applied sciences

MDPI

Article

AI-Driven Particulate Matter Estimation Using Urban CCTV: A Comparative Analysis Under Various Experimental Conditions

Woochul Choi, Hongki Sung and Kyusoo Chong *

Department of Future & Smart Construction Research, KICT (Korea Institute of Civil Engineering and Building Technology), Goyang-si 10223, Republic of Korea; wcchoi@kict.re.kr (W.C.); sunghongki@kict.re.kr (H.S.)
* Correspondence: ksc@kict.re.kr

Abstract: Despite the high public interest in particulate matter (PM), a key determinant for indoor and outdoor activities, the current PM information provided by monitoring stations (e.g., data per administrative district) is insufficient. This study employed the closed-circuit television (CCTV) cameras densely installed within a city to explore the spatial expansion of PM information. It conducted a comparative analysis of PM estimation effects under diverse experimental conditions based on AI image recognition. It also fills a gap by providing an optimal analysis framework that comprehensively considers the combination of variables, including the sun's position, day and night settings, and the PM distribution per class. In the deep learning model structure and process comparison experiment, the hybrid DL-ML model using ResNet152 and XGBoost showed the highest predictive power. The classification model was better than the ResNet regression model, and the hybrid DL-ML model with the post-processed XGBoost was better than the single ResNet152 model regarding AI prediction of PM. All four experiments that excluded the nighttime, added the solar incidence angle variable, applied the distribution of PM per class, and removed the outlier removal algorithm showed high predictive power. In particular, the final experiment that satisfied all conditions, including the exclusion of nighttime, addition of solar incidence angle variable, and application of outlier removal algorithm, derived predictive values that are expected to be commercialized.

Keywords: fine particulate matter (PM2.5); deep learning; machine learning; CNN; ResNet; XGBoost; meteorological data; solar incidence angle; CCTV

check for updates

Citation: Choi, W.; Sung, H.; Chong, K. AI-Driven Particulate Matter Estimation Using Urban CCTV: A Comparative Analysis Under Various Experimental Conditions. *Appl. Sci.* **2024**, *14*, 9629. https://doi.org/10.3390/app14219629

Academic Editor: Dimitris Kaskaoutis

Received: 20 September 2024
Revised: 16 October 2024
Accepted: 18 October 2024
Published: 22 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The leisure activities of citizens have significantly increased due to the recent rise in the GDP, the implementation of the 52 h workweek system, and consumption patterns to enhance the quality of life. In line with increased leisure activities, particulate matter (PM) information has long been a key determinant for outdoor or indoor activities. In particular, in Korea, considering that PM issues have been regarded as a social disaster and that public interest in such information is very high, checking the PM status/information has become a daily routine before partaking in indoor or outdoor activities. The Ministry of Environment manages an air pollution warning system that activates a level 1 warning when the concentration of fine dust based on PM2.5 exceeds 75 $\mu g/m^3$ for more than 2 h and a level 2 warning when it exceeds 150 $\mu g/m^3$. Through this, it is responding to and regulating high-concentration fine dust, such as banning outdoor classes and implementing a two-part vehicle system. However, the PM information provided to citizens by the Ministry of Environment is featured by area-based data (e.g., data per administrative district) from a dot-type PM monitoring station closest to the current locations. In other words, providing dense/detailed PM information within the citizens' living area is difficult. It is necessary to install additional PM monitoring stations to secure dense/detailed PM information, but there are burdens from covering installation costs and difficulties in

operation and management. Even if PM monitoring stations are installed, it is insufficient to reflect all the citizens' rights to live. In this regard, there are studies predicting PM based on AI with images from existing closed-circuit television (CCTV) that attempt to expand PM information spatially [1–4].

South Korea has CCTV control centers in the 226 basic local governments. Based on this, 1,767,894 public CCTVs thoroughly controlled urban areas in 2023 to monitor traffic, crime prevention, and disasters [5]. Considering that 530 national monitoring stations are currently measuring urban air quality levels [6], the AI analysis based on CCTV infrastructure can provide citizens with more detailed PM information. Rather than monitoring the local PM2.5 concentrations with a high accuracy, it is necessary to increase the spatial density of observations using urban CCTVs. The current research utilizes video images but mostly conducts methodological experiments on deep learning models to estimate PM. Research on deep learning models also focuses on comparing the predictive power of each model, and experimental research under various conditions, such as model structure and application of machine learning, is insufficient. There are a few studies that used meteorological variables as input data. However, no study has suggested an optimal analysis framework that comprehensively considers the combination of variables, including the sun's position, day and night settings, and the PM distribution per class.

In this sense, this study attempts to find the optimal modeling combination for PM image analysis by comparing and analyzing the model structure, the types within the ResNet series, and the application and configuration of machine learning based on ResNet of the representative deep learning model series, the Convolutional Neural Network (CNN) series, to propose an optimal process for PM image analysis. Furthermore, this study aims to propose an optimal PM analysis framework based on AI by comparing the predictive power of PM under different experimental conditions, such as excluding the nighttime, adding the angle of sun incidence, distributing the PM per class, and applying an outlier removal algorithm. To this end, we review previous studies in Section 2 and propose new research materials and methods in Section 3. In Section 4, we perform an empirical analysis of the different experimental conditions, followed by a discussion in Section 5 and conclusions in Section 6, pointing out the implications for research.

2. Literature Review

Research on PM estimation based on images can be divided into analytical, empirical, and experimental studies depending on the choice of experimental conditions and variables. The analytical and empirical studies are as follows: The recent studies based on images obtained by cameras mainly employ CNNs. CNNs were first introduced in 1989 in Lecun's presentation on "Backpropagation applied to handwritten zip" [7]. It was specifically designed in 2003 in Behnke's presentation on "Hierarchical Neural Network for image interpretation", which enabled us to overcome the problem of overfitting with a huge number of parameters and large computational load depending on the image resolution [8]. Since then, various CNN-based models (AlexNet, ZFNet, GoogleNet, VGGNet, and ResNet) have been studied to improve the extraction and classification of data features with geometric and textual associations between the data [9].

Chakma et al. [10] applied a CNN model to analyze PM2.5 based on images stage by stage; images were analyzed according to the PM2.5 concentrations based on two types of main transfer learning strategies: random forests based on CNN fine tuning and CNN features. There have been other studies that have predicted various PM levels based on numerous CNNs [2,3,11–14]. Bo et al. [15] employed ResNet and SVR, both types of CNNs, to improve the index estimation of PM2.5 for input images. Silva et al. [16] utilized the ResNet-101 model to estimate dust concentrations from vehicle traffic on unpaved roads. Li and Mao [17] employed an Image Pyramid and an improved ResNet technique to detect real-time dust patterns. Yin et al. [18] combined data augmentation, particle size constraints, and ResNet techniques to improve the accuracy of PM2.5 concentration estimation. Atreya

and Mukherjee [19] also proposed an efficient ResNet model for atmospheric visibility classification. It was found that ResNet techniques were mainly used to analyze PM images. In addition, there are studies using various deep learning algorithms, such as LSTM (Long Short-Term Memory), or combined machine learning models, DNNs (Deep Neural Networks), and RNNs (Recurrent Neural Networks) [20–27].

More studies have been conducted that apply various influencing variables and experimental conditions to improve the predictive power regarding PM. Most studies have indicated a close correlation between PM and meteorological factors [28–31]. Kim and Moon [32] utilized meteorological and PM data and derived influencing variables on PM generation through multiple regression analysis. They suggested a PM prediction model using machine learning, such as linear regression and a Support Vector Machine. Won et al. [1,2] experimented with learning with all or partial CCTV images and compared analysis results when different meteorological variables, such as humidity and wind velocity, were applied. They also performed experiments during the nighttime and presented the predictability of PM based on CCTV images. Hong and Lee [3] calculated the edge intensity of the region of interest (ROI) and analyzed its correlation with PM prediction. Thus, in the prediction of PM based on images, the model selection, variables, and experimental conditions need to be considered to provide public services. Based on these previous studies, this study proposes an optimal framework for the prediction of PM based on AI.

3. Materials and Methods

3.1. Image Data

Based on the review of the location selection of CCTVs for image analysis [4], the researcher installed CCTV equipment in the research institute to which the researcher belongs (Goyang-si, Gyeonggi-do, Republic of Korea) to efficiently collect image data. Before installing the CCTV equipment, the researcher received legal consultations on personal information and security issues to minimize the issues that may occur while securing the images. The data acquired from the equipment include the latitude–longitude coordinates of the points where CCTVs were installed, image data, and data related to the date and time of image recording. The acquired data, including image data, were collected through Network Video Record (NVR, XRN-410S v2.44). An NVR server is a special computer system with a software program that records video in digital format on a high-capacity storage device. The CCTV is Hanwha Techwin PNP-9200RH (Seongnam city, Korea), the focal length is 4.8~96 mm (20×) zoom lens, and the frame rate per second can be set up to 30 fps. The image's resolution is 8 megapixels at a maximum of 30 fps, and in this study, it was set to 1920 × 1080, considering the resolution of public CCTV images. The CCTV installation point is latitude 37.67102804, longitude 126.7394964, and it was installed at a height of 15 m from the ground with an elevation angle of 0 degrees (horizontal) to simultaneously observe the ground and the atmosphere for fine dust analysis. The viewpoint of CCTV was set to 44.5 degrees north latitude as the azimuth toward the downtown area. The CCTV installation building and camera locations are as shown in Figure 1.

The image collection period was from 21 April 2021 to 14 February 2022 from 8:00 to 18:00 (from 8:00 to 22:00 for some images), and the image data were utilized in line with the concentration distribution and analysis purpose of PM. The images used for analysis are not all CCTV images but separately stored images of fixed infrastructure (e.g., buildings, facilities, etc.) that is unaffected by seasons and weather conditions and can be continuously analyzed for a long time. Image data processing includes the following processes: First, the recording of images from CCTVs; second, the collection of image data by the NVR; third, data storage on the DB server; fourth, data analysis on the analysis server; and fifth, they are used as learning data on the learning server. Table 1 shows the meta-information about the CCTV image data used for the experiment, and Figure 2 shows the difference between the images of the CCTVs for the experiment according to the deviation of PM.

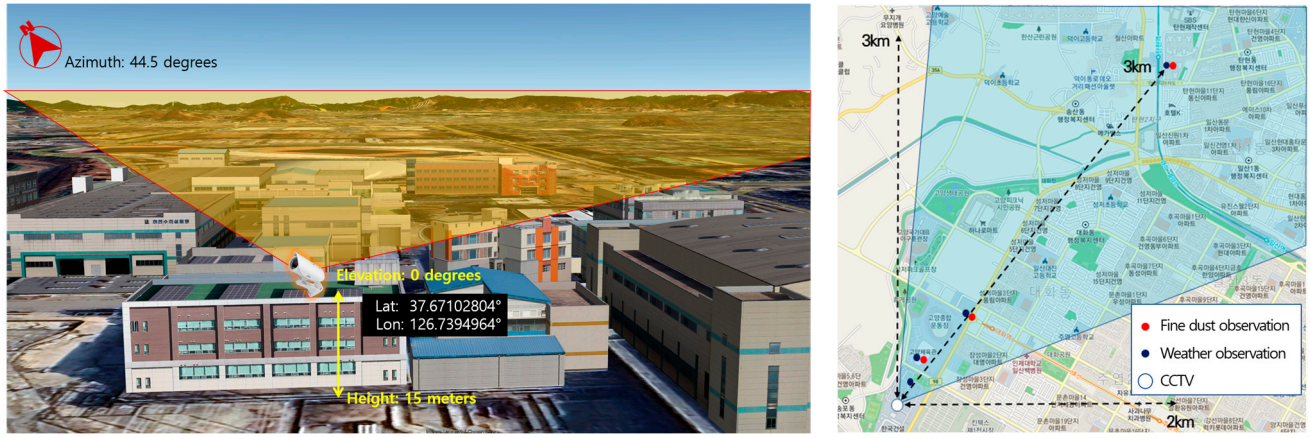


Figure 1. The CCTV installation building and camera locations.

Table 1. Meta-information on CCTV images.

	Category	Data Name	Data Type	Data Value	Collection Methods
CCTV image data for an experiment	ID	DeviceID	Varchar	KICT_1	NVR
	Position coordinates (latitude)	Latitude	Float8	37.67102804	
	Position coordinates (longitude)	Longitude	Float8	126.7394964	
	Image data	Video	Video	Image data	
	Date/time	Date_Time	Date	21 April 2022 20:17:05	

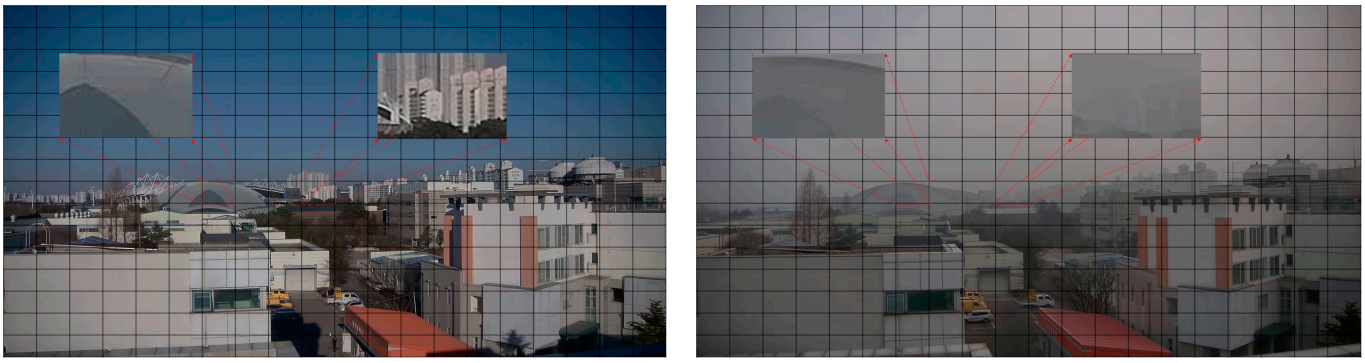


Figure 2. Differences in CCTV images based on PM deviations.

3.2. Weather Observation Data

We collected weather observation data to construct highly reliable learning/training data. Based on previous studies on influencing factors on PM, the weather observation elements were set as the PM (PM2.5 and PM10), temperature, humidity, wind direction, and wind velocity [33]. Regarding the meteorological observation equipment, PM and environmental observation equipment were installed at each ROI. As for the PM observation equipment, six conditions such as performance rating evaluation (Grade 1), accuracy (higher than 90%), mobility, convenience (retention of own power), data acquisition (automatic transmission and reception of communication network), and others (simultaneous observation of weather conditions, such as temperature/humidity, and illuminance) were set, and simple PM observation equipment (AM100) satisfying the conditions was created. The simple PM observation equipment is a light-scattering method using optical characteristics, such as rotation and reflection, that appear by shining light on fine dust particles, and it has the advantage of being easy to move because it can be attached and detached.

The PM observation equipment consists of PM2.5 and PM10 data, and collecting the data through API connections at 30 s intervals is possible.

The weather observation devices can collect temperature, humidity, and wind speed data. To ensure the reliability of the equipment, the GMX500 (integrated sensor) and CR300 (data logger) have been installed, which are products that have been awarded the CE DoC (Declaration of Conformity) certification for data loggers by the EU and have passed ISO9001 certification-based tests. The weather condition monitoring equipment consists of air temperature, humidity, illuminance, wind direction, and wind velocity data that can be collected via API connections with a data collection cycle of 1 min. Table 2 is the meta-information of the weather condition observation data.

Table 2. Meta-information of weather observation data.

Division (Unit)	Data Name	Format	Data Value	Generation Cycle	Collection Methods
PM ($\mu g/m^3$)	FineDust_PM2.5	Float4	35	30 s	Simple PM observation equipment
PM ($\mu g/m^3$)	FineDust_PM10	Float4	50		
Position coordinates (latitude)	Latitude	Float8	37.67102804		
Position coordinates (longitude)	Longitude	Float8	126.7394964		
Atmospheric temperature (°C)	Out_Temperature	Float4	25.9	1 min	RWIS observation
Humidity (%)	Out_Humidity	Float4	84.7		
Illuminance (lx)	Illumination	Float8	600		
Wind direction	Wind_Direction	Varchar	"north, east, and northwest"		
Wind velocity (m/s)	Wind_Speed	Float4	2.5		

3.3. Weather Public Data

In order to ensure the reliability of the learning data and the expandability of the results of this study, additional public weather data were created. To this end, we utilized the data from the web platforms of the Korea Environment Corporation (AirKorea) [6] and Korea Meteorological Administration (Open MET Data Portal) [34]. The fine dust monitoring station operated by AirKorea uses the beta-ray absorption method, which measures the mass concentration by collecting dust on a filter paper and using the difference in beta rays absorbed and dissipated when passing through it. It was possible to collect PM values (PM2.5, PM10) from AirKorea and meteorological data (e.g., atmospheric temperature, wind velocity, humidity, atmospheric pressure, and precipitation) from Open MET Data Portal. As for the data generation frequency, atmospheric pressure data are available every three hours, cloud amount data are available at 1-day intervals, and all other meteorological factors are available at 1 h intervals. Table 3 indicates the meta-information on the public weather data.

Table 3. Meta-information on public weather data.

Division (Unit)	Data Name	Format	Data Value	Generation Cycle	Collection Methods
PM ($\mu g/m^3$)	FineDust_PM2.5	Float4	35	1 h	AirKorea
PM ($\mu g/m^3$)	FineDust_PM10	Float4	50	1 h	
Location information	Location	Varchar	"Juyeop-dong and Burim-dong"	1 h	

Table 3. *Cont.*

Division (Unit)	Data Name	Format	Data Value	Generation Cycle	Collection Methods
Meteorological conditions	Weather	Varchar	"Sunny, cloudyand rainy"	1 h	Open MET Data Portal
Atmospheric temperature (°C)	Temperature	Float4	25.9	1 h	
Wind velocity (m/s)	Wind_Speed	Float4	2.5	1 h	
Humidity (%)	Out_Humidity	Float4	84.7	1 h	
Atmospheric pressure (hpa)	Air_Pressure	Float4	1016.0	3 h	
Precipitation (mm/h)	Rain_Strength	Float4	10.032	1 h	
Cloud amount (decile)	Cloud_Amount	Float4	9.5	1 day	

3.4. Data Preprocessing

We first aligned the collection cycle of CCTV images, weather observation, and public weather data with one minute for data preprocessing. The preprocessing for image analysis consists of image data processing, an image data task scheduler, frame-by-frame extraction and ROI setting, and meteorological and image data matching. Specifically, the image data were first divided and processed into small units of image data on the DB server, and the image data were converted to frame-by-frame image data using the scheduler. The image data are reprocessed into the set ROI area within the converted image data. The location and visual data of public weather, sensing, and image data are combined via linear stretching and then calibrated according to each data generation/collection cycle to construct one dataset according to a certain cycle.

The preprocessing process in the program is as follows: starting with the library call, the preprocessing process is completed in the following steps: accessing the database, browsing the image area, retrieving atmospheric sensor data, retrieving weather sensor data, retrieving public information data, separating the data per sensor, combining data frames, converting timestamp attributes to string attributes, generating columns of data frame file names, checking and interpolating missing values, rounding PM values, matching data between file lists and data frames, and exporting CSV.

Finally, we developed an ROI derivation algorithm based on correlation analysis between PM observations and CCTV images and selected an ROI optimized for PM estimation. In detail, after constructing converted images for CCTV images, we derived the correlation between each pixel value of the converted images and the corresponding PM observation values. This allowed us to select the pixels with relatively high correlations as optimal ROIs for constructing a deep learning model.

3.5. ResNet-Based Methods for Image Classification and PM2.5 Prediction

In this study, we adopted a CNN model, which is a representative technique for deep learning in images. The main process of a CNN requires the creation of a feature map through convolution and the down/sub-sampling of the feature map via pooling. This process can create multiple feature maps with diverse features at smaller scales than the original images. The generated features are then fed into the Fully Convolutional Network for Semantic Segmentation (FCN) to classify the main features of the images. After the convolutional processes and pooling, it can respond more flexibly to the overfitting problem, which is the main problem of the FCN, and effectively reduce the computational cost of learning. The CNN is mainly used for extracting and classifying data features with geometric and structural associations between the data and has been further developed and optimized technologically.

This study used ResNet, a representative CNN model, to predict PM. ResNet is a popular image classification technique in the vision field due to its efficient architecture. ResNet takes images as input, recognizes diverse features of images accurately, and can

be predicted after learning. In this study, we analyzed ResNet and experimented with the model structures, number of ResNet layers, machine learning composition, and application status. First, as for the model structure, we built a ResNet regression model and a classification model, respectively. Regarding the regression model, the last FC layer (Fully Connected Layer) is replaced with a single neuron, and Sigmoid is replaced with Softmax; regarding the classification model, the number of FC outputs of the last ResNet is changed from 1000 to 201 in consideration of the maximum and distribution ranges of PM2.5. Second, ResNet50 and ResNet152 are constructed for comparison according to the number of learning layers of ResNet. By applying the layer structure Conv(49-Layer) of ResNet50 and the layer structure Conv(151-Layer) of ResNet152 differently, we try to compare the results. Third, machine learning and deep learning consist of SVR and XGBoost, respectively. SVR is a model that defines the decision boundary, which is a basis for classification. XGBoost is an ensemble technique that combines multiple weak decision trees, weights them, and sequentially incorporates the learning errors of the weak prediction models into the next learning model. Fourth, we construct a model that combines deep learning and machine learning (DL-ML) and a single deep learning (DL) model. In the hybrid DL-ML model, the deep learning inputs consist of RGB and weather observation data; the machine learning inputs consist of PM2.5 prediction values via deep learning, public weather data, and the sun incidence angle. The single DL model was set to simultaneously input RGB data, weather observation data, public weather data, and the solar incidence angle to the deep learning input. The image analysis process for the single DL model can be found in Figure 3.

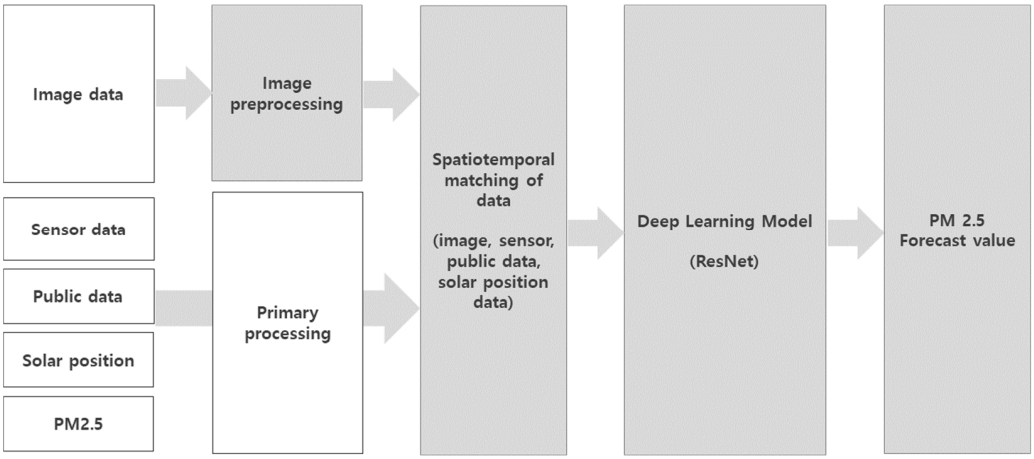


Figure 3. Single DL model process.

We conducted experiments under various conditions. The model structure mainly utilized to control the external influencing factors was the hybrid DL-ML model. The hybrid DL-ML model improves PM prediction values by combining the image-specific PM prediction variables calculated using deep learning analysis with post-processing machine learning techniques in which PM-influencing variables, such as meteorological variables and solar incidence angle variables, are used as inputs. As for the post-processing machine learning technique, Support Vector Regression (SVR) and XGBoost were mainly utilized. SVR is constructed based on a Support Vector Machine (SVM), and it is known to generate similar results to general regression analysis when the response variable is continuous [35]. It is also characterized by finding a line with the maximum gap between observations belonging to different categories and connecting them with a line, and it is able to predict both binary and continuous values. Assuming from the beginning that the data have noise, they do not perfectly correspond to the actual values with noise to a particular model [36]. Therefore, differences between the actual and predicted values within a reasonable range are allowed. XGBoost is a decision tree algorithm that employs several Classification and Regression Trees (CARTs), reduces the error values, and finds the optimal tree [37]. XGBoost adjusts the complexity of the tree to minimize data loss during training while

avoiding the problem of overfitting results. The depth of the tree starts at zero and goes through a pruning process, where information gathering is calculated one by one, and branches are removed until the information gathering points are negative. This iteration finally creates a model by combining the trees with the highest scores [38]. In other words, it has the advantages of the decision tree technique, i.e., simple data preprocessing and relatively fast computational speed [39]. The image analysis process of the hybrid DL-ML model can be found in Figure 4.

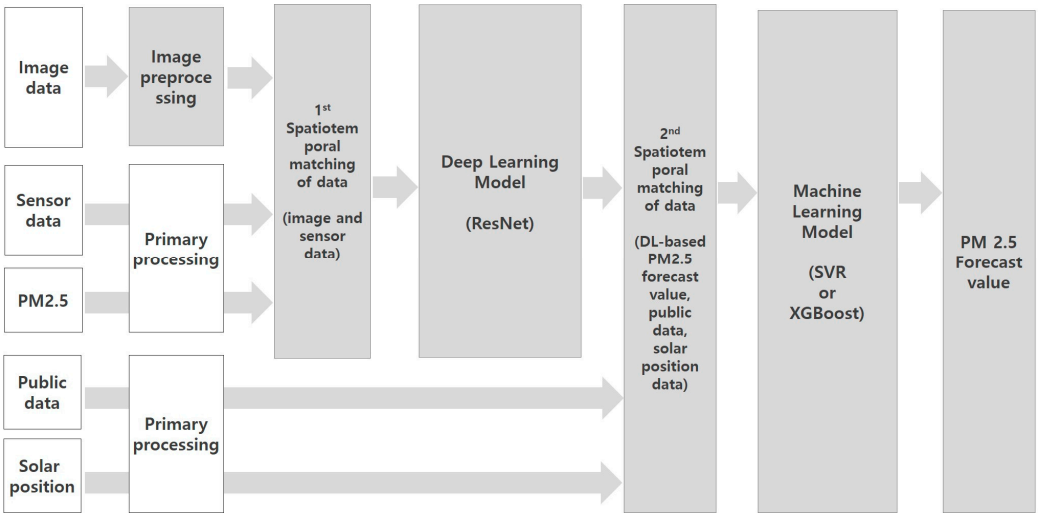


Figure 4. Hybrid DL-ML model process.

4. Empirical Analysis

4.1. Data Analysis Design

In this study, we designed the data analysis frame by classifying various experimental conditions, including the deep learning model structure and variable settings, to estimate PM concentrations in the images by combining image and weather condition data. The dependent variable is PM2.5, and the independent variables are the field weather observation data, public weather data, and solar incidence angle. As for the field weather observation data, the factors influencing the PM concentrations (i.e., humidity, temperature, wind direction, and wind velocity) and illuminance, making it possible to estimate the weather status, were simultaneously input. We utilized the humidity, temperature, wind velocity, wind direction, and precipitation as input data for public weather data. The data collection period was from 21 April 2021 to 14 February 2022 (between 8:00 and 22:00), and the data composition varies depending on the purpose of the experiment.

The deep learning model uses ResNet, which is a CNN series mainly utilized for PM image analysis, and the hyperparameter values were set using repeated comparisons for optimal modeling. In terms of epochs, as a result of conducting a comparative analysis after applying values such as 10, 100, 200, and 300 with batch sizes of 32, 128, and 256, the optimal hyperparameter setting value of this model was derived as follows: epochs of 100 and a batch size of 256. Additionally, the data used for analysis were set to 60% for learning data, 20% for verification data, and 20% for test data. The analysis results of the experiments are interpreted with the RMSE (Root Mean Square Error), which can be considered PM errors, and the Mean Absolute Percentage Error (MAPE), which is easy to understand relatively between performances, is utilized.

We conducted four experiments on deep learning model structures. In all four experiments, only the deep learning analysis structures were set differently depending on the purpose of the experiments; all other conditions were controlled. First, we compared the PM predictive power of the ResNet regression model and the classification model. Second, we compared the predictive power of ResNet50, in which the number of ResNet layers was set as 49, and of ResNet152, with 151 layers set. Third, we compared the predictive

power of the hybrid DL-ML model with the post-processed machine learning based on ResNet152 by applying the machine learning techniques, including SVR and XGBoost, respectively. Fourth, we compared the predictive power of the single DL model and the hybrid DL-ML model.

We conducted four additional experiments with different learning variables and observation conditions following the deep learning analysis. First, we compared the PM predictive power of the daytime dataset (08:00–18:00), excluding the nighttime between 18.00 and 22.00, with that of the entire dataset (08.00–22.00). This experiment aimed to identify how resolution degradation from a lack of nighttime light affects predictive power. Second, we compared the predictive power with or without applying the solar incidence angle variable. The variable angle of incidence is important because it results in differences in the illuminance and shadow position depending on the position of the sun. The sun azimuth (right and left positions) and the sun elevation (upper and lower elevations) were calculated and entered as inputs to vary the position of the sun. The sun azimuth is the angle between the projection of the sun's rays and a line to the south and can be calculated by measuring the angle clockwise from north (0 degrees) (see Equation (1)). The position of the sun is the angle between the sun and the Earth's surface, which is used to determine the height of the sun. As shown in Equation (2), it can be calculated from the hour angle, the declination of the sun, and the local latitude at which the height of the sun can be observed.

$$\cos\varphi = \frac{\sin\delta\cos L - \cos\omega\cos\delta\sin L}{\sin\theta} \quad (1)$$

Here, φ: solar azimuth; L: local latitude; δ: solar declination; θ: solar azimuth angle; ω: hour angle.

$$\sin\beta = \cos L\cos\delta\cos\omega + \sin L\sin\delta \quad (2)$$

Here, β: solar altitude; L: local latitude; δ: solar declination; ω: hour angle.

Third, we confirmed the difference in the predictive power of the datasets with the distribution ratio of each PM class in actual air conditions. To minimize the damage to public health caused by air pollution, the Ministry of Environment classifies PM concentrations into four classes (good, moderate, bad, and very bad) for immediate understanding of outdoor activity recommendations and precautions. Based on PM2.5 values, South Korea categorizes the range between 0 and 15 $\mu g/m^3$ as "good", between 16 and 35 $\mu g/m^3$ as "moderate", between 36 and 75 $\mu g/m^3$ as "bad", and higher than 76 $\mu g/m^3$ as "very bad". To prevent data bias, this study applied the distribution ratio per PM class in 2021 at the national monitoring station in Juyeop-dong, Goyang-si, Gyeonggi-do, and constructed a dataset that simulates an analysis environment similar to the air condition. The total number of existing datasets is 178,316, with distribution patterns of 54.4% (97,056) in Class 1 (good), 21.0% (37,474) in Class 2 (moderate), 22.6% (40,325) in Class 3 (bad), and 1.94% (3461) in Class 4 (very bad). The total number of actual datasets with the distribution ratio per PM class is 96,500, with distribution patterns of 46.6% (44,969) in Class 1, 38.8% (37,442) in Class 2, 12.7% (12,352) in Class 3, and 1.8% (1737) in Class 4. For reference, the average PM2.5 in 2021 from the Goyang City monitoring station was 21 $\mu g/m^3$, the month with the worst fine dust was March (34 $\mu g/m^3$), and the month with the best fine dust was September (12 $\mu g/m^3$) [6].

Fourth, we confirmed the difference in the predictive power of the datasets with outlier removal algorithms. We applied a local outlier removal algorithm based on Median Absolute Deviation (MAD), which removes sudden information in the images (e.g., vehicles and images with poor focus) and noise in each pixel. After selecting n images at one-second intervals before and after the target image taken at time t, the i-th pixel values (DN) of the n images were divided into outliers and intruders. The values above 3 MADs scaled from the median value are classified as outliers, and the average of the pixel values classified as outliers is calculated and assigned as the new pixel value of the corresponding pixel in the preprocessed images. The calculation method can be found in Equation (3). Additionally,

when PM2.5 exceeded PM10, or suddenly high or low observed values (e.g., smoking or street cleaning) were identified as outliers using the sliding window technique. In the above experiments, except for the experimental conditions and variables that change, since all other conditions (e.g., AI model, number of data, and variables) are controlled, the effect of direct predictive power improvement can be confirmed.

$$\mathrm{MAD} = \mathrm{c} \times \mathrm{median}(\mathrm{abs}(\mathrm{DN} - \mathrm{median}(\mathrm{DN}))) \quad (3)$$

$$\mathrm{c} = \frac{1}{\sqrt{(2)} \times \mathrm{erfcinv}(3/2)} \approx 1.4826$$

Here, median: median value; erfcinv: Inverse Complementary Error Function.

4.2. Comparison of Analysis Results Using Deep Learning Models

The data collection for the deep learning model comparative analysis took place from 21 April 2021 to 14 February 2022 from 08:00 to 18:00 and with the same number of data points (178,316). In Experiment 1, we compared the predictive power of regression and classification models based on ResNet. As a result of the analysis, the RMSE of the regression model was 23.1, and the MAPE was 271.71%; the RMSE of the classification model was 5.93, and the MAPE was 32.41%, indicating that the predictive power of the classification model was overwhelmingly higher. In Experiment 2, we compared the predictive powers of ResNet50 and ResNet152 based on the hybrid DL-ML model according to the number of learning layers in ResNet. As a result of the analysis, the model with ResNet50 had an RMSE of 6.26 and MAPE of 36.06%, whereas the model with ResNet152 had an RMSE of 6.18 and MAPE of 33.75%, showing that the predictive power of ResNet152 was slightly higher. In this sense, we conducted other experiments using ResNet152 for analysis. In Experiment 3, we compared the predictive powers of SVR and XGBoost in the hybrid DL-ML model. The analysis results indicate that the model with SVR had an RMSE of 7.33 and MAPE of 35.90%, whereas the model with XGBoost had an RMSE of 3.32 and MAPE of 25.52%, indicating that the predictive power of the model with XGBoost was higher. In Experiment 4, we compared the ResNet152-based single DL model's predictive power and the hybrid DL-ML model post-processed with XGBoost. The single DL model had an RMSE of 5.66 and MAPE of 34.04%, whereas the hybrid DL-ML model had an RMSE of 3.32 and MAPE of 25.52%, confirming that the hybrid DL-ML model had a higher predictive power. Table 4 shows the comparative analysis results according to the deep learning model.

Table 4. Comparison of analysis results using deep learning models.

Experimental Condition	Deep Learning Model	Data Collection Date	Time	No. of Data Points	Variable				Application of Distribution Per Class	Application of Outlier Removal Algorithm	Analysis Results	
					Dependent Variable	Input Variables						
						Observation Data	Public Data	Solar Incidence Angle			RMSE	MAPE
Experiment 1. Comparison of model structures (regression and classification)	ResNet regression model	21 Apr. 2021~14 Feb. 2022	08–18	178,316	PM2.5	o	x	x	x	x	23.1	271.71
	ResNet classification model	21 Apr. 2021~14 Feb. 2022	08–18	178,316	PM2.5	o	x	x	x	x	5.93	32.41
Experiment 2. Comparison of the number of ResNet layers	ResNet50 + SVR	21 Apr. 2021~14 Feb. 2022	08–18	178,316	PM2.5	o	o	o	x	x	6.26	36.06
	ResNet152 + SVR	21 Apr. 2021~14 Feb. 2022	08–18	178,316	PM2.5	o	o	o	x	x	6.18	33.75
Experiment 3. Machine learning model comparison	ResNet152 + SVR	21 Apr. 2021~14 Feb. 2022	08–18	178,316	PM2.5	o	o	o	x	x	7.33	35.90
	ResNet152 + XGBoost	21 Apr. 2021~14 Feb. 2022	08–18	178,316	PM2.5	o	o	o	x	x	3.32	25.52

Table 4. *Cont.*

Experimental Condition	Deep Learning Model	Data Collection Date	Time	No. of Data Points	Variable				Application of Distribution Per Class	Application of Outlier Removal Algorithm	Analysis Results	
					Dependent Variable	Input Variables						
						Observation Data	Public Data	Solar Incidence Angle			RMSE	MAPE
Experiment 4. DL and DL-ML model comparison	ResNet152	21 Apr. 2021~14 Feb. 2022	08–18	178,316	PM2.5	o	o	o	x	x	5.66	34.04
	ResNet152 + XGBoost	21 Apr. 2021~14 Feb. 2022	08–18	178,316	PM2.5	o	o	o	x	x	3.32	25.52

"o and x" indicate whether variables or experimental conditions are used.

4.3. Comparison of Analysis Results Using Learning Variables and Observation Conditions

We conducted four experiments to find the optimal combination of learning variables and observation conditions for PM image analysis. The data collection period for Experiments 1 and 2 was from 21 April 2021 to 19 July 2021, and that for Experiments 3 and 4 was from 21 April 2021 to 14 February 2022. The data collection time was from 8:00 to 22:00 only in Experiment 1, while it was from 8:00 to 18:00 for the other experiments. The datasets varying from 95,039 to 178,316 were constructed for each experiment to apply distributions per time and PM class. All conditions were controlled except for the key factors per the purpose of each experiment. The experiment with a deep learning model was also conducted based on ResNet, as in Section 4.2, but a single DL model per experiment was employed, and the hybrid DL-ML model had SVR and XGBoost applied differently.

In Experiment 1, we compared the predictive power of the dataset for the entire time between 8.00 and 22.00 and the dataset for the daytime (8.00–18.00), excluding the nighttime between 18.00 and 22.00. As a result, the dataset for the daytime had an RMSE of 3.12 and MAPE of 38.67%, and the entire time (08.00–22.00) had an RMSE of 3.30 and MAPE of 41.02%. In other words, the daytime dataset analysis slightly improved the predictive power over the entire dataset analysis. In Experiment 2, we compared the predictive power of the addition status of solar incidence angle variables (i.e., solar azimuth and solar altitude values) to the input. As a result, the dataset with solar incidence angle added had an RMSE of 2.94 and MAPE of 36.21%, slightly higher than the existing dataset with an RMSE of 3.12 and MAPE of 38.67%. In Experiment 3, we compared the predictive power of status by adding a dataset with a distribution ratio per PM class to simulate real atmospheric conditions. The dataset with the distribution ratio per PM class had an RMSE of 4.18 and a MAPE of 25.19%, slightly more than the existing dataset with an RMSE of 4.42 and a MAPE of 27.43%. In Experiment 4, we compared the predictive power of the addition status of a dataset with the outlier removal algorithm. The dataset with the outlier removal algorithm had an RMSE of 3.61 and a MAPE of 19.74%, slightly higher than the existing dataset with an RMSE of 4.01 and a MAPE of 21.70%. Table 5 shows the results of the analysis by learning variable and observation condition.

Table 5. Comparison of analysis results based on learning variables and observation conditions.

Experimental Condition	Deep Learning Model	Data Collection Dates	Time	No. of Data	Variables				Application of Distribution Per Class	Application of Outlier Removal Algorithm	Analysis Results	
					Dependent Variable	Input Variables						
						Observation Data	Public Data	Solar Incidence Angle			RMSE	MAPE
Experiment 1. Exclusion of nighttime	ResNet152 + SVR	21 Apr. ~19 Jul. 2021	08–22	143,504	PM2.5	o	o	x	x	x	3.30	41.02
	ResNet152 + SVR	21 Apr. ~19 Jul. 2021	08–18	95,039	PM2.5	o	o	x	x	x	3.12	38.67
Experiment 2. Addition of solar incidence angle variable	ResNet152 + SVR	21 Apr. ~19 Jul. 2021	08–18	95,039	PM2.5	o	o	x	x	x	3.12	38.67
	ResNet152 + SVR	21 Apr. ~19 Jul. 2021	08–18	95,039	PM2.5	o	o	o	x	x	2.94	36.21

Table 5. *Cont.*

Experimental Condition	Deep Learning Model	Data Collection Dates	Time	No. of Data	Variables				Application of Distribution Per Class	Application of Outlier Removal Algorithm	Analysis Results	
					Dependent Variable	Input Variables						
						Observation Data	Public Data	Solar Incidence Angle			RMSE	MAPE
Experiment 3. Application of distribution per PM class	ResNet152	21 Apr. 2021~14 Feb. 2022	08–18	178,316	PM2.5	o	x	x	x	x	4.42	27.43
	ResNet152	21 Apr. 2021~14 Feb. 2022	08–18	96,500	PM2.5	o	x	x	o	x	4.18	25.19
Experiment 4. Application of outlier removal algorithm	ResNet152 + XGBoost	21 Apr. 2021~14 Feb. 2022	08–18	100,000	PM2.5	o	o	o	o	x	4.01	21.70
	ResNet152 + XGBoost	21 Apr. 2021~14 Feb. 2022	08–18	100,000	PM2.5	o	o	o	o	o	3.61	19.74

'o, x' indicates whether variables and experimental conditions are used.

5. Discussion

5.1. Analysis Results Based on Deep Learning Models

This paper discusses the results of four experiments based on the ResNet model, which is a representative CNN series model frequently utilized as a PM image analysis model. First, it examines the model structure required for PM prediction research through Experiments 1 and 4. As a result of conducting a comparative analysis of classification and regression models based on ResNet in Experiment 1, the predictive power of the classification model was found to be higher with a very large difference because the number of FC outputs was changed from 1000 to 201 in consideration of the actual maximum PM2.5 and the distribution range of 0~200 $\mu g/m^2$. It can be seen that selecting a model structure that aligns with the study's purpose and goal is an important contributor to higher predictive power. As a result of comparing and analyzing the single DL model based on ResNet152 and the hybrid DL-ML model post-processed with XGBoost in Experiment 4, the hybrid DL-ML model showed better predictive power. Although the two-step configuration of separating the input variables and applying them to DL and ML complicates the entire model process, the difference in the RMSE, which is considered PM error, is 2.34 (MAPE is 8.52%), which is not negligible. In this study, weather observation data, public weather data, solar incidence angles, and numerous other influencing variables were used as inputs to improve the predictive power. With many input values, it is assumed that applying them separately to DL and ML models is more effective than applying them all at once to DL models. In the practical application phase, it seems necessary to conduct further comparative experiments with the single DL model and the hybrid DL-ML model based on the number of influential variables to be input into the model, as fewer input values than the influential variables used in this study can be used in this phase.

Next, we will discuss the PM prediction models that should be considered in Experiments 2 and 3. As a result of comparing the predictive power of the ResNet learning layers in Experiment 2, ResNet152 was slightly higher than ResNet50. The number of layers does not guarantee an absolute improvement in the predictive power, and the higher number of layers may have an overfitting problem to the computing power, namely processing power and capacity. ResNet152 with more layers has a slight advantage in this PM estimation study, but further experiments are needed for different research purposes. When machine learning algorithms such as SVR and XGBoost were applied in the hybrid DL-ML model under Experiment 3, the predictive power of XGBoost was found to be high. It depends on the data characteristics and input structure, but in this PM estimation study, the model with XGBoost was confirmed to be superior. However, considering that there are cases where SVR showed stable predictive power in other experiments, such as Experiments 1 and 2 in Section 4.3, SVR is not an inappropriate model. In this sense, further experiments are assumed to be needed for this purpose.

5.2. Comparison of Analysis Results in Line with Learning Variables and Observation Conditions

This section discusses the analysis results based on the classification of observation conditions and model application (learning variables and outlier removal algorithm application). As for the observation conditions, in Experiment 1, we compared the result from the application of the dataset excluding the nighttime, and in Experiment 3, we compared the result from applying a distribution per PM class to simulate a real air pollution environment. In Experiment 1, the dataset's predictive power, excluding nighttime hours, was slightly higher. The assumption that the quality of CCTV images decreases at night due to the lack of light and that the PM predictive power also decreases as a result was reflected in the experiment. However, there is a point that should not be neglected. While daytime is when people are most active outside, diverse activities occur at night (e.g., exercise and outdoor performance). In this sense, PM analysis is also required during nighttime hours, not just during the day. It is assumed that various attempts should be made to improve the techniques and reliability of analyzing the nighttime. Experiment 3 showed a slightly higher predictive power of the dataset with the distribution per PM class. The predictive power of the dataset that simulates an actual air pollution environment was higher, but it is assumed to not be an absolute factor. This experiment is assumed to be meaningful in that it attempted to construct a dataset in line with the actual air pollution condition rather than focusing on improving the predictive power.

Next, we discuss the results of the comparative analysis based on the model application, such as the learning variables and the application of the outlier removal algorithm. Experiment 2 showed the high predictive power of the dataset with the solar incidence angle variable. The assumption in the solar incidence angle directly affects the chroma and brightness of the image and causes shadow positions to be reflected. Image analysis based on RGB cameras has the limitation that the sun's position highly influences it. The same result was found in Experiment 1, which excluded the nighttime, as in Experiment 2. It is expected that the limitations of sun-related RGB camera-based PM image analysis will be overcome to some extent if future research continues to analyze not only the position of the sun but also the specific illuminance differences according to time of day and the CCTVs in each direction. In Experiment 4, we found that the predictive power of the dataset with the outlier removal algorithm was improved. The reliability of data was secured by removing information on unexpected events (e.g., vehicles, birds, and poor focus), PM error values (inversion of PM2.5 to PM10), and values with large sudden drops (e.g., smoking and street cleaning) applied to outlier removal algorithms. In any study, outlier removal is important to improve predictive power. In this study, the main algorithm was to remove local outliers based on MAD, but the inversion of PM2.5 and values with large sudden drops (e.g., smoking and street cleaning) were added, leading to higher PM predictive power. It is assumed to be important to consider not only the optimal variables but also the targeting of outliers that align with the study's purpose.

In this study, the hybrid DL-ML model with the machine learning algorithm XGBoost based on ResNet152 had the highest predictive power, with an RMSE of 3.32 and a MAPE of 25.52% in the comparison experiment of the deep learning model structure. When comparing learning variables and observation conditions, the RMSE for the dataset with the added sun incidence angle was the lowest (2.94), and the MAPE was 19.74% for the dataset with the outlier removal algorithm, showing the highest relative predictive power. Overall, Experiment 4, which was based on the hybrid DL-ML model with ResNet152 and XGBoost, the sun incidence angle variable added, the distribution of PM per class applied, and the outlier removal algorithm applied, was hypothesized to be the optimal condition for PM prediction research. Although the experimental results may differ by preparing or limiting additional experimental conditions, the prediction values of an RMSE of 3.61 and MAPE of 19.74%, calculated through Experiment 4, give us great expectations for the potential for commercializing this study.

6. Conclusions

We compared and analyzed the effects of various experimental conditions, such as the deep learning model structure, machine learning application and configuration, exclusion of nighttime hours, the addition of solar incidence angle variable, the application of distributions per PM class, and the use of the outlier removal algorithm, to suggest the optimal process for PM image analysis. In the deep learning model structure and process comparison experiment, the hybrid DL-ML model using ResNet152 and XGBoost showed the highest predictive power, with an RMSE of 3.32 and MAPE of 25.52%. We concluded that the classification model was better than the ResNet regression model, and the hybrid DL-ML model with post-processed XGBoost is better than the single ResNet152 model in terms of the prediction of PM based on AI. In the comparison experiment, according to learning variables and observation conditions, all four experiments excluded the nighttime, added the solar incidence angle variable, applied the distribution per PM class, and removed the outlier removal algorithm, which showed high predictive power. In particular, Experiment 4, which satisfied all experimental conditions, was able to derive predictive values that are expected to be commercialized, with an RMSE of 3.61 and MAPE of 19.74%. This study's findings suggest the possibility and framework of CCTV image-based PM analysis to provide dense PM spatial distribution as a public service.

This study has the following limitations: First, we controlled the analysis conditions of each experiment but could not control all eight experiments due to the nature of the experiments. However, phased experiments with controlled analysis conditions may have suggested a clearer PM prediction framework. Second, methodologically, the experiments were conducted only on the CNN series, especially ResNet models. The selection was made due to the frequent utilization of the ResNet technique in the PM prediction research and the focus on other experiments. In future studies, different deep learning models, such as LSTM, RNN, and CNN, should be investigated for their strength in predicting time series data. Third, the vertical values of the ROI were not considered. It is judged that there will be a variation in fine dust depending on the vertical value, but this experiment was carried out ground-based fine dust measurement and estimation. For more precise vertical distribution and three-dimensional estimation, it is judged that fine dust estimation analysis via elevation should be performed. Fourth, we did not analyze the different CCTV channels. In particular, the CCTV direction and time of day must be analyzed, as these may vary significantly in reproducing the sunlight value. Future studies should address the analysis of PM images, such as analyzing different deep learning models, CCTV directions, and additional high-quality datasets. Additionally, continuous research should be conducted on developing PM control systems for citizen services, such as spatial informatization of PM predictions and three-dimensional interpolating measurement blind spots. Such attempts will enable the provision of dense PM information that reflects citizens' rights to live based on the existing CCTV infrastructure, which is expected to contribute to establishing a social safety net to protect people's health from PM.

Author Contributions: Conceptualization, W.C., H.S. and K.C.; methodology, W.C. and K.C.; software, W.C.; validation, H.S. and K.C.; formal analysis, W.C.; investigation, W.C. and H.S.; data curation, W.C.; writing—original draft preparation, W.C.; writing—review and editing, W.C. and K.C.; visualization, W.C.; supervision, K.C. All authors have read and agreed to the published version of the manuscript.

Funding: The research for this paper was carried out under the KICT Research Program (project no. 20240437-001) funded by the Ministry of Science and ICT.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Won, T.; Eo, Y.D.; Sung, H.K.; Chong, K.S.; Youn, J. Effect of the Learning Image Combinations and Weather Parameters in the PM Estimation from CCTV Images. *J. Korean Soc. Surv. Geod. Photogramm. Cartogr.* **2020**, *38*, 573–581. [CrossRef]
2. Won, T.; Eo, Y.D.; Jo, S.M.; Song, J.; Youn, J. Estimation of PM concentrations at night time using CCTV images in the area around the road. *J. Korean Soc. Surv. Geod. Photogramm. Cartogr.* **2021**, *39*, 393–399. [CrossRef]
3. Hong, S.; Lee, J. A Method for Inferring Fine Dust Concentration Using CCTV. *J. Korea Inst. Inf. Commun. Eng.* **2019**, *23*, 1234–1239. [CrossRef]
4. Choi, W.C.; Cheong, K.S. A Study on Selection of Image Analyzing CCTV Location for Local Fine Dust Observation based on Fuzzy AHP. *J. Korea Plann. Assoc.* **2022**, *57*, 150–160. [CrossRef]
5. Korean Statistical Information Service. Available online: https://kosis.kr/ (accessed on 8 June 2024).
6. Air Korea. Available online: https://www.airkorea.or.kr/web/ (accessed on 10 June 2024).
7. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jacke, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neur. Comput.* **1989**, *1*, 541–551. [CrossRef]
8. Behnke, S. *Hierarchical Neural Networks for Image Interpretation, Lecture Notes in Computer Science, 2766*; Springer: Heidelberg/Berlin, Germany, 2003.
9. Song, A.R.; Kim, Y.I. Deep Learning-based Hyperspectral Image Classification with Application to Environmental Geographic Information Systems. *Korean J. Rem. Sens.* **2017**, *33*, 1061–1073. (In Korean with English Abstract)
10. Chakma, A.; Vizena, B.; Cao, T.; Lin, J.; Zhang, J. Image-based air quality analysis using deep convolutional neural network. In Proceedings of the 2017 IEEE International Conferences on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 3949–3952. [CrossRef]
11. Ko, K.K.; Jung, E.S. Improving Air Pollution Prediction System through Multimodal Deep Learning Model Optimization. *Appl. Sci.* **2022**, *12*, 10405. [CrossRef]
12. Karen, S.; Andrew, Z. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015; pp. 1–14.
13. Wang, Z.; Zheng, X.; Li, D.; Zhang, H.; Yang, Y.; Pan, H. A VGGNet-like approach for classifying and segmenting coal dust particles with overlapping regions. *Comp. Industr.* **2021**, *132*, 103506. [CrossRef]
14. Wang, Z.; Li, D.; Zheng, X.; Xie, D. A Novel Coal Dust Characteristic Extraction to Enable Particle Size Analysis. *IEEE* **2021**, *70*, 9513512. [CrossRef]
15. Bo, Q.; Yang, W.; Rijal, N.; Xie, Y.; Feng, U.; Zhang, J. Particle Pollution Estimation from Images Using Convolutional Neural Network and Weather Features. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 3433–3437. [CrossRef]
16. Silva, A.; Ranasinghe, R.; Sounthararajah, A.; Haghighi, H.; Kodikara, J. Beyond Conventional Monitoring: A Semantic Segmentation Approach to Quantifying Traffic-Induced Dust on Unsealed Roads. *Sensors* **2024**, *24*, 510. [CrossRef]
17. Li, B.; Mao, B. Real Time Dust Detection with Image Pyramid and Improved ResNeSt. *IEEE* **2023**, *2023*, 152–155. [CrossRef]
18. Yin, S.; Li, T.; Cheng, X.; Wu, J. Remote sensing estimation of surface PM2.5 concentrations using a deep learning model improved by data augmentation and a particle size constraint. *Atmosph. Environ.* **2022**, *287*, 119282. [CrossRef]
19. Atreya, Y.; Mukherjee, A. Efficient Resnet Model for Atmospheric Visibility Classification. In Proceedings of the 2021 2nd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 1–3 October 2021; pp. 1–5. [CrossRef]
20. Wang, S.; Yin, J.; Zhou, Z. A comprehensive evaluation method for dust pollution: Digital image processing and deep learning approach. *J. Hazard. Mater.* **2024**, *475*, 134761. [CrossRef] [PubMed]
21. Kim, S.; Yu, H.; Yoon, J.; Park, E. Micro-Locational Fine Dust Prediction Utilizing Machine Learning and Deep Learning Models. *Comput. Syst. Sci. Eng.* **2024**, *48*, 413–429. [CrossRef]
22. Jeon, S.; Son, Y.S. Prediction of fine dust PM10 using a deep neural network model. *Korean J. Appl. Stat.* **2018**, *31*, 265–285. [CrossRef]
23. Jang, S.B. Dust Prediction System based on Incremental Deep Learning. *J. Converg. Cult. Technol.* **2023**, *9*, 301–307. [CrossRef]
24. Li, L.; Zhang, R.; Sun, J.; He, Q.; Kong, L.; Liu, X. Monitoring and prediction of dust concentration in an open-pit mine using a deep-learning algorithm. *J. Environ. Health Sci. Eng.* **2021**, *19*, 401–414. [CrossRef]
25. Xiong, R.; Tang, P. Machine learning using synthetic images for detecting dust emissions on construction sites. *Smart Sustain. Built Environ.* **2021**, *10*, 487–503. [CrossRef]
26. Lee, K.H.; Hwang, W.S.; Choi, M.R. Design of a 1-D CRNN Model for Prediction of Fine Dust Risk Level. *J. Digit. Converg.* **2021**, *19*, 215–220. [CrossRef]
27. Hwang, C.H.; Shin, K.W. CNN-LSTM Combination Method for Improving Particular Matter Contamination (PM2.5) Prediction Accuracy. *J. Korea Inst. Inf. Commun. Eng.* **2020**, *24*, 57–64. [CrossRef]
28. Lou, C.; Liu, H.; Li, Y.; Peng, Y.; Wang, J.; Dai, L. Relationships of relative humidity with PM 2.5 and PM 10 in the Yangtze River Delta, China. *Environ. Monit. Assess.* **2017**, *189*, 582. [CrossRef] [PubMed]
29. Tai, A.P.K.; Mickley, L.J.; Jacob, D.J. Correlations between fine particulate matter (PM2.5) and meteorological variables in the United States: Implications for the sensitivity of PM2.5 to climate change. *Atmosph. Environ.* **2010**, *44*, 3976–3984. [CrossRef]
30. Park, C.S. Variations of PM10 concentration in Seoul during 2015 and relationships to weather condition. *J. Assoc. Korean Photo-Geogr.* **2017**, *27*, 47–64.

31. Lee, J.H.; Kim, Y.M.; Kim, Y.K. Spatial panel analysis for PM2.5 concentration in Korea. *J. Korean Data Inf. Sci. Soc.* **2017**, *28*, 473–481.
32. Kim, H.L.; Moon, T.H. Machine learning-based Fine Dust Prediction Model using Meteorological data and Fine Dust data. *J. Korean Assoc. Geogr. Inf. Stud.* **2021**, *24*, 92–111. [CrossRef]
33. Choi, W.H.; Cheong, K.S. Analysis of the Factors Affecting Fine Dust Concentration Before and After COVID-19. *J. Korean Soc. Hazard Mitig.* **2021**, *21*, 395–402. [CrossRef]
34. Open MET Data Portal. Available online: https://data.kma.go.kr/ (accessed on 10 June 2024).
35. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
36. Awad, M.; Khanna, R. *Support Vector Regression, Efficient Learning Machines*; Springer: NewYork, NY, USA, 2015; pp. 67–80. Available online: https://link.springer.com/chapter/10.1007/978-1-4302-5990-9_4 (accessed on 15 June 2024).
37. Sung, S.H.; Kim, S.; Ryu, M.H. A Comparative Study on the Performance of Machine Learning Models for the Prediction of Fine Dust: Focusing on Domestic and Overseas Factors. *Innov. Stud.* **2020**, *15*, 339–357. [CrossRef]
38. Chen, T.; Giestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016. [CrossRef]
39. Hwang, H.; Kim, S.; Song, G. XGBoost Model to Identify Potential Factors Improving and Deteriorating Elderly Cognition. *J. Korean Inst. Next Gener. Comput.* **2018**, *14*, 16–24.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.