Electronic Supplementary Material (ESI) for Lab on a Chip.

Supplemental information

Microfluidic ultrafine particle dosimeter using electrical detection method with machine-learning-aided algorithm for real-time monitoring of particle density and size distribution

Tae-Hoon Lee, Hong-Beom Kwon, Woo-Young Song, Seung-Soo Lee, Yong-Jun Kim*

School of Mechanical Engineering, Yonsei University, Seoul, 03722, Republic of Korea

* To whom correspondence should be addressed: yjk@yonsei.ac.kr

[Lifespan evaluation of the PCB-based chip]

In atmospheric UFPs, ionic matter (e.g., Na^+, NO^{3-}, NH^{4+}, Cl^-, S_4O^{2-}) are important components.[1-3] Thus, we used NaCl UFPs as test aerosol. The mean diameter and number concentration of NaCl UFPs were settled to be 50 nm and 10^4 N·cm^{-3} because these are mean values in urban regions.[4]

UFPs that enter our sensor are mostly deposited on the impaction plate in the PCB-based UFP analysis chip, which causes the current fraction of the aerodynamic particle sizer ($R_{a,2}$) to deviate from the initial measurement value ($R_{a,2,ini}$) by changing the structure of the aerodynamic particle sizer (i.e., overloading). In that case, the PCB-based UFP analysis chip must be replaced. Thus, the elapsed time, which $R_{a,2}$ deviation from $R_{a,2,ini}$ becomes 5 % was decided to be the lifespan of the chip (τ_{life}).

Figure S1(a) shows the long-term monitoring results of the currents from the aerodynamic particle sizer ($I_{a,2}$) and the Faraday cup electrometer (I_f). NaCl UFPs, whose concentration was approximately 13 times higher than the mean values in urban areas (1.3×10^5 N cm^{-3}), were injected. Because the elapsed time that $R_{a,2}$ showed over 5 % deviation was 35 hr, τ_{life} was determined to be 19 days (=13×35 hr).

Figure S1(b) shows the UFPs deposition on the impaction plates in 2nd stage after the monitoring. The overloading of NaCl UFPs lead to a changing of impaction plate structure and deviation of $R_{a,2}$. In the previous studies, deviation from overloading of the impactor has been reported.[5]

As shown in Figure S1(a), overloading of the impaction plate did not result in a notable decrease in the reliability of the current values. Furthermore, it was found that $R_{a,2}$ returned to $R_{a,2,ini}$, because the collected UFPs probably fell off from the impaction plate. Because of these reasons, if moderate accuracy is acceptable, the PCB-based UFP analysis chip is predicted to be usable for a longer time, even if the usage time that can ensure $R_{a,2}$ measurement with high accuracy was set as τ_{life}.

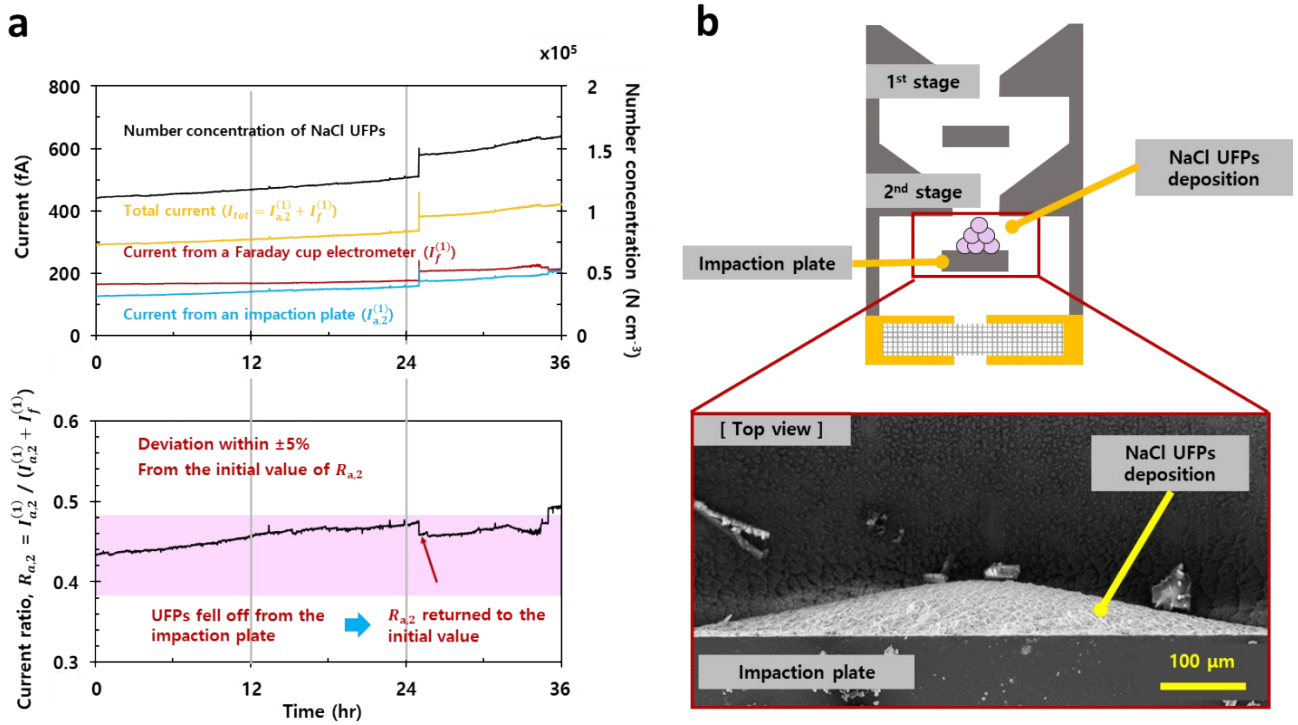
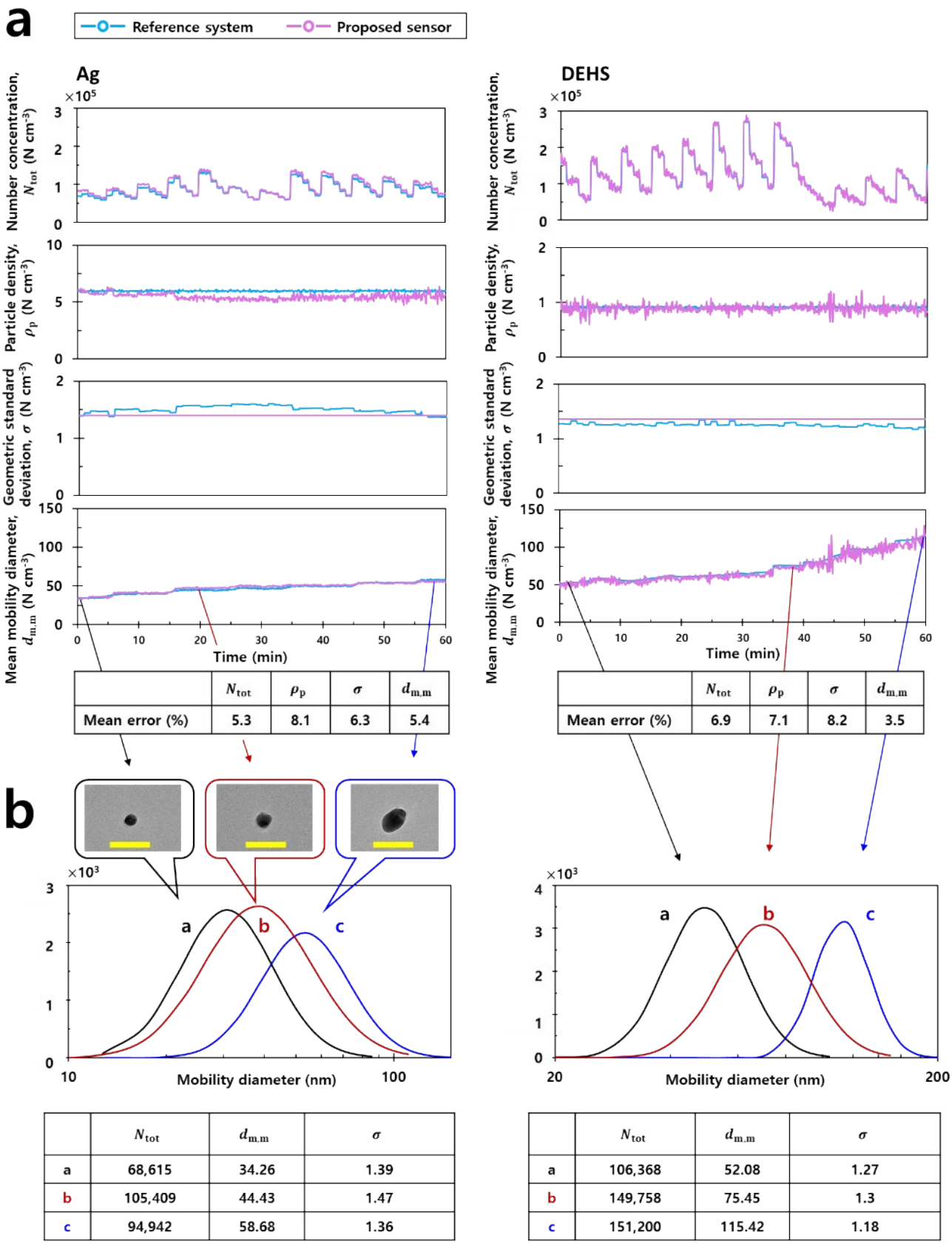


Figure S1. (a) Long-term measurement result from our device when 50-nm-sized NaCl UFPs were introduced; (b) SEM images of the deposition pattern on the impaction plate of the 2nd stage after the experiment.

[Test aerosol information obtained from the reference system]



	N_{tot}	ρ_p	σ	$d_{m,m}$
Mean error (%)	5.3	8.1	6.3	5.4

	N_{tot}	ρ_p	σ	$d_{m,m}$
Mean error (%)	6.9	7.1	8.2	3.5

	N_{tot}	$d_{m,m}$	σ
a	68,615	34.26	1.39
b	105,409	44.43	1.47
c	94,942	58.68	1.36

	N_{tot}	$d_{m,m}$	σ
a	106,368	52.08	1.27
b	149,758	75.45	1.3
c	151,200	115.42	1.18

Figure S2. (a) The four parameters (N_{tot}, , ρ_p, σ, $d_{m,m}$,) of the test aerosols (Ag, DEHS) measured by the reference system (SMPS+ELPI); (b) the size distribution of test aerosols. The inset figure is the TEM images of Ag particles (scale bars equal 100nm).

[One possible way to measure the geometric standard deviation (σ)]

We suggest that measuring σ_g can be achieved by refining our device: 2-stage cascade aerodynamic particle sizer (Figure S3). Owing to the semiconductor manufacturing process, the aerodynamic particle sizer stage can be easily added by modifying the microchannel structures of the microfluidic chip. In the circuit, only one additional channel of the electrometer is required.

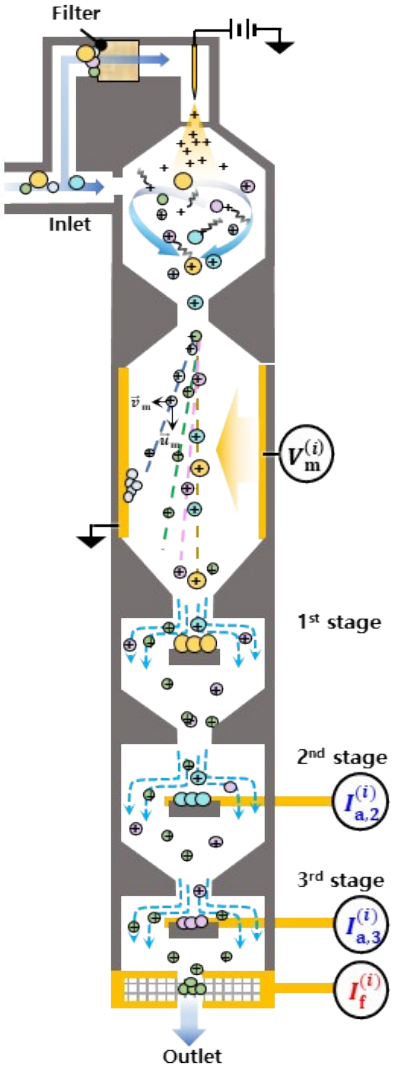


Figure S3. Modified version of our device: adding an additional stage of the aerodynamic particle sizer.

The particle size distribution was assumed as a lognormal distribution in the algorithm. The measured current data set (R_m, $R_{a,2}$, $R_{a,3}$, I_{tot}) will be related to the four parameters of UFPs (i.e., effective density (ρ_p), median mobility diameter ($d_{m,m}$), total number concentration (N_{tot}), and geometric standard deviation (σ_g)), which can be expressed as

$$I_{tot} = I_{a,2}^{(1)} + I_{a,3}^{(1)} + I_f^{(1)} = eQ_s \int \bar{q}(d_m)\left[1 - \eta_{a,1}(d_a)\right] \cdot N_{tot} \cdot D_{log}(d_{m,m}, \sigma, d_m) d\log d_m \ . \quad \text{(S1)}$$

$$R_m = \frac{I_{tot} - \left[I_{a,2}^{(2)} + I_{a,3}^{(2)} + I_f^{(2)}\right]}{I_{tot}} = \frac{\int \bar{q}(d_m) \cdot \left[1 - \eta_{a,1}(d_a)\right] \cdot \eta_m(d_m) \cdot D_{log}(d_{m,m,}\sigma,d_m) d\log d_m}{\int \bar{q}(d_m) \cdot \left[1 - \eta_{a,1}(d_a)\right] \cdot D_{log}(d_{m,m,}\sigma,d_m) d\log d_m}. \quad \text{(S2)}$$

$$R_{a,2} = \frac{I_{a,2}^{(1)}}{I_{tot}} = \frac{\int \bar{q}(d_m) \cdot \left[1 - \eta_{a,1}(d_a)\right] \cdot \eta_{a,2}(d_a) \cdot D_{log}(d_{m,m,}\sigma,d_m) d\log d_m}{\int \bar{q}(d_m) \cdot \left[1 - \eta_{a,1}(d_a)\right] \cdot D_{log}(d_{m,m,}\sigma,d_m) d\log d_m} \quad \text{(S3)}$$

$$R_{a,3} = \frac{I_{a,3}^{(1)}}{I_{tot}} = \frac{\int \bar{q}(d_m) \cdot \left[1 - \eta_{a,1}(d_a)\right] \cdot \left[1 - \eta_{a,1}(d_a)\right] \cdot \eta_{a,3}(d_a) \cdot D_{log}(d_{m,m,}\sigma,d_m) d\log d_m}{\int \bar{q}(d_m) \cdot \left[1 - \eta_{a,1}(d_a)\right] \cdot \left[1 - \eta_{a,2}(d_a)\right] \cdot D_{log}(d_{m,m,}\sigma,d_m) d\log d_m}, \quad \text{(S4)}$$

In the same way like the early algorithm, the four equations can be used to make four databases: $d_{m,m} = f_1(\sigma_g, R_m)$ using equation S2, $\rho_{p,1} = f_2(\sigma_g, d_{m,m}, R_{a,1})$ and $\rho_{p,2} = f_3(\sigma_g, d_{m,m}, R_{a,2})$ using equation S2-4, $N_{tot} = I_{tot}/f_4(\sigma_g, d_{m,m})$ using equation S1. If σ_g and $d_{m,m}$ are retrieved accurately, the output of the 2nd and 3rd databases should be the same ($\rho_{p,1} \approx \rho_{p,2}$). To achieve this, we suggest one of the possible retrieval algorithms (Figure S4). In the algorithm, subscript c indicates parameters are candidate values. $\sigma_{g,c}$ are scanned from the settled range, for example, 1 to 3. Then, subsequent candidate values ($d_{m,m,c}$, $\rho_{p,c}$ $\rho_{p,1}$, $\rho_{p,2}$) and residual error between $\rho_{p,1}$ and $\rho_{p,2}$ (Err_c) are obtained. The algorithm can resolve $d_{m,m,c}$, $\sigma_{g,c}$, $\rho_{p,c}$, and $N_{tot,c}$ at the minimum Err_c, which will be printed as $d_{m,m}$, σ_g, ρ_p and N_{tot}.

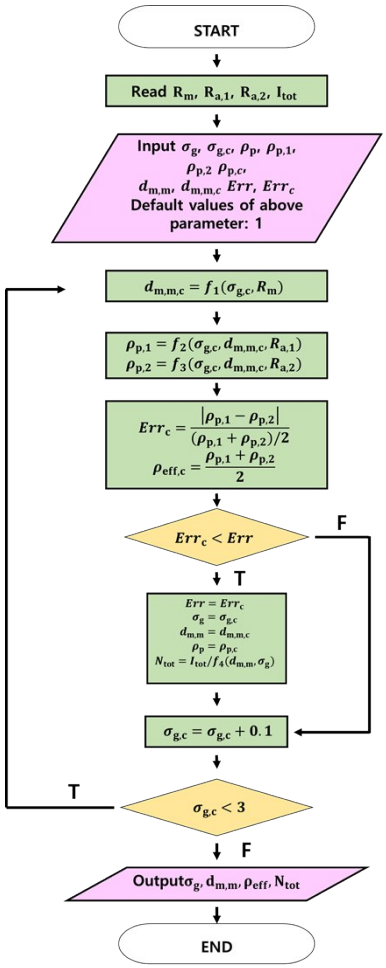


Figure S4. Possible retrieval algorithms for measuring geometric standard deviation.

[Effect on hydroscopic particle growth]

Figure S5 shows the experimental setup for evaluating the effect on hydroscopic particle growth. N_2 (99.9 % purity) carrier gas was supplied to the particle generation and humidty control sections using three mass flow controllers (MFCs; VIC-D220, MFC Korea, KR). In the particle generation section, NaCl particles were generated from the atomizer (model 9302, TSI Inc., USA) where NaCl solution with a concentration of 0.15 mg L^{-1} were splitted into nano droplets by being impacted on the impaction plate at a near-sonic speed. These droplets were dehydrated by passing through the diffusion dryer and became NaCl particles. The relative humidity of Flow #1 that carried NaCl particles was approximately 60 %RH. Meanwhile, Flow #2 was fully dehydrated (0 %RH), and Flow #3 was fully saturated (100 %RH) by passing through a humidifier. In the mixing valve, these three flows were mixed together. The corresponding relative humidty can be estimated by using

$$\%\ RH = \frac{60 \cdot Q_1 + 100 \cdot Q_3}{Q_1 + Q_2 + Q_3}, \tag{S5}$$

where Q_x is the volumetric flow rate of Flow #x. Thus, by adjusting the volumetric flow rates, the relative humidity of the mixed flow entering the measurement section can be controlled.

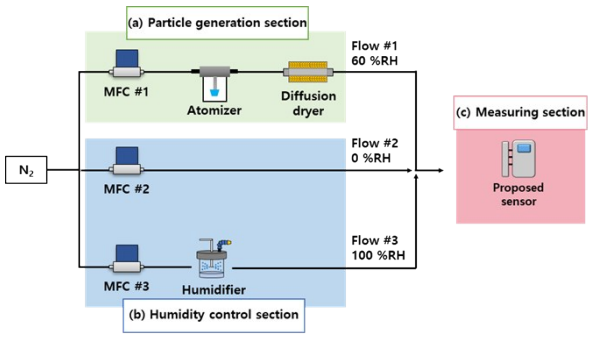


Figure S5. Experimental setup for evaluating the effect on hydroscopic particle growth, which consists of (a) particle generation, (b) humidity control and (c) measurement sections.

Figure S6 shows the real-time measurement results of measuring NaCl particles in a humidity-changing environment. When relative humidity was 16.7 %RH (lower than the efflorescence humidity = 45 %RH), the mean mobility diameter ($d_{m,m}$) and particle density (ρ_p) were respectively 40.7 nm and 1.7 g cm^{-3}, which were close to, but slightly lower than the bulk density of NaCl.[6] The reason for this is that since NaCl particles are brittle, they presumably bounced off the impaction plate and collected in the filter,[7] leading to a lower current fraction of the aerodynamic particle sizer ($R_{a,2}$), and hence lower ρ_p. When relative humidity was 80 %RH (higher than the deliquescence humidity = 75 %RH), they grew into spherical droplets via heterogeneous nucleation. The corresponding $d_{m,m}$ and ρ_p were respectively 54.6 nm and 1.3 g cm^{-3}, which approached to the bulk density of water (1 g cm^{-3}). The above results supported the previous studies, which reported hydroscopic particles in ambient air, mainly composed of ionic matter, tend to grow into droplets by absorbing water vapor in humid environments.[8, 9]

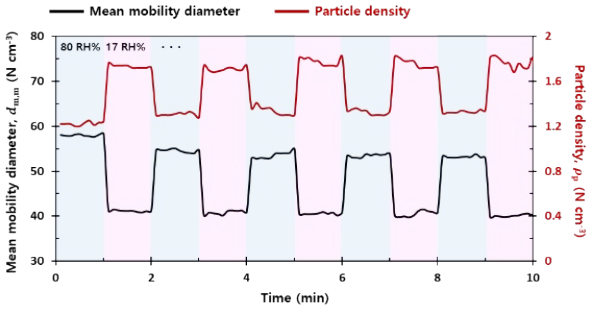


Figure S6. The real-time measurement results of measuring NaCl particles in a humidity-changing environment.

[Properties of the machine learning algorithm]

The machine learning algorithm consists of 7 layers and 30 neurons per layer. This is the number of optimized layers and neurons showing the lowest error rate. As the activation function, the function with the lowest error was used after comparing and evaluating the four functions of sigmoid, tanh, ReLU, and Leaky ReLu. Each function has different characteristics.[10]

The sigmoid function has been frequently used since it naturally activates the values of individual neurons in the form of an S-shaped curve. It can be expressed as

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{S6}$$

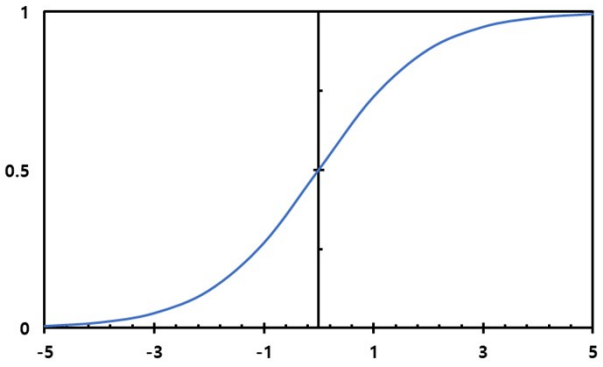


Figure S7. Sigmoid function

Since the sigmoid function range is 0<x<1, the weight(w) is all positive or all negative according to the gradient(δ) and learning data(x), so the learning direction is limited. As a result, it causes low learning speed and convergence rate.

To solve this problem, tanh came out. It is to adjust the size and position of the sigmoid function. It can be defined as

$$\tanh(x) = 2\sigma(2x) - 1 = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{S7}$$

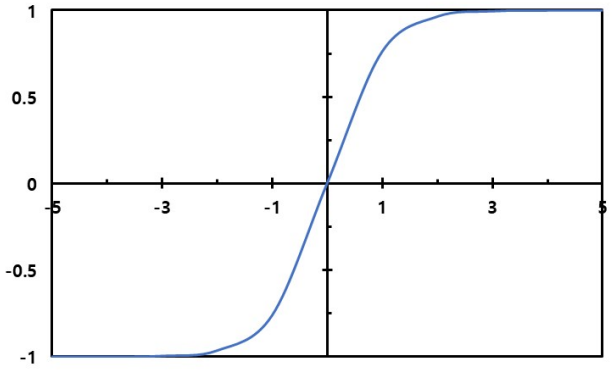


Figure S8. tanh function

Since the range of tanh is -1<x<1, learning speed and convergence speed are faster than sigmoid.

However, sigmoid and tanh function may cause the vanishing gradient problem, which converges weight to zero and the learning speed is still slow. To solve these problems, ReLU, which has no vanishing gradient problem and is faster than the previous activation function more than 6 times, has been developed. It can be express as

$$f(x) = max(0,x) \tag{S8}$$

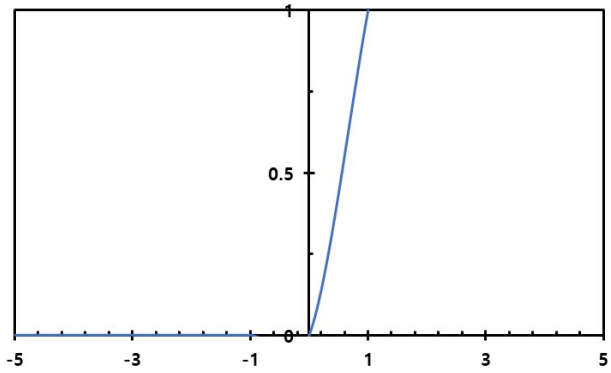


Figure S9. ReLU function

When x is positive, the gradient is constant at 1 and differentiation is convenient, so computational complexity is low. Therefore, the learning speed is 6 times faster than sigmoid or tanh. However, if x is negative, the gradient is unconditionally zero. To overcome this shortcoming, Leaky ReLU was devised and can be expressed as

$$f(x) = max(0.01x,x) \tag{S9}$$

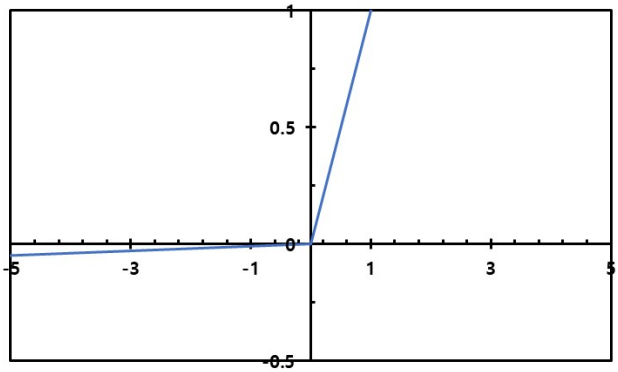


Figure S10. Leaky ReLU function

It has the same characteristics as ReLU except the gradient is 0.01 when x is negative.

Although it is true that later functions are more advanced, an activation function suitable for each situation differs.[11, 12] As a result of comparing and evaluating all four functions, the ReLU function was selected because of showing slightly higher performance than others.

As a method to minimize the loss, we use adam optimizer, which showed good performance in various situations. [13, 14]

Since the problem that the neural network is trying to solve is non-convex optimization, it may not be possible to find the optimal point depending on the starting point. Therefore, it is very important to set initial values for learning parameters such as weights and biases of each layer. Thus, we use the HE uniform, which is known to be the most suitable initializing function for ReLU. Moreover, there is a risk of overfitting, which only adapts to the learning data as the learning period increases, resulting in poor performance. To prevent overfitting, early stopping, which is a method of adjusting epoch numbers, was used. As a result, the upper limit epoch was set at 10,000, but the epoch may vary depending on learning by early stopping. Also, L1 and L2 regularization, which is a method of adjusting weight decays, were used.

The machine learning algorithm was constructed as shown in Table S1 and the result of the comparative evaluation as shown in Table S2

Training set	Test set	Layer	Neuron	Parameter optimization	Weight initialization	Epoch	Regularization
78,000	**11,000**	**7**	**30**	**Adam**	**HE uniform**	**10,000**	**Early stopping, L1, L2**

Table S1. Properties of machine learning algorithm

Activation function	Sigmoid	Tanh	ReLU	Leaky ReLU
Error rate (%)	**9.39**	**8.22**	**3.56**	**6.75**

Table S2. Error rate according to activation function

[References]

1. R. Rengarajan, M. Sarin and A. Sudheer, *Journal of Geophysical Research: Atmospheres*, 2007, **112**.
2. R. L. Mariani and W. Z. de Mello, *Atmospheric Environment*, 2007, **41**, 2887-2892.
3. D. Voutsa, C. Samara, E. Manoli, D. Lazarou and P. Tzoumaka, *Environmental Science and Pollution Research*, 2014, **21**, 4995-5006.
4. P. K. Saha, E. S. Robinson, R. U. Shah, N. Zimmerman, J. S. Apte, A. L. Robinson and A. A. Presto, *Environmental science & technology*, 2018, **52**, 6798-6806.
5. C. Van Gulijk, J. Marijnissen, M. Makkee and J. Moulijn, *Journal of aerosol science*, 2003, **34**, 635-640.
6. J. S. Torrecilla, T. Rafione, J. García and F. Rodríguez, *Journal of Chemical & Engineering Data*, 2008, **53**, 923-928.
7. H. Kuuluvainen, A. Arffman, E. Saukko, A. Virtanen and J. Keskinen, *Journal of aerosol science*, 2013, **55**, 104-115.
8. C. Peng and C. K. Chan, *Atmospheric Environment*, 2001, **35**, 1183-1192.
9. P. Saxena, L. M. Hildemann, P. H. McMurry and J. H. Seinfeld, *Journal of Geophysical Research: Atmospheres*, 1995, **100**, 18755-18770.
10. S. Sharma, *towards data science*, 2017, **6**.
11. H. Gao, L. Cai and S. Ji, 2020.
12. S. Mastromichalakis, *arXiv preprint arXiv:2012.07564*, 2020.
13. S. Mehta, C. Paunwala and B. Vaidya, 2019.
14. I. K. M. Jais, A. R. Ismail and S. Q. Nisa, *Knowl. Eng. Data Sci.*, 2019, **2**, 41-46.