TRAINING

We incrementally updated the model using the previous data to fit the time-series data, as shown in Figure 7. As the iteration goes on, the data during the lookback period are joined to the input. In the first step, data as long as the lookback period is used as training data, validation and test data are formed with the data of the next periods.
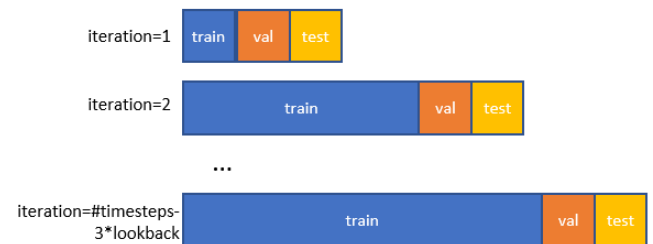

FIGURE 7. Training/validation/test data split for sequential data.

For the first and second models, the validation and test data are as long as 7 days. For the third model, the lookback period is set to a day, which is formed with 144 length on 10-minute basis.

We compared 3 algorithms in predicting PEFR. First, the input is encoded as a 2D tensor of size (timesteps, the number of input features), and current input is iteratively feeding to RNN over the timesteps after chucking with the lookback length for RNNs.

To find the optimal structure of RNN, we varied the number of RNN and dense layers. For the DNN model, we also explored the structures and found the optimal one.

We used the Adam optimizer with the learning rate of 0.001 to fit the model with data, and the call-back option was adopted to find the optimal epoch.

The regression model was evaluated using RMSE. RMSE is a measure commonly used when dealing with the difference between the predicted and actual values of the model. The RMSE formula is shown in Equation 1. One of the big advantages of using RMSE is that the size of the error value is proportional to the actual value because the difference between the actual value and the predicted value is squared, and then the square root is taken during the calculation process. We also adopted the mean absolute percentage error (MAPE) for intuitive interpretation.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (y_t - \widehat{y_t})^2} \qquad (1)$$

where, y_t is the actual value of PEFR at time t and $\widehat{y_t}$ is the estimated PEFR by the model at time t, n is the total number of time stamps in the observation period.

$$MAPE = \sqrt{\frac{100}{n} \sum_{t=1}^{n} (\frac{y_t - \hat{y}}{y_t})^2} \qquad (2)$$

III. RESULT

Figure 8 shows the time-series of peak expiratory flow rate (PEFR) per person. IQA includes PEFR, PM2.5, humidity, temperature and CO2. Figure 9 shows the time-series of IQA during a day of the subject of IH-001.

The number of days and the number of data points that met our criteria (5 days at least) used in the experiment are listed in Table 1. We built an individual model using 60% of data as individual's training data, 20% as validation data, and 20% as test data.

TABLE 1. The number of days and data points for an individual.

ID	IH-001	IH-002	IH-003	IH-004	IH-005	IH-007	IH-008	IH-010	IH-011	IH-012
Days	23	25	14	16	16	31	11	8	16	24
Data	3312	3600	2016	2304	2304	4464	1584	1152	2304	3456
ID	IH-013	IH-019	IH-020	IH-022	IH-024	IH-025	IH-027	IH-029	IH-030	total
Days	22	15	31	12	5	15	10	23	24	-
Data	3168	2160	4464	1728	720	2160	1440	3312	3456	26496

The number of subjects was 30, but the quality of the data differs from the subjects. We found many missing values, and the discontinuity in lines is caused by these missing values. The final data set consisted of 19 subjects, excluding subjects whose data points were less than 6 days.

The RNN model to predict PEFR after 7 days had the best RMSE of 42.5. The comparison of the 3 models is shown in Table 2. The results show that the granular IAQ model performs the best. This indicates that precise examination of
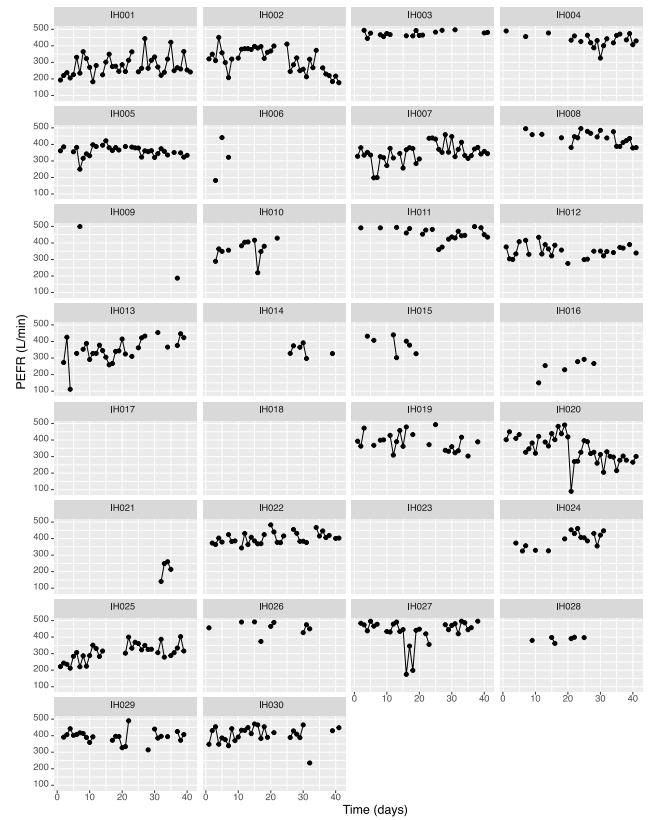

FIGURE 8. Time-series plot of the daily PEFR of the subjects.

TABLE 2. Validation results in terms of RMSE of 3 models.

	PEFR-only	Daily IAQ	Granular IAQ
RMSE	269.8	225.5	42.5

the IAQ will improve the performance of PEFR forecasting. The comparison of PEFR-only and the daily IAQ tells that the IAQ has significant effects on PEFR.

We checked the performance of the algorithm for individuals because we developed a personalized model. Table 3 shows the individual RMSE. We can see that there are some differences in each person's results with the best performance of 21.6 and the worst performance of 60.3. For intuitive understanding, MAPE is also represented in Table 3. Overall, MAPE is limited to 14%.

In Figure 10, we demonstrate a case of comparison of actual and predicted PEFR. The red line represents the actual value, and the green line represents the predicted value. The overall pattern of the predicted values was consistent with the actual values.

The optimal RNN model is a model consisting of 3 GRU layers and 1 fully connected layer. In the fully connected layer, the activation function was set to a real function. Instead, we apply a fully connected layer with 1 unit in the last layer. Since the fully connected layer at this time was only for dimension reduction, the activation function was not used.
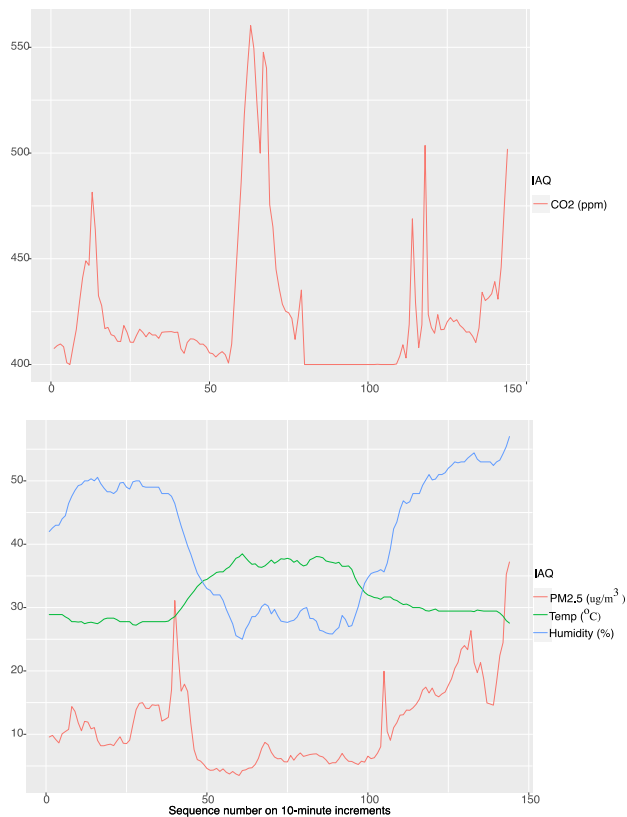
J. Woo *et al.*: Forecasting Effects of Real-Time Indoor PM2.5 on PEFR of Asthmatic Children in Korea: Deep Learning Approach

IEEE *Access*

FIGURE 9. Time-series plot of IAQ of IH-001.

FIGURE 10. The comparison between the predicted and actual values of PEFR.

To investigate whether the consideration of the sequence of IAQ in predicting PEFR was taken into account, a DNN model was constructed to predict the peak expiratory flow

TABLE 3. Individual result of the granular IAQ model.

Subject ID	RMSE	MAPE	Subject-specific average PEFR (L/min)
IH-001	44.6	12.5	287.1
IH-002	38.9	13.7	329.0
IH-003	42.8	7.5	476.4
IH-004	40.9	14.0	432.6
IH-005	47.0	15.4	356.7
IH-007	47.3	19.2	354.2
IH-008	35.1	10.9	440.8
IH-010	53.0	12.4	344.9
IH-011	21.6	10.3	454.1
IH-012	35.4	15.5	352.3
IH-013	33.8	10.6	358.6
IH-019	60.3	20.3	385.6
IH-020	59.6	26.7	345.9
IH-022	29.6	9.0	402.1
IH-024	58.8	12.8	356.4
IH-025	30.2	13.4	333.8
IH-027	51.3	10.0	467.3
IH-029	37.2	15.7	399.9
IH-030	40.5	16.0	407.6
Mean	42.5	14.0	383.4
SD	10.8	4.5	52.2

	PEFR_1	...	PEFR_7	PM2.5_1	...	PM2.5_7	Temp_1	...	Temp_7	Humidity_1	...	Humidity_7	CO2_1	...	CO2_7
IH-001	238		238.0	9.3		8.7	28.9		28.9	41.0		41.2	417.6		417.7
IH-002	239		238.0	8.7		8.6	28.9		28.9	41.2		41.4	417.7		415.5
...															
IH-003	206		206.0	13.6		13.6	27.8		27.8	48.5		48.5	429.5		429.5

FIGURE 11. Data feed to DNN.

rate of the next day with input data from the previous week. To build the DNN model, the data were reshaped into 2d tensors (sample axis, feature axis). The DNN consists of 4 fully connected layers. 'Fully connected' refers to that all neurons between 2 adjacent layers are fully pairwise connected. Adam was used as an optimizer to make the learning speed fast and stable. In addition, a drop-out layer was used to avoid overfitting. The data format feed to DNN is shown in Figure 11.

Comparing 2 algorithms, RNN that remembered the sequence showed better performance than DNN. DNN considered all the previous sequences equally and exhibited low performance, as shown in Table 4. This implied that the sequence of IAQ of the previous day affected the PEFR of the next day significantly.

The best models for RNN and DNN are mentioned in Table 5. Three layers of GRU for RNN and three layers of DNN were placed gradually decreasing the number of filter units from 100 to 30. This implies that, at the lower level,
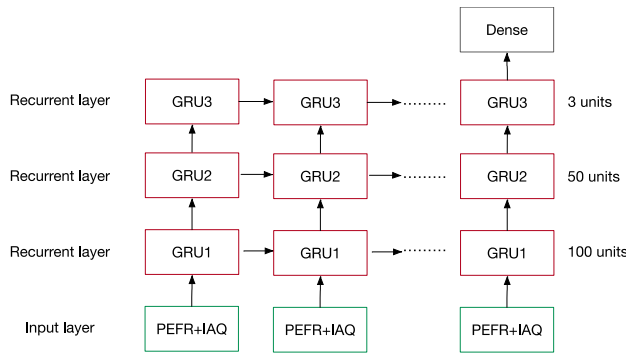
IEEE Access

J. Woo *et al.*: Forecasting Effects of Real-Time Indoor PM2.5 on PEFR of Asthmatic Children in Korea: Deep Learning Approach

FIGURE 12. The best architecture of RNN model for PEFR prediction.

TABLE 4. Comparison of the performance of prediction with RNN models with that of DNN ones.

	RNN	DNN
RMSE of Granular IAQ	42.5	80.8

TABLE 5. Specification of RNN and DNN models with the best performance.

Model	Settings
layer 1: gru or dense	unit100, relu activation
layer 2: gru or dense	unit50, relu activation
layer 3: gru or dense	unit30, relu activation
layer 4: dense	unit 1
optimizer	adam(learning rate=0.0001, beta1=0.9, beta2=0.999, decay=0.001)

the model extensively focuses on 3 feature extractions. First, the Adam optimizer is used for parameter optimization with a learning rate of 0.0001 with a decay rate of 0.001, beta1 (the exponential decay rate for the first moment estimates) of 0.9 and beta2 (the exponential decay rate for the second-moment estimates) of 0.999.

IV. CONCLUSION

Despite the sizable studies on the link between indoor air pollution and the incidence or exacerbation of allergic diseases, there is still limited availability of big real-time data and accurate predictive models. In this study, the PEFR data at 4 months from asthmatic children were matched with the PM2.5 concentrations in real-time and better prediction of respiratory symptoms was observed using deep learning models.

We developed a personalized model to predict the peak expiratory flow rate the next day, considering indoor air quality data including PM2.5, humidity, temperature, and CO_2 level in previous days. In addition, we developed 3 models that differ from the lookback method to observe the past data and derive patterns from them. These models were developed to investigate whether IAQ affects future PEFR. The first

model used the daily PEFR only, the second model used the daily IAQ, and the third model used every 10-minute IAQ. The comparison between the second and third models showed that fine granular examination of IAQ is useful for prediction of PEFR. We found that the previous experience under certain air conditions would influence the PEFR of the following days. In addition, the influence estimation became more accurate when examining the fine granular IAQ rather than the aggregated daily IAQ.

Data forecasting and real-time indoor air monitoring using deep learning have shown to produce vital information that would alert patients to take necessary medications to prevent falling sick and assist them in improving their indoor environments.

In this current study, we adopted a RNN with GUR unit. As a future study, we will compare the more advanced model 1-Dimensional Convolutional Neural Network (1D CNN) and WaveNet that incorporates long-term sequences. Using the long-term sequence model, we will expand the lookback period for 10-minute model from a day to several days.

This kind of prediction is beneficial to public health in reducing the burden of asthma disease and making better use of limited resources to treat asthma patients [46]. In addition, this approach could play a cutting-edge role for scientific data-driven medical decision-making, as well as asthma exacerbation prevention activities, after successfully clarifying deep learning algorithms using a larger sample size for whom pulmonary function and real-time personal exposure to PM data are available.

REFERENCES

[1] D. Mannino and A. Buist, "Global burden of COPD: Risk factors, prevalence, and future trends," *Lancet*, vol. 370, no. 9589, pp. 765–773, 2007, doi: 10.1016/s0140-6736(07)61380-4.

[2] GBD 2015 Chronic Respiratory Disease Collaborators, "Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: A systematic analysis for the global burden of disease study 2015," *Lancet. Respiratory Med.*, vol. 5, no. 9, pp. 691–706, 2017, doi: 10.1016/s2213-2600(17)30293-x.

[3] *Vital Signs: Asthma in Children—United States, 2001–2016*, Centers Disease Control Prevention (CDC), Washington, DC, USA, 2018.

[4] J. Y. Jang, "Korean guideline for asthma," *Korean J. Med.*, vol. 90, no. 4, pp. 298–306, 2016, doi: 10.3904/kjm.2016.90.4.298.

[5] Y. M. Cho, C.-B. Kim, K. N. Yeon, E. S. Lee, and K. Kim, "Trends in the prevalence of childhood asthma in Seoul metropolitan city, Korea: The Seoul atopy · Asthma-friendly school project," *J. Preventive Med. Public Health*, vol. 51, no. 6, pp. 275–280, 2018, doi: 10.3961/jpmph.18.090.

[6] I. Jhun, J. M. Gaffin, B. A. Coull, M. F. Huffaker, C. R. Petty, W. J. Sheehan, S. N. Baxi, P. S. Lai, C.-M. Kang, J. M. Wolfson, D. R. Gold, P. Koutrakis, and W. Phipatanakul, "School environmental intervention to reduce particulate pollutant exposures for children with asthma," *J. Allergy Clin. Immunol., Pract.*, vol. 5, no. 1, pp. 154–159, 2017, doi: 10.1016/j.jaip.2016.07.018.

[7] E. Lee, A. Kim, Y.-M. Ye, S.-E. Choi, and H.-S. Park, "Increasing prevalence and mortality of asthma with age in Korea, 2002–2015: A nationwide, population-based study," *Allergy Asthma Immunol. Res.*, vol. 12, no. 3, pp. 467–484, May 2020, doi: 10.4168/aair.2020.12.3.467.

[8] S. Ram, W. Zhang, M. Williams, and Y. Pengetnze, "Predicting asthma-related emergency department visits using big data," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 4, pp. 1216–1223, Jul. 2015, doi: 10.1109/JBHI.2015.2404829.

J. Woo *et al.*: Forecasting Effects of Real-Time Indoor PM2.5 on PEFR of Asthmatic Children in Korea: Deep Learning Approach

IEEE*Access*

[9] D. Price, A. M. Wilson, A. Chisholm, A. Rigazio, A. Burden, M. Thomas, and C. King, "Predicting frequent asthma exacerbations using blood eosinophil count and other patient data routinely available in clinical practice," *J. Asthma Allergy*, vol. 9, pp. 1–12, Jan. 2016, doi: 10.2147/jaa.S97973.

[10] M. S. Harkins, K. Fiato, and G. K. Iwamoto, "Exhaled nitric oxide predicts asthma exacerbation," *J. Asthma, Off. J. Assoc. Care Asthma*, vol. 41, no. 4, pp. 471–476, Jan. 2004, doi: 10.1081/jas-120033990.

[11] M. Xu, K. G. Tantisira, A. Wu, A. A. Litonjua, J.-H. Chu, B. E. Himes, A. Damask, and S. T. Weiss, "Genome wide association study to predict severe asthma exacerbations in children using random forests classifiers," *BMC Med. Genet.*, vol. 12, no. 1, pp. 1–8, Dec. 2011, doi: 10.1186/1471-2350-12-90.

[12] R. J. B. Loymans, P. J. Honkoop, E. H. Termeer, J. B. Snoeck-Stroband, W. J. J. Assendelft, T. R. J. Schermer, K. F. Chung, A. R. Sousa, P. J. Sterk, H. K. Reddel, J. K. Sont, and G. T. Riet, "Identifying patients at risk for severe exacerbations of asthma: Development and external validation of a multivariable prediction model," *Thorax*, vol. 71, no. 9, pp. 838–846, Sep. 2016, doi: 10.1136/thoraxjnl-2015-208138.

[13] R. Sato, K. Tomita, H. Sano, H. Ichihashi, S. Yamagata, A. Sano, T. Yamagata, T. Miyara, T. Iwanaga, M. Muraki, and Y. Tohda, "The strategy for predicting future exacerbation of asthma using a combination of the asthma control test and lung function test," *J. Asthma*, vol. 46, no. 7, pp. 677–682, Jan. 2009, doi: 10.1080/02770900902972160.

[14] C. Kim, Y. Son, and S. Youm, "Chronic disease prediction using character-recurrent neural network in the presence of missing information," *Appl. Sci.*, vol. 9, no. 10, p. 2170, May 2019, doi: 10.3390/app9102170.

[15] D. H. Arnold, M. R. Sills, and C. G. Walsh, "The asthma prediction rule to decrease hospitalizations for children with asthma," *Current Opinion Allergy Clin. Immunol.*, vol. 16, no. 3, pp. 201–209, Jun. 2016, doi: 10.1097/ACI.0000000000000259.

[16] A. H. YoussefAgha, W. P. Jayawardene, D. K. Lohrmann, and G. S. El Afandi, "Air pollution indicators predict outbreaks of asthma exacerbations among elementary school children: Integration of daily environmental and school health surveillance systems in Pennsylvania," *J. Environ. Monit.*, vol. 14, no. 12, pp. 3202–3210, 2012, doi: 10.1039/C2EM30430A.

[17] J. Woo, G. Rudasingwa, and S. Kim, "Assessment of daily personal PM$_{2.5}$ exposure level according to four major activities among children," *Appl. Sci.*, vol. 10, no. 1, p. 159, Dec. 2019, doi: 10.3390/app10010159.

[18] K. Isiugo, R. Jandarov, J. Cox, P. Ryan, N. Newman, S. A. Grinshpun, R. Indugula, S. Vesper, and T. Reponen, "Indoor particulate matter and lung function in children," *Sci. Total Environ.*, vol. 663, pp. 408–417, May 2019, doi: 10.1016/j.scitotenv.2019.01.309.

[19] S. J. Galli, "Toward precision medicine and health: Opportunities and challenges in allergic diseases," *J. Allergy Clin. Immunol.*, vol. 137, no. 5, pp. 1289–1300, 2016, doi: 10.1016/j.jaci.2016.03.006.

[20] R. Khalili, S. M. Bartell, X. Hu, Y. Liu, H. H. Chang, C. Belanoff, M. J. Strickland, and V. M. Vieira, "Early-life exposure to PM$_{2.5}$ and risk of acute asthma clinical encounters among children in Massachusetts: A case-crossover analysis," *Environ. Health*, vol. 17, no. 1, p. 20, Dec. 2018, doi: 10.1186/s12940-018-0361-6.

[21] J. Wu, T. Zhong, Y. Zhu, D. Ge, X. Lin, and Q. Li, "Effects of particulate matter (PM) on childhood asthma exacerbation and control in Xiamen, China," *BMC Pediatrics*, vol. 19, no. 1, pp. 1–11, Dec. 2019, doi: 10.1186/s12887-019-1530-7.

[22] H. Lim, H.-J. Kwon, J.-A. Lim, J. H. Choi, M. Ha, S.-S. Hwang, and W.-J. Choi, "Short-term effect of fine particulate matter on children's hospital admissions and emergency department visits for asthma: A systematic review and meta-analysis," *J. Preventive Med. Public Health*, vol. 49, no. 4, pp. 205–219, Jul. 2016, doi: 10.3961/jpmph.16.037.

[23] S. Kim, J. Lee, S. Park, G. Rudasingwa, S. Lee, S. Yu, and D. H. Lim, "Association between peak expiratory flow rate and exposure level to indoor PM$_{2.5}$ in asthmatic children, using data from the escort intervention study," *Int. J. Environ. Res. Public Health*, vol. 17, no. 20, p. 7667, Oct. 2020, doi: 10.3390/ijerph17207667.

[24] M. Hulin, D. Caillaud, and I. Annesi-Maesano, "Indoor air pollution and childhood asthma: Variations between urban and rural areas," *Indoor Air*, vol. 20, no. 6, pp. 502–514, Dec. 2010, doi: 10.1111/j.1600-0668.2010.00673.x.

[25] E. Cleary, M. Asher, R. Olawoyin, and K. Zhang, "Assessment of indoor air quality exposures and impacts on respiratory outcomes in River Rouge and Dearborn, Michigan," *Chemosphere*, vol. 187, pp. 320–329, Nov. 2017, doi: 10.1016/j.chemosphere.2017.08.091.

[26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[27] I. M. Baytas, C. Xiao, X. Zhang, F. Wang, A. K. Jain, and J. Zhou, "Patient subtyping via time-aware LSTM networks," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Halifax, NS, Canada, Aug. 2017, pp. 65–74, doi: 10.1145/3097983.3097997.

[28] C. Xiao, E. Choi, and J. Sun, "Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1419–1428, 2018, doi: 10.1093/jamia/ocy068.

[29] L. Liu, W. Liu, Y. Zheng, H. Ma, and C. Zhang, "Third-eye: A mobilephone-enabled crowdsensing system for air quality monitoring," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–26, Mar. 2018, doi: 10.1145/3191752.

[30] I. K. Ahani, M. Salari, and A. Shadman, "An ensemble multi-step-ahead forecasting system for fine particulate matter in urban areas," *J. Cleaner Prod.*, vol. 263, Aug. 2020, Art. no. 120983, doi: 10.1016/j.jclepro.2020.120983.

[31] K. S. Harishkumar, K. M. Yogesh, and I. Gad, "Forecasting air pollution particulate matter (PM$_{2.5}$) using machine learning regression models," *Proc. Comput. Sci.*, vol. 171, pp. 2057–2066, Jan. 2020, doi: 10.1016/j.procs.2020.04.221.

[32] W. G. Cobourn, "An enhanced PM$_{2.5}$ air quality forecast model based on nonlinear regression and back-trajectory concentrations," *Atmos. Environ.*, vol. 44, no. 25, pp. 3015–3023, Aug. 2010, doi: 10.1016/j.atmosenv.2010.05.009.

[33] R. Wongsathan and I. Seedadan, "A hybrid ARIMA and neural networks model for PM-10 pollution estimation: The case of Chiang Mai city moat area," *Proc. Comput. Sci.*, vol. 86, pp. 273–276, Jan. 2016, doi: 10.1016/j.procs.2016.05.057.

[34] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 4–21, Jan. 2017, doi: 10.1109/JBHI.2016.2636665.

[35] C. Kim, Y. Son, and S. Youm, "Chronic disease prediction using character-recurrent neural network in the presence of missing information," *Appl. Sci.*, vol. 9, no. 10, p. 2170, May 2019, doi: 10.3390/app9102170.

[36] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019, doi: 10.1016/j.ymssp.2018.05.050.

[37] A. I. Messinger, R. R. Deterding, and S. J. Szefler, "Bringing technology to day-to-day asthma management," *Amer. J. Respiratory Crit. Care Med.*, vol. 198, no. 3, pp. 291–292, Aug. 2018, doi: 10.1164/rccm.201805-0845ED.

[38] Y. Cheng, F. Wang, P. Zhang, and J. Hu, "Risk prediction with electronic health records: A deep learning approach," in *Proc. SIAM Int. Conf. Data Mining*, Jun. 2016, pp. 432–440, doi: 10.1137/1.9781611974348.49.

[39] K. M. Gulla and S. K. Kabra, "Peak expiratory flow rate as a monitoring tool in asthma," *Indian J. Pediatrics*, vol. 84, no. 8, pp. 573–574, 2017, doi: 10.1007/s12098-017-2398-x.

[40] G. Luo, B. L. Stone, B. Fassl, C. G. Maloney, P. H. Gesteland, S. R. Yerram, and F. L. Nkoy, "Predicting asthma control deterioration in children," *BMC Med. Informat. Decis. Making*, vol. 15, no. 1, Dec. 2015, Art. no. 84, doi: 10.1186/s12911-015-0208-9.

[41] *Global Strategy for Asthma Management and Prevention*, Global Initiative Asthma, Fontana, CA, USA, 2019. Accessed: Jul. 31, 2020. [Online]. Available: https://ginasthma.org/wp-content/uploads/2019/06/GINA-2019-main-report-June-2019-wms.pdf

[42] *Spirometry in Practice: A Practical Guide to Using Spirometry in Primary Care*, Brit. Thoracic Soc. (BTS) COPD Consortium, London, U.K., 2005.

[43] Smart One. (2018). *FEV1 and Peak Flow on Your Smart Phone*. Accessed: Jan. 28, 2021. Available:. [Online]. Available: https://www.mirsmartone.com/wp-content/uploads/2018/11/SmartOne_ENG_201705-1.pdf

[44] S. Kim, S. Park, and J. Lee, "Evaluation of performance of inexpensive laser based PM$_{2.5}$ sensor monitors for typical indoor and outdoor hotspots of South Korea," *Appl. Sci.*, vol. 9, no. 9, p. 1947, May 2019, doi: 10.3390/app9091947.

[45] Korea Law Translation Center. (2019). *Special Act on the Reduction and Management of Fine Dust*. [Online]. Available: https://elaw.klri.re.kr/kor_mobile/viewer.do?hseq=50524&type=part&key=39

[46] K. L. Khatri and L. S. Tamil, "Early detection of peak demand days of chronic respiratory diseases emergency department visits using artificial neural networks," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 285–290, Jan. 2018, doi: 10.1109/JBHI.2017.2698418.

IEEE Access

J. Woo *et al.*: Forecasting Effects of Real-Time Indoor PM2.5 on PEFR of Asthmatic Children in Korea: Deep Learning Approach

JIYOUNG WOO received the B.S., M.S., and Ph.D. degrees in industrial engineering from KAIST in 2000, 2002, and 2006, respectively. From 2008 to 2010, she was a Researcher with the Artificial Intelligence Laboratory, The University of Arizona, USA. From 2010 to 2016, she was a Research Professor at the Graduate School of Information Security, Korea University. She is currently a Professor with the Department of Big Data Engineering, Soonchunhyang University. Her research interests include data mining on biomedical data and sensor data.

JI-HYUN LEE received the B.Sc. degree from the Department of Mathematics, Soonchunhyang University, Asan-si, South Korea, in 2018, and the M.Sc. degree from the SCH Media Laboratories, Department of Future Convergence Technology, Soonchunhyang University, in 2020.

She is currently a Medical AI Research Engineer with the Department of Radiology, Seoul National University College of Medicine, SMG-SNU Boramae Medical Center, Seoul, Republic of Korea. Her research interests include artificial intelligence, deep learning, and medical image.

YEONJIN KIM received the Bachelor of Science degree in environmental health sciences from Soonchunhyang University, South Korea, where she is currently pursuing the master's degree with the Department of Environmental Health System ICT, Graduate School. Her research interests include environmental health impact assessment and air quality.

GUILLAUME RUDASINGWA received the bachelor's degree in medicine and health science from the Department of Environmental Health Sciences, University of Rwanda, in 2015, and the master's degree in public health (MPH) majoring in chronic disease epidemiology from Seoul National University, South Korea, in 2019. He is currently working with the Integrated Research Center for Risk Assessment, Soonchunhyang University. His research interests include cohort and nested case-control studies, chronic disease epidemiology, and infectious disease epidemiology.

DAE HYUN LIM received the bachelor's degree from the Korea University College of Medicine, Seoul, and the master's and doctorate degrees in pediatrics from the School of Medicine, Inha University, Incheon, South Korea. He was a Visiting Scientist at the Allergy Division, Department of Internal Medicine, University of California at San Diego, San Diego, USA, from 2003 to 2004. He is currently a Professor with the Department of Pediatrics, School of Medicine, Inha University. His research interests include the areas of environments, especially air quality, and pediatric allergy such as asthma, rhinitis, dermatitis, and respiratory diseases.

SUNGROUL KIM received the Ph.D. degree from the Department of Environmental Health Science and Engineering, Johns Hopkins Bloomberg School of Public Health (JHSPH), Baltimore, MD, USA. He completed the certificate program ''Risk Sciences and Public Policy'' at JHSPH, where he was a Research Associate Professor at the Department of Epidemiology, from 2007 to 2009. Currently, he is teaching at the BK21 Department, Graduate School, Soonchunhyang University, South Korea. He has been a principal investigator of research projects establishing platforms for ICT-based air and health symptom monitoring among asthmatics funded by the Korea Environmental Industry and Technology Institute, Ministry of Environment, South Korea, since 2016.

● ● ●