Article

Assessment and Calibration on Low-cost PM2.5 Sensor using Machin Learning (Hybrid-LSTM Neural Network): Feasibility Study to Build Air Quality Monitoring System

Donggeun Park *, †**, Geon-Woo Yoo †, Seong-Ho Park † and Jong-Hyeon Lee†**

Research Institute of Environmental Health and Safety, 410, Jeongseojin-ro, Seo-gu, Incheon 404-844, Korea
* Correspondence: dgpark94@pusan.ac.kr
† These authors contributed equally to this work.

Abstract: Although commercially-available low-cost air quality sensors have low accuracy, the sensor system are being used to collect the data for the regulation of PM2.5 emission caused by industrial activities or to estimate the personal exposure for PM2.5. In this work, to solve the accuracy problem of low-cost PM sensor, we developed a new PM22.5 calibration model by combining the deep neural network (DNN) optimized in calibration problem and a LSTM optimized in time-dependent characteristics. First, two datasets were generated to test the accuracy performance and generalization performance of the PM2.5 calibration machine learning (ML) model. The PM2.5 concentrations, temperature and humidity by low-cost sensor and gravimetric-based PM2.5 measuring instrument were sampled for a sufficiently long time. The proposed model was compared with benchmark (multiple linear regression model) and low-cost sensor results. For root mean square error (RMSE) for PM2.5 concentrations, the proposed model reduced 41-60% of error compared to the raw data of low-cost sensor, and reduced 30-51% of error compared to the benchmark model. R^2 of ML model, MLR and raw data were 93, 80 and 59 %. Also, the developed model still showed consistent calibration performance when calibrated with new sensors in different locations. Low-cost sensors combined with ML model not only can improve the calibration performance of benchmark, but also can be applied to the sensor monitoring systems for various epidemiologic investigations and regulatory decisions.

Keywords: machine learning; deep learning; calibration; air quality; low-cost sensors; exposure assessment

1. Introduction

Air pollution caused by industrialization and urbanization is causing serious environmental and health problems. For example, fine particulate matter (PM) is generated from various emission sources of industrial activities such as industry, transportation and combustion. In particular, fine dust with a diameter of less than 2.5 μm (PM2.5) causes various diseases such as cardiovascular diseases, asthma, and neurotoxicity because it is directly exposed to the lungs and circulatory system. Therefore, it is very important to get the data for regulation on industrial emission by monitoring the PM2.5 concentration generated by the emission activity [1].

In South Korea, gravimetric-based PM2.5 measuring instrument has been used as national reference method (NRM) to monitor the PM2.5 concentrations. However, installing NRM equipment at each of short-distance sampling locations is expensive (>$10,000). This limits obtaining PM2.5 information at the community level.

Light-scattering low-cost PM2.5 sensors can make a possibility to solve the cost problem. Since the low-cost sensor can obtain the PM concentration in real-time, it has been used in various studies such as personal exposure assessment [3,4], indoor exposure estimation [5] and outdoor monitoring [6,7,8]. However, low-cost sensors are sensitive to environmental variables such as temperature and humidity due to their light scattering

method. Badura et al. [9] conducted the validation test to evaluate the reliability of the low-cost sensors in outdoor field during a long period using the national standard measuring equipment. Above 80% relative humidity, raw data by low-cost sensors observed an apparent overestimation of PM2.5 concentration measurements.

Vogt et al. [10] performed the comparison of three models of low-cost PM2.5 sensors (Plantower 5003, Sensirion SPS30 and Alphasense OPC-N3) against the gravimetric device in outdoor field. Among the low-cost sensors, the SPS 30 sensor showed high accuracy in PM2.5 concentration measurement and high correlation among individual sensors. However, it has been shown that the low-cost sensor has lower accuracy than the national standard measurement network equipment for a certain measurement period due to the limitations of the physical characteristics of the sensor.

Zusmana et al. [11] developed metropolitan region-specific calibration models based on the multi-linear regression method (MLR) and the time-series data by various low-cost sensors (PM2.5, temperature and humidity) and the NRM network equipment (PM2.5) to solve the sensitivity problem driven by environmental variables. The calibration model confirmed the possibility of applying a low-cost sensor at the community-level by solving the accuracy degradation caused by the physical characteristics of low-cost sensor. However, the metropolitan region-specific calibration model still showed low accuracy (R2 = 0.67-0.84) in a specific data period.

In this study, we develop a low-cost PM2.5 calibration model using machine learning (ML) methodology to solve the limitation of accuracy degradation due to the light-scatter physical characteristics of low-cost sensors. This work has two purposes; (1) development on a low-cost calibration machine learning (ML) model with higher accuracy and comparison with the MLR modeling method and proposed ML model. (2) A generalization test to show if the PM2.5 calibration machine learning algorithm is applicable to the new spatial-temporal field condition.

The process of this study is shown in Figure 1. First, low-cost Sensirion SPS30 and NRM equipment were colocated to develop the ML model. If high concentrations of PM2.5 are not sampled, incorrect performance evaluation results may be obtained [11]. Therefore, the experiment is carried out until PM2.5 samples are obtained more than 50 $\mu g \, / \, m^3$. ML model and MLR model are developed based on the obtained data sample (train set), and model performance is compared using the other independent data not used for model development (test set). The MLR is used as a benchmark for evaluating the newly developed calibration ML model. Finally, generalization tests are performed to confirm that the optimized ML model is applicable in the new location-period.
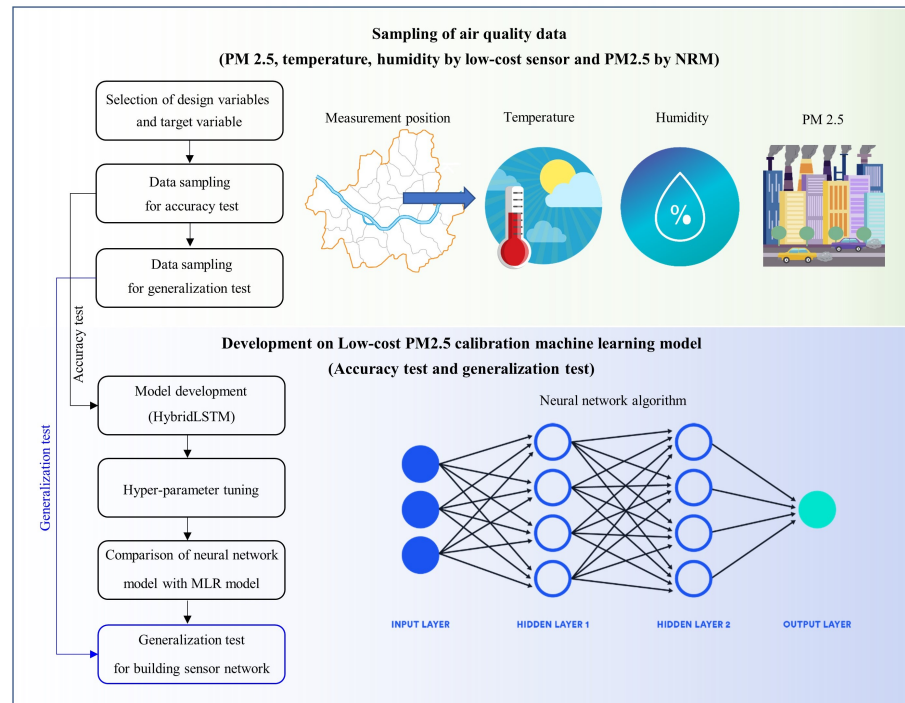
Figure 1. Research flow chart. This work is performed in two step; (1) Data collection for PM2.5, temperature and humidity by low-cost sensor and PM2.5 by gravimetric instrument with high accuracy. (2) Machine learning model development based on the collected data set. Calibration performance from the developed model is compared with raw data by low-cost sensor and calibration results by benchmark method (multi-linear regression method). Also, the proposed machine learning method extends the new sensor to test model generalization.

The novelty of this study is divided into three categories; (1) Improvement of the calibration performance against the existing calibration model (MLR) by using a new machine learning algorithm. it is very important because it does not only provide reliable community-level monitoring, but also can help exposure assessment in epidemiological studies. (2) Development on low-cost sensor ML model using sensor data and a gravimetric sampling device for PM2.5. To our knowledge, there is very scarce literature calibrating low-cost sensor systems against reference gravimetric methods. (3) Generalization test for whether the developed ML model is applicable at the new measurement location and time. Most of papers for the low-cost sensor calibration confirm the validity of the calibration model at the only same location and time, but not at the new locations and times. Generalization test is essential to build a monitoring sensor-network that provides PM2.5 concentrations at various locations.

2. Methods

2.1. Data sampling to develop calibration machine learning (ML) model

2.1.1. Air quality measurement instruments

In general, ML algorithm functions the relationship between input variables and output variables. In this study, input variables were set as PM2.5 by Sensirion low cost SPS 30 (<$50) and temperature and humidity by Sensirion SHT85 (< $30) to model the complex relation for environmental variables and PM2.5 of light-scatter method.

It is very important to have consistent precision among low-cost sensors in order to build a monitoring sensor network system by ML model. Because the low-cost Sensirion SPS 30 has excellent inter-sensor precision [10], the ML model based on SPS 30 have possibility of maintaining the consistent performance even with new sensors. Therefore, the

PM2.5 measurement results by the SPS 30 sensor is set as the input variable. Environmental variables such as temperature and humidity have an effect on decreasing the accuracy of low-cost sensor based on light-scatter method [12]. Therefore, two environmental variables were also set as input conditions to model the complex physical characteristics among PM$_{2.5}$, temperature and humidity.

The target variable is PM$_{2.5}$ concentration measured from the gravimetric instrument. The quality of target variable plays an important role to develop ML model with high calibration performance. The gravimetric method is based on TEOM (Tapered Element Oscillating Microbalance) technology, which intakes the atmospheric air through a filter, heats it, continuously measures the filter weight, and calculates the mass concentration of PM in near real time. It has been used in many countries to monitor PM2.5 concentrations in field because TEOM has high accuracy in field test compared the various air quality devices [13].

2.1.2. Dataset for calibration machine learning (ML) modeling

Dataset is required to develop and test the ML model. Dataset includes the labeled time series type by the aforementioned input variables (PM$_{2.5}$ of SPS 30, temperature and humidity of SHT 85) and target variables (TEOM). In general, the dataset is divided into a training set for estimating the ML model parameters and a test set for evaluating the ML model calibration performance. In this study, two datasets were generated to test the accuracy performance and generalization performance of the PM2.5 calibration ML model.

For validating the accuracy performance of ML model, the test set must have different combinations of variables compared with the train set. In addition, ML model must consider adequate design space between training set and test set. For example, if a variable with a higher concentration range than the training set used for model development is input to the developed ML model, the calibration performance has possibility of deteriorate [14]. In other words, the training set must contain a sufficiently high concentration of PM2.5. Also, data set with a small concentration range of PM2.5 may give incorrect evaluations of certain metrics, such as R2 [11].

In this study, Low-cost Sensirion SPS30 and SHT85and NRM equipment were colocated to develop the ML model as shown Figure 2.
The sampled data of time series type (PM$_{2.5}$ of SPS 30, temperature and humidity of SHT 85 and PM$_{2.5}$ of TEOM) are shown in Figure 3. The data sampling period was measured over 110 days. 77 days were designated as training set and 33 days were designated as test set. The maximum PM2.5 concentrations in the train set and test set were sampled for a sufficiently long time to include PM2.5 data higher than 50 $\mu g / m^3$ at least, which is scenario of high concentration determined by world health organization (WHO). Maximum PM$_{2.5}$ concentration measured by Gravimetric method is 115 $\mu g / m^3$. Therefore, the concentration of PM$_{2.5}$ sampled in this work is enough high to validate model calibration performance.
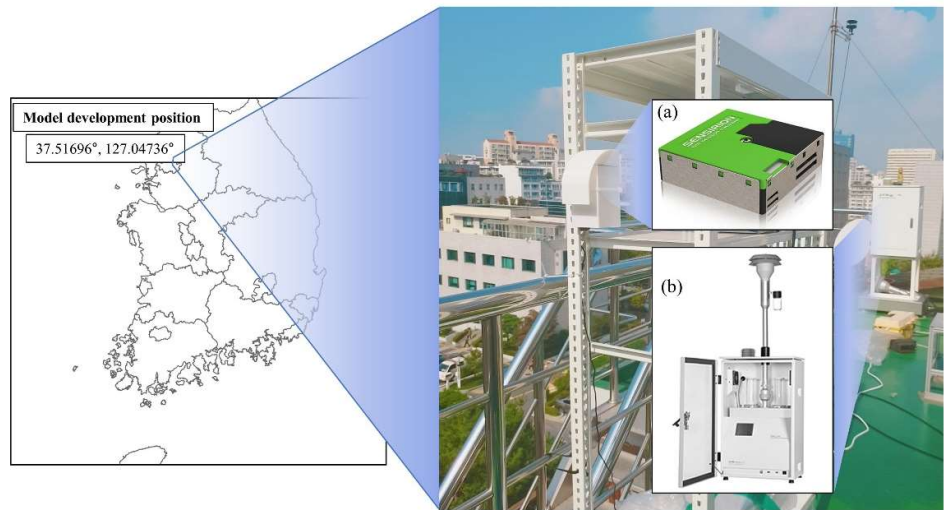
Figure 2. Measurement position and field test setup; (a) low-cost sensor for PM$_{2.5}$, temperature and humidity (b) gravimetric instruments.
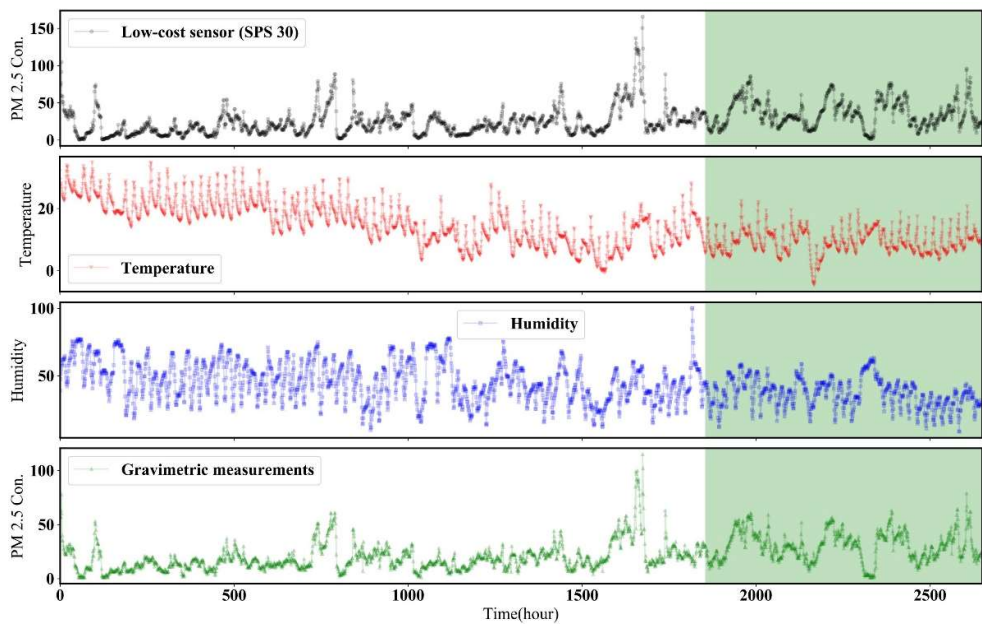

Figure 3. Results of collected data set; From top to bottom, PM$_{2.5}$ of SPS 30, temperature and humidity of SHT 85 and PM$_{2.5}$ of TEOM. Region of white box represents the train set that is used to develop the calibration model. Region of green box represents the test set that is used to evaluate the developed calibration models.

Table 1 represents the statistical information (maximum, minimum, average and standard deviation for the collected dataset) of the corresponding dataset. We correlate the complexity between input variables (data of low cost sensors) and target variable (data of high accuracy device) through ML method based on the data set.

Table 1. Results of statistical information for data set sampled from low cost sensor and gravimetric instrument.

Variables	Minimum	Maximum	Average	Standard deviation
Temperature	-4.713	34.76	13.96	6.96
Humidity	8.55	99.99	43.51	19.45
Low-cost PM$_{2.5}$	0.39	165.56	27.11	19.45
Gravimetric PM$_{2.5}$	1	115	22.15	14.21

In addition, additional dataset was sampled with new sensors in different locations to validate the generalized performance of ML model. Generalization test is essential to build a monitoring sensor-network that provides PM2.5 concentrations at various locations.

2.2. Machine learning algorithm

The measured PM2.5 concentration and environmental variable data (temperature and humidity) have time series characteristics. That is, air quality data has a time-dependent characteristic, which is a relationship between past data and current data. Among machine learning algorithms, the long short-term memory (LSTM) neural network is an algorithm optimized for time-dependent characteristics. In this study, in order to calibrate the low-cost PM2.5 sensors, we develop a new PM2.5 calibration model by customizing the deep neural network (DNN) optimized in calibration problem and a LSTM optimized in time-dependent characteristics. The overall system architecture of the PM2.5 calibration HybridLSTM algorithm proposed in Figure 4.
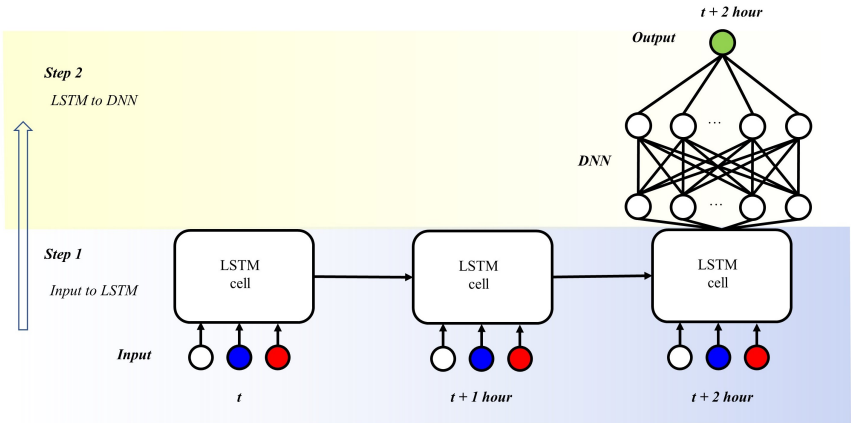

Figure 4. HybridLstm model architecture.

(Step 1) PM2.5, temperature, and humidity of the low-cost sensor data described in Section 2.1 are input to the network in the form of a time series including historical trends for 24 hours. Values entered with historical data provide time-dependent properties between time series data through LSTM cells.

(Step 2) The time-dependent values are passed to the DNN architecture, and the neural network parameters are trained to minimize the differences between the values of the target variables (TEOM PM2.5) and results predicted by the model. The key to the HybridLSTM algorithm is to approach the calibration problem differently from the application of conventional LSTM approach. For example, HybridLSTM algorithm is to make the time series of the target variable (TEOM PM2.5) the same as the last time series of the input variables (low cost PM2.5, temperature and humidity). The results feed the DNN to get the complex non-linearity between input and target. Since the theory for the time-dependent of LSTM [15,16] and the process of calculation of complex nonlinearity between input and output of DNN [17,18] have been well described in many studies, in this study, the description of theory is omitted in order to avoid unnecessary repetition.

Since calibration accuracy by ML model varies depending on the combination of hyper-parameters (such as learning rate, network architecture, batch size, optimization function, etc.), it is very important to find an optimized hyper-parameter. however, there is a limit to comparing a huge number of combinations. Therefore, many studies determine the hyper-parameter by trial and error methods [17,18]. In this study, hyper-parameters with optimal ML calibration performance were determined by changing various hyper-parameter combinations. The hyper-parameter optimization results were summarized in Table 2.

Table 2. Results of hyper-parameter optimization for HybridLstm.

Optimized parameters	Values
Callback	24
Number of layer	5
DNN Node	8/12/24/12/4
Learning rate	0.0065
Batch size	15
Epoch	100
Optimization algorithm	Adam

Callback is a parameter for how long the LSTM cell gives time-dependency, and the number of DNN layer and node are parameters that determine the degree of nonlinearity between the input variable and the output variable. Too many layers cause overfitting and deteriorate the calibration performance of new input data. The learning rate is that the neural network reduces the loss between the input and output. A learning rate having too large prevents the solution from convergence. Batch size represents the size divided among the entire train set for training the neural network. Epoch refers to the number of iterations to train a neural network. The Adam algorithm was used to optimize the neural network because Adam method showed high convergence and accuracy among many algorithms in regression problem [17].

2.3. Benchmark method

The multi-linear regression (MLR) method, which showed high correction performance in the previous studies [11], was used as a benchmark to evaluate the performance of the hybridLSTM model proposed in this study. Benchmark method is the same as the equation below;

$$y = \sum_{i=1}^{3} \omega_i \cdot x_i + b \tag{1}$$

where, x_1, x_2 and x_3 are SPS30 $PM_{2.5}$, temperature and humidity, and y is the result by TEOM equipment $PM_{2.5}$. w and b are parameters optimized by the dataset described above. The benchmark model was developed using the same training data used to develop HybridLSTM model, and the model performance is evaluated using the same test data.

The metrics used for model development and evaluation were R^2 and root mean square error (RMSE). In general, many performance metrics are used to evaluate regression models, but in evaluating sensor calibration performance, two indicators can be sufficiently explained [11]. The metrics for R^2 and RMSE are expressed as follows;

$$R^2 = 1 - \sum_{i=1}^{n} \frac{(t_i - y_i)^2}{(t_i - \bar{t})^2} \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{t}_i)^2} \tag{3}$$

Where, t_i, y_i, $\bar{t}$ and n represent the TEOM sample, the corrected result by the model, the average of the TEOM samples, and the total number of samples, respectively.

3. Results

3.1. Comparison of accuracy among proposed model, benchmark and low-cost sensor

Figure 5 shows the learning process of the ML model during 10 training experiment using the aforementioned optimal hyper-parameter. The validation set determines whether the model have an overfitting problem for new PM2.5 data. The proposed model for the loss of the validation set and the train set sufficiently was converged during the repeated training experiment as shown Figure 5. In other words, the developed model represents a robust model without overfitting. Therefore, researchers can develop and use a consistent model using the hyper parameter optimized in this study to calibrate the low-cost PM2.5 sensors.
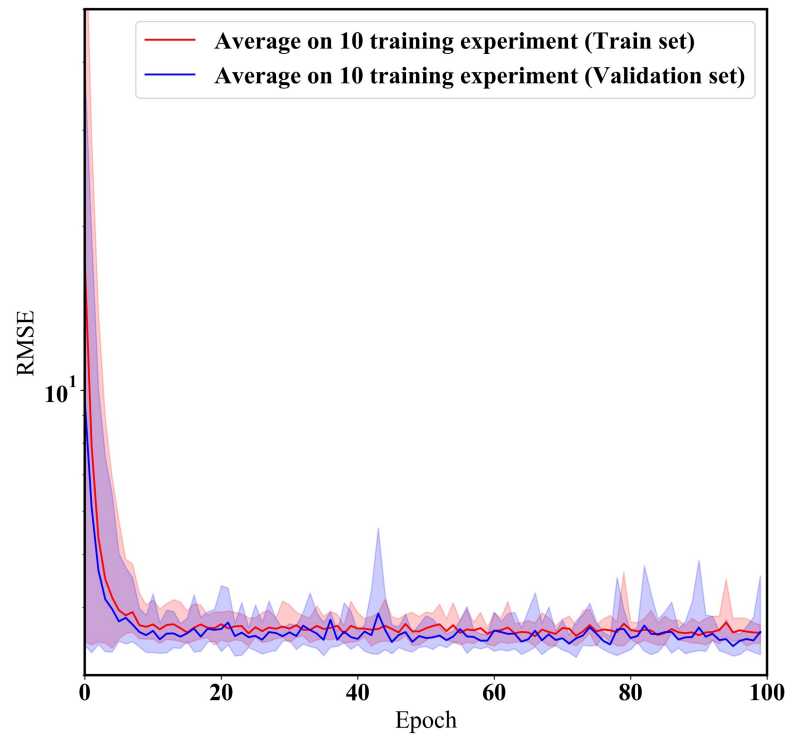

Figure 5. Results of robust model test. The robust test is performed on 10 repeat training experiments to validate that a ML model have the consistent calibration performance.

Figure 6 show the RMSE as a result of calibrating the test set at 7-day intervals using the optimized model, benchmark, and SPS-30 sensor. The proposed model with time-dependent characteristics showed higher calibration performance for all periods than the benchmark model and raw data. The quantitative error reduction rate for each of periods is shown in Table 3. The proposed model reduced 41-60% of error compared to the

raw data (low-cost sensor), and reduced 30-51% of error compared to the calibration results by benchmark model.
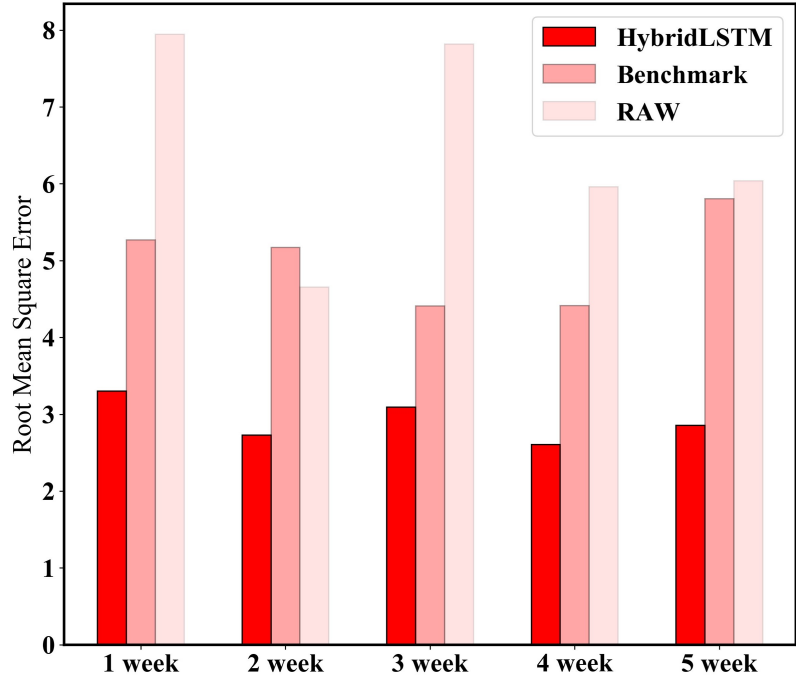

Figure 6. Results of error comparison of HybridLstm, benchmark and raw data at 1-week intervals based on the aforementioned test set.

Table 3. Results of comparison of calibration result of HybridLstm with benchmark(MLR) and low-cost sensor (RAW) in term of decrease rate of RMSE.

Error decrease rate of RMSE	1 week	2 week	3 week	4 week	5 week
$\dfrac{\text{MLR} - \text{HybridLSTM}}{\text{MLR}}$	37.33 %	47.27 %	29.88 %	43.33 %	50.86 %
$\dfrac{\text{RAW} - \text{HybridLSTM}}{\text{RAW}}$	58.45 %	41.4 %	60.46 %	58.05 %	52.76 %

Figure 7 represents the comparison results of the developed model, benchmark model, and raw sensor against all samples of TEOM. Raw data by low-cost PM2.5 sensors showed a huge overestimation in more than 50 um/m3. The incorrect monitoring in high-concentration situations not only leads to incorrect exposure assessment, but also leads to errors in determining government regulation. The benchmark method underestimated compared to the gravimetric method. On the other hand, the proposed model showed small variation results when compared to the TEOM results in the high concentration as well as in the low concentration section. HybridLSTM had the most similar results to TEOM, and R² was about 93 %.
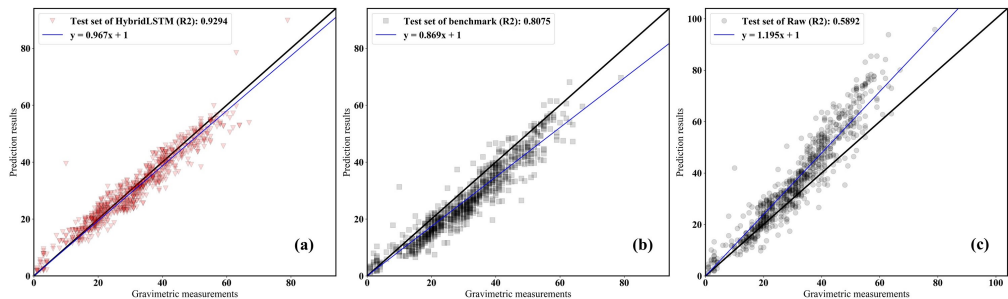

Figure 7. Results of scatter plot for the HybridLSTM (a), benchmark (b) and raw data (c) versus gravimetric measurements.

Figure 8 shows the time series comparison results of the developed model, benchmark model, and raw sensor against TEOM at 1-week intervals. The dot line represents the high PM2.5 concentrations. Based on the dot line, the low-cost sensor results showed the overestimation compared to the gravimetric measurement results. Also, the benchmark method represents the underestimation compared to the gravimetric instrument under dot line. However, the calibration algorithm proposed in this work can calibrate not only the PM2.5 of high concentration but also PM2.5 of low concentration in terms of high accuracy.
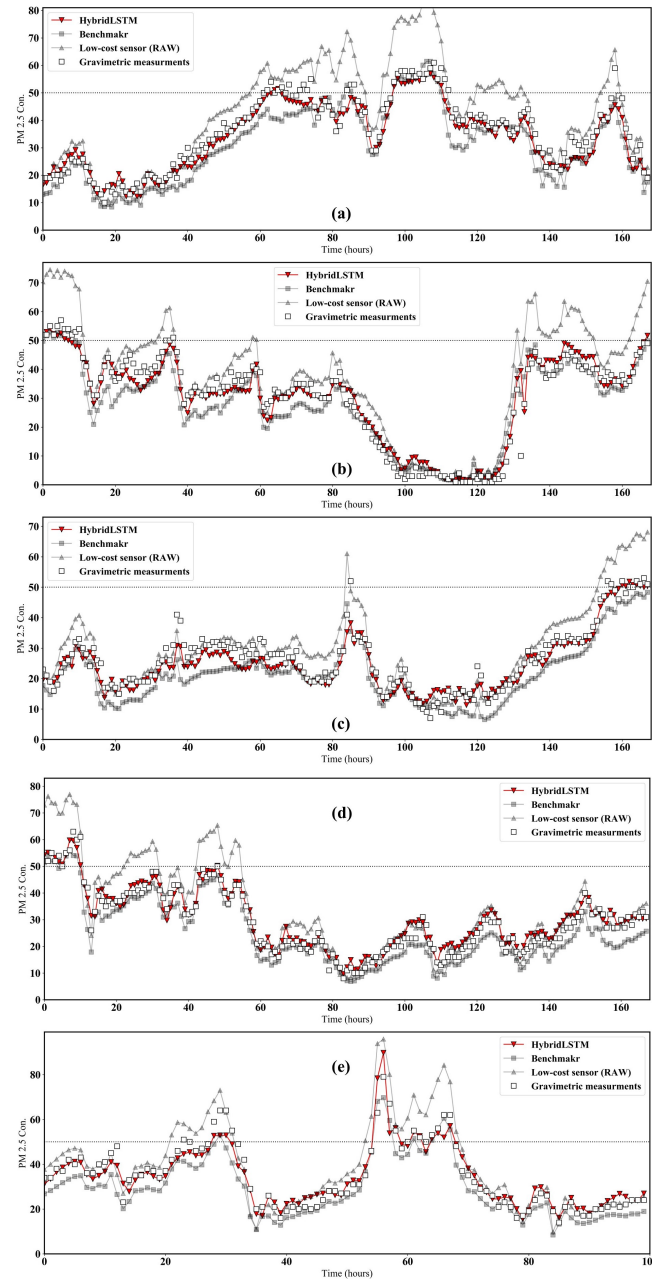

Figure 8. Results of time series comparison results for the HybridLSTM, benchmark and raw data versus gravimetric measurements at 1-week intervals; (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5 weeks. The dot line represents the high PM2.5 concentrations.

3.2. Generalization test of develop machine learning model

A high-reliability sensor monitoring network can be established only when the developed PM calibration model shows generalized calibration performance for new locations and devices. Therefore, in this study, a model generalization test was performed to see if the verified model with high calibration performance consistently maintains high performance in new locations and sensors.

The generalization test was performed at a new location by using deep learning model developed from the previous accuracy evaluation process as shown Figure 9. The two new positions are approximately 10 km apart and are located at the school (Figure 9 (a)) and near the road (Figure 9 (b)), respectively. Samples were collected until the peak PM2.5 concentration reached higher than 70 μg / m³ for correct calibration performance evaluation. The number of samples for generalization test set is 504. The statistical information (minimum, maximum, average and standard deviation) of test set are shown in Figure 9. In near load, the variation of low-cost sensor was higher than the position of school due to the emission of transportation. However, the performance (R2 > 90 %) of the ML model of is maintained despite the two different positional characteristics and new sensor equipment. It was possible to improve the accuracy of low-cost PM2.5 sensors as well as build a reliable sensor monitoring network using the machine learning methodology proposed in this study.
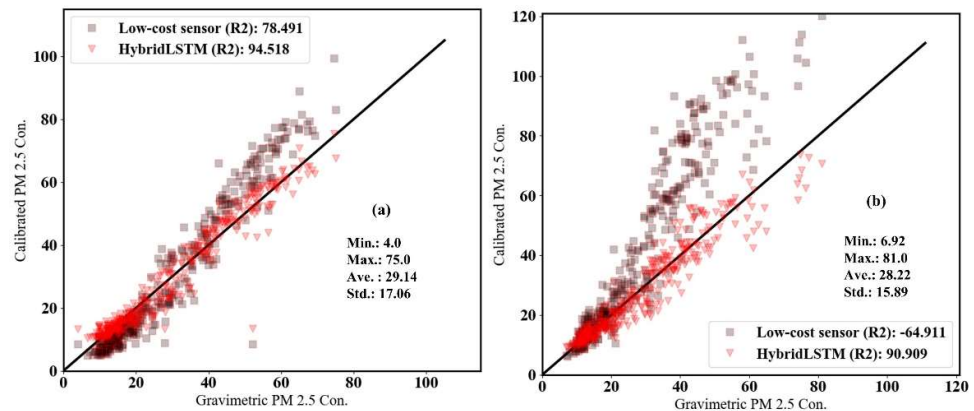

Figure 9. Results of generalization test of optimized ML model at the new location; school (a) and near load (b).

4. Discussion and conclusion

In this study, a new PM2.5 machine learning calibration model (HybridLSTM) was developed and the calibration performance was compared with the existing metropolitan region-specific calibration model and raw data. Also, generalized performance test was performed for validating the possibility of establishing a sensor monitoring network. The results performed are summarized as follows.

(1) HybridLSTM PM2.5 calibration model with time-dependent characteristics showed optimal performance in improving the accuracy of low-cost PM2.5 sensors. For RMSE, the proposed model reduced 41-60% of error compared to the raw data of low-cost sensor, and reduced 30-51% of error compared to the benchmark model. Raw data by low-cost PM2.5 sensors showed a huge overestimation compared to the gravimetric method in samples of high PM2.5 concentrations. The benchmark method showed the underestimated calibration results compared to the measurement results by the gravimetric method. The incorrect monitoring in high-concentration situations not only leads to incorrect exposure assessment, but also leads to errors in determining government regulation. The proposed model showed the small variation with NRM method.

(2) Most of papers for the low-cost sensor calibration confirm the validity of the calibration model at the only same location and time, not the new location and time. When

the developed model was applied to a new position and sensor, it showed consistent calibration performance. Low-cost sensors combined with machine learning models not only exceed the performance of existing benchmarks, but also sensor monitoring systems for various epidemiologic investigations and regulatory decisions can produce higher reliability results.

The proposed model solves the existing accuracy limitations of low-cost sensor and can provide results with high reliability not only for monitoring but also for research in various environmental fields. Although generalized performance was shown in two locations in this study, the method proposed in this study needs to be verified in more locations to build a more reliable sensor monitoring network. Therefore, in future work, we plan to test whether low-cost PM2.5 sensors combined with machine learning at various locations and time including different seasons can be applied to sensor network construction. When constructing a sensor network with high resolution based on high accuracy, we will test the possibility of providing air quality information to areas where sensors are not installed through the interpolation method.

Author Contributions: D.P. was responsible for developing machine learning algorithm, analyzing the results of calibration by models and oversaw the research and provided initial research ideas. G.W.Y. researched the machine learning papers, analyzed results from developed machine learning. S.H.P. collected the data set and prepared the manuscript. J.H.L. provided initial research ideas and editing the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this work are available on request from the corresponding author.

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

References

1. Conti, M.E.; Ciasullo, R.; Tudino, M.B.; Matta, E.J. The industrial emissions trend and the problem of the implementation of the Industrial Emissions Directive (IED). *Air Qual. Atmos. Heal.* **2015**, *8*, 151–161, doi:10.1007/s11869-014-0282-7.
2. Morawska, L.; Thai, P.K.; Liu, X.; Asumadu-Sakyi, A.; Ayoko, G.; Bartonova, A.; Bedini, A.; Chai, F.; Christensen, B.; Dunbabin, M.; et al. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? *Environ. Int.* **2018**, *116*, 286–299, doi:10.1016/j.envint.2018.04.018.
3. Steinle, S.; Reis, S.; Sabel, C.E.; Semple, S.; Twigg, M.M.; Braban, C.F.; Leeson, S.R.; Heal, M.R.; Harrison, D.; Lin, C.; et al. Personal exposure monitoring of PM2.5 in indoor and outdoor microenvironments. *Sci. Total Environ.* **2015**, *508*, 383–394, doi:10.1016/j.scitotenv.2014.12.003.
4. Li, J.; Mattewal, S.K.; Patel, S.; Biswas, P. Evaluation of nine low-cost-sensor-based particulate matter monitors. *Aerosol Air Qual. Res.* **2020**, *20*, 254–270, doi:10.4209/aaqr.2018.12.0485.
5. Tahsiin, F.; Anggraeni, L.; Chandra, I.; Salam, R.A.; Bethaningtyas, H. Analysis of Indoor Air QualityBased on Low-Cost Sensors. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2020**, *10*, 2627–2633, doi:10.18517/ijaseit.10.6.12989.
6. Jiang, Y.; Zhu, X.; Chen, C.; Ge, Y.; Wang, W.; Zhao, Z.; Cai, J.; Kan, H. On-field test and data calibration of a low-cost sensor for fine particles exposure assessment. *Ecotoxicol. Environ. Saf.* **2021**, *211*, 111958, doi:10.1016/j.ecoenv.2021.111958.
7. Kim, S.; Park, S.; Lee, J. Evaluation of performance of inexpensive laser based PM2.5 sensor monitors for typical indoor and outdoor hotspots of South Korea. *Appl. Sci.* **2019**, *9*, doi:10.3390/app9091947.
8. Shi, J.; Chen, F.; Cai, Y.; Fan, S.; Cai, J.; Chen, R.; Kan, H.; Lu, Y.; Zhao, Z. Validation of a light-scattering PM2.5 sensor monitor based on the long-term gravimetric measurements in field tests. *PLoS One* **2017**, *12*, 1–13, doi:10.1371/journal.pone.0185700.
9. Badura, M.; Batog, P.; Drzeniecka-Osiadacz, A.; Modzel, P. Evaluation of low-cost sensors for ambient PM2.5 monitoring. *J. Sensors* **2018**, *2018*, doi:10.1155/2018/5096540.
10. Vogt, M.; Schneider, P.; Castell, N.; Hamer, P. Assessment of low-cost particulate matter sensor systems against optical and gravimetric methods in a field co-location in norway. *Atmosphere (Basel).* **2021**, *12*, doi:10.3390/atmos12080961.

11. Zusman, M.; Schumacher, C.S.; Gassett, A.J.; Spalt, E.W.; Austin, E.; Larson, T. V.; Carvlin, G.; Seto, E.; Kaufman, J.D.; Sheppard, L. Calibration of low-cost particulate matter sensors: Model development for a multi-city epidemiological study. *Environ. Int.* **2020**, *134*, 105329, doi:10.1016/j.envint.2019.105329.

12. Załuska, M.; Gładyszewska-Fiedoruk, K. Regression model of PM2.5 Concentration in a single-family house. *Sustain.* **2020**, *12*, doi:10.3390/su12155952.

13. Gilliam, J.H.; Hall, E.S. Reference and Equivalent Methods Used to Measure National Ambient Air Quality Standards (NAAQS) Criteria Air Pollutants vol.1. **2016**, *I*, 2016–2017, doi:10.13140/RG.2.1.3471.8329.

14. Kim, Y.; Kim, Y.; Yang, C.; Park, K.; Gu, G.X.; Ryu, S. Deep Learning Framework for Material Design Space Exploration using Active Transfer Learning and Data Augmentation. *npj Comput. Mater.*, doi:10.1038/s41524-021-00609-2.

15. Wei, J.; Yang, F.; Ren, X.C.; Zou, S. A short-term prediction model of PM2.5 concentration based on deep learning and mode decomposition methods. *Appl. Sci.* **2021**, *11*, doi:10.3390/app11156915.

16. Yang, G.; Lee, H.M.; Lee, G. A hybrid deep learning model to forecast particulate matter concentration levels in Seoul, South Korea. *Atmosphere (Basel).* **2020**, *11*, doi:10.3390/atmos11040348.

17. Park, D.; Cha, J.; Kim, M.; Go, J.S. Multi-objective optimization and comparison of surrogate models for separation performances of cyclone separator based on CFD, RSM, GMDH-neural network, back propagation-ANN and genetic algorithm. *Eng. Appl. Comput. Fluid Mech.* **2020**, *14*, 180–201, doi:10.1080/19942060.2019.1691054.

18. Park, D.; Go, J.S. Design of cyclone separator critical diameter model based on machine learning and cfd. *Processes* **2020**, *8*, 1–13, doi:10.3390/pr8111521.