Journal of Hazardous Materials 511 (2026) 142172

Contents lists available at ScienceDirect

Journal of Hazardous Materials

journal homepage: www.elsevier.com/locate/jhazmat

A multi-view machine learning approach for estimating $PM_{2.5}$ concentrations from smartphone photographs

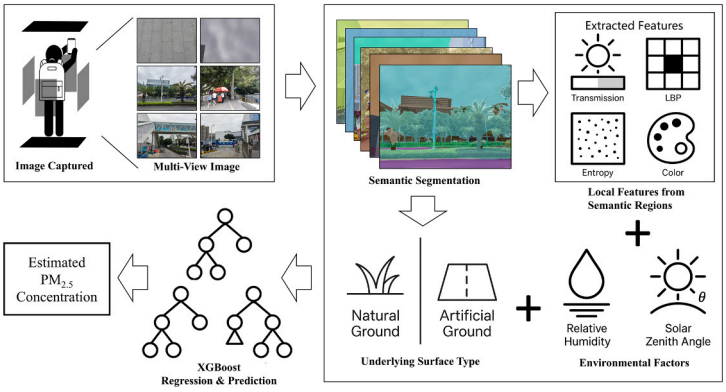
Jianzheng Liu [a,b,c], Zurong Zheng [a], Fei Yao [d,*], Weifeng Li [e,f,**]

[a] School of Public Affairs, Xiamen University, Xiamen, Fujian 361005, China
[b] Fujian Key Laboratory of Urban Intelligent Sensing and Computing, Xiamen, Fujian 361005, China
[c] Xiamen Key Laboratory of Integrated Application of Intelligent Technology for Architectural Heritage Protection, Xiamen, Fujian 361005, China
[d] China-UK Low Carbon College, Shanghai Jiao Tong University, Shanghai 201306, China
[e] Department of Urban Planning and Design, The University of Hong Kong, Pokfulam, Hong Kong SAR, China
[f] Urban Systems Institute, The University of Hong Kong, Hong Kong SAR, China

HIGHLIGHTS

- We construct a spatiotemporally aligned multi-view image-$PM_{2.5}$ dataset.
- The dataset addresses spatiotemporal mismatch issues in priori studies.
- We develop a multi-view machine learning approach for image-based $PM_{2.5}$ estimation.
- The approach reduces sensitivity to visual variations in single-view methods.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Smartphone Photograph
$PM_{2.5}$
XGBoost
Mobile Monitoring
Image Feature Extraction

ABSTRACT

Accurate and efficient estimates of fine particulate matter ($PM_{2.5}$) concentrations and associated exposure from smartphone photographs can provide the public with personalized risk information, helping to raise environmental risk awareness and reduce health risks. Existing studies are limited to estimating $PM_{2.5}$ concentrations from single-view smartphone photographs, where the spatiotemporal mismatch between photo capture and $PM_{2.5}$ measurements during data collection often introduces systematic bias that could otherwise be avoided. To address this, we construct a rigorously spatiotemporally registered benchmark dataset of $PM_{2.5}$ measurements and multi-view smartphone photographs for two Chinese provinces (Hebei and Fujian) between 2022 and 2024. For each province, we use the XGBoost model to relate $PM_{2.5}$ measurements to image features that we extract from multiple semantic segmentation regions within the multi-view smartphone photographs. The proposed method achieves R^2 values of 0.96 and 0.89, and RMSE values of 6.82 and 3.63 $\mu g/m^3$ on the independent test sets for the Hebei and Fujian provinces, respectively, which consistently outperforms alternative methods that

* Corresponding author.
** Corresponding author at:Department of Urban Planning and Design, The University of Hong Kong, Pokfulam, Hong Kong SAR, China.
E-mail addresses: jzliu@xmu.edu.cn (J. Liu), zhengzurong2023@163.com (Z. Zheng), fei.yao@sjtu.edu.cn (F. Yao), wfli@hku.hk (W. Li).

https://doi.org/10.1016/j.jhazmat.2026.142172
Received 20 December 2025; Received in revised form 11 April 2026; Accepted 20 April 2026
Available online 26 April 2026

rely on single-view smartphone photographs or overall image features. Key predictive features include image color extracted from vegetation regions, as well as color and local binary pattern extracted from sky regions. Overall, our work advances image-based air quality monitoring by providing a multi-view machine learning approach and the first rigorously spatiotemporally registered dataset of $PM_{2.5}$ measurements and smartphone photographs, which enables more reliable $PM_{2.5}$ estimates and highlights the role of semantic image features in predictive performance.

1. Introduction

Accurate air quality information is crucial for enabling the public to take effective measures to protect their health. The public often obtains air quality information from publicly released data that are underpinned by ground-based monitoring stations. However, the limited spatial coverage of these monitoring stations restricts their data to representing only average regional air pollution levels, thus failing to capture the specific air quality conditions to which individuals are exposed to [1]. Therefore, a discrepancy exists between publicly released air quality information and actual air pollution exposure levels within microenvironments, resulting in inaccurate perceptions of air pollution. This highlights the urgent need for a low-cost and easily accessible method that can provide the public with personalized air pollution exposure information. Such environmental monitoring and feedback about exposure levels can equip people with useful information for precautionary behavior [2–4]. To address this need, image-based methods have emerged as a promising solution, with previous studies demonstrating a strong correlation between image features and air pollutant concentrations [5]. The widespread use of mobile cameras further enhances the feasibility of these methods by enabling large-scale, cost-effective data collection.

The primary focus of this study is fine particulate matter ($PM_{2.5}$), which has well-documented adverse health impacts [6–9]. Recently, a number of studies have proposed methods that use mobile camera images to estimate $PM_{2.5}$ concentrations (e.g., [10–17]), yielding significant advancements. However, a key limitation of these studies is their reliance on matching images with $PM_{2.5}$ data from the nearest air quality monitoring station, a strategy that lacks precise spatiotemporal registration. A notable distance often exists between the image's location and the monitoring station, meaning that the station's data may not accurately represent the air quality at the precise location where the photo was taken. This discrepancy may introduce additional uncertainty and potential errors into the estimation models during the training phase. To address this mismatch, some studies have adopted vehicle-based monitoring [18–20]. However, such methods are costly, and its measurements are sensitive to driving speed, as airflow caused by the vehicle's movement can distort pollutant readings, leading to measurement uncertainties [21,22]. In addition, since sampling routes typically follow urban main roads, spatial coverage remains limited. It largely ignores areas where people commonly engage in outdoor activities, such as residential communities and public squares.

Additionally, previous methods largely rely on single-view images for estimating $PM_{2.5}$ concentrations (e.g., [23,24]). The key challenge of such an approach is that visual information is captured from only one direction, which in turn makes $PM_{2.5}$ estimates highly susceptible to varying illumination conditions along the camera's line of sight [25,26]. For example, images taken facing the sun can differ markedly from those taken with the sun behind the camera, even under the same ambient $PM_{2.5}$ concentration. Multi-view approaches can theoretically mitigate the issue by integrating information from various directions, moving beyond the narrow context of single-view approaches [27]. However, such approaches remain underexplored due to the lack of a rigorously spatiotemporally registered dataset of $PM_{2.5}$ measurements and multi-view smartphone photographs, not to mention the unresolved question of whether simply increasing the number of views can consistently improve model performance.

The objective of this study is to overcome the spatiotemporal mismatch in training data and the visual limitations of single-view approaches in current image-based $PM_{2.5}$ estimation studies. The core contributions of our study primarily lie in two aspects: the novelty of our data collection and the innovation of our methodological framework. Specifically:

(1) We establish a systematic data collection workflow and construct two rigorously spatiotemporally registered datasets of $PM_{2.5}$ measurements and multi-view smartphone photographs for two Chinese provinces (Hebei and Fujian) between 2022 and 2024. These datasets contain synchronized measurements of $PM_{2.5}$ concentrations and associated multi-view images, including sky, ground, and four horizontal directions (front, rear, left, and right). They represent the first-ever datasets of their kind to combine rigorously spatiotemporally registered $PM_{2.5}$ measurements with comprehensive multi-view smartphone photographs in the field of image-based $PM_{2.5}$ estimation, thereby mitigating the spatiotemporal mismatch in training data in previous studies and providing a highly reliable starting point for future research in this field.
(2) We design a novel multi-view methodological framework. The proposed method offers clear advantages over single-view methods, as it mitigates the impacts of varying illumination conditions and other factors on the accuracy of $PM_{2.5}$ concentration estimates. In addition, by leveraging ubiquitous smartphone technology, our approach has great potential to provide personalized air pollution feedback directly to individuals, fostering greater health awareness and promoting behavioral adaptations [3,28,29].

2. Materials and methods

Accurately estimating $PM_{2.5}$ concentrations from smartphone images presents a multifaceted challenge, requiring not only a high-quality, spatiotemporally registered dataset but also effective feature extraction strategies and modeling techniques. We propose a novel workflow that (1) establishes a high-quality, spatiotemporally registered dataset through rigorous data collection and quality control protocols, (2) extracts a set of predictive features using a region-wise strategy, and (3) utilizes these predictors to build, optimize, and evaluate a multi-view machine learning approach for estimating $PM_{2.5}$ concentrations. Each stage of the proposed workflow is detailed in the subsequent sections.

2.1. Dataset

Here we describe how we collect $PM_{2.5}$ concentration measurements and multi-view smartphone photographs and how we screen these data to create a rigorously spatiotemporally registered dataset of outdoor $PM_{2.5}$ paired with multi-view imagery for two Chinese provinces (Hebei and Fujian) between 2022 and 2024.

2.1.1. Data collection

Fig. 1 illustrates the locations and procedure of data collection, along with a representative example. Specifically, data are collected using a handheld aerosol monitor (TSI DustTrak II 8532), a centimeter-level handheld GPS receiver (HowayGIS T38P), and a smartphone. Data

collectors either walk or cycle along predetermined routes to collect data samples every few hundred meters. Various smartphone brands, such as Xiaomi, Huawei, and Apple, are used during the data collection process to improve the model's robustness to differences in device types. These routes are deliberately chosen to cover different levels of traffic, various types of urban scenes, and a wide range of residential areas. We space sampling points several hundred meters apart to prevent potential data leakage from training to test sets that may artificially inflate model performance, as excessive proximity could generate highly similar images and undermine their independence [30]. For a given sampling point, images are captured from six directions (front, rear, left, right, sky, and ground) using a smartphone with preset settings (e.g., using default auto-exposure and auto white balance, with the flash, HDR, and artificial filters disabled). It should be noted that the four horizontal images simply capture the scenes relative to the photographer's facing direction (i.e., front, rear, left, and right), rather than aligning with fixed geographical compass directions (e.g., North, South, East, West). The geographic coordinates are logged by the GPS receiver, and the $PM_{2.5}$ concentrations are measured by the DustTrak monitor.

Ultimately, we create two high-quality multi-view datasets with rigorous spatiotemporal registration: the Hebei dataset and the Fujian dataset. Hebei Province and Fujian Province were selected as they represent typical regions in northern and southern China. Hebei features a temperate monsoon climate with four distinct seasons and cold, dry winters. As a major heavy industry base, it frequently experiences severe $PM_{2.5}$ pollution during winter. In contrast, Fujian is located on the southeast coast and has a subtropical monsoon climate characterized by warmth, high humidity, and abundant rainfall. Benefiting from its extensive forest coverage, Fujian consistently ranks among the provinces with the best air quality in China. Monitoring campaigns were conducted for the Fujian dataset in April–May 2022, December 2022, and December 2024, and for the Hebei dataset in July 2022 and November 2024. In total, 9966 images were collected at 1661 sampling points. The resolution of these images varies from 3264 × 2448 to 4032 × 3024 pixels. Detailed statistics of the datasets and the distribution of data collection conditions are summarized in Tables S1 and S2 of the Supplementary Materials.

Considering the relationship among visual degradation, $PM_{2.5}$ concentrations, and aerosol hygroscopic growth [31,32], as well as the potential influence of illumination angle on the images, we also include solar zenith angle and relative humidity as auxiliary features to further improve $PM_{2.5}$ estimation from smartphone images. The solar zenith angle is calculated based on latitude, longitude, and photo capture time, while relative humidity is sourced from the Meteorological and Energy Fluxes Data provided by the NASA POWER website [33].

2.1.2. Data quality check

Given the complexities of field data collection, the raw $PM_{2.5}$ concentration measurements and images require rigorous curation to ensure data quality. Building on the data quality check protocols proposed in previous research [34,35], we develop a tailored quality assurance and control protocol with modifications to address the unique characteristics of the current datasets.

Aerosol monitors may exhibit measurement deviations. Thus, we calibrate our monitor against nearby government air quality stations before data collection. Calibration is performed through collocation, which involves operating our monitor directly alongside the reference station to compare simultaneous $PM_{2.5}$ readings. If any systematic deviations are detected, the instrument is recalibrated to ensure measurement accuracy.

To ensure that the multi-view smartphone photographs and $PM_{2.5}$ concentration measurements are rigorously spatiotemporally registered, we first assign location data to each photo. Each photo is matched to the temporally nearest GPS instrument reading, provided the time difference is within five seconds; if no corresponding instrument reading is available, we use the GPS coordinates recorded in the photo's EXIF metadata as a fallback. The entire six-photo set is discarded when location data are absent from both sources. We then require that the six photos constituting a sampling point be captured within a 60-second interval and that the spatial distance between any two of them not exceed 10 m. Any set of photos that fails to meet these criteria is removed. Finally, rather than relying on a single instantaneous $PM_{2.5}$ reading, we define the $PM_{2.5}$ concentration at each sampling point using a dynamic time-window average. Specifically, we identify the earliest and latest photo capture times among the set of images at a given location. All valid $PM_{2.5}$ readings within this time window are extracted

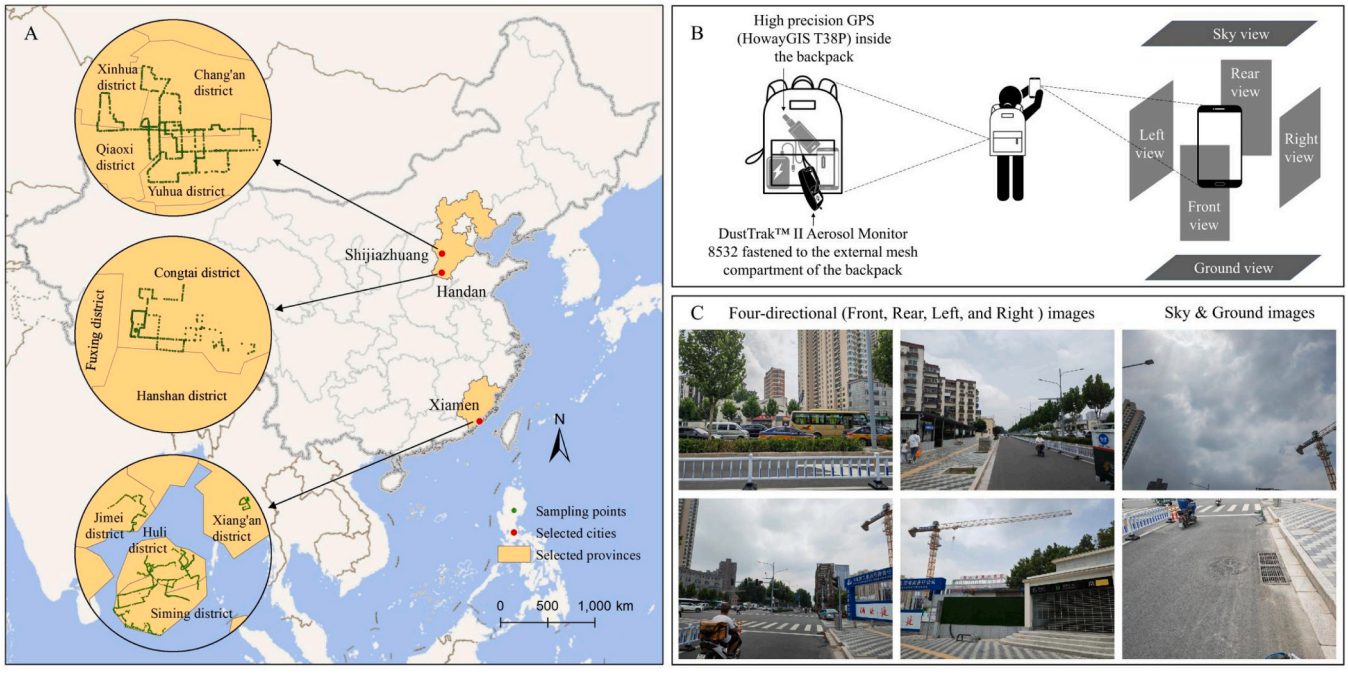


Fig. 1. Data collection. (A) Map of study areas and sampling points. (B) Illustration of data collection procedure. (C) An example of multi-view images collected using the data collection procedure in this study.

and averaged to represent the final $PM_{2.5}$ value. In rare instances where no valid $PM_{2.5}$ reading is captured within this dynamic window, we apply a brief tolerance as a fallback. Specifically, we link the photo set to a valid measurement provided that at least one photo timestamp falls within a 10-second window of that measurement. If no such measurement is found even with this tolerance, the air quality data for that sampling point are marked as missing.

While we mention in Section 2.1.1 that we space sampling points several hundred meters apart to prevent potential data leakage from training to test sets that may otherwise artificially inflate model performance, data collection in practice may still yield some exceptions due to the labour-intensive nature of fieldwork. Therefore, we enforce a minimum spatial separation of 100 m between any two sampling points, systematically removing one if the distance falls below this threshold. This procedure, together with the measures implemented during the data collection, jointly ensures the independence between the training and testing datasets, thereby contributing to the validity of model evaluation.

Finally, yet importantly, we remove potential outliers by setting $PM_{2.5}$ concentrations below 0 $\mu g/m^3$ or above 1000 $\mu g/m^3$ to NaN, and eliminate indoor samples by manually reviewing all photos. The threshold of 0–1000 $\mu g/m^3$ is consistent with United States Environmental Protection Agency [36] for air sensor technologies. Indoor and outdoor air pollution can exhibit substantially different characteristics and may therefore require tailored modelling strategies; in this study, we focus on outdoor $PM_{2.5}$ pollution and, accordingly, exclude indoor samples.

2.2. Feature engineering

Here we describe how we apply a region-wise feature extraction strategy to construct model predictors from different semantic regions of multi-view images we prepared in Section 2.1. As illustrated in Fig. 2, this strategy involves two steps: image segmentation and feature extraction from each segmented region. This region-wise strategy also enables a subsequent analysis of the relative importance of features from different semantic regions (e.g., sky, vegetation, and buildings) as visual indicators for $PM_{2.5}$ concentration estimation.

2.2.1. Image segmentation

We adopt the widely used DeepLabv3 + [37] deep learning approach for image segmentation, implemented via the open-source MMSegmentation framework.[1] The detailed configuration files are available in the OpenMMLab repository. Specifically, DeepLabv3 + employs a ResNet-50 backbone with an output stride of 8 (R-50-D8). Its encoder-decoder architecture and Atrous Spatial Pyramid Pooling (ASPP) effectively handle objects of varying sizes while preserving sharp segmentation boundaries. The model was trained on the Cityscapes dataset for 40,000 iterations with a crop size of 769 × 769, achieving an mIoU of 80.46% on the validation set. Because the scenes in the Cityscapes dataset closely resemble those in our collected outdoor smartphone photographs, the model's proven accuracy provides a reliable foundation for our subsequent region-wise feature extraction.

The algorithm performs image segmentation by classifying each pixel of an image into one of 19 predefined categories: terrain (i.e., soil or sand), vegetation (i.e., trees), sky, wall, building, road, traffic sign, traffic light, sidewalk, fence, pole, bus, train, truck, car, bicycle, motorcycle, rider, and person. We further refine these semantic segmentation classes for the purpose of our study. Specifically: (1) for sky images, we exclude all non-sky elements to isolate the clear sky region; (2) for ground images, we merge "road" and "sidewalk" into an "artificial ground" category and "terrain" and "vegetation" into a "natural ground" category; and (3) for four-directional images, we retain the original "vegetation" class while consolidating other categories into "buildings" (building, wall, fence, pole, traffic light, traffic sign) and "ground" (road, sidewalk, terrain) regions. The remaining mobile objects, including "person" and "train", are excluded to focus feature extraction on the primary, stable components of the image scene.

2.2.2. Feature extraction

Following the image segmentation, we proceed with feature extraction from different segmented regions. This includes the sky region extracted from the sky images, as well as the building, vegetation, and ground regions extracted from the four-directional images. We consider image features including color, transmission, information entropy, and local binary pattern (LBP), aiming to capture not only the direct visual cues of air pollution but also the microenvironment context of each sampling point. Image color can capture changes in tone, contrast, and saturation resulting from $PM_{2.5}$-induced light scattering [23]. Transmission directly reflects the scattering and attenuation of light caused by particles [17,38], making it a widely used indicator for $PM_{2.5}$ estimation. LBP identifies the loss of sharp details and the blurring of textures, such as edges and contours, that occur in hazy conditions [39]. Information entropy measures the overall complexity of an image, which tends to decrease as fine details are obscured by higher $PM_{2.5}$ concentrations [40]. Detailed mathematical definitions of the extracted features are provided in Text S1 in the Supplementary Materials.

Next, for each sampling point, we determine the underlying surface type by analyzing the corresponding ground image. Specifically, we compare the pixel proportions of artificial and natural ground regions within the ground image and classify the surface type according to the region with the highest proportion. The underlying surface provides valuable information about the microenvironment surrounding each sampling point and has important implications for estimating $PM_{2.5}$ concentrations. A higher density of nearby emission sources is associated with higher $PM_{2.5}$ concentrations [41]. Artificial ground generally indicates a man-made urban environment filled with vehicles, heating, cooking, and other human activities that contribute to $PM_{2.5}$ emissions, while natural ground typically corresponds to green spaces that help absorb particles.

Finally, we concatenate the underlying surface type, the four image features extracted from the sky and four-directional images, relative humidity, and solar zenith angle into a single feature vector, each of which corresponds to a measured $PM_{2.5}$ concentration. We split the full dataset into a training set and a test set, which contain 80% and 20% of the data, respectively. To avoid data leakage, all multi-view images from the same sampling point are kept together in the same dataset split. Feature vectors and corresponding $PM_{2.5}$ concentration measurements are used to train a machine learning model to learn the complex patterns between features and $PM_{2.5}$ concentrations. The trained model is then applied to the feature vectors from the test set, with the resulting predictions used to evaluate the model's performance.

2.3. Model building and evaluation

We employ the XGBoost algorithm [42] to estimate $PM_{2.5}$ concentrations from multi-view images for two reasons: its high predictive accuracy and its intrinsic capability to assess feature importance.

The predictive strength of the XGBoost model stems from its advanced gradient boosting framework, in which decision trees are sequentially built to correct the errors of preceding trees. In this process, hyperparameters including the number of trees, the fraction of samples used for fitting each tree, the maximum depth of each tree, and learning rate control the trade-off between model complexity and generalization performance. We optimized the hyperparameters of the XGBoost model using grid search with 10-fold cross-validation on the training set. The search space included the learning rate (eta) ranging from 0.01 to 0.20,

[1] The detailed configuration files can be accessed via the open-mmlab repository: https://github.com/open-mmlab/mmsegmentation/tree/main/configs/deeplabv3plus

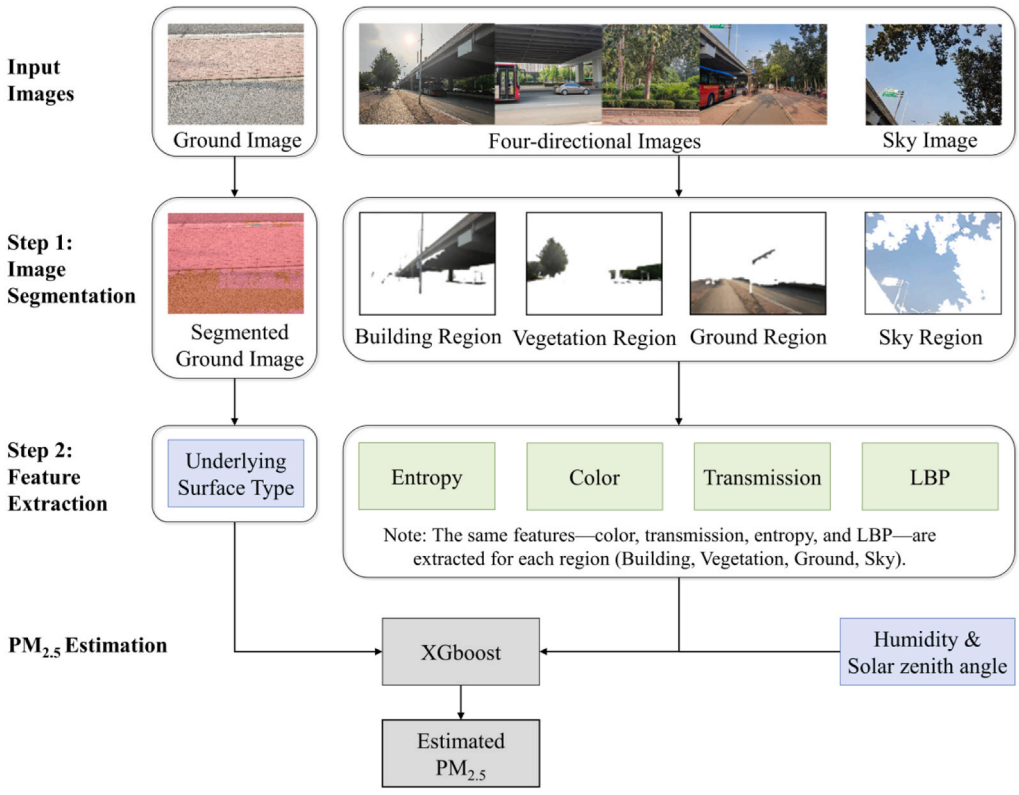


Fig. 2. Flow chart of feature engineering.

maximum tree depth from 4 to 10, subsample ratio from 0.5 to 1.0, and the number of trees from 200 to 700 (in increments of 100). In addition, gamma and lambda were fixed at their default values of 0 and 1, respectively. The grid search identified the optimal hyperparameters as 600 trees, a maximum depth of 5, a subsample ratio of 0.6, and a learning rate of 0.01.

The intrinsic capability of the XGBoost model to assess feature importance comes from its built-in gain metric, which measures the average improvement in the objective function when a given feature is used for splitting across the trees in an ensemble. Compared to permutation-based methods [43] or SHAP (Shapley Additive exPlanations) values [44], the gain metric identifies the most influential variables in a similar manner, while being readily available and computationally efficient. In this study, we use the built-in gain metric to examine the relative importance of features extracted from different semantic regions, including vegetation, ground, and buildings, and to provide insight into which features are most predictive of $PM_{2.5}$ concentrations.

We train and evaluate separate XGBoost models for the Hebei and Fujian datasets, given the unique characteristics of each dataset. This approach also allows us to validate the effectiveness and broader applicability of the proposed method. To describe the trained model performance on the test set, we employ two widely used statistical metrics: the coefficient of determination (R^2) and the root mean squared error (RMSE). R^2 quantifies the proportion of variance in $PM_{2.5}$ concentrations within the test set that is explained by the model, whereas RMSE measures the absolute magnitude of the prediction errors. A higher R^2 and a smaller RMSE indicate better model performance.

3. Results

3.1. Optimal model configuration

To investigate how different combinations of views and feature extraction strategies affect prediction accuracy, we design eight models to evaluate and compare their model performance on the Hebei dataset, with the aim of identifying the optimal model configuration (see Table 1). These models differ in three aspects. The first aspect is the Horizontal Views involved, either "1 (Front)" representing a single horizontal direction, or "4 (All-directions)" representing the integration of photographs from all four horizontal directions. The second aspect is the Vertical Views involved, which determines whether vertical perspectives are included ("Sky & Ground") or excluded ("None"). The third aspect is the Feature extraction strategy, which involves either an "Image-wise" approach for global features or a "Region-wise" strategy designed for a fine-grained analysis of various major semantic regions. In addition, for this comparative analysis, all eight models are developed using only image-based features; meteorological variables, namely relative humidity and solar zenith angle, are not included, as they would overshadow the role of image-based features, making it difficult to identify the optimal model configuration.

Table 1
Performance comparison between different model configurations.

Model	Horizontal view setting	Vertical view inclusion (sky & ground images)	Feature extraction strategy	R^2	RMSE (μg/m³)
1	1 (Front)	None	Image-wise	0.69	19.25
2	1 (Front)	None	Region-wise	0.75	16.33
3	1 (Front)	Sky & Ground	Image-wise	0.81	14.01
4	**1 (Front)**	**Sky & Ground**	**Region-wise**	**0.82**	**13.90**
5	4 (All directions)	None	Image-wise	0.74	16.14
6	4 (All directions)	None	Region-wise	0.80	14.12
7	4 (All directions)	Sky & Ground	Image-wise	0.79	15.07
8	4 (All directions)	Sky & Ground	Region-wise	0.79	14.22

The results show that the multi-view method generally outperforms those based on a single-view image. This is reflected by two aspects. First, a comparison between Models 1 and 5, and between Models 2 and 6, suggests that the inclusion of four directional images significantly improves the model's accuracy by providing additional complementary information. Second, a comparison between Models 1 and 3, Models 2 and 4, and Models 5 and 7, suggests the inclusion of vertical view (sky and ground) images plays a similar role. However, it is not always the case that including more images always improves the model's accuracy, as shown by comparisons between Model 3–7, Model 4–8, and Model 6–8. This suggests that multi-directional information can be adequately captured with a certain number of multi-view images, and that adding more than this number provides only marginal, negligible, or even negative benefits. Regarding the feature extraction method, the region-wise strategy provides a substantial improvement for models using a single directional image (Model 1 vs. 2), and also boosts model performance when using all four directional images (Model 5 vs. 6). However, its role becomes negligible once sky and ground images are part of the model predictors (Model 3 vs. 4; Model 7 vs. 8). In summary, while incorporating multi-view images improves model performance, adding too many of them may provide little additional benefit or even have adverse effects. Meanwhile, the region-wise feature extraction strategy offers a significant advantage when the number of input images is limited.

We select Model 4, which achieves the highest R^2 and lowest RMSE, as the base model, onto which we further add solar zenith angle and relative humidity to form the final model for estimating $PM_{2.5}$ concentrations. Table 2 shows that solar zenith angle and relative humidity individually increase the model's R^2 by 0.0214 and 0.1268, respectively, while decreasing the model's RMSE by 0.8358 and 6.1662 μg/m^3, respectively. Collectively, they improve the final model to have a high R^2 of 0.9557 and a low RMSE of 6.8247 μg/m^3. The larger role of relative humidity as compared to solar zenith angle is consistent with previous studies [12,15,39,45].

3.2. Model performance and applicability as compared to previous methods

We apply the optimal model configuration derived from the Hebei dataset to the Fujian dataset, achieving an R^2 of 0.89 and an RMSE of 3.63 μg/m^3. Compared with the model performance on the Hebei dataset, the R^2 is slightly lower but remains above 0.8, indicating that most of the $PM_{2.5}$ variance has still been effectively captured despite the direct use of the optimal model configuration transferred from the Hebei dataset. In contrast, the RMSE values are considerably smaller, which may be attributed to the relatively lower $PM_{2.5}$ concentration levels in Fujian (29.01 ± 11.31, as reported in our measurements after data quality check) as opposed to those in Hebei (43.92 ± 33.63, as reported in our measurements after data quality check). In short, the strong model performance on both the Hebei and Fujian datasets demonstrates the capability and generalizability of our proposed multi-view approach for estimating $PM_{2.5}$ concentrations from smartphone photographs.

In addition, while the sampling strategy and data quality checks already minimize spatial data leakage by enforcing a minimum distance between samples, we conduct a more stringent spatial cross-validation to rigorously evaluate the model's spatial transferability. This validation scheme assesses the model's performance under a challenging real-world scenario, in which predictions must be made for geographic regions entirely absent from the training data. Specifically, we divide the study area into grid cells of 1000 × 1000 m, and all observations located within the same grid cell are treated as a single group. These groups were then randomly assigned to either the training set or the test set. A 10-fold spatial cross-validation confirms the model's robust performance, yielding an R^2 of 0.91 (std. dev. 0.03), RMSE of 7.71 (std. dev. 1.98) on the Hebei dataset, and an R^2 of 0.86 (std. dev. 0.09), RMSE of 3.61 (std. dev. 1.20) on the Fujian dataset. These results underscore the model's practical utility and its ability to perform reliably in geographical distinct areas unrepresented in the training data.

Table 3 further compares the accuracy of our proposed multi-view approach with various machine and deep learning models from previous studies that rely on single-view images. These single-view approaches exhibit inherent limitations in capturing comprehensive and holistic visual information, often resulting in restricted predictive accuracy, with R^2 values ranging from 0.71 to 0.94. In contrast, our proposed multi-view method consistently achieves superior estimation performance, with R^2 values close to or exceeding 0.9. In addition, because the images used in this study were collected using mobile cameras rather than fixed cameras, our approach maintains excellent performance under complex environmental, lighting, and shooting conditions.

Finally, Table 4 presents a performance comparison between our proposed method and several competitive baseline models evaluated on the Hebei dataset. These baselines include simple linear models (Linear Regression and ElasticNet) using the exact same extracted feature set as our proposed model, as well as classical end-to-end deep learning regression models (ResNet-50 and DenseNet-121). For the multi-view deep learning models, a direct feature concatenation strategy is adopted to fuse the visual information from multiple views.

The experimental results demonstrate the necessity of the proposed method. The simple linear and ElasticNet baseline models perform poorly, with R^2 values only around 0.5, indicating that the underlying relationship between smartphone image features and $PM_{2.5}$ ground truth is complex and non-linear, making advanced tree-boosting algorithms like XGBoost indispensable.

While single-view CNNs produced reasonable estimates (R^2 up to 0.8746), their accuracy remains lower than that of the proposed multi-view XGBoost model. In addition, applying CNNs to multi-view data results in poorer performance compared to single-view CNNs. This decline can be attributed to the fact that these networks typically

Table 2
Model performance under different meteorological feature combinations.

Model	R^2	RMSE (μg/m^3)
Model 4	0.8164	13.8986
Model 4 + solar zenith angle	0.8378	13.0628
Model 4 + relative humidity	0.9432	7.7324
Model 4 + solar zenith angle + relative humidity	0.9557	6.8247

Table 3
Comparative performance evaluation of air quality prediction models: The proposed method vs. Single view image-based models from recent research.

Method	Study Area	Performance metric
Met-EfficientNet-B1-BiLSTM [12]	Delhi, India	Pearson r: 0.94, MAE: 4.06 μg/m^3
VGG-LSTM [24]	Shanghai, China	R^2: 0.94, RMSE: 5.11 μg/m^3
PE-ResNet18 [46]	Beijing, China	LCC: 0.91, RMSE: 13.47 μg/m^3
Regression-Based CNN [47]	Delhi, India	R^2: 0.94, NMSE: 0.03 μg/m^3
En3C-AQI-Net [48]	Delhi, India	R^2: 0.86, Accuracy: 89.28%
Patch-wise GBDT [39]	Shanghai, China	R^2: 0.88, RMSE: 10.42 μg/m^3
DCCN-ALSTM [49]	Shanghai, China	R^2: 0.71, RMSE: 14.07 μg/m^3
AQE-Net [50]	Karachi, Pakistan	Accuracy: 70.1%
MIFFN [15]	Beijing, China	R^2: 0.85, RMSE: 40.78 μg/m^3
SVR [17]	Beijing, China	Pearson r: 0.92, RMSE: 29.90 μg/m^3
MIFF [51]	Beijing, China	LCC: 0.85
AQC-Net [52]	Lanzhou, China	Accuracy: 74%
CNN-GBM [53]	Shanghai, China	R^2: 0.85, RMSE: 10.02 μg/m^3
Proposed Method in this study	**Hebei Province, China**	**R^2: 0.96, RMSE: 6.82 μg/m^3**
	Fujian Province, China	**R^2: 0.89, RMSE: 3.63 μg/m^3**

Table 4
Performance comparison between the proposed method and competitive baseline models.

Method	Backbone	R^2	RMSE(μg/m³)
Single-view	Linear Regression	0.5503	21.0382
Multi-view	Linear Regression	0.4848	25.065
Single-view	ElasticNet	0.5549	20.9311
Multi-view	ElasticNet	0.5632	23.0794
Single-view	Resnet-50	0.8746	11.6165
Multi-view	Resnet-50	0.8129	14.792
Single-view	Densenet-121	0.8585	15.8236
Multi-view	Densenet-121	0.7953	15.4732
Multi-view	**The proposed method**	**0.9557**	**6.8247**

contain a massive number of trainable parameters and require huge amounts of training data. Our dataset contains several thousand images, which is sufficient for traditional machine learning algorithms, but is relatively limited for deep learning models. Furthermore, concatenating high-dimensional visual data from multiple views significantly expands the feature space. Under the current sample size, this high dimensionality makes CNNs prone to overfitting. In contrast, XGBoost requires fewer computational resources and provides model interpretability. This interpretability is a key advantage, as it makes the feature importance analysis in Section 3.3 below possible.

3.3. Region-wise feature importance analysis

3.3.1. Importance analysis for different types of features

Fig. 3 demonstrates the importance of features from different semantic regions of multi-view images. Color and LBP are the most important features for predicting $PM_{2.5}$ concentrations in both Hebei and Fujian datasets. The important role of these features is closely related to the variations in weather conditions (such as sunny, cloudy, and overcast) during data collection. Different weather conditions correspond to different patterns of natural light, which directly alter the brightness, color, and edge texture in smartphone photographs [54,55]. For example, under conditions with poor lighting (such as overcast), the overall image contrast decreases, which may affect the visual cues of $PM_{2.5}$ and make image-based estimation more challenging. Consequently, the model relies heavily on Color and LBP, as they are able to capture these visual variations under complex lighting conditions.

The underlying surface type shows a significantly higher importance score in the Fujian dataset than in the Hebei dataset, likely because the Fujian dataset features a more balanced distribution of samples between "natural ground" and "artificial ground" categories (artificial vs. natural ground proportions: 77.2%/22.8% in Fujian; 98.4%/1.6% in Hebei). This more even split in Fujian makes the binary classification a more informative and consistently useful predictor for estimating $PM_{2.5}$ concentrations. In contrast, in the Hebei dataset, samples from the "artificial ground" category overwhelmingly outnumber those from the "natural ground" category, reducing the discriminative power of the underlying surface type and limiting its contributions to the model.

The feature importance analysis of the meteorological variables is consistent with the findings in Table 2, confirming again that relative humidity is the most influential factor. The rationale behind such a high importance score lies in the physical process of aerosol hygroscopic growth [56,57]. When ambient humidity rises, the water-soluble components within $PM_{2.5}$ take up water and expand in volume. This size expansion substantially enhances the particles' light scattering and extinction. As a result, smartphone photographs captured under high humidity conditions often exhibit reduced visibility and image contrast, thus making image-based estimation more challenging if humidity is ignored. Therefore, consistent with previous research [23,39,45,53], incorporating relative humidity is necessary, as it allows the model to account for the impact of ambient moisture on visual degradation, thereby estimating the true $PM_{2.5}$ concentration and leading to substantial improvements in model performance. Due to its role in accounting for aerosol hygroscopic growth, relative humidity holds high feature importance.

In contrast, the solar zenith angle contributes only marginally to the model performance. This may be because the information provided by the solar zenith angle has already been captured by the multi-view method used in this study. Specifically, solar zenith angle is a direct reflection of the time of day, which strongly affects the direction of ambient illumination [54]. Therefore, previous single-view studies have sought to incorporate the specific time of photo capture to account for these lighting variations [39,53]. In our study, however, the multi-view approach uses images from multiple directions rather than relying on a single view. As a result, the lighting information associated with solar zenith angle is already captured by the multi-view images, making solar zenith angle less informative as an additional predictor. The relatively low feature importance of solar zenith angle in this study therefore suggests that the effect of time of day on our model is limited.

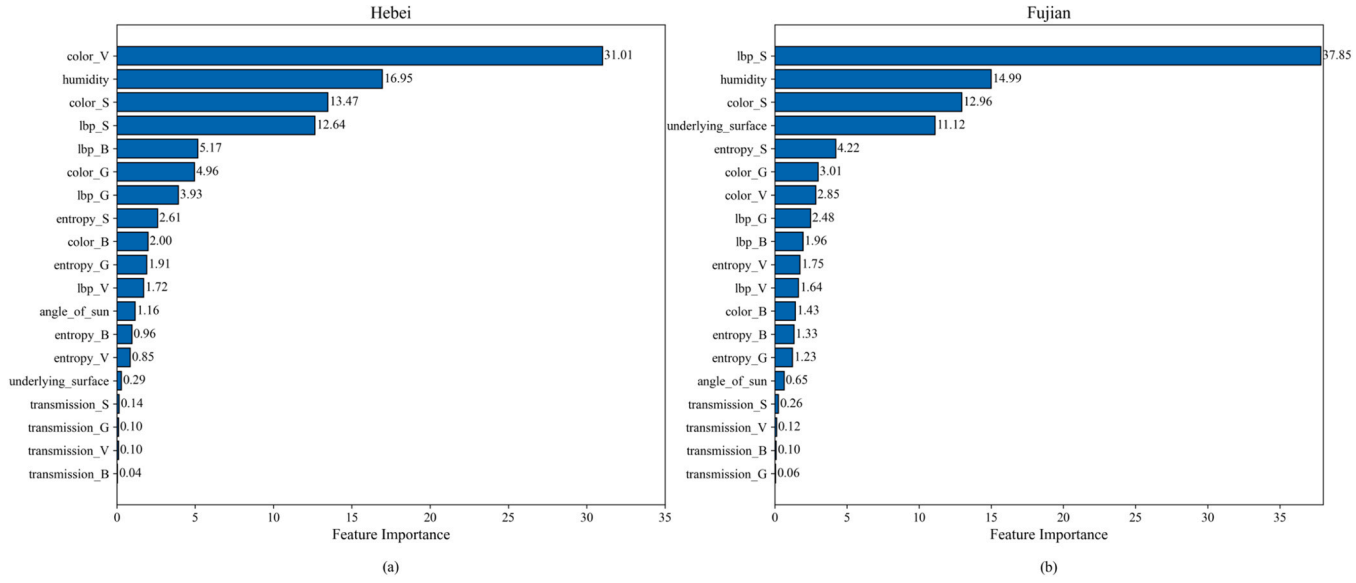


Fig. 3. Feature importance analysis. The suffix "V" indicates features extracted from the vegetation semantic region, "S" denotes features from the sky region, "B" refers to features from the buildings region, and "G" represents features extracted from the ground region.

Notably, the importance of transmission is relatively low in our analysis, which differs from previous findings [39]. This discrepancy likely arises because earlier studies used images captured by fixed cameras, where the positions of objects and the shooting angles remained constant, making image clarity the main source of variation. Under such controlled conditions, transmission could serve as an effective indicator of $PM_{2.5}$ concentration. In contrast, our study uses photographs from mobile monitoring, which introduce substantial variations in image illumination conditions. This variability reduces the usefulness of transmission for $PM_{2.5}$ estimation, while enhancing the relevance of other image features.

3.3.2. Importance analysis of different semantic regions

Fig. 3 also shows that features extracted from the sky region consistently demonstrate high importance. This contrasts with findings from previous fixed-camera studies, where features derived from building regions were typically found to be more critical for $PM_{2.5}$ estimation than those from the sky region [39,53].

A plausible explanation is the much greater variation in viewpoint and scene composition introduced by mobile monitoring. In fixed-camera settings, buildings serve as static references for assessing visual cues related to $PM_{2.5}$ concentration, whereas the sky region is prone to cloud disturbances, introducing instability to the extracted features. Furthermore, the building region contains richer visual details than the sky, potentially offering more visual information [39]. However, in our study, images are collected across different times and locations. Under these circumstances, buildings and the ground vary considerably in architectural style, color, material, and camera distance. Additionally, because the ground is in close proximity to the lens, the optical path is too short for $PM_{2.5}$-induced scattering to produce perceptible visual degradation. As a result, building and ground features, such as color and edge texture, become noisy and unsuitable as consistent predictors. In contrast, the sky provides a more stable reference in mobile settings. Because the sky tends to exhibit broadly similar visual characteristics across different locations at any given time, subtle variations in its features can more reliably reflect changes in $PM_{2.5}$ concentrations. While still affected by cloud disturbances, the sky background remains relatively stable when images are captured from different angles at the same location compared to the constantly changing buildings. Therefore, the color of the sky (color_S) and the haziness of sky edges (lbp_S), the latter captured by the local binary pattern of the sky region, act as stable indicators of $PM_{2.5}$-induced light extinction, explaining their prominent importance in our study.

Nonetheless, information from the sky alone is insufficient for tracking $PM_{2.5}$ variations, as the features derived from it exhibit only subtle changes, as mentioned earlier. It is thus essential to incorporate features extracted from other semantic regions. Our analysis reveals that the importance of features from other semantic regions varies significantly across the two datasets. While features extracted from the vegetation region are exceptionally dominant in the Hebei dataset, their importance in the Fujian dataset is comparable to that of buildings and the ground. A possible explanation is that geographical differences between the two study areas, including variations in natural environments, urban characteristics, and pollution levels, lead to different data distributions, which in turn alter the relative importance of predictive features. Specifically, in the Hebei dataset, vegetation color (color_V) is the most important feature. This pattern aligns with the province's temperate climate and the sampling periods of July and November. Between these months, vegetation appearance changes markedly from summer foliage to sparse or leafless conditions in late autumn, coinciding with seasonal spikes in $PM_{2.5}$ concentrations during the heating season. As a result, vegetation color exhibits a stronger association with $PM_{2.5}$ variations in Hebei. In contrast, in the Fujian dataset, vegetation color has a lower importance score, while the edge texture of the sky (lbp_S) is the most prominent feature. This is because Fujian has a subtropical climate and widespread evergreen vegetation, meaning that vegetation appearance remains relatively stable across sampling months.

Furthermore, Fujian has one of the highest vegetation coverage rates in China, and vegetation is widely distributed in urban environments. Consequently, vegetation-related features show minimal spatial variation across locations, providing limited discriminative power for machine learning models. In the absence of significant seasonal or spatial variations in vegetation that could serve as a predictive proxy, the model relies on the direct optical degradation caused by aerosol extinction. This visual degradation is best captured in the unobstructed sky region. This explains why sky texture, combined with relative humidity to account for aerosol hygroscopic growth, becomes the primary indicator for estimating $PM_{2.5}$ in this evergreen southern province.

4. Discussion

A major limitation of image-based methods is their insufficient robustness and stability, as their predictive results are often susceptible to changes in camera angles that significantly affect image content and illumination conditions. Our study addresses this critical gap by introducing a multi-view approach and incorporating meteorological information, demonstrating that integrating information from multiple perspectives substantially improves model accuracy and robustness. In addition, the proposed method shows superior performance over multiple competitive baseline models. This suggests that users can achieve a more reliable $PM_{2.5}$ estimation simply by capturing two additional photos from different angles at a given location. In practice, therefore, this multi-view approach improves estimation accuracy without substantially increasing the burden on users.

Another critical challenge in existing image-based $PM_{2.5}$ estimation research is the spatiotemporal mismatch within training data. To address this issue and demonstrate the effectiveness of our multi-view approach, we construct two high-quality datasets across diverse urban environments in northern and southern China. These benchmark datasets contain rigorously spatiotemporally registered $PM_{2.5}$ measurements and multi-view images. They overcome the limitations of prior training datasets by eliminating spatiotemporal mismatch during data collection, while also expanding spatial coverage beyond urban main roads to include diverse everyday microenvironments. By doing so, our work serves as a foundational starting point for the field to construct larger-scale datasets and more robust models. Leveraging the datasets, we develop a multi-view machine learning approach to estimate $PM_{2.5}$ concentrations from smartphone photographs, further quantifying the relative contributions of features from different semantic regions within images to inform subsequent research.

Encouraging health-protective behaviors has become an important direction in current air pollution management. Personalized air pollution exposure feedback plays a pivotal role in promoting such behaviors [1,4,58,59]. However, previous studies have predominantly used air pollution sensors to provide such information. The purchase and maintenance costs of these sensors create a barrier to entry, thereby limiting their widespread adoption and feasibility for large-scale public implementation. In addition, although many studies have attempted to combine personalized air pollution exposure data with interventions such as personalized health advice, communication with participants, group discussions and customized exposure reports [29,60,61], such intervention methods often rely on manual operation and human support. While they may be feasible in small-scale experimental studies, they are likely difficult to scale to larger populations. In contrast, our proposed method can be deployed on smartphones. The ubiquity and flexibility of smartphones allow for the integration of multiple intervention methods at a minimal marginal cost.

Furthermore, when deployed through crowdsourcing or citizen science approaches, the scalability of the proposed method allows individual observations to be aggregated into a large-scale, high-resolution dataset in which each data point is paired with a set of multi-view

images. This approach stands in contrast to traditional sensor networks, whose widespread adoption is often hampered by their costs of purchase, maintenance, and calibration [62-64]. Our method, by contrast, leverages ubiquitous smartphones to substantially lower the barrier to participation, enabling crowdsourced monitoring on a scale previously difficult to achieve. This process not only fosters public awareness and community engagement, but also offers the fine-grained evidence needed to identify areas of environmental injustice, evaluate the effectiveness of local interventions, and ultimately support more inclusive and responsive evidence-based ecological and environmental governance [65–69]. Therefore, the proposed method has the promising potential to help raise public awareness of air pollution, promoting health-protective behaviors and supporting the precise management of environmental risks.

5. Limitations

Overall, we introduce a multi-view machine learning approach for estimating $PM_{2.5}$ concentrations from smartphone photographs. Our primary contributions lie in eliminating the spatiotemporal mismatch during the training data construction phase and proposing a novel multi-view methodological framework. While our proposed framework demonstrates superior performance and outperforms existing baseline models such as CNNs, these findings are observed within the constraints of the current relatively small dataset size. Moving forward, expanding the dataset and exploring more advanced multi-view learning architectures (such as attention-based fusion or transformer networks) remain highly promising direction for advancing image-based $PM_{2.5}$ estimation.

Our primary goal in this study was to demonstrate the feasibility of the multi-view approach using a small but rigorously curated dataset. While the strong performance on both the Hebei and Fujian datasets validates the reliability of our proposed method, this relatively limited dataset size inevitably introduces some limitations, alongside those inherent to mobile monitoring itself. For instance, because portable aerosol monitors like the DustTrak lack a heated drying function, ambient moisture may introduce measurement bias. In addition, our dataset does not cover rainy or extreme conditions. Furthermore, as our dataset focuses on specific months and daylight hours across two provinces, the direct generalizability of the trained models is geographically limited. Directly applying these models to geographical regions with different climates or landscapes may therefore involve substantial domain shifts, thereby degrading model performance. In fact, overcoming domain shifts remains an unresolved challenge across many machine learning research fields [70,71], including image-based $PM_{2.5}$ estimation. Accordingly, the current models should be used with caution under complex weather conditions and when applied to other regions with different climates and landscapes.

Importantly, while the specific models are region-dependent, our proposed methodological framework is transferable. To apply our approach in a new area, researchers can simply follow our workflow to collect local data and retrain the model. Therefore, the established framework still provides a reliable foundation for scaling up crowdsourced monitoring in diverse environments. Ultimately, our work serves as a meaningful starting point for the field, enabling future research to build larger-scale datasets and better models. To address these remaining gaps and better overcome cross-regional domain shifts, we recommend that future research expand the dataset to encompass more diverse locations, seasons, and complex weather conditions. Developing advanced models with domain generalization capabilities will also be an important direction to fully address this challenge.

6. Conclusions

Existing image-based $PM_{2.5}$ estimation methods often suffer from two shortcomings: the spatiotemporal mismatch in training data and the limited visual information of single-view images. Our study contributes to overcoming these limitations. Specifically, the core contributions of this study lie in the novelty of our data collection and the innovation of our methodological framework. We establish a systematic workflow to construct two rigorously spatiotemporally registered benchmark datasets, effectively overcoming spatiotemporal mismatch and providing a reliable starting point for future research. A novel multi-view methodological framework is also proposed. Models independently trained and evaluated on each dataset demonstrate that this framework effectively mitigates the visual biases in traditional single-view approaches, improving estimation accuracy and model robustness. Furthermore, our feature importance analysis yields valuable insights into the physical and environmental interpretability of the model.

However, the current models are still constrained by the restricted geographic and temporal representativeness of the datasets and by cross-region domain shifts caused by variations in climate, urban environments, and vegetation. Future research should expand data collection to more diverse climates, seasons, and complex weather conditions to build larger-scale datasets and better models. Moreover, developing models that can better generalize across domains will be essential for mitigating the impact of geographical domain shifts.

Ultimately, by leveraging ubiquitous smartphone technology, our approach shows strong potential to provide personalized, low-cost air quality feedback during daily activities. This capability complements existing monitoring networks by filling critical spatiotemporal gaps and offering individuals more accurate exposure information. In the long term, portable monitoring systems integrated into smartphones enabled by algorithms such as our multi-view image-based approach could enhance public awareness, encourage protective behaviors, and support policymakers in developing more precise and sustainable air quality management strategies.

Environmental Implication

Fine particulate matter ($PM_{2.5}$) is a hazardous air pollutant with adverse health effects. Estimating $PM_{2.5}$ concentrations from smartphone photographs can provide personalized exposure information to help individuals reduce health risks. Existing methods, however, largely rely on single-view images matched to nearby $PM_{2.5}$ monitoring data, a strategy that introduces systematic spatiotemporal misalignment in the training data and leaves estimations highly susceptible to variations in illumination and camera angles. We propose a multi-view machine learning approach that addresses such bias and leverages complementary visual information to substantially improve $PM_{2.5}$ estimation, with the potential to advance environmental justice by expanding access to localized, reliable air quality information.

CRediT authorship contribution statement

Jianzheng Liu: Conceptualization, Methodology, Writing - Review & Editing, Supervision, Funding Acquisition, Project Administration, Visualization, Investigation, Resources, Data Curation. **Zurong Zheng:** Formal Analysis, Writing - Original Draft, Visualization, Investigation, Software. **Fei Yao:** Writing - Review & Editing, Supervision, Validation. **Weifeng Li:** Supervision, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 42101199) and Fujian Provincial Social Science

Foundation General Research Fund (Grant No.: FJ2025B163). The meteorological data used in this study were obtained from the NASA Langley Research Center POWER Project funded through the NASA Earth Science Directorate Applied Science Program.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jhazmat.2026.142172.

Data availability

Data will be made available on reasonable request.

References

[1] McCarron, A., Semple, S., Braban, C.F., Swanson, V., Gillespie, C., Price, H.D., 2023. Public engagement with air quality data: using health behaviour change theory to support exposure-minimising behaviours. J Expo Sci Environ Epidemiol 33 (3), 321–331. https://doi.org/10.1038/s41370-022-00449-2.

[2] Helbig, C., Ueberham, M., Becker, A.M., Marquart, H., Schlink, U., 2021. Wearable sensors for human environmental exposure in urban settings. Curr Pollut Rep 7 (3), 417–433. https://doi.org/10.1007/s40726-021-00186-4.

[3] Becker, A.M., Masson, T., Helbig, C., Mohamdeen, A., Schlink, U., 2023. Wearable sensors increase perceived environmental health threat in cyclists and pedestrians: a randomized field study. J Transp & Health 32, 101660. https://doi.org/10.1016/j.jth.2023.101660.

[4] Sater, R.A., Perona, M., Huillery, E., Chevallier, C., 2024. The power of personalised feedback: evidence from an indoor air quality experiment. Behav Public Policy. https://doi.org/10.1017/bpp.2024.46.

[5] Eldering, A., Hall, J.R., Hussey, K.J., Cass, G.R., 1996. Visibility model based on satellite-generated landscape data. Environ Sci & Technol 30 (2), 361–370. https://doi.org/10.1021/es940717k.

[6] Huang, X., Steinmetz, J., Marsh, E.K., Aravkin, A.Y., Ashbaugh, C., Murray, C.J.L., Yang, F., Ji, J.S., Zheng, P., Sorensen, R.J.D., Wozniak, S., Hay, S.I., McLaughlin, S. A., Garcia, V., Brauer, M., Burkart, K., 2025. A systematic review with a Burden of Proof meta-analysis of health effects of long-term ambient fine particulate matter (PM2.5) exposure on dementia. Nat Aging 5 (5), 897–908. https://doi.org/10.1038/s43587-025-00844-y.

[7] Lelieveld, J., Evans, J.S., Fnais, M., Giannadaki, D., Pozzer, A., 2015. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature 525 (7569), 367–371. https://doi.org/10.1038/nature15371.

[8] Strak, M., Weinmayr, G., Rodopoulou, S., Chen, J., de Hoogh, K., Andersen, Z.J., Atkinson, R., Bauwelinck, M., Bekkevold, T., Bellander, T., Boutron-Ruault, M.-C., Brandt, J., Cesaroni, G., Concin, H., Fecht, D., Forastiere, F., Gulliver, J., Hertel, O., Hoffmann, B., Samoli, E., 2021. Long term exposure to low level air pollution and mortality in eight European cohorts within the ELAPSE project: pooled analysis. BMJ 374, n1904. https://doi.org/10.1136/bmj.n1904.

[9] Wei, Y., Feng, Y., Danesh Yazdi, M., Yin, K., Castro, E., Shtein, A., Qiu, X., Peralta, A.A., Coull, B.A., Dominici, F., Schwartz, J.D., 2024. Exposure-response associations between chronic exposure to fine particulate matter and risks of hospital admission for major cardiovascular diseases: population based cohort study. BMJ 384, e076939. https://doi.org/10.1136/bmj-2023-076939.

[10] Feng, C., Tian, Y., Gong, X.Y., Que, X.R., Wang, W.D., 2018. MCS-RF: mobile crowdsensing-based air quality estimation with random forest. Int J Distrib Sens Netw 14 (10). https://doi.org/10.1177/1550147718804702.

[11] Gu, K., Qiao, J.F., Li, X.L., 2019. Highly efficient picture-based prediction of PM2.5 concentration. IEEE Trans Ind Electron 66 (4), 3176–3184. https://doi.org/10.1109/tie.2018.2840515.

[12] Kamble, A., Aramkul, S., Champrasert, P., 2025. A mobile image-driven PM2.5 estimation framework using deep learning techniques. Ieee Access 13, 16196–16207. https://doi.org/10.1109/access.2024.3521966.

[13] Liu, X.Y., Song, Z., Ngai, E., Ma, J., Wang, W.D., 2015. PM2.5 monitoring using images from smartphones in participatory sensing. *34th*. IEEE Conf Comput Commun (INFOCOM) 630–635. https://doi.org/10.1109/INFCOMW.2015.7179456.

[14] Rijal, N., Gutta, R.T., Cao, T.T., Lin, J., Bo, Q.R., Zhang, J., & Ieee. (2018). Ensemble of Deep Neural Networks for Estimating Particulate Matter from Images. *3rd IEEE International Conference on Image, Vision and Computing (ICIVC)*, 733-738. https://doi.org/10.1109/ICIVC.2018.8492790.

[15] Wang, F., Yao, S.Q., Luo, H.W., Huang, B., 2022. Estimating high-resolution PM2.5 concentrations by fusing satellite AOD and smartphone photographs using a convolutional neural network and ensemble learning, 21, Article Remote Sens 14 (6), 1515. https://doi.org/10.3390/rs14061515.

[16] Wu, L.J., Liu, X.S., Zhang, X., Wang, R., Guo, Z.H., 2025. End-to-end deep learning for pollutant prediction using street view images, 11, Article Urban Clim 60, 102368. https://doi.org/10.1016/j.uclim.2025.102368.

[17] Yao, S.Q., Wang, F., Huang, B., 2022. Measuring PM2.5 concentrations from a single smartphone photograph, 18, Article Remote Sens 14 (11), 2572. https://doi.org/10.3390/rs14112572.

[18] Feldman, A., Kendler, S., Marshall, J., Kushwaha, M., Sreekanth, V., Upadhya, A.R., Agrawal, P., Fishbain, B., 2023. Urban air-quality estimation using visual cues and a deep convolutional neural network in Bengaluru (Bangalore), India. Environ Sci & Technol 58 (1), 480–487. https://doi.org/10.1021/acs.est.3c04495.

[19] Xu, J.S., Zhang, M.Q., Ganji, A., Mallinen, K., Wang, A., Lloyd, M., Venuta, A., Simon, L., Kang, J.W., Gong, J.M., Zamel, Y., Weichenthal, S., Hatzopoulou, M., 2022. Prediction of short-term ultrafine particle exposures using real-time street-level images paired with air quality measurements. Environ Sci & Technol 56 (18), 12886–12897. https://doi.org/10.1021/acs.est.2c03193.

[20] Zhong, H., Chen, D., Wang, P., Wang, W., Shen, S., Liu, Y., Zhu, M., 2025. Predicting on-road air pollution coupling street view images and machine learning: a quantitative analysis of the optimal strategy. Environ Sci & Technol 59 (7), 3582–3591. https://doi.org/10.1021/acs.est.4c08380.

[21] Hassani, A., Castell, N., Watne, Å.K., Schneider, P., 2023. Citizen-operated mobile low-cost sensors for urban PM2.5 monitoring: field calibration, uncertainty estimation, and application. Sustain Cities Soc 95, 104607. https://doi.org/10.1016/j.scs.2023.104607.

[22] Munir, M.M., Adrian, M., Saputra, C., Lestari, P., 2022. Utilizing low-cost mobile monitoring to estimate the PM2.5 inhaled dose in urban environment. Aerosol Air Qual Res 22 (6), 220079. https://doi.org/10.4209/aaqr.220079.

[23] Lin, T.C., Wang, S.Y., Kung, Z.Y., Su, Y.H., Chiueh, P.T., Hsiao, T.C., 2023. Unmasking air quality: a novel image-based approach to align public perception with pollution levels. Environ Int 181, 108289. https://doi.org/10.1016/j.envint.2023.108289.

[24] Wang, X.C., Wang, M.Z., Liu, X.J., Mao, Y., Chen, Y., Dai, S.S., 2024. Surveillance-image-based outdoor air quality monitoring. Environ Sci Ecotechnology 18, 100319. https://doi.org/10.1016/j.ese.2023.100319.

[25] Guérin, J., Gibaru, O., Nyiri, E., Thieryl, S., Boots, B., 2018. Semantically meaningful view selection. IEEE/RSJ Int Conf Intell Robots Syst (IROS) 2018, 1061–1066. https://doi.org/10.1109/IROS.2018.8593524.

[26] Alzahrani, M., Usman, M., Jarraya, S.K., Anwar, S., Helmy, T., 2024. Deep models for multi-view 3D object recognition: a review. Artif Intell Rev 57 (12), 323. https://doi.org/10.1007/s10462-024-10941-w.

[27] Yu, Z.W., Dong, Z.Y., Yu, C.C., Yang, K.X., Fan, Z.W., Chen, C.L.P., 2025. A review on multi-view learning. Front Comput Sci 19 (7), 197334. https://doi.org/10.1007/s11704-024-40004-w.

[28] Bales, E., Nikzad, N., Quick, N., Ziftci, C., Patrick, K., Griswold, W.G., 2019. Personal pollution monitoring: mobile real-time air quality in daily life. Pers Ubiquitous Comput 23 (2), 309–328. https://doi.org/10.1007/s00779-019-01206-3.

[29] Park, Y.M., Chavez, D., Sousan, S., Figueroa-Bernal, N., Alvarez, J.R., Rocha-Peralta, J., 2023. Personal exposure monitoring using GPS-enabled portable air pollution sensors: a strategy to promote citizen awareness and behavioral changes regarding indoor and outdoor air pollution. J Expo Sci Environ Epidemiol 33 (3), 347–357. https://doi.org/10.1038/s41370-022-00515-9.

[30] Ramos P., Ramos R., Garcia N. Data leakage in visual datasets. arXiv preprint arXiv:2508.17416, 2025. doi:10.48550/arXiv.2508.17416.

[31] Malm, W.C., Day, D.E., 2001. Estimates of aerosol species scattering characteristics as a function of relative humidity. Atmos Environ 35 (16), 2845–2860. https://doi.org/10.1016/S1352-2310(01)00077-2.

[32] Malm, W.C., Sisler, J.F., Huffman, D., Eldred, R.A., Cahill, T.A., 1994. Spatial and seasonal trends in particle concentration and optical extinction in the United States. J Geophys Res Atmospheres 99 (D1), 1347–1370. https://doi.org/10.1029/93JD02916.

[33] NASA. NASA, 2025. Prediction of Worldwide Energy Resource (POWER) Project Data. NASA Langley Research Center.

[34] Liu, J., Li, W., Li, J., 2016. Quality screening for air quality monitoring data in China. Environ Pollut 216, 720–723. https://doi.org/10.1016/j.envpol.2016.06.037.

[35] Liu, J., Yao, F., Chen, H., Zhao, H., 2025. Quantifying the source-receptor relationships of PM2.5 pollution and associated health impacts among China, South Korea, and Japan: a dual perspective and an interdisciplinary approach. Environ Health Perspect 133 (3-4), 047011. https://doi.org/10.1289/EHP14550.

[36] United States Environmental Protection Agency. Peer Review and Supporting Literature Review of Air Sensor Technology Performance Targets (EPA/600/R-14/143), 2018. ⟨https://19january2021snapshot.epa.gov/sites/static/files/2018-10/documents/peer_review_and_supporting_literature_review_of_air_sensor_technology_performance_targets.pdf⟩.

[37] Chen L.C.E., Zhu Y.K., Papandreou G., Schroff F., Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. 15th European Conference on Computer Vision (ECCV) 2018;11211:833–851. doi:10.1007/978-3-030-01234-2_49.

[38] Li Y., Huang J., Luo J. Using user generated online photos to estimate and monitor air pollution in major cities. Proceedings of the 7th International Conference on Internet Multimedia Computing and Service 2015. doi:10.1145/2808492.2808564.

[39] Wang, X.C., Wang, M.Z., Liu, X.J., Zhang, X.X., Li, R.C., 2022. A PM2.5 concentration estimation method based on multi-feature combination of image patches. Environ Res 211, 113051. https://doi.org/10.1016/j.envres.2022.113051.

[40] Sun, K.Z., Tang, L.J., Huang, S.F., Qian, J.S., 2022. A photo-based quality assessment model for the estimation of PM2.5 concentrations. Iet Image Process 16 (4), 1008–1016. https://doi.org/10.1049/ipr2.12201.

[41] Lee, C., 2020. Impacts of multi-scale urban form on PM2.5 concentrations using continuous surface estimates with high-resolution in US metropolitan areas. Landsc Urban Plan 204, 103935. https://doi.org/10.1016/j.landurbplan.2020.103935.

[42] Chen T.Q., Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD) 2016:785–794. doi:10.1145/2939672.2939785.
[43] Breiman, L., 2001. Random Forests. Mach Learn 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.
[44] Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Proc 31st Int Conf Neural Inf Process Syst 4768–4777. https://doi.org/10.48550/arXiv.1705.07874.
[45] Li, J., Garshick, E., Hart, J.E., Li, L., Shi, L., Al-Hemoud, A., Huang, S., Koutrakis, P., 2021. Estimation of ambient PM2.5 in Iraq and Kuwait from 2001 to 2018 using machine learning and remote sensing. Environ Int 151, 106445. https://doi.org/10.1016/j.envint.2021.106445.
[46] Fang, X.Q., Li, Z., Yuan, B., Chen, Y.H., 2024. Image-based PM2.5 estimation from imbalanced data distribution using prior-enhanced neural networks. Ieee Sens J 24 (4), 4677–4693. https://doi.org/10.1109/jsen.2023.3343080.
[47] Sarkar, P., Saha, D.D.V., Saha, M., 2024. Real-time air quality index detection through regression-based convolutional neural network model on captured images, 9, Article Environ Qual Manag 34 (1), e22276. https://doi.org/10.1002/tqem.22276.
[48] Mohan, A.S., Abraham, L., 2024. An ensemble deep learning approach for air quality estimation in Delhi, India. Earth Sci Inform 17 (3), 1923–1948. https://doi.org/10.1007/s12145-023-01210-5.
[49] Zhang, B., Geng, Z.Y., Zhang, H.W., Pan, J.G., 2022. Densely connected convolutional networks with attention long short-term memory for estimating PM2.5 values from images, 10, Article J Clean Prod 333, 130101. https://doi.org/10.1016/j.jclepro.2021.130101.
[50] Ahmed, M., Shen, Y.L., Ahmed, M., Xiao, Z.M., Cheng, P., Ali, N., Ghaffar, A., Ali, S., 2022. AQE-Net: a deep learning model for estimating air quality of Karachi City from mobile images. Remote Sens 14 (22), 5732. https://doi.org/10.3390/rs14225732.
[51] Wang, G.C., Shi, Q., Wang, H., Sun, K.Z., Lu, Y.X., Di, K.X., 2022. Multi-modal image feature fusion-based PM2.5 concentration estimation, 9, Article Atmos Pollut Res 13 (3), 101345. https://doi.org/10.1016/j.apr.2022.101345.
[52] Zhang, Q., Fu, F.C., Tian, R., 2020. A deep learning and image-based model for air quality estimation, 11, Article Sci Total Environ 724, 138178. https://doi.org/10.1016/j.scitotenv.2020.138178.
[53] Luo, Z.Y., Huang, F.F., Liu, H., 2020. PM2.5 concentration estimation using convolutional neural network and gradient boosting machine. J Environ Sci 98, 85–93. https://doi.org/10.1016/j.jes.2020.04.042.
[54] Hold-Geoffroy, Y., Athawale, A., & Lalonde, J.-F. (2019). Deep Sky Modeling for Single Image Outdoor Lighting Estimation. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6920-6928.
[55] Zhu Y., Wang T., Fu X., Yang X., Guo X., Dai J., Qiao Y., Hu X. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2023:21747–21758. doi:10.1109/CVPR52729.2023.02083.
[56] Ma, N., Zhao, C.S., Chen, J., Xu, W.Y., Yan, P., Zhou, X.J., 2014. A novel method for distinguishing fog and haze based on PM2.5, visibility, and relative humidity. Sci China-Earth Sci 57 (9), 2156–2164. https://doi.org/10.1007/s11430-014-4885-5.
[57] Chen, J., Li, Z., Lv, M., Wang, Y., Wang, W., Zhang, Y., Wang, H., Yan, X., Sun, Y., Cribb, M., 2019. Aerosol hygroscopic growth, contributing factors, and impact on haze events in a severely polluted region in northern China. Atmos Chem Phys 19 (2), 1327–1342. https://doi.org/10.5194/acp-19-1327-2019.
[58] Becker, A.M., Marquart, H., Masson, T., Helbig, C., Schlink, U., 2021. Impacts of Personalized sensor feedback regarding exposure to environmental stressors. Curr Pollut Rep 7 (4), 579–593. https://doi.org/10.1007/s40726-021-00209-0.
[59] Luo, X., Wang, S., Liu, C., Fu, Q., Wang, H., Yi, M., Guo, X., Wang, Q., Fu, Y., 2025. Concentration distribution and group disparity of traffic-derived NO_2 exposure in Baoshan District. Carbon Footprints 4, 14. https://doi.org/10.20517/cf.2024.53.
[60] Dobson, R., O'Donnell, R., De Bruin, M., Turner, S., Semple, S., 2017. Using air quality monitoring to reduce second-hand smoke exposure in homes: the AFRESH feasibility study. Tob Prev & Cessat 3. https://doi.org/10.18332/tpc/74645.
[61] Boso, A., Alvarez, B., Oltra, C., Garrido, J., Muñoz, C., Hofflinger, A., 2020. Out of sight, out of mind: participatory sensing for monitoring indoor air quality, 15, Article Environ Monit Assess 192 (2), 104. https://doi.org/10.1007/s10661-019-8058-z.
[62] deSouza, P.N., 2022. Key concerns and drivers of low-cost air quality sensor use. Sustainability 14 (1), 584. https://doi.org/10.3390/su14010584.
[63] Williams, D.E., 2019. Low cost sensor networks: how do we know the data are reliable? Acs Sens 4 (10), 2558–2565. https://doi.org/10.1021/acssensors.9b01455.
[64] Wang, R., Zhang, Y., Zhao, S., Wang, X., 2024. Open-data-based city-scale gridded carbon dioxide emission inventory: supporting urban carbon monitoring in Chengdu, China. Carbon Footprints 3, 14. https://doi.org/10.20517/cf.2024.19.
[65] Lu, T.J., Garcia, D.A., Garcia, A., Liu, Y.S., 2023. Leveraging crowd-sourced environmental data to assess air pollution exposure disparity: a case of Los Angeles County, 13, Article Int J Appl Earth Obs Geoinf 125, 103599. https://doi.org/10.1016/j.jag.2023.103599.
[66] Mahajan, S., Chung, M.K., Martinez, J., Olaya, Y., Helbing, D., Chen, L.J., 2022. Translating citizen-generated air quality data into evidence for shaping policy. Humanit & Soc Sci Commun 9 (1), 122. https://doi.org/10.1057/s41599-022-01135-2.
[67] Mahajan, S., Kumar, P., Pinto, J.A., Riccetti, A., Schaaf, K., Camprodon, G., Smári, V., Passani, A., Forino, G., 2020. A citizen science approach for enhancing public understanding of air pollution. Sustain Cities Soc 52, 101800. https://doi.org/10.1016/j.scs.2019.101800.
[68] Mahajan, S., Luo, C.H., Wu, D.Y., Chen, L.J., 2021. From do-it-yourself (DIY) to do-it-together (DIT): reflections on designing a citizen-driven air quality monitoring framework in Taiwan, 12, Article Sustain Cities Soc 66, 102628. https://doi.org/10.1016/j.scs.2020.102628.
[69] O'Regan, A.C., Grythe, H., Hellebust, S., Lopez-Aparicio, S., O'Dowd, C., Hamer, P. D., Santos, G.S., Nyhan, M.M., 2024. Data fusion for enhancing urban air quality modeling using large-scale citizen science data, 16, Article Sustain Cities Soc 116, 105896. https://doi.org/10.1016/j.scs.2024.105896.
[70] Liu, C., Wang, L., Lyu, L., Sun, C., Wang, X., Zhu, Q., 2023. DEJA VU: continual model generalization for unseen domains. Proc Int Conf Learn Represent (ICLR).
[71] Qu, S., Zou, T., Röhrbein, F., Lu, C., Chen, G., Tao, D., & Jiang, C. (2023). Upcycling models under domain and category shift. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,