



Image-based estimation of ambient particulate matter: a scoping review

Daniele Oxoli¹ · Afshin Moazzam¹ · Maria Antonia Brovelli¹ · Songnian Li² · Francesco Pirotti³ · Shishuo Xu⁴

Received: 14 March 2026 / Revised: 30 June 2026 / Accepted: 24 July 2026
© The Author(s) 2026

Abstract

Air quality is a major public health concern, while particulate matter (PM) measurements remain limited by the spatial coverage and high operational costs of reference monitoring stations. This scoping review synthesises the literature on image-based PM estimation from outdoor photos as a potential low-cost approach to increase PM concentration sampling coverage. Methods are classified by input data requirements and modelling strategy, with summaries of performance and implementation resources. Results suggest that deep learning models generally exhibit the strongest predictive capability, although cross-study comparisons remain indicative, as they are limited by the availability of comparable metrics and by differences in geographic application contexts. Machine learning and physics-based methods are generally less accurate but more interpretable and computationally efficient, while hybrid approaches offer a promising trade-off between accuracy and operability. Nevertheless, most applications remain below reference regulatory accuracy levels. Training datasets are often limited in size and coverage, benchmarks are lacking, and code and data are not always available. Future work should prioritise multi-site datasets, improved hybrid models, and open resources to support reproducible model assessment and reuse.

Keywords Air pollution monitoring · Particulate matter · PM_{2.5} · Outdoor pictures · Image processing

1 Introduction

Air quality monitoring is a key component of public health and environmental protection, supported by scientific evidence on the negative effects of air pollution on health and ecosystems [1]. Among these pollutants, particulate matter (PM) is associated with a burden of respiratory and cardiovascular disease morbidity and mortality. Fine particulate

matter with an aerodynamic diameter below 2.5 µm (PM_{2.5}) is of particular concern because it can penetrate deeply into the respiratory system and be transported over long distances in the lower atmosphere. For these reasons, accurate and high-resolution monitoring of PM concentrations is important for health risk mitigation strategies and policy support [2–4].

Traditional monitoring networks rely on ground-based stations, which provide regulatory-grade measurements but have high costs and often limited spatial coverage [5, 6]. Satellite products, such as aerosol optical depth (AOD), improve regional coverage but often lack the spatial and temporal resolution needed for health risk assessment applications [7, 8]. This gap motivates complementary approaches that can increase spatial detail at lower cost, including image-based methods that leverage fixed webcams, surveillance cameras, and mobile photos [4–6].

Image-based methods for PM estimation have gained attention in the last decade. These methods exploit visual characteristics in outdoor photographs (e.g., sky colour, contrast, haze, and visibility of distant objects) as indirect indicators of ambient PM. The wide availability of images from

✉ Daniele Oxoli
daniele.oxoli@polimi.it

¹ Department of Civil and Environmental Engineering, Politecnico Di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy

² Department of Civil Engineering, Toronto Metropolitan University, 350 Victoria Street, Toronto, ON M5B 2K3, Canada

³ Interdepartmental Research Center in Geomatics, Department of Land, Environment, Agriculture and Forestry, University of Padua, Via Dell'Università 16, 35020 Legnaro, Italy

⁴ Beijing University of Civil Engineering and Architecture, 15 Yongyuan Road, Daxing District, Beijing 102616, China

fixed webcams, surveillance cameras, and mobile devices enables dense spatial and temporal sampling without specialised instrumentation [5, 6]. When positional metadata are available, image-based estimates can easily be integrated with ground measurements, satellite observations, and reanalysis model estimates in geographic information systems (GIS) to support mapping and near real-time visualisation [7, 8].

Different methodological approaches have been proposed in the literature. Physics-based models use atmospheric optics to link image properties with PM [9]. Data-driven approaches based on machine learning (ML) and deep learning (DL) predict PM directly from pixels or learned features using, for example, convolutional neural networks (CNN). Hybrid solutions combine physics-informed features derived from images with ML or DL to improve prediction accuracy. Reported performance varies across geography, season, image acquisition settings, and availability of PM measurements co-located in time and space with the images. In general, reported performance ranges from moderate to high, with lower errors in controlled settings and higher errors in heterogeneous urban contexts [5, 10].

Open challenges remain. Many studies are limited to specific regions, which makes generalisation complex. Performance can degrade under variable lighting, weather, and visibility. Replicability is also constrained with only a few authors openly publishing datasets and source code [6, 11]. Furthermore, only a few recent literature reviews on image-based PM estimation are available [12, 13]. These generally focus on a limited set of studies and do not explicitly report implementation resource availability, including open source code, nor systematically assess the geographic coverage of test cases.

In view of the above, this review presents an analysis of the literature related to image-based estimation of ambient PM, with a focus on $PM_{2.5}$, with the following objectives: (i) identify all the relevant studies available on this emerging topic; (ii) to compare the proposed methodological approaches and identify the most promising directions; and (iii) to assess current applicative limitations, also with respect to authoritative PM monitoring, and to document the availability of resources (datasets, code, benchmarking protocols) that can support reproducibility and future improvements.

The remainder of this paper is organised as follows: Sect. 2 describes the review methodology. Section 3 presents a synthesis of the review results, including the geographic and temporal distribution of the studies, objectives, data sources, modelling strategies, reported performance, and code and data availability. Section 4 discusses the main methodological trends, reproducibility issues, and potential performance gaps with respect to regulatory PM monitoring.

Section 5 concludes with a summary of the main findings and directions for future work.

2 Methods

This literature review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework to promote transparency and reproducibility of the study. The literature search was conducted across major bibliographic databases, including Web of Science and Scopus, targeting peer-reviewed journal articles and conference proceedings from 2014 to 2024. The bibliographic database search combined pollutant-specific keywords (“ $PM_{2.5}$ ”, “particulate matter”, “air quality”) with imaging-related keywords (“image-based”, “photograph”, “camera”, “visibility”, “haze detection”).

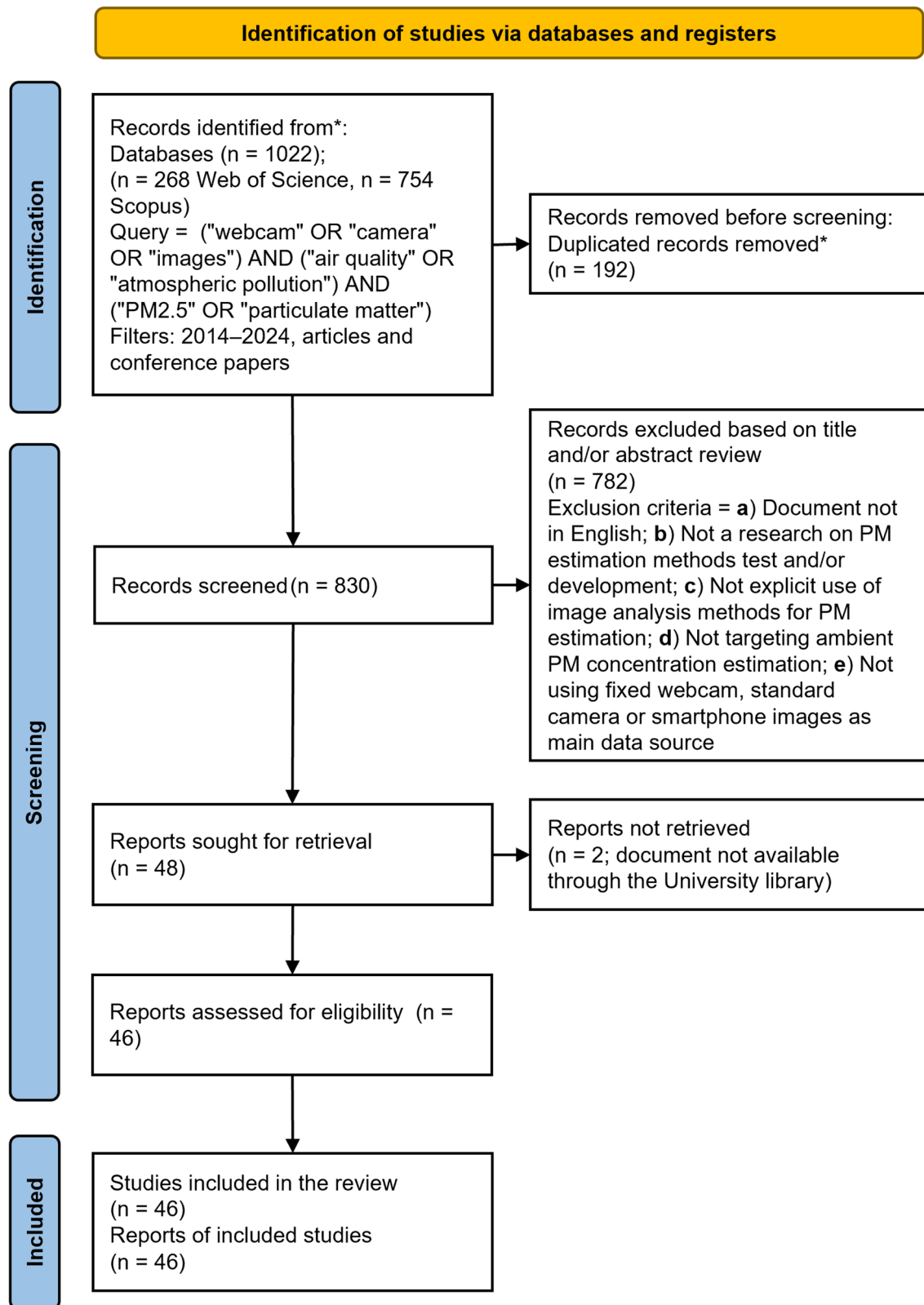
The initial search yielded 1022 records. Artificial Intelligence (ChatGPT) was used to support duplicate screening from the bibliographic tables exported from the selected databases. The excluded duplicates were then manually cross-checked on a sample basis to verify the correctness of the procedure. After removing duplicates (192), titles and abstracts of the remaining 830 unique records were screened for relevance, resulting in the exclusion of 782 records. Exclusion criteria included language, lack of content connected to the development or test of image-based ambient PM estimation methods, and use of main input data different from fixed webcam, standard camera, or smartphone photos.

Following title and abstract screening and subsequent verification of actual accessibility from the authors’ library resources, 46 articles were selected for full-text review. The PRISMA flow diagram summarising this process is presented in Fig. 1.

The stringent selection criteria adopted produce a relatively small corpus of studies, highlighting the modest extent of the literature specifically addressing the topic. While the size of the corpus does not fully support a systematic review, it is sufficiently broad to conduct a scoping review aiming at identifying key methodological trends, providing indications on modelling performances, and assessing the actual availability of resources to inform future studies.

Data from these studies are extracted using a standardised template capturing bibliographic information, study objectives, geographic location, type of input data, type of algorithmic approach, and performance metrics. Limitations and proposed improvements are elaborated from the interpretation of the articles content.

Quantitative information is summarised in charts (e.g., by year of publication, study region, input data type, and model category) to investigate trends. Where available,



*AI used to parse record lists and remove duplicates

Fig. 1 PRISMA flow diagram

performance metrics are recorded in their original units, with values reported directly by the authors of the studies. For consistency, the reported performance values refer, whenever available, to validation, test-set, or cross-validation results rather than training performance. When studies reported multiple metrics, model variants, or data splits, we selected the value corresponding to the best-performing configuration for $PM_{2.5}$ estimation, prioritising the main model proposed by the authors and preserving the metric type used in the original paper.

Qualitative synthesis focuses instead on identifying common challenges, such as lighting variability and camera calibration, and innovative solutions, such as feature fusion and physics-informed neural networks, proposed in the literature. To improve readability, citation lists in the main text are kept selective and representative, while the complete mapping of reviewed studies to categories, data sources, methods, and performance metrics is reported in Table 1 of Appendix A.

3 Results

3.1 Geographic and temporal distribution

The geographic distribution presented here refers to the test sites considered in the reviewed studies, which is not necessarily linked to the nationalities or institutional affiliations of the respective authors.

The reviewed studies show a strong geographic bias toward East Asia, especially China, which is associated with

28 out of the 46 selected studies (Fig. 2). Many works focus on major metropolitan regions using single-city datasets, while others include multi-site analyses across two or more provinces and/or regions [4, 7, 10, 14–16].

Taiwan is the second most represented location, with 5 studies [5, 9, 10, 17, 18]. One of these [10] evaluates sites in both Taiwan and mainland China. India follows with 4 studies focusing on highly urbanised and industrial regions [19–22]. Among these, two studies jointly evaluate sites in India and, respectively, China [21] and Thailand [19] to test models robustness and transferability across different atmospheric and geographical conditions.

Other contributions from Asia include South Korea [23] and Bangladesh [8]. Outside Asia there are isolated studies from the USA [2, 4] (with [4] evaluating also a site in China), Austria [24], Poland [3], Australia [25], and Greece [26]. In Europe, one study [27] integrates datasets from both Germany and Czechia. One study does not indicate the study region [28], focusing on model development without targeting a particular location.

Finally, the year distribution of the selected studies is reported in Fig. 3, showing a moderate increase in publication frequency from 2018 onwards.

3.2 Objectives of the reviewed studies

The reviewed studies present a variety of objectives that reflect differences in application context, methodological scope, and target air pollutants. However, comparable workflows are used to implement most of these studies (see Fig. 4 for a generalised workflow). The majority of works

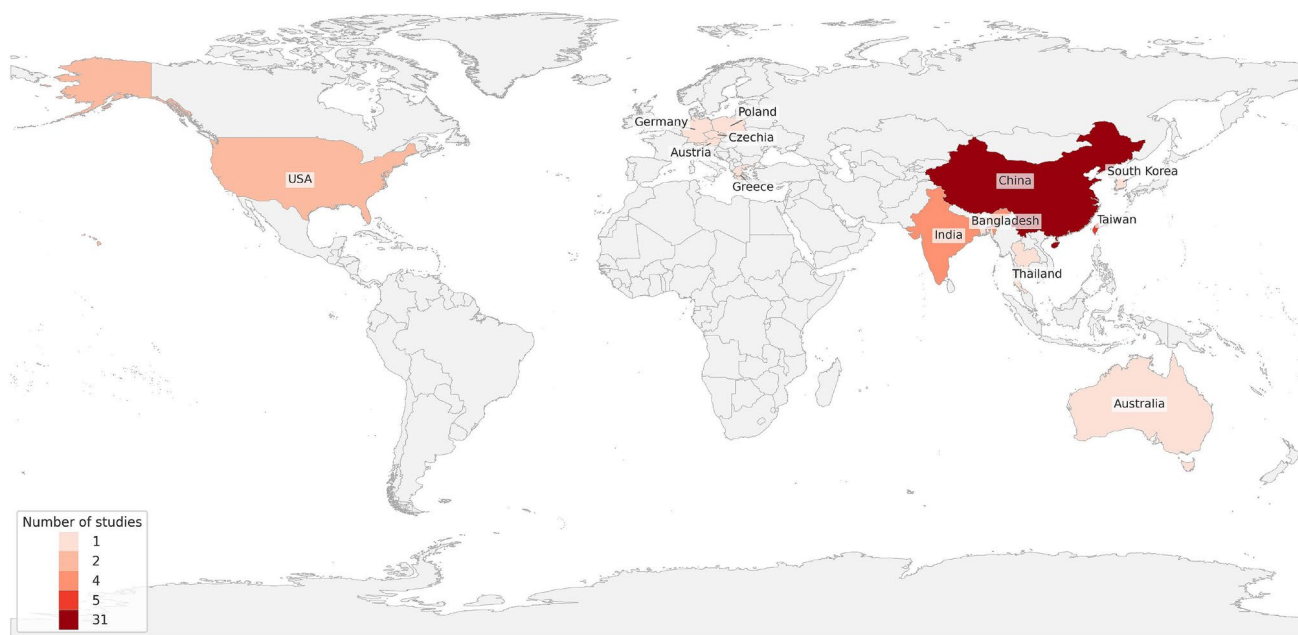


Fig. 2 Reviewed studies distribution by country (multi-country studies are counted once for each country involved)

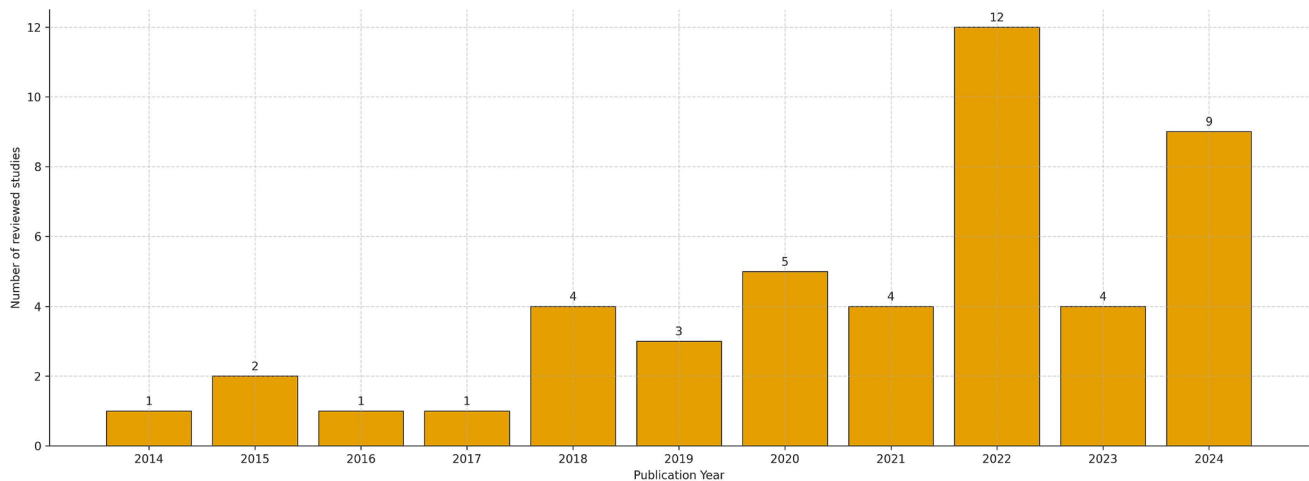


Fig. 3 Reviewed studies distribution by publication year

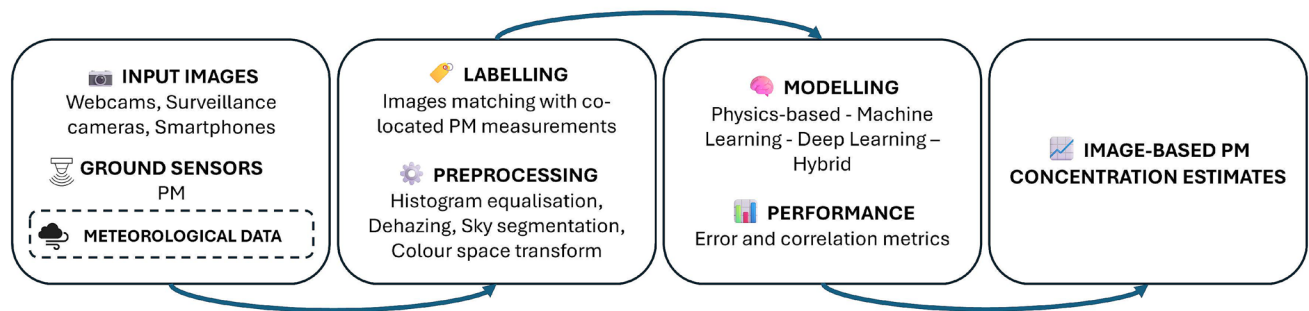


Fig. 4 Generalised workflow of image-based methods for PM estimation

focused on estimating $PM_{2.5}$ concentrations from red, green, blue (RGB) images [4, 5, 11, 25, 29–31], while a smaller subset extended the analysis to multiple pollutants, including PM_{10} , nitrogen dioxide (NO_2), and ozone (O_3), see e.g., [3, 9, 10, 24, 27, 32]. Some works framed the problem as a classification task, predicting categorical air quality indices (AQI) or discrete PM level classes rather than continuous concentration values [6, 8, 22, 28, 33, 34].

Early contributions were largely proof-of-concept demonstrations, designed to establish that outdoor image characteristics could be linked to haze and PM concentrations under controlled or semi-controlled conditions [4, 31, 34–36]. These studies often relied on small, local datasets and relatively simple empirical or physics-based models.

During the considered period, research increasingly shifted towards real-world operational scenarios, using heterogeneous image sources and explicitly addressing atmospheric variability. Public surveillance and traffic camera networks are the most used data sources [5, 7, 10, 26, 30, 32, 37–39], enabling continuous and automated monitoring in urban areas. Other works explore participatory sensing and user-contributed imagery, such as mobile-based volunteer monitoring [8, 36], crowdsourcing [31], or user-generated

online photos [34], to potentially expand geographic coverage of applications.

Some recent studies incorporated temporal prediction into image-based PM estimation, moving beyond single-episode analysis to forecasting short-term pollutant levels. These approaches typically leverage sequences of images acquired over time and employ architectures such as hybrid CNN combined with long short-term memory (LSTM) layers or attention-based mechanisms [10, 14, 30, 32, 37–39], capturing temporal dynamics of PM concentrations.

Several studies aimed to improve modelling capabilities by combining image-derived features with auxiliary environmental data. Examples included fusion with AOD retrievals from satellite or ground-based instruments [7], and integration of meteorological variables such as humidity, wind speed, and temperature [10, 19, 20, 40, 41]. This integration was generally reported to increase predictive accuracy and reduce the sensitivity of models to exogenous biasing factors.

Finally, a subset of works explicitly prioritised either deployment feasibility or interpretability. Lightweight architectures, optimised for real-time execution on mobile devices or edge-computing platforms, were investigated

in [42, 43], enabling in-field applications without complex hardware equipment. Other studies focused on explainability, linking model-learned features to physical atmospheric properties such as scattering, visibility, or perceived haze, also intending to promote user trust in automated predictions [2, 9, 18, 24, 25, 36].

3.3 Data sources and modelling strategies

Image-based approaches for PM estimation consider a range of data sources (Fig. 5) and modelling strategies (Fig. 6), reflecting a moderate heterogeneity across inputs and analysis methods.

Most studies relied on RGB imagery as the primary input, occasionally transformed into alternative colour spaces such as HSV, Lab, or YCbCr to better isolate haze-sensitive channels and minimise illumination effects [8, 20, 21, 33, 35–37, 42]. Other works derived colour-channel ratios, such as the Blue-to-Red index, which provide haze-sensitive features less affected by absolute illumination [6, 9, 11, 24, 44].

Among the different sources, fixed camera systems or webcams were the most used, appearing in 35 of the reviewed works. Representative examples include public webcam networks, traffic or surveillance cameras, and fixed research camera installations [4, 5, 7, 9, 10, 25, 26]. Operated by public agencies, research institutions, or open-access monitoring networks, these systems provide stable viewpoints for continuous observation, enabling consistent monitoring over long time periods. Complementing these,

16 studies employed mobile or crowdsourcing imagery, either exclusively or in combination with fixed cameras. These included smartphone photographs from participatory sensing campaigns and images extracted from social media or other user-contributed sources [8, 19, 31, 34, 36]. While these broaden spatial coverage by capturing diverse graphical contexts, they also introduce greater variability in acquisition conditions. A single study explored direct integration of satellite imagery-derived AOD [7]. This suggests that, to date, satellite remote sensing products still lack the spatial granularity and methodological transferability required for reliable ground-level PM concentration estimation and health-risk monitoring applications, due also to the indirect nature of their measurements [45].

Air-quality ground sensors appear in all reviewed studies and serve as the essential reference for labelling images with co-located PM_{2.5} concentrations, providing ground-truth measurements for models training, calibration and validation. Several studies extended their scope to multi-pollutant estimation, incorporating PM₁₀, NO₂, O₃, sulfur dioxide (SO₂), and carbon monoxide (CO) measurements [8, 10, 22, 27, 46]. Weather data from ground sensors, such as temperature, humidity, wind speed, and pressure, were often integrated to account for meteorological influences on PM formation and diffusion, with representative examples including both local meteorological observations and atmospheric reanalysis model outputs [3, 7, 9, 10, 19, 38].

Pre-processing included radiometric calibration [23, 25, 36], histogram equalisation [11, 35, 44, 47, 48], sky

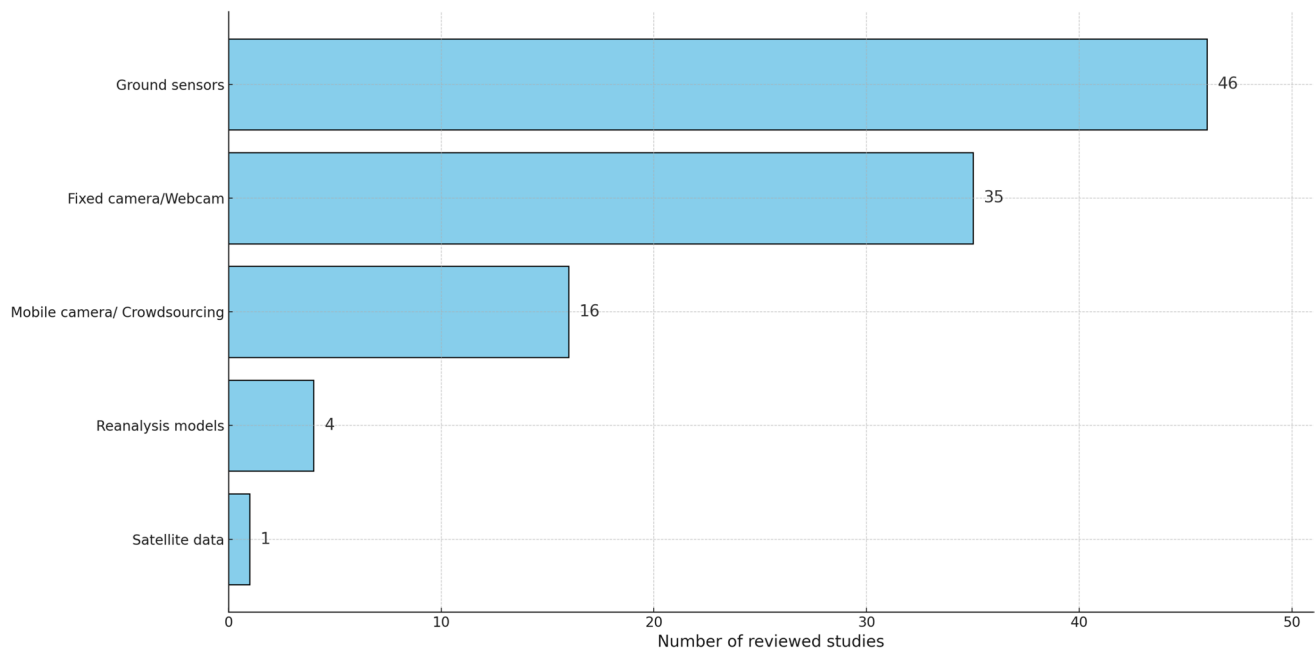


Fig. 5 Distribution of input data types across the reviewed studies

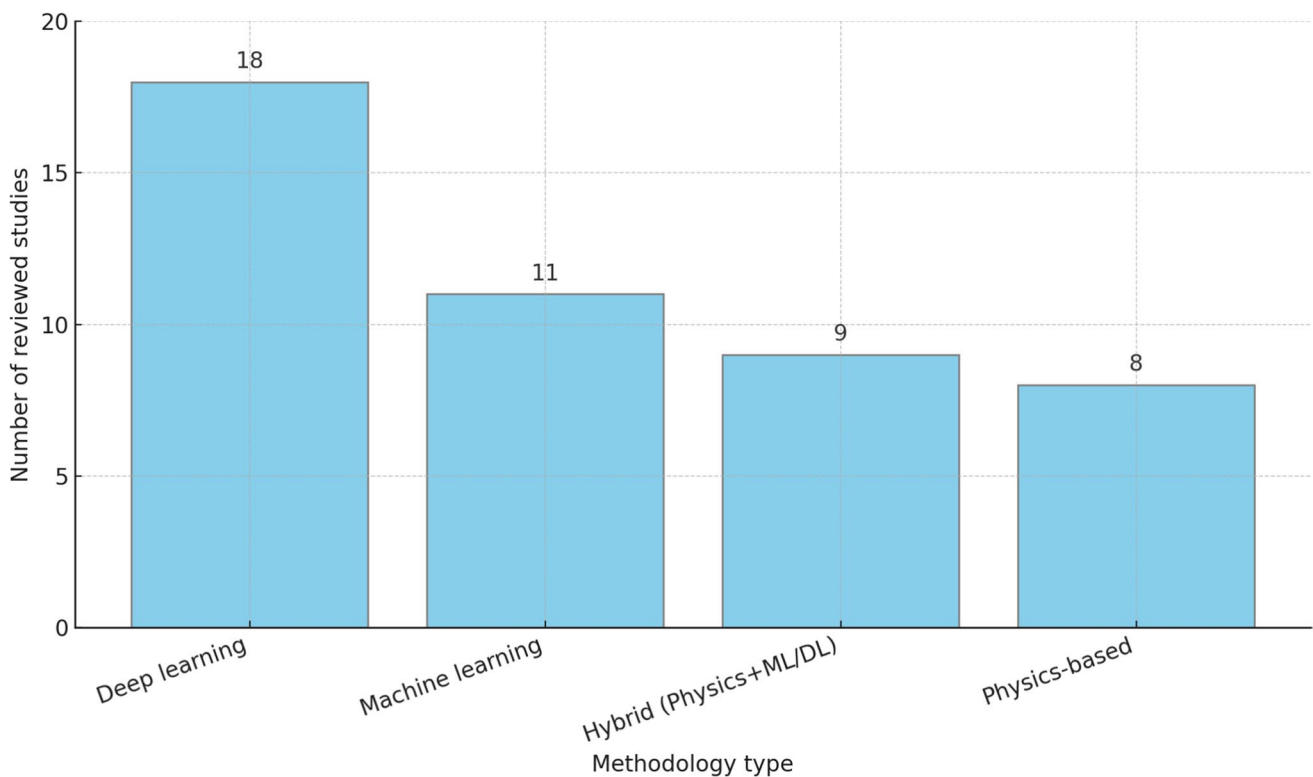


Fig. 6 Distribution of methodology types across the reviewed studies

segmentation [8], or dehazing [15, 21, 33, 35]. Geometric corrections [16] were also applied in some studies to stabilise or rectify input imagery.

The modelling strategies applied in the reviewed studies are grouped, for convenience, into four main methodology types: Physics-based, ML, DL, and Hybrid approaches combining elements of the first three. A summary of the distribution of modelling strategies is shown in Fig. 6, while a more detailed breakdown by sub-types is presented in Fig. 7. This classification is indicative and partly subjective. Since many studies tested multiple methods, we assigned each paper to the type that we believed best represented its main contribution. To reduce ambiguity, the final category was assigned by considering both the terminology used by the authors and the main model tested and discussed as the study contribution. When several models were compared, baseline or comparison models were not considered for classification; instead, the label reflected the principal approach proposed by the authors. Ambiguous cases were resolved through manual re-checking of the full paper, although a residual degree of subjectivity remains unavoidable for complex architectures combining multiple modelling components.

Physics-based models are grounded in atmospheric optics, linking PM concentrations to visibility through

extinction and attenuation formulations such as the Beer–Lambert law and Koschmieder’s equation, scattering theory, or colour ratios [2, 6, 9, 11, 23, 24, 35, 44]. The main strengths of these approaches rely on their interpretability and low computational requirements, but they are generally sensitive to lighting variability and haze in input images.

ML models rely on regression applied often to image-extracted features such as texture descriptors, including the Gray Level Co-occurrence Matrix (GLCM), the Gray Tone Difference Matrix (GTDM), and entropy [3, 47, 49], and visibility proxies such as the dark channel prior (DCP) [17]. Core algorithms include Support Vector Regression (SVR) [4, 17, 28, 31], tree-based models such as decision trees (DT), random forests (RF), and extremely randomised trees (ETR) [18, 21, 48], and artificial neural networks or multilayer perceptrons structures (ANN-MLP) [3, 42, 47, 49]. Some of the papers tested and compared a combination of ML regressors (e.g. [18, 28, 48]).

DL approaches have become dominant in recent years. CNN, including architectures such as VGG, ResNet, DenseNet, and EfficientNet, are trained end-to-end for haze-to-PM estimation [5, 8, 20, 22, 25, 29, 38, 43]. Many studies instead adopted recurrent neural networks (RNN), LSTM and gated recurrent unit (GRU) layers, to capture temporal

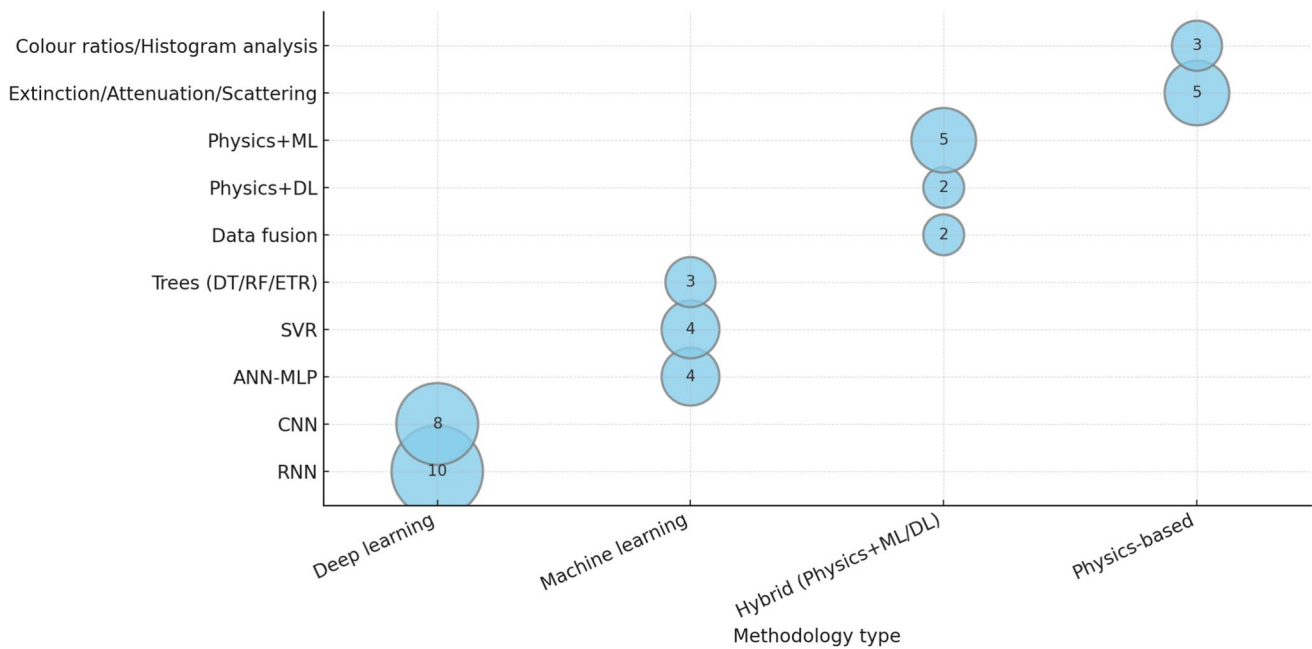


Fig. 7 Distribution of methodology sub-types across the reviewed studies, organised by the four main methodology types (Deep learning, Machine learning, Hybrid, and Physics-based models). The

marker size is proportional to the number of studies employing a specific methodology sub-types, and the overlaid value indicates the exact count

dependencies of pollution concentrations [10, 14, 19, 27, 30, 32, 37, 39, 41, 50]. Attention mechanisms and ensemble strategies were also explored [14, 25, 38]. These models achieve high predictive performance but are generally more demanding in terms of tuning and computation time. Overall, temporal architectures are mainly used to improve the continuity of image-based PM estimation, near-real-time monitoring, or short-term forecasting from sequential images. By contrast, fewer studies explicitly connect these temporal estimates with spatialisation strategies or spatial-context variables aimed at extending predictions from individual camera locations to broader monitoring fields.

Finally, hybrid approaches combine physics-based proxies with ML or DL frameworks. In Physics+ML, features such as extinction indices or transmission maps are used as inputs to regressors like SVR or tree-based models [16, 33, 36, 40, 46]. In Physics+DL, these features are embedded into CNN to guide learning with physics-informed priors [15, 34]. Data fusion strategies integrate imagery with meteorological variables, ground-based sensors, or satellite AOD [7, 26]. Some of these papers can actually fit more than one hybrid model sub-type (e.g., [7, 34, 36]), proposing multiple hybridisation strategies. The most representative sub-type based on the author's interpretation is selected for the proposed categorisation. By combining interpretability from physics with the flexibility of learning architectures,

hybrid models are proposed both to improve robustness under heterogeneous atmospheric and illumination conditions, and across geographic contexts.

3.4 Declared models performance

Performance reporting across the reviewed studies is heterogeneous, with authors adopting different evaluation metrics depending on their methodological approach, target pollutant, dataset structure, and validation design.

Among prediction error metrics, the most widely reported indicator is the root mean square error (RMSE), declared in 25 studies. Representative examples include studies based on different image sources and modelling settings [4, 5, 7, 10, 18, 20, 25]. Other error-based indicators include the mean absolute error (MAE), used in 15 studies including [8, 16, 19, 30, 38, 39], followed by the mean absolute percentage error (MAPE) [19, 30, 39, 43] and symmetric MAPE (SMAPE) [19, 30, 39], each appearing in only a few cases. Although MAPE and SMAPE would provide useful dimensionless indicators for cross-study comparison, they are reported only by a small subset of studies and mostly within DL approaches. Therefore, they were not considered in this work for comparisons due to the imbalance.

Correlation-based indicators are also widely used. The coefficient of determination (R^2) is the most frequently

reported correlation metric, computed in 23 studies, including [2, 4, 5, 9, 10, 15, 18, 25]. Other indicators such as Pearson's r [3, 19, 30, 39, 48] and rank-based measures such as Spearman's ρ or Kendall's τ coefficients [3, 21, 33, 34, 44, 48] are occasionally reported alongside error metrics to capture linear or monotonic associations.

The reported RMSE values span a broad range across studies, from below 1 to nearly $90 \mu\text{m}^3$, depending on dataset characteristics and validation experiments. However, RMSE is directly influenced by the absolute concentration range of the test dataset and is therefore difficult to compare across studies conducted in different regions, seasons, and pollution regimes. An equivalent consideration applies to R^2 . Although it is dimensionless, it depends on the variance structure of the test dataset and may favour settings with wider PM concentration ranges, while penalising models tested under narrower concentration variability. Given these limitations, R^2 is retained in this review only as a necessary descriptive indicator, because it is the most consistently reported metric across methodology types. Its use is therefore intended to support a qualitative synthesis of reported model fit, rather than a true quantitative comparison of predictive performance across studies.

In total, R^2 was reported in 23 studies, distributed across methodologies as follows: 8 DL, 6 ML, 5 hybrid (Physics+ML/DL), and 4 physics-based. The R^2 aggregated statistics, reported in Fig. 8, are therefore used only as a descriptive and indicative synthesis of reported model fit. DL models show the highest reported values, with a median of 0.94 and an interquartile range (IQR) of 0.81–0.98. ML methods report a median R^2 of 0.8 and a wider IQR of 0.71–0.90, spanning a broad overall range from 0.68 to 0.97. Hybrid approaches (Physics+ML/DL) yield consistently high performance, with a median R^2 of 0.89 and IQR 0.87–0.91. Physics-based models report slightly lower values, with a median R^2 of 0.86 and IQR 0.81–0.88.

3.5 Code and data availability

Code and data availability are moderately variable across the reviewed studies. Results are summarised in Fig. 9. Of the 46 papers, only a small subset (4) explicitly provided both datasets and implementation source code [6, 8, 11, 42]. A larger subset made only the datasets available. This includes studies (7) where datasets were publicly released via repositories such as Figshare or GitHub [4, 7, 14–16, 34,

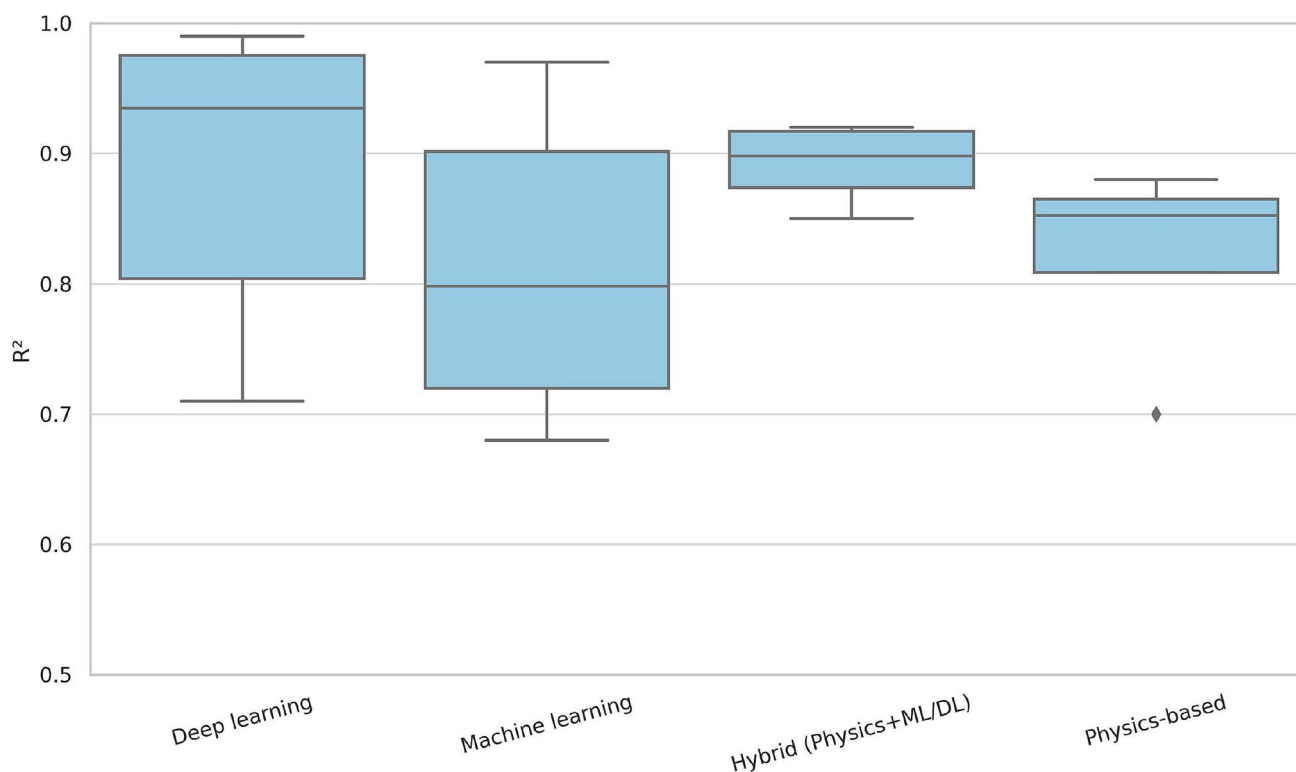
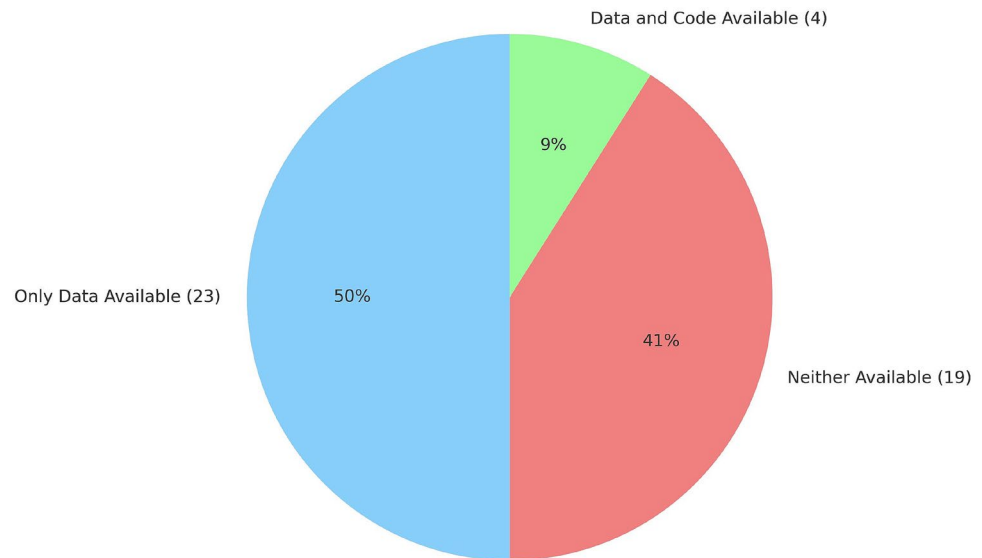


Fig. 8 Box-plots of reported R^2 values by methodology type

Fig. 9 Data and code availability in reviewed studies. The “Only Data Available” group includes studies where datasets were publicly shared by the authors through open repositories, made available upon request to the authors, or not directly shared but derived from referenced open data repositories



40], datasets available on request to the authors (8) [10, 18, 22–25, 38, 41], and datasets not directly shared but based on referenced open data repositories (8) [3, 5, 26, 28, 33, 44, 48, 50]. 19 studies did not report either data or code, often because of privacy concerns, government ownership of surveillance images, or reliance on third-party datasets that could not be redistributed. Examples of such cases are reported in studies using private surveillance, mobile, or locally collected image datasets [9, 17, 19, 29, 30, 32, 36].

4 Discussion

The 46 studies show methodological trends and trade-offs between accuracy, complexity, and operational aspects. DL methods usually achieve the best performance, with R^2 values typically above 0.85 in several studies. Their performance also suggests high stability across different test conditions and validation experiments. At the same time, DL models depend on large training datasets, which may still be lacking for many regions. Furthermore, training and deployment often rely on high-performance graphics processing units (GPUs) to handle many parameters and long training times, which may further limit real-time or low-resource applications. ML approaches reach intermediate results, with a median R^2 value around 0.80 and a broader range compared to DL. They are generally less demanding than DL in terms of both data and computational resources, and can operate with fewer labelled samples and modest hardware, which makes them suitable in contexts where large annotated datasets or high-performance computing

infrastructure are not available. Physics-based models are generally easy to interpret and light to run, since they often rely on simple analysis of image characteristics without the need for auxiliary data or complex model training. However, these models normally report lower R^2 values than DL. Hybrid solutions, combining physical models with ML or DL, represent a compromise. Reported values show that these approaches achieve consistently high R^2 close to 0.9. The integration of image-derived features, transmittance maps, or additional variables often improves both accuracy and robustness, while retaining some level of interpretability. Some of the DL studies report generally low error values also for percentage-based metrics [19, 30, 39, 43]. However, such performance metrics are reported heterogeneously across the corpus, and MAPE and SMAPE values are available only for a small subset of studies, mostly within DL approaches. This limits comparison with other modelling strategies using percentage-based error metrics, which would otherwise provide a less biased indication of performance because they are less directly affected by absolute concentration levels, like for RMSE and R^2 . Since datasets also differ substantially in PM concentration ranges, image sources, geographic settings, and validation procedures, direct quantitative comparison across methodology types remains limited.

The type of input image source is also important. Fixed webcams are the most common because they guarantee a stable view over time. Smartphone, crowd-sourced, and surveillance images enlarge the spatial coverage but bring more variability in image acquisition conditions. The inclusion of ancillary data is frequent to reduce noise due to the

atmospheric or weather conditions. Preprocessing operations such as radiometric calibration, histogram equalisation, sky region segmentation, and dehazing are often necessary and suggested to improve the quality of outputs.

Performance varies with different geographic, atmospheric, and image acquisition conditions. High performance is often reported in limited areas or under stable atmospheric conditions. Weather and visibility can degrade accuracy (e.g., cloudy, foggy, or rainy scenes) [4]. Sensor and measurement constraints (e.g., spatial granularity or calibration) also introduce variability, suggesting benefits from precise space–time co-location of image with ground PM measurements [30]. Finally, algorithmic design choices (e.g., architectures or activation functions) can affect error levels [14, 20].

From a geospatial perspective, most reviewed studies treat image-based PM estimates as local or point-based observations associated with a camera, smartphone, or nearby monitoring station, while only a limited number explicitly address their spatialisation into continuous PM fields. The most direct strategies are based on data fusion with spatially continuous predictors, such as satellite AOD, meteorological, and geographical variables [7, 31]. However, explicit modelling of spatial autocorrelation in residuals is not reported in the reviewed studies. This indicates that research has focused more on temporal continuity for near-real-time estimation and forecasting from fixed camera streams, which is a necessary step toward operational PM_{2.5} exposure assessment, than on integrating image-based estimates into spatially explicit monitoring frameworks. As these methods mature, a promising research direction will be to combine temporally continuous camera-based estimates with denser and better distributed sensing networks, in order to increase spatial coverage and support the reconstruction of PM fields over extended areas. This perspective is consistent with recent guidance identifying low-cost sensor networks as a way to fill gaps in conventional monitoring and improve spatial coverage [51].

By commenting on the distribution of papers by year, study region, and methodology, some temporal trends can be observed. Chinese studies span the entire timeframe, from early physics-based methods (2014–2015) to recent DL and hybrid approaches, often supported by large multi-city datasets. Non-Chinese studies appear mainly from 2020 onwards. Methodologically, the overall progression moves from simpler physical models in the early years to machine learning, hybrid solutions, and finally deep learning approaches in the most recent period. The predominance

of studies based in China can be partly attributed to the country's rapid economic growth and the consequent deterioration of air quality in recent decades, which has triggered intensive national research efforts on air pollution. This trend aligns with other bibliometric analyses on air quality and PM research, which identify China as the leading contributor to the global literature since the mid-2010s [52]. This geographic bias may also have methodological implications. Models developed in East Asian urban settings may not transfer directly to regions with different aerosol composition, seasonality, visibility, urban morphology, and image-acquisition conditions. This is expected to reasonably affect both physics-based calibration and DL generalisation. A few studies have begun to address this issue through cross-regional or multi-country designs, such as China and the USA [4], Thailand and India [19], and India and China [21]. Broader geographic expansion will therefore require benchmark datasets covering different climatic and aerosol regimes, together with explicit cross-regional validation and transfer-learning protocols.

Concerning the availability of code and data resources, around half of the papers provide data either publicly, on request, or clearly point to open data sources. When available, datasets varied in scope and accessibility. Some were small, locally collected image sets tied to specific monitoring stations (e.g., [22, 24]), while others offered multi-regional or multi-season image collections designed to support generalisation (e.g., [5, 15, 16]). When the reasons for limited data sharing are discussed, they are primarily related to ownership or privacy constraints affecting surveillance and traffic-camera imagery [10, 37]. Beyond data sharing itself, reproducibility is also constrained by the limited availability of public synchronised image–PM datasets and by spatial mismatches between camera viewpoints and monitoring stations [16].

Studies releasing code relied on common Python deep learning frameworks (TensorFlow, PyTorch, Keras) but also MATLAB scripts (e.g., [6, 11]), which, while functional, reduced adaptability for integration into modern workflows. Despite more than half of the reviewed studies providing access to datasets in some form, the availability of implementation code remains very limited. Explicit reasons for not releasing code are rarely provided in the reviewed studies, suggesting that code sharing is often left to individual author choices. This imbalance constrains full reproducibility and benchmarks definition, suggesting that broader adoption of data and code sharing practices could further strengthen the field.

A final relevant point to consider is the comparison between image-based PM estimates and authoritative ground-based monitoring. When compared with authoritative ground-based measurements, typically provided by monitoring stations based on gravimetric or high-precision optical methods [53], most image-based estimation methods remain below the accuracy thresholds indicated for regulatory use. According to established technical references [54], which follow the U.S. Environmental Protection Agency guidance for PM_{2.5} monitoring, two methods or sensors can be considered comparable when their agreement reaches $R^2 \geq 0.90$, while values between 0.80 and 0.90 indicate moderate concordance. More recent guidelines for low-cost PM_{2.5} sensors recommend a multi-metric evaluation framework including bias, error, precision, and site-specific collocation performance [51, 55]. In this review, we rely mainly on R^2 values for comparability assessment, since other indicators were not consistently available across the analysed papers. The $R^2 \geq 0.90$ threshold is therefore interpreted as a conservative historical benchmark for sensor comparability, rather than as a standalone operative regulatory standard.

Among the studies that reported R^2 , only a few DL approaches reach or exceed the $R^2 \geq 0.90$ threshold, and usually under controlled conditions or limited datasets. The majority of ML, hybrid, and physics-based models show lower R^2 values, especially when tested in heterogeneous environments or across diverse image sources. This indicates a gap between image-based estimation and regulatory-grade monitoring. Further progress requires larger and more diverse multi-site datasets and methodological improvements to increase robustness and comparability across atmospheric conditions and geographic contexts.

5 Conclusions

This review highlights the growing potential of image-based approaches for estimating PM concentrations, while noting that several challenges remain before regulatory applicability can be achieved. DL techniques generally report the strongest predictive performance, whereas ML and physics-based models provide lower computational complexity and greater interpretability. Hybrid methods that combine physical insights with ML or DL represent a promising compromise between accuracy, robustness, and operational feasibility. These findings should be interpreted cautiously because the available comparisons are affected by heterogeneous datasets, metrics, and validation procedures.

The reviewed studies also show that data diversity, stable image acquisition, and ancillary meteorological information remain important factors for improving model robustness.

From a geographic perspective, the corpus of reviewed studies is unbalanced. Most studies are based in East Asia, with a large share conducted in China, while other regions are sparsely represented. Moreover, the number of relevant papers is limited (46 in total for 2014–2024) due to the tight inclusion criteria adopted (focus on ambient PM estimation from standard RGB images with explicit methodological testing). The yearly distribution is also uneven, with a significant increase in relevant publications in recent years compared to the early period of the review time window.

Despite recent advances, most image-based methods remain below the accuracy typically expected in regulatory frameworks, especially when tested across diverse regions, atmospheric conditions, and uncontrolled image acquisition settings. Reproducibility remains limited because implementation code is rarely shared, even though more than half of the reviewed studies provide access to datasets in some form.

Future research should focus on the development of large multi-site datasets that cover multiple seasonal and atmospheric conditions, advances in hybrid and physics-informed model architectures, and open sharing of data and source code. Lightweight models suitable for edge devices and real-time applications are also a relevant aspect to investigate. In addition, future studies should consistently report concentration-independent error metrics, such as normalised mean absolute error and mean bias error, to enhance comparability between models as well as with ground-based sensor measurements. With these developments, image-based approaches could become a useful complement to traditional monitoring networks and provide stronger support for air quality risk assessment.

Appendix A

Appendix A reports a summary table (Table 1) with the main features of the reviewed studies. These include bibliographic reference, year, country of application, input data used, methodology type and subtype (with a short description of methodological procedures), availability of data and code (with links to Web sources) and performance metrics. For the latter, only RMSE, MAE, MAPE, SMAPE, R^2 , Pearson's r , Spearman's ρ or Kendall's τ metrics are included when available; other metrics are not reported.

Table 1 Summary table of the main features of the studies considered for the review, sorted by publication year

(Reference) Title (year)	Country	Data input	Methodology	Data and code availability	Validation/test performance
[35] Real-Time Air Quality Estimation Based on Color Image Processing (2014)	China (Beijing)	ground sensors; fixed camera/webcam; mobile camera/crowdsourcing	Physics-based—(Extinction/Attenuation/Scattering): Applies depth-related contrast loss and colour cast relationships to infer visibility and PM	not reported	others
[36] PM2.5 Monitoring Using Images From Smartphones in Participatory Sensing (2015)	China (Beijing)	ground sensors; mobile camera/crowdsourcing	Hybrid (Physics + ML/DL)—(Physics + ML): Participatory smartphone sensing with device/scene indicators (dark-channel, visibility characteristics) and uses ML regressors for PM estimation	not reported	MAE = 14.54 $\mu\text{g}/\text{m}^3$; $R^2 = 0.8982$
[34] Using User Generated Online Photos to Estimate and Monitor Air Pollution in Major Cities (2015)	China (Beijing)	ground sensors; mobile camera/crowdsourcing	Hybrid (Physics + ML/DL)—(Physics + DL): Processes crowdsourced and synthetic photos with haze/contrast indicators using a DL classifier	data available (link (https://www.cs.rochester.edu/u/yl/haze-level.html), accessed on Oct. 8, 2025)	Spearman's $\rho = 0.89$ + others
[4] Particle Pollution Estimation Based on Image Analysis (2016)	China (Beijing, Shanghai) + USA (Phoenix)	ground sensors; fixed camera/webcam	Machine learning—(SVR): Trains city-specific SVR models on fixed-camera datasets with solar/time/geolocation covariates	data available (link (https://figshare.com/articles/figure/Particle_pollution_estimation_based_on_image_analysis/1603556), accessed on Oct. 8, 2025)	RMSE = 40.43 $\mu\text{g}/\text{m}^3$; $R^2 = 0.68$
[31] Inferring Fine-Grained PM2.5 with Bayesian-Based Kernel Method for Crowdsourcing System (2017)	China	ground sensors; mobile camera/crowdsourcing	Machine learning—(SVR): Uses > 4 k crowdsourced smartphone photos with metadata (GPS/lens/orientation) to train an SVR, together with ground sensors measurements to estimate PM levels	not reported	MAE = 28 $\mu\text{g}/\text{m}^3$; $R^2 = 0.97$
[29] End-to-End Learning for Image-Based Air Quality Level Estimation (2018)	China (Beijing)	ground sensors; fixed camera/webcam	Deep learning—(CNN): End-to-end single-image CNN mapping visual appearance directly to air-quality level/PM proxy without explicit intermediate variables	not reported	MAE = 0.309 $\mu\text{g}/\text{m}^3$
[50] Learn to Predict PM2.5 Concentration with Image Contrast-Sensitive Features (2018)	China	ground sensors; mobile camera/crowdsourcing	Deep learning—(RNN): RNN trained on contrast-sensitive image indicators; benchmarks against regression baselines	data not shared but available from open repositories	RMSE = 35.85 $\mu\text{g}/\text{m}^3$; $R^2 = 0.8052$
[26] Towards Improved Air Quality Monitoring Using Publicly Available Sky Images (2018)	Greece (Thessaloniki)	ground sensors; fixed camera/webcam; mobile camera/crowdsourcing; reanalysis models	Hybrid (Physics + ML/DL)—(Data fusion): Detects and localizes sky regions, extracts optical sky characteristics, and fuses them with meteorology/reanalysis to estimate air quality	data not shared but available from open repositories	others
[11] Highly Efficient Picture-Based Prediction of PM2.5 Concentration (2018)	China	ground sensors; mobile camera/crowdsourcing	Physics-based—(Colour ratios/Histogram analysis): Uses RGB/HSV ratios and histogram/contrast metrics summarised with logistic or parametric mapping to estimate PM levels	data and code available (link (https://kegu.netlify.app), accessed on Oct. 8, 2025)	RMSE = 51.88 $\mu\text{g}/\text{m}^3$ + others

Table 1 (continued)

(Reference) Title (year)	Country	Data input	Methodology	Data and code availability	Validation/test performance
[44] Effective and Efficient Photo-Based PM _{2.5} Concentration Estimation (2019)	China (Beijing)	ground sensors; fixed camera/webcam	Physics-based—(Extinction/Attenuation/Scattering): Uses a compact set of saturation, gradient, and contrast metrics linked to extinction to regress PM levels	data not shared but available from open repositories	RMSE = 57.69 $\mu\text{g}/\text{m}^3$; Spearman's $\rho = 0.7832$; Kendall's $\tau = 0.5802$
[46] Estimation of Multiple Atmospheric Pollutants through Image Analysis (2019)	China (Beijing, Shanghai)	ground sensors; fixed camera/webcam	Hybrid (Physics+ML/DL)—(Physics+ML): Extracts extinction/contrast from fixed cameras and combines with tree/boosting or SVR regressors; addresses multiple pollutants	not reported	RMSE = 8.68 $\mu\text{g}/\text{m}^3$; $R^2 = 0.917$
[15] Discern Depth Under Foul Weather: Estimate PM _{2.5} for Depth Inference (2019)	China	ground sensors; mobile camera/crowdsourcing	Hybrid (Physics+ML/DL)—(Physics+DL): Uses physics-inspired depth indicators together in a CNN; physics constrains features while DL estimates PM _{2.5} concentrations	data available (link (https://www.moji.com), accessed on Oct. 8, 2025)	$R^2 = 0.92$
[43] A Deep Learning and Image-Based Model for Air Quality Estimation (2020)	China	ground sensors; fixed camera/webcam	Deep learning—(CNN): Single-image convolutional regression for PM/AQI	not reported	RMSE = 0.5854 $\mu\text{g}/\text{m}^3$; MAE = 0.2919 $\mu\text{g}/\text{m}^3$; MAPE = 18.6633% + others
[2] Estimating PM _{2.5} From Photographs (2020)	USA (Chicago)	ground sensors; fixed camera/webcam	Physics-based—(Extinction/Attenuation/Scattering): Applies Beer–Lambert extinction model and visibility–PM relationships on webcam imagery using depth-dependent contrast decay	not reported	$R^2 = 0.88$
[47] Prediction of PM _{2.5} Concentration Based on Weighted Bagging and Image Contrast-Sensitive Features (2020)	China	ground sensors; fixed camera/webcam	Machine learning—(ANN-MLP): Contrast-sensitive image statistics feed an ensemble of MLP learners combined via weighted bagging for PM estimation	not reported	RMSE = 33.812 $\mu\text{g}/\text{m}^3$; MAE = 34.515 $\mu\text{g}/\text{m}^3$ + others
[9] Quantifying Source Apportionment for Ambient Haze: An Image Haze Extraction Approach with Air Quality Monitoring Data (2020)	Taiwan	ground sensors; fixed camera/webcam	Physics-based—(Extinction/Attenuation/Scattering): Derives haze metrics and analyses relationships with ground sensors PM/O ₃ /NO ₂ and humidity measurements	not reported	$R^2 = 0.8452$
[30] ResNet-LSTM for Real-Time PM _{2.5} and PM ₁₀ Estimation Using Sequential Smartphone Images (2020)	China (Hong Kong)	ground sensors; fixed camera/webcam; mobile camera/crowdsourcing	Deep learning—(RNN): ResNet encodes sequential smartphone frames; LSTM aggregates over time with meteorological covariates to estimate PM _{2.5} /PM ₁₀	not reported	MAE = 5.31 $\mu\text{g}/\text{m}^3$; SMAPE = 0.247%; Pearson's $r = 0.79$
[3] Application of Photo Texture Analysis and Weather Data in Assessment of Air Quality in Terms of Airborne PM ₁₀ and PM _{2.5} Particulate Matter (2021)	Poland	ground sensors; fixed camera/webcam; reanalysis models	Machine learning—(ANN-MLP): Uses texture descriptors (e.g., GLCM and entropy metrics) with meteorology in an MLP	data not shared but available from open repositories	Pearson's $r = 0.89$; Spearman's $\rho = 0.87$
[16] HVAQ: A High-Resolution Vision-Based Air Quality Dataset (2021)	China (Hangzhou)	ground sensors; fixed camera/webcam; mobile camera/crowdsourcing	Hybrid (Physics+ML/DL)—(Physics+ML): Provides a benchmark dataset and uses gradient-boosting baselines on synchronised images and sensor measurements for PM estimation and performance assessment	data available (link (https://github.com/implicitDeclaration/HVAQ-dataset), accessed on Oct. 8, 2025)	MAE = 1.45 $\mu\text{g}/\text{m}^3$

Table 1 (continued)

(Reference) Title (year)	Country	Data input	Methodology	Data and code availability	Validation/test performance
[42] Performance Evaluation of Photographic Measurement in the Machine-Learning Prediction of Ground PM _{2.5} Concentration (2021)	China (Beijing)	ground sensors; fixed camera/webcam; reanalysis models	Machine learning—(ANN-MLP): Uses sky/ground photos together with ERA5 variables (boundary-layer height, wind, precipitation, temperature) in an MLP model for PM _{2.5}	data and code available (link (https://github.com/Limin-Feng1993/Machine_Learning_of_PM2.5_by_Photographic_Measurement), accessed on Oct. 8, 2025)	RMSE = 29 $\mu\text{g}/\text{m}^3$; $R^2 = 0.7$
[41] PM _{2.5} Monitoring: Use Information Abundance Measurement and Wide and Deep Learning (2021)	China	ground sensors; fixed camera/webcam	Deep learning—(RNN): Combines an information-abundance score with shallow predictors and a recurrent sequential branch for PM regression	data available on request	others
[49] An Improved Picture-Based Prediction Method of PM _{2.5} Concentration (Beijing) (2022)	China (Beijing)	ground sensors; fixed camera/webcam	Machine learning—(ANN-MLP): Uses colour ratios, contrast, and texture statistics from images as inputs to an MLP trained with co-located PM _{2.5} ground sensors measurements	not reported	RMSE = 48.53 $\mu\text{g}/\text{m}^3$ + others
[48] A Photo-Based Quality Assessment Model for the Estimation of PM _{2.5} Concentrations (2022)	China	ground sensors; fixed camera/webcam	Machine learning—(Trees (DT/RF/ETR)): Decision-tree ensemble using visibility/contrast and image-quality indicators	data not shared but available from open repositories	RMSE = 51.60 $\mu\text{g}/\text{m}^3$; Pearson's $r = 0.8082$; Spearman's $\rho = 0.8177$; Kendall's $\tau = 0.6115$
[14] Densely Connected Convolutional Networks with Attention Long Short-Term Memory for Estimating PM _{2.5} Values From Images (2022)	China (Beijing, Shanghai)	ground sensors; fixed camera/webcam	Deep learning—(RNN): DenseNet image encoder with attention LSTM aggregates sequences to estimate hourly PM from webcam datasets	data available (link (https://figshare.com/articles/gure/Particle_pollution_estimation_based_on_image_analysis/1603556), accessed on Oct. 8, 2025)	RMSE = 14.07 $\mu\text{g}/\text{m}^3$; $R^2 = 0.71$
[7] Estimating High-Resolution PM _{2.5} Concentrations by Fusing Satellite AOD and Smartphone Photographs Using a Convolutional Neural Network and Ensemble Learning (2022)	China (Beijing)	ground sensors; fixed camera/webcam; mobile camera/crowdsourcing; reanalysis models; satellite data	Hybrid (Physics + ML/DL)—(Data fusion): Combines haze/visibility indicators and CNN descriptors with satellite AOD and ERA5; fuses via fuzzy neural network and gradient-boosting to regress PM _{2.5}	data available (link (https://figshare.com/articles/gure/Particle_pollution_estimation_based_on_image_analysis/1603556), accessed on Oct. 8, 2025)	RMSE = 40.78 $\mu\text{g}/\text{m}^3$; $R^2 = 0.85$
[23] Estimation of Aerosol Extinction Coefficient Using Camera Images and Application in Mass Extinction Efficiency Retrieval (2022)	South Korea (Daejeon, Busan)	ground sensors; fixed camera/webcam	Physics-based—(Extinction/Attenuation/Scattering): Derives aerosol extinction coefficients from image contrast/radiative transfer characteristics and maps extinction to PM _{2.5} via mass-extinction efficiency	data available on request	$R^2 = 0.70$

Table 1 (continued)

(Reference) Title (year)	Country	Data input	Methodology	Data and code availability	Validation/test performance
[17] Estimation of PM _{2.5} Concentration Based on Support Vector Regression with Improved Dark Channel Prior and High Frequency Information in Images (2022)	Taiwan	ground sensors; fixed camera/webcam	Machine learning—(SVR): Enhances dark-channel prior and high-frequency details to capture haze; regression is used to estimate PM concentrations	not reported	RMSE = 7.769 $\mu\text{g}/\text{m}^3$; R ² = 0.816
[32] Image-Based Air Quality Forecasting through Multi-Level Attention (2022)	China (Shanghai)	ground sensors; fixed camera/webcam	Deep learning—(RNN): Attention-augmented LSTM on webcam sequences for multi-hour PM forecasting	not reported	RMSE = 23.57 $\mu\text{g}/\text{m}^3$; MAE = 15.84 $\mu\text{g}/\text{m}^3$
[40] Measuring PM _{2.5} Concentrations From a Single Smartphone Photograph (2022)	China (Beijing, Shanghai)	ground sensors; mobile camera/crowdsourcing	Hybrid (Physics + ML/DL)—(Physics + ML): Extracts haze/sky indicators from smartphone photos and blends them with supervised predictors to estimate PM _{2.5} from ground sensors paired images	data available (link (https://figshare.com/articles/publication/Particle_pollution_estimation_based_on_image_analysis/1603556), accessed on Oct. 8, 2025)	RMSE = 10.003 $\mu\text{g}/\text{m}^3$; R ² = 0.874
[28] Multi-Kernel Support Vector Machine for Particulate Matter Estimation (2022)	Not specified country	ground sensors; fixed camera/webcam	Machine learning—(SVR): Uses multi-kernel fusion across heterogeneous visual descriptors from outdoor images to estimate PM levels/classes	data not shared but available from open repositories	others
[33] Multi-Modal Image Feature Fusion-Based PM _{2.5} Concentration Estimation (2022)	China	ground sensors; mobile camera/crowdsourcing	Hybrid (Physics + ML/DL)—(Physics + ML): Fuses multiple visual feature families (e.g., colour ratios, dark-channel priors, CNN descriptors) in a regression framework	data not shared but available from open repositories	Spearman's ρ = 0.8211 + others
[5] Real-Time Image-Based Air Quality Estimation by Deep Learning Neural Networks (2022)	Taiwan (Kaohsiung City)	ground sensors; fixed camera/webcam	Deep learning—(CNN): Single-image CNN trained with ground sensors-labelled webcam frames for real-time PM estimation	data not shared but available from open repositories	RMSE = 5.4 $\mu\text{g}/\text{m}^3$; R ² = 0.8
[24] Relating Atmospheric Aerosol Amounts to Blue-to-Red Ratio and Grayscale Contrast Fluctuations Using Digitalization of Routine Webcam Photographs Taken in the Urban Environment of Vienna (2022)	Austria (Vienna)	ground sensors; fixed camera/webcam	Physics-based—(Colour ratios/Histogram analysis): Relates blue-to-red ratios and grayscale contrast fluctuations from webcam imagery to aerosol loading/PM via correlation-driven models	data available on request	R ² = 0.86
[6] Detection of Air Pollution in Urban Areas Using Monitoring Images (2023)	China (Beijing, Hangzhou, Kunming, Lasa, Shijiazhuang, Tianjin, Wuhan)	ground sensors; fixed camera/webcam	Physics-based—(Colour ratios/Histogram analysis): Computes correlations between color ratios/contrast indicators and PM; includes an AQI classifier from image features	data and code available (link (https://pan.baidu.com/s/1wLu1MxXor4FTFWBV99hfZw?pwd=DBCp#list/path=%2F), password = DBCP, accessed on Oct. 8, 2025)	others

Table 1 (continued)

(Reference) Title (year)	Country	Data input	Methodology	Data and code availability	Validation/test performance
[39] Personalized Ambient Pollution Estimation Based on Stationary Camera Taken Images Under Cross-Camera Information Sharing in Smart City (2023)	China (Hong Kong)	ground sensors; fixed camera/webcam	Deep learning—(RNN): Sequence model with cross-camera knowledge sharing to customise per-camera PM estimators from stationary webcams	not reported	MAE = 5.311 $\mu\text{g}/\text{m}^3$; SMAPE = 0.264%; Pearson's $r = 0.792$
[38] Real-Time Image-Based Air Quality Forecasts Using a 3D-CNN Approach with an Attention Mechanism (2023)	China (Shanghai)	ground sensors; fixed camera/webcam	Deep learning—(CNN): 3D-CNN over image stacks with attention to peculiar frames/regions for short-term PM forecasting	data available on request	MAE = 17.13 $\mu\text{g}/\text{m}^3$ + others
[18] Unmasking Air Quality: A Novel Image-Based Approach to Align Public Perception with Pollution Levels (2023)	Taiwan	ground sensors; fixed camera/webcam	Machine learning—(Trees (DT/RF/ETR)): Gradient-boosting/trees on visibility/appearance features to infer hourly PM; calibrated on two co-located ground sensors	data available on request	RMSE = 8.31 $\mu\text{g}/\text{m}^3$; $R^2 = 0.78$
[19] A Mobile Image-Driven PM2.5 Estimation Framework Using Deep Learning Techniques (2024)	Thailand + India	ground sensors; mobile camera/crowdsourcing	Deep learning—(RNN): Extracts per-frame CNN features from smartphone sequences and integrates with Bi-LSTM for temporal modelling; incorporates meteorological variables	not reported	MAE = 4.06 $\mu\text{g}/\text{m}^3$; SMAPE = 8.3%; Pearson's $r = 0.94$
[20] An Improved Activation Function in Convolution Neural Network to Estimate the Hazardous Air Pollutant Based on Images (2024)	India	ground sensors; mobile camera/crowdsourcing	Deep learning—(CNN): Convolutional backbone with a custom activation to stabilise haze-like regression; incorporates meteorological variables	not reported	RMSE = 28.15 $\mu\text{g}/\text{m}^3$; MAE = 18.94 $\mu\text{g}/\text{m}^3$; $R^2 = 0.93$
[25] Applying Traffic Camera and Deep Learning-Based Image Analysis to Predict PM2.5 Concentrations (2024)	Australia (Melbourne)	ground sensors; fixed camera/webcam	Deep learning—(CNN): Learns spatial patterns and visibility characteristics from traffic-camera imagery; augments with meteorological covariates to estimate PM2.5 concentrations	data available on request	RMSE = 0.76 $\mu\text{g}/\text{m}^3$; $R^2 = 0.98$
[21] Can an Image Tell the Tale: Looking Beyond the Haze to Determine PM2.5 Concentration (2024)	India + China	ground sensors; fixed camera/webcam	Machine learning—(Trees (DT/RF/ETR)): Aggregates region-of-interest indicators (e.g., colour contrast) in an ERT for PM regression	not reported	RMSE = 12.56 $\mu\text{g}/\text{m}^3$; $R^2 = 0.93$; Kendall's $\tau = 0.90$
[27] High-Precision Microscale Particulate Matter Prediction in Diverse Environments Using a Long Short-Term Memory Neural Network and Street View Imagery (2024)	Germany + Czech Republic	ground sensors; fixed camera/webcam	Deep learning—(RNN): Uses street-view image features with LSTM-based temporal modelling for microscale PM prediction co-located with dense ground sensor measurements	not reported	RMSE = 1.39 $\mu\text{g}/\text{m}^3$; MAE = 0.38 $\mu\text{g}/\text{m}^3$; $R^2 = 0.99$
[22] Real-Time Air Quality Index Detection through Regression-Based Convolutional Neural Network Model on Captured Images (2024)	India (Bangalore, Delhi, Tamil Nadu)	ground sensors; fixed camera/webcam	Deep learning—(CNN): CNN regression using captured RGB and additional pollutant covariates to predict AQI/PM indices	data available on request	MAE = 0.1386 $\mu\text{g}/\text{m}^3$ + others
[37] Surveillance-Image-Based Outdoor Air Quality Monitoring (2024)	China (Shanghai)	ground sensors; fixed camera/webcam	Deep learning—(RNN): Extracts CNN features from surveillance camera frames and models temporal PM dynamics with an RNN trained with ground sensors measurements	not reported	RMSE = 5.11 $\mu\text{g}/\text{m}^3$; $R^2 = 0.94$

Table 1 (continued)

(Reference) Title (year)	Country	Data input	Methodology	Data and code availability	Validation/test performance
[10] Time-Series Forecasting of PM2.5 and PM10 Concentrations Based on the Integration of Surveillance Images (2024)	China (Shanghai)+Taiwan (Kaohsiung)	ground sensors; fixed camera/webcam	Deep learning—(RNN): Trained on surveillance imagery with ERA5/pollutant inputs and transfers an RNN model across cities (pretrain-fine-tune) for PM forecasting	data available on request	RMSE = 1.81 $\mu\text{g}/\text{m}^3$; $R^2 = 0.9739$
[8] Uncovering Local Aggregated Air Quality Index with Smartphone Captured Images Leveraging Efficient Deep Convolutional Neural Network (2024)	Bangladesh (Dhaka)	ground sensors; mobile camera/crowdsourcing	Deep learning—(CNN): CNN applied to smartphone sky images to infer AQI/PM classes	data and code available (data: link (https://zenodo.org/records/7878371); code: link (https://github.com/lepotatoguy/aqi), accessed on Oct. 8, 2025)	RMSE = 42.22 $\mu\text{g}/\text{m}^3$; MAE = 29.68 $\mu\text{g}/\text{m}^3$ + others

Validation/Test performance values refer to validation, test-set, or cross-validation results. Reported metrics include the most recurrent indicators such as RMSE, R^2 , MAE, MAPE/SMAPe, Pearson's r , Spearman's ρ , and Kendall's τ when available. 'others' indicates less recurrent metrics reported only in individual studies

Funding Open access funding provided by Politecnico di Milano within the CRUI-CARE Agreement. International Society for Photogrammetry and Remote Sensing (ISPRS), Scientific Initiatives 2025, Daniele Oxoli, Ministero dell'Università e della Ricerca, 202258ACSL—PE10, Francesco Pirotti.

Data availability The data supporting the findings of this review are derived entirely from published literature cited in the manuscript.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Hoffmann, B., et al. (2021). WHO air quality guidelines 2021—aiming for healthier air for all: A joint statement by medical, public health, scientific societies and patient representative organisations. *International Journal of Public Health*, 66, 1604465. <https://doi.org/10.3389/ijph.2021.1604465>
- Pudasaini, B., et al. (2020). Estimating PM2.5 from photographs. *Atmospheric Environment: X*, 5, 100063. <https://doi.org/10.1016/j.aeoa.2020.100063>
- Chuchro, M., Sarlej, W., Grzegorzczak, M., & Nurzyńska, K. (2021). Application of photo texture analysis and weather data in assessment of air quality in terms of airborne PM10 and PM2.5 particulate matter. *Sensors*, 21(16), 5483. <https://doi.org/10.3390/s21165483>
- Liu, C., Tsow, F., Zou, Y., & Tao, N. (2016). Particle pollution estimation based on image analysis. *PLoS ONE*, 11(2), e0145955. <https://doi.org/10.1371/journal.pone.0145955>
- Kow, P.-Y., Hsia, I.-W., Chang, L.-C., & Chang, F.-J. (2022). Real-time image-based air quality estimation by deep learning neural networks. *Journal of Environmental Management*, 307, 114560. <https://doi.org/10.1016/j.jenvman.2022.114560>
- Chu, Y., Chen, F., Fu, H., & Yu, H. (2023). Detection of air pollution in urban areas using monitoring images. *Atmosphere*, 14(5), 772. <https://doi.org/10.3390/atmos14050772>
- Wang, F., Yao, S., Luo, H., & Huang, B. (2022). Estimating high-resolution PM2.5 concentrations by fusing satellite AOD and smartphone photographs using a convolutional neural network and ensemble learning. *Remote Sensing*, 14(6), 1515. <https://doi.org/10.3390/rs14061515>
- Mondal, J. J., et al. (2024). Uncovering local aggregated air quality index with smartphone captured images leveraging efficient deep convolutional neural network. *Scientific Reports*, 14(1), 1627. <https://doi.org/10.1038/s41598-023-51015-1>
- Lin, F.-C., Lu, S.-H., Liang, C.-J., Chen, Y.-H., & Liang, J.-J. (2020). Quantifying source apportionment for ambient haze: An

- image haze extraction approach with air quality monitoring data. *Environmental Research*, 184, 109216. <https://doi.org/10.1016/j.envres.2020.109216>
10. Wu, Y., Wang, X., Wang, M., Liu, X., & Zhu, S. (2024). Time-series forecasting of PM_{2.5} and PM₁₀ concentrations based on the integration of surveillance images. *Sensors*, 25(1), 95. <https://doi.org/10.3390/s25010095>
 11. Gu, K., Qiao, J., & Li, X. (2018). Highly efficient picture-based prediction of PM_{2.5} concentration. *IEEE Transactions on Industrial Electronics*, 66(4), 3176–3184. <https://doi.org/10.1109/TIE.2018.2840515>
 12. Elbestar, M., Aly, S. G., & Ghannam, R. (2024). Advances in air quality monitoring: A comprehensive review of algorithms for imaging and sensing technologies. *Advanced Sensor Research*, 3(11), 2300207.
 13. Marzagui, K., Ait Abdelouahad, A., Dahmouni, A., & El Mali, A. D. (2024). A review on image quality assessment for air quality index estimation, In *2024 11th international conference on wireless networks and mobile communications (WINCOM)*, IEEE, pp. 1–6.
 14. Zhang, B., Geng, Z., Zhang, H., & Pan, J. (2022). Densely connected convolutional networks with attention long short-term memory for estimating PM_{2.5} values from images. *Journal of Cleaner Production*, 333, 130101. <https://doi.org/10.1016/j.jclepro.2021.130101>
 15. Li, K., et al. (2019). Discern depth under foul weather: Estimate PM_{2.5} for depth inference. *IEEE Transactions on Industrial Informatics*, 16(6), 3918–3927. <https://doi.org/10.1109/TII.2019.2943631>
 16. Chen, Z., Zhang, T., Chen, Z., Xiang, Y., Xuan, Q., & Dick, R. P. (2021). HVAQ: A high-resolution vision-based air quality dataset. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–10. <https://doi.org/10.1109/TIM.2021.3104415>
 17. Hsieh, C.-H., Chen, K.-Y., Jiang, M.-Y., Liaw, J.-J., & Shin, J. (2022). Estimation of PM_{2.5} concentration based on support vector regression with improved dark channel prior and high frequency information in images. *IEEE Access*, 10, 48486–48498. <https://doi.org/10.1109/ACCESS.2022.3172468>
 18. 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. *Environment International*, 181, 108289. <https://doi.org/10.1016/j.envint.2023.108289>
 19. Kamble, A., Aramkul, S., & Champrasert, P. (2024). A mobile image-driven PM_{2.5} estimation framework using deep learning techniques. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3521966>
 20. Bhimavarapu, U. (2024). An improved activation function in convolution neural network to estimate the hazardous air pollutant based on images. *Wireless Personal Communications*, 135(4), 2401–2420. <https://doi.org/10.1007/s11277-024-11174-4>
 21. Chakraborty, S., Palit, S., & Bhandari, H. (2024). Can an image tell the tale: Looking beyond the haze to determine PM_{2.5} concentration, In *2024 international joint conference on neural networks (IJCNN)*, IEEE, pp. 1–8. <https://doi.org/10.1109/IJCNN60899.2024.10651248>
 22. 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. *Environmental Quality Management*, 34(1), e22276. <https://doi.org/10.1002/tqem.22276>
 23. Shin, J., Kim, D., & Noh, Y. (2022). Estimation of aerosol extinction coefficient using camera images and application in mass extinction efficiency retrieval. *Remote Sensing*, 14(5), 1224. <https://doi.org/10.3390/rs14051224>
 24. Frisch-Niggemeyer, A., Weihs, P., Revesz, M., Schreier, S. F., & Richter, A. (2022). Relating atmospheric aerosol amounts to blue to red ratio and grayscale contrast fluctuations using digitalization of routine webcam photographs taken in the urban environment of Vienna. *Atmospheric Environment*, 290, 119345. <https://doi.org/10.1016/j.atmosenv.2022.119345>
 25. Liu, Y., et al. (2024). Applying traffic camera and deep learning-based image analysis to predict PM_{2.5} concentrations. *Science of the Total Environment*, 912, 169233. <https://doi.org/10.1016/j.scitotenv.2023.169233>
 26. Spyromitros-Xioufis, E., et al. (2018). Towards improved air quality monitoring using publicly available sky images. *Multimedia tools and applications for environmental & biodiversity informatics* (pp. 67–92). Springer. https://doi.org/10.1007/978-3-319-76445-0_5
 27. Liu, X., et al. (2024). High-precision microscale particulate matter prediction in diverse environments using a long short-term memory neural network and street view imagery. *Environmental Science & Technology*, 58(8), 3869–3882. <https://doi.org/10.1021/acs.est.3c06511>
 28. Gaur, D., Mehrotra, D., & Singh, K. (2022). Multi kernel support vector machine for particulate matter estimation. *Asian Journal of Water, Environment and Pollution*, 19(5), 89–95. <https://doi.org/10.3233/AJW220075>
 29. Zhang, C., Yan, J., Li, C., Wu, H., & Bie, R. (2018). End-to-end learning for image-based air quality level estimation. *Machine Vision and Applications*, 29(4), 601–615. <https://doi.org/10.1007/s00138-018-0919-x>
 30. Song, S., Lam, J. C., Han, Y., & Li, V. O. (2020). ResNet-LSTM for real-time PM_{2.5} and PM₁₀ estimation using sequential smartphone images. *IEEE Access*, 8, 220069–220082. <https://doi.org/10.1109/ACCESS.2020.3042278>
 31. Li, S., Xi, T., Tian, Y., & Wang, W. (2017). Inferring fine-grained PM_{2.5} with bayesian based kernel method for crowdsourcing system, In *GLOBECOM 2017 - 2017 IEEE global communications conference*, pp. 1–6. <https://doi.org/10.1109/GLOCOM.2017.8254421>
 32. Zhang T., & Dick, R. P. (2022). Image-based air quality forecasting through multi-level attention, In *2022 IEEE international conference on image processing (ICIP)*, pp. 686–690. <https://doi.org/10.1109/ICIP46576.2022.9897241>
 33. Wang, G., Shi, Q., Wang, H., Sun, K., Lu, Y., & Di, K. (2022). Multi-modal image feature fusion-based PM_{2.5} concentration estimation. *Atmospheric Pollution Research*, 13(3), 101345. <https://doi.org/10.1016/j.apr.2022.101345>
 34. Li, Y., Huang, J. & Luo, J. (2015). Using user generated online photos to estimate and monitor air pollution in major cities, In *Proceedings of the 7th international conference on internet multimedia computing and service*, pp. 1–5. <https://doi.org/10.1145/2808492.2808564>
 35. Wang, H., Yuan, X., Wang, X., Zhang, Y., & Dai, Q. (2014). Real-time air quality estimation based on color image processing, In *2014 IEEE visual communications and image processing conference*, IEEE, pp. 326–329. <https://doi.org/10.1109/VCIP.2014.7051572>
 36. Liu, X., Song, Z., Ngai, E., Ma, J., & Wang, W. (2015). PM_{2.5} monitoring using images from smartphones in participatory sensing, In *2015 IEEE conference on computer communications workshops (INFOCOM WKSHPs)*, IEEE, pp. 630–635. <https://doi.org/10.1109/INFCOMW.2015.7179456>
 37. Wang, X., Wang, M., Liu, X., Mao, Y., Chen, Y., & Dai, S. (2024). Surveillance-image-based outdoor air quality monitoring. *Environmental Science and Ecotechnology*, 18, 100319. <https://doi.org/10.1016/j.ese.2023.100319>
 38. Elbaz, K., Shaban, W. M., Zhou, A., & Shen, S.-L. (2023). Real time image-based air quality forecasts using a 3D-CNN approach with an attention mechanism. *Chemosphere*, 333, 138867. <https://doi.org/10.1016/j.chemosphere.2023.138867>

39. Song, S., Li, V. O. K., Lam, J. C. K., & Wang, Y. (2023). Personalized ambient pollution estimation based on stationary-camera-taken images under cross-camera information sharing in smart city. *IEEE Internet of Things Journal*, 10(17), 15420–15430. <https://doi.org/10.1109/JIOT.2023.3263949>
40. Yao, S., Wang, F., & Huang, B. (2022). Measuring PM2.5 concentrations from a single smartphone photograph. *Remote Sensing*, 14(11), 2572. <https://doi.org/10.3390/rs14112572>
41. Gu, K., Liu, H., Xia, Z., Qiao, J., Lin, W., & Thalmann, D. (2021). PM2.5 monitoring: Use information abundance measurement and wide and deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, 32(10), 4278–4290. <https://doi.org/10.1109/TNNLS.2021.3105394>
42. Feng, L., Yang, T., & Wang, Z. (2021). Performance evaluation of photographic measurement in the machine-learning prediction of ground PM2.5 concentration. *Atmospheric Environment*, 262, 118623. <https://doi.org/10.1016/j.atmosenv.2021.118623>
43. Zhang, Q., Fu, F., & Tian, R. (2020). A deep learning and image-based model for air quality estimation. *Science of the Total Environment*, 724, 138178. <https://doi.org/10.1016/j.scitotenv.2020.138178>
44. Yue, G., Gu, K., & Qiao, J. (2019). Effective and efficient photo-based PM2.5 concentration estimation. *IEEE Transactions on Instrumentation and Measurement*, 68(10), 3962–3971. <https://doi.org/10.1109/TIM.2018.2886091>
45. Fowlie, M., Rubin, E., & Walker, R. (2019). Bringing satellite-based air quality estimates down to Earth, In *AEA papers and proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, pp. 283–288.
46. Zhang, T., & Dick, R. P. (2019). “Estimation of multiple atmospheric pollutants through image analysis, In *2019 IEEE international conference on image processing (ICIP)*, IEEE, pp. 2060–2064. <https://doi.org/10.1109/ICIP.2019.8803130>.
47. Qiao, J., He, Z., & Du, S. (2020). Prediction of PM2.5 concentration based on weighted bagging and image contrast-sensitive features. *Stochastic Environmental Research and Risk Assessment*, 34(3), 561–573. <https://doi.org/10.1007/s00477-020-01787-z>
48. Sun, K., Tang, L., Huang, S., & Qian, J. (2022). A photo-based quality assessment model for the estimation of PM2.5 concentrations. *IET Image Processing*, 16(4), 1008–1016. <https://doi.org/10.1049/ipr2.12201>
49. Chen, Q., Chen, W., & Pan, G. (2022). An improved picture-based prediction method of PM2.5 concentration. *IET Image Processing*, 16(11), 2827–2833. <https://doi.org/10.1049/ipr2.12204>
50. He, Z., Ye, X., Gu, K., & Qiao, J. (2018). Learn to predict PM2.5 concentration with image contrast-sensitive features, In *2018 37th chinese control conference (CCC)*, pp. 4102–4106. <https://doi.org/10.23919/ChiCC.2018.8482979>.
51. Malings, C., Amegah, K., Basart, S., Diez, S., Rosales, C. M. F., & Zimmerman, N. (2024). *Integrating low-cost sensor systems and networks to enhance air quality applications* (p. 293). World Meteorological Organization.
52. Han, M., Yang, F., & Sun, H. (2021). A bibliometric and visualized analysis of research progress and frontiers on health effects caused by PM2.5. *Environmental Science and Pollution Research*, 28(24), 30595–30612.
53. Mahajan, S., & Helbing, D. (2025). Dynamic calibration of low-cost PM2.5 sensors using trust-based consensus mechanisms. *npj Climate and Atmospheric Science*, 8(1), 257. <https://doi.org/10.1038/s41612-025-01145-2>
54. Watson, J. G., Chow, J. C., Moosmüller, H., Green, M., & Frank, N. (1998). Guidance for using continuous monitors in PM2.5 monitoring networks, Nevada Univ. System, Desert Research Inst., Reno, NV (United States); <https://www3.epa.gov/ttnamti1/files/ambient/pm25/r-98-012.pdf>.
55. Williams, R., Kaufman, A., Hanley, T., Rice, J. & Garvey, S. (2014). Evaluation of field-deployed low cost PM sensors, EPA/600/R-14/464 (NTIS PB 2015–102104).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.