Geospatial Identification Of Air Pollution Risk Zones In Urban India Using Hybrid Machine Learning Approaches

[1*]Rakesh Kumar Jha, [2]Dr. Brajesh Kumar

[1*]Research Scholar School of Computer Science and Engineering, Sandip University, Sijoul, Madhubani, Bihar
Email ID- jha.rkjha23@gmail.com
[2]Associate Professor School of Computer Science and Engineering, Sandip University, Sijoul, Madhubani, Bihar
Email ID- brajeshkumar@sandipuniversity.edu.in

Abstract
The spatial heterogeneity of urban air pollution in India is high because of the high urbanization rate, traffic, industrialization, construction dust, biomass burning, and mixed land-use. The traditional monitoring networks and city-wide averages are not enough to detect the exposure to pollution at the neighbourhood level and the risk areas. New possibilities to incorporate pollutant concentration, meteorological variables, mobile monitoring, and hybrid machine learning methods provide geospatial technologies with new opportunities to consider the vulnerability factors, population exposure, and land-use indicators into high-resolution risk maps. The review is a synthesis of existing evidence on GIS-based mapping, spatial interpolation, hot spot analysis, land-use regression, satellite-derived indicators, and machine learning models to assess air pollution in urban settings. Special focus is put on hybrid modelling methods that integrate ground monitoring, satellite observations, spatial predictors, and deep learning or ensemble algorithms to enhance prediction and risk classification. The review highlights that risk-zone identification must go beyond prediction of the pollutants and include exposure load, socioeconomic vulnerability, health sensitivity, proximity of the sources, seasonal persistence, and model uncertainty. An integrated risk-zone mapping conceptual framework is suggested to support Indian city risk-zone mapping. These strategies have the potential to enhance focused pollution management, optimization of monitoring networks, community-health warnings, and urban environmental planning.

Keywords: Geospatial analysis; air pollution risk zones; urban India; hybrid machine learning; $PM_{2.5}$; exposure mapping.
How to cite this article: Jha RK, Kumar B. Geospatial Identification Of Air Pollution Risk Zones In Urban India Using Hybrid Machine Learning Approaches. Int J Drug Deliv Technol. 2026;16(53s): 568-579. DOI: 10.25258/ijddt.16.53s.106
Source of support: Nil.
Conflict of interest: None.

1. Introduction

Urban air pollution is among the gravest environmental and public-health issues in India, where urbanization, motorization, industrialization, construction expansion, biomass burning, road dust, and waste burning have increased the levels of pollutants in most cities. The spatial complexity of Indian urban air pollution is due to the fact that the sources of emissions are numerous and interdependent: traffic corridors, industrial clusters, commercial centres, construction areas, dense residential areas, and peri-urban expansion belts are all sources of emissions. Thus, the traditional air-quality averages and fixed monitoring stations at the city level are commonly inadequate to measure the pollution variation at the neighbourhood level and to identify the population in high-risk situations.

The trend of air-quality modelling has shifted to data-driven and hybrid modelling systems, compared to the traditional deterministic and statistical methods. The traditional models can be used to understand the processes of the atmosphere, yet they might not be able to capture the complex nonlinear interactions between meteorology, land use, traffic, pollutant transport, seasonal variation, and human activity. Karroum et al. (2020) emphasized the role of air-quality modelling in comprehending the behaviour of pollutants and in environmental management. More recent research indicates that machine learning can be used to predict air-quality better, since it can handle large, nonhomogeneous datasets and identify nonlinear relationships that are hard to model with traditional statistical techniques (Méndez et al., 2023).

Artificial intelligence and machine learning are becoming more applicable in India to air-pollution management. Rautela and Goyal (2024) pointed out that AI-based systems have the potential to enhance monitoring, forecasting, decision support, and pollution-control planning under Indian conditions. But the difficulty does not lie in only predicting the concentration of pollutants but also to indicate risk areas of air pollution. A risk zone is not merely an area with high $PM_{2.5}$, PM_{10}, NO_2, or AQI values. It is a region where population exposure is coupled with high pollution, social vulnerability, sensitive land uses, and frequent exceedance events, and low capacity to mitigate exposure.

This issue revolves around the geospatial aspect. The pollution rates are drastically different in cities due to uneven distribution of roads, industries, vegetation, built-up density, settlements, schools, hospitals, and meteorological conditions. The local exposure disparities can be concealed by a citywide air-quality measure. As an illustration, a roadside neighbourhood, an industrial-neighbourhood, or a school close to a high-traffic corridor might be at significantly higher risk than the average in the city. Pollutant data can be related to

*Author for Corresponden ;jha.rkjha23@gmail.com the location-specific exposure and vulnerability indicators through geospatial tools including GIS, spatial interpolation, hotspot analysis, land-use regression, and remote sensing.

The hybrid machine learning methods are particularly applicable to urban air-pollution risk mapping since no data source can be used to capture all the urban pollution processes. Ground monitoring offers precise measurements of the pollutants but typically has a small area of coverage. Satellite measurements have a wider coverage but need to be calibrated and can be influenced by retrieval uncertainty. Dispersion and accumulation are explained by meteorological data, whereas the proximity and exposure of sources and land-use, road-network, industrial, and population data are identified. These datasets can be combined into hybrid ML models to produce high-resolution pollution surfaces, predict pollution events, identify hotspots, and aid in risk-zone classification.

Recent surveys indicate that machine learning and deep learning are becoming popular in spatiotemporal air-quality prediction. According to Agbehadji and Obagbuwa (2024), sophisticated ML and DL models have the capability to combine time series, space, sensor data, and environmental predictors to forecast air-quality. Likewise, Méndez et al. (2023) demonstrated that forecasting models are relevant in predicting high-pollution events and aiding timely responses. In the case of urban India, these models can only be useful when they go beyond the accuracy of predictions and help in identifying risks in a practical way. The purpose of this review is to synthesize geospatial and hybrid machine learning methods to detect air pollution risk areas in urban India. It analyzes ground surveillance, satellite remote surveillance, weather parameters, land-use parameters, traffic parameters, population exposure, and vulnerability parameters. It also discusses spatial interpolation, hotspots analysis, land-use regression, ensemble learning, deep learning, and hybrid modelling. The review is valuable in that it moves the emphasis of the prediction of pollutants to the combined, exposure-sensitive, and vulnerability-conscious risk-zone mapping of Indian urban settings.

2. Urban Air Pollution in India: Sources, Patterns, and Risk Dimensions

2.1 Major Sources and Pollutants in Urban India

In India, urban air pollution is a combination of multiple interacting sources, such as vehicular emissions, industrial activities, construction dust, road dust, biomass burning, domestic fuel use, open waste burning, and regional transport of pollutants. The major pollutants include $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and O_3. These pollutants are spatially varying, based on the density of traffic, the location of industries, the land use, the intensity of constructions, and the meteorological conditions. Recent research on the Indian cities indicates that the patterns of air-quality vary across regions and cannot be assessed using generalization at the city level but need to be done spatially (Dar and Shaik, 2025). Table 1 shows the key air pollutants in the major urban areas, their sources, spatial applicability, health effects, and contribution to the risk-zone mapping.

Table 1. Major urban air pollutants, sources, spatial relevance, and health implications in Indian cities

Pollutant	Major Sources	Spatial Relevance	Health Implications	Role in Risk Mapping
$PM_{2.5}$	Traffic, industry, biomass burning	Traffic corridors, industrial zones, dense areas	Respiratory and cardiovascular disease	Core exposure-risk pollutant
PM_{10}	Road dust, construction, industry	Roads and construction zones	Respiratory irritation	Dust-risk indicator
NO_2	Vehicles, combustion	Roads and intersections	Asthma, airway inflammation	Traffic-pollution indicator
SO_2	Industry, coal combustion	Industrial and power-plant zones	Respiratory problems	Industrial-risk indicator
CO	Vehicles, combustion, generators	Congested roads and enclosed corridors	Reduced oxygen delivery	Combustion exposure indicator
O_3	Photochemical reactions	Regional and seasonal variation	Lung irritation	Multi-pollutant risk indicator

2.2 Spatial and Seasonal Pollution Patterns

Indian cities have a very uneven air pollution. The high concentrations are usually located along the traffic routes, industrial belts, construction sites, business districts, landfills, and high population residential settlements. Seasonal fluctuation is significant as well. The winter months have been found to have a greater level of particulate pollution due to low wind speed, shallow boundary-layer height, and pollutant concentration, particularly in the cities of north India. The New Delhi PM 2.5 prediction work hybrids indicate the significance of seasonal and temporal variation modelling in understanding the behaviour of urban pollution (Masood et al., 2023).

2.3 Exposure and Vulnerability Dimensions

The risk of pollution is based not just on concentration but also on the exposure and the vulnerability of the population. Outdoor workers, street vendors, children, elderly people, traffic police, and informal-settlement residents, as well as people with respiratory or cardiovascular diseases, are more vulnerable to polluted air. Risk-zone mapping should thus consider schools, hospitals, roadside settlements and communities near industries as sensitive receptors. Predictive environmental health-risk mapping can be used to bridge the gap between pollutant concentrations and vulnerable groups and the priorities of the population-health (Rajesh et al., 2025).

2.4 From Pollution Hotspots to Risk Zones

A pollution hot spot is a place where the concentration of pollutants is great and a risk zone is a combination of the intensity of pollution and exposure, vulnerability and frequent exceedance. The significance of this difference is that a highly-polluted site does not necessarily pose the highest-public-health risk when there is low population exposure. Conversely, a more densely populated, yet moderately polluted roadside or informal settlement can be a more risky area. Geospatial PM 0 hotspots modelling shows that it is possible to use pollution mapping to inform targeted interventions (Yasin et al., 2025).

3. Geospatial Data Sources and Risk-Zone Indicators

3.1 Ground Monitoring, Mobile Monitoring, and Sensor Data

Ground-based monitoring data are essential for air pollution risk-zone mapping because they provide direct measurements of pollutants such as $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and O_3. In India, data on monitoring are usually available in fixed monitoring stations maintained by pollution-control agencies. Nevertheless, fixed stations can be rather sparse, and can fail to resolve street-level or neighbourhood-level variation. Spatial coverage can be enhanced through mobile monitoring and low-cost sensor networks, which can record pollution levels all over traffic corridors, residential areas, industrial zones, and other micro-environment. Mobile monitoring can be especially helpful in detecting fine-scale spatial variation of urban pollutant exposure (Xu et al., 2022).

3.2 Satellite Remote Sensing and Earth Observation Data

Satellite remote sensing has extensive spatial coverage and can be applied in areas where ground monitoring stations are sparse. Satellite-based indicators in common use are Aerosol Optical Depth, Sentinel-5P pollutants, MODIS products, TROPOMI observations, land surface temperature, NDVI, built-up indices, and night-time lights. These data sets are used to estimate the distribution of pollutants, identify patterns of emissions, and aid spatial gap-filling. Google Earth Engine has gained significance to process extensive satellite data and produce air-quality indicators in urban and regional regions as well (Halder et al., 2023).

3.3 Meteorological, Land-Use, and Urban-Form Indicators

The significance of meteorological variables is that they affect the dispersion, accumulation, and transformation of pollutants. The movement and persistence of pollutants in urban areas are influenced by temperature, wind speed, wind direction, rainfall, relative humidity, and boundary-layer height. The land-use and urban-form variables that include road density, traffic intensity, industrial land use, built-up area, vegetation cover, elevation, and landfill proximity are used to explain spatial differences in pollutant levels. These are often employed as predictors in machine learning models to estimate PM 2.5 and map air quality (Thapa et al., 2025).

3.4 Population, Exposure, and Vulnerability Indicators

The concentration of pollution is not a sufficient measure of the risk zones. The indicators of population exposure and vulnerability are needed to transform pollution maps into the public-health risk maps. Significant indicators are population density, children, elderly, schools, hospitals, informal settlements, roadside communities, outdoor workers, and individuals with existing respiratory or cardiovascular conditions. These indicators can be used to determine areas where exposure to pollution can result in increased health effects. The exposure to traffic is particularly significant in urban settings since the populations at risk tend to live, work, or travel close to large roads (Dheekwal et al., 2025). Table 2 shows the key data types and indicators of geospatial air pollution risk-zone mapping.

Table 2. Data categories and indicators used for geospatial air pollution risk-zone mapping

Data Category	Indicator Examples	Application	Strengths	Limitations
Ground monitoring	$PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, O_3	Calibration and validation	Direct pollutant measurement	Sparse spatial coverage
Mobile monitoring	Street-level pollutant data	Fine-scale exposure mapping	Captures local variability	Requires repeated surveys
Satellite data	AOD, NO_2, LST, NDVI, night-time lights	Spatial prediction and gap filling	Wide spatial coverage	Retrieval uncertainty
Meteorology	Wind, rainfall, humidity, temperature	Dispersion and seasonal modelling	Explains temporal variation	Resolution mismatch
Land use	Roads, industries, built-up areas, vegetation	Source-proximity mapping	Useful for intra-urban analysis	Requires updated spatial layers

Population and vulnerability	Density, schools, hospitals, informal settlements	Exposure-risk assessment	Links pollution with health risk	Data may be incomplete

4. Geospatial Methods for Air Pollution Risk-Zone Identification

4.1 GIS-Based Mapping and Spatial Overlay

Mapping based on GIS is among the most common methods of determining spatial variation of air pollution in urban areas. It enables the data on concentration of pollutants to be integrated with land use, traffic networks, industrial locations, population density, schools, hospitals, and vulnerable settlements. They are common GIS operations that involve thematic mapping, buffer analysis, proximity analysis, weighted overlay, and spatial layering. Such techniques are practical since they transform pollution information into risk maps that are easy to interpret and visualize. In the case of urban India, overlay using GIS is particularly applicable since the sources of pollution and the vulnerable populations are frequently in close proximity. Recent surveys on spatial air-quality modelling indicate that GIS-based approaches are still significant in converting complex air-quality information into decision-support tools (Badawy, 2025).

4.2 Spatial Interpolation and Geostatistical Methods

The estimation of pollutant concentrations in those regions where the monitoring data are not directly available is done by spatial interpolation. The most commonly used interpolation techniques are Inverse Distance Weighting, ordinary kriging, universal kriging, co-kriging and spline interpolation. The IDW estimates values using the assumption that the observations that are closer are more similar than those that are far apart. Kriging techniques employ spatial autocorrelation to make statistically informed predictions. Co-kriging may incorporate some related variables like meteorology, land use or elevation. The techniques can be applied to construct continuous pollutant surfaces based on point monitoring data. Their performance is however largely dependent on the density and spatial distribution of monitoring-stations. Interpolation might not be sufficient to resolve the variability in pollution at the neighbourhood scale in Indian cities with sparse monitoring networks (Karroum et al., 2020).

4.3 Hotspot Analysis and Spatial Autocorrelation

Hotspot analysis identifies locations where high pollutant values are spatially clustered. Methods such as Getis–Ord Gi*, Local Moran's I, and cluster-outlier analysis are useful for distinguishing statistically significant pollution hotspots from random high values. Getis–Ord Gi* identifies clusters of high or low values, while Local Moran's I helps detect local clusters and spatial outliers. These tools are important for risk-zone mapping because they show where pollution is not only high but also spatially persistent or clustered. In geospatial PM_{10} mapping, hotspot analysis can guide targeted interventions by identifying priority zones for pollution control and monitoring expansion (Yasin et al., 2025).

4.4 Land-Use Regression and Exposure Mapping

Land-use regression is used to relate the pollutant concentrations to the spatial predictors, including the density of roads, traffic intensity, the proximity of industries, built-up area, vegetation cover, elevation, population density, and the distance between the sources of emissions. It is applicable in intra-urban exposure assessment since it describes the effects of local environmental and urban-form variables on the levels of pollutants. Exposure mapping builds on this method by superimposing the surfaces of pollutants with population and vulnerability data. This can be used to determine the areas where high pollution coincides with high population density, schools, hospitals, or vulnerable communities. The spatial and temporal resolution of the input data is significant since risk-zone classification is not reliable with too coarse or inconsistent datasets (Zhalehdoost & Taleai, 2025). Table 3 provides a comparative evaluation of key geospatial techniques to map air pollution risks.

Table 3. Comparative assessment of geospatial methods for air pollution risk mapping

Method	Main Application	Strengths	Limitations	Suitability for Urban India
GIS overlay	Combining pollution, exposure, and vulnerability layers	Simple, visual, policy-friendly	Sensitive to layer quality and weighting	High
Buffer analysis	Assessing risk near roads, industries, landfills, and construction zones	Useful for source-proximity analysis	Does not capture pollutant dispersion fully	High
IDW	Estimating pollutant values between monitoring stations	Easy to apply and interpret	Weak when monitoring stations are sparse	Moderate
Kriging	Spatial prediction using autocorrelation	Statistically robust and spatially informed	Requires adequate monitoring density	Moderate to high

Co-kriging	Interpolation using pollutant and auxiliary variables	Can improve prediction with supporting data	Requires reliable correlated variables	Moderate
Getis–Ord Gi*	Identifying statistically significant hotspots	Detects high-value clusters	Does not explain causes of hotspots	High
Local Moran's I	Detecting clusters and spatial outliers	Useful for local spatial patterns	Sensitive to spatial scale and neighbourhood definition	High
Land-use regression	Predicting pollution using spatial predictors	Strong for intra-urban exposure mapping	Requires detailed and updated predictor data	High
Exposure overlay	Linking pollution with population and vulnerability	Converts concentration maps into risk maps	Depends on quality of demographic data	High

5. Machine Learning and Hybrid Modelling Approaches

5.1 Conventional Machine Learning Models

Traditional machine learning models are popular in air-quality prediction since they are able to address nonlinear interactions among pollutants, meteorology, traffic, land use, and urban features. Popular models are: Random Forest, Support Vector Machine, Decision Tree, Gradient Boosting, XGBoost, Artificial Neural Networks, and statistical learning models. These models are useful for predicting AQI, $PM_{2.5}$, PM_{10}, and other pollutants in urban areas. In Indian cities, optimized machine learning models have shown potential for improving AQI prediction and supporting air-quality management (Natarajan et al., 2024).

5.2 Deep Learning Models for Air-Quality Prediction

The use of deep learning models in air-quality prediction is on the rise due to their ability to handle large and complex data. CNN, LSTM, GRU, CNN-LSTM, ConvLSTM, and encoder-decoder networks are models that are helpful in capturing spatial patterns and temporal trends. Time-series forecasting can be done with LSTM and GRU models, and CNN-based models can be used to extract spatial features of gridded or image-like data. Hybrid CNN-LSTM models are particularly applicable to fine-scale urban pollution estimation due to their ability to combine spatial feature extraction with a time prediction potential (Zhang et al., 2022).

5.3 Hybrid Machine Learning Approaches

Hybrid machine learning methods are modelling methods or datasets that are used together to enhance prediction accuracy and spatial representation. These methods can combine satellite data with ground surveillance, GIS layers with machine learning algorithms, or deep learning with optimization algorithms. They can be satellite-ground fusion models, ensemble learning, decomposition-based models, optimization-based ML, GIS-ML integration, and spatial-autocorrelation-based deep learning. Hybrid models are particularly practical in the Indian context since the number of monitoring stations is usually restricted, and the patterns of pollution are extremely dynamic. The hybrid extreme learning machine with snake optimization has been used on PM2.5 prediction in New Delhi, demonstrating the usefulness of optimization-based modelling in air-quality prediction in urban areas (Masood et al., 2023).

5.4 Model Inputs, Outputs, and Performance Metrics

Air pollution mapping machine learning models need a wide range of input variables. These usually consist of pollutant levels, Aerosol Optical Depth, meteorological parameters, road density, traffic density, industrial land use, built-up area, vegetation cover, elevation, population density, and time-related indicators. The products can be pollutant concentration maps, AQI predictions, hotspots maps, exposure surfaces, and risk-zone categories. R^2, RMSE, MAE, MAPE, accuracy, precision, recall, F1-score, and cross-validation are typically used to measure model performance. Spatial validation is also critical in the context of geospatial risk-zone mapping since a model can be statistically effective and yet fail to capture local trends of pollution. Surveys of air-quality forecasting indicate that the choice of model and its evaluation heavily relies on the type of data, the scale of prediction, the type of pollutant, and the purpose of its use (Méndez et al., 2023). Table 4 shows the key machine learning and hybrid modelling methods applied to air pollution mapping.

Table 4. Machine learning and hybrid modelling approaches for air pollution mapping

Model Category	Examples	Common Inputs	Typical Outputs	Strengths	Limitations
Conventional ML	RF, SVM, XGBoost, ANN	Pollutants, meteorology, land use	Concentration prediction, AQI forecast	Handles nonlinear relationships	May ignore spatial dependency
Deep learning	CNN, LSTM, GRU, ConvLSTM	Time-series and spatial data	Forecasting and spatial pattern detection	Captures complex temporal patterns	Requires large datasets

Ensemble models	RF + GBM, stacking, boosting	Multi-source predictors	Improved prediction surfaces	Higher prediction stability	Lower interpretability
Satellite–ground fusion	AOD + monitoring data	Satellite and station data	Gridded $PM_{2.5}$ estimates	Improves spatial coverage	Needs strong calibration
Optimization-based ML	Optimized ANN, ELM-SO	Tuned model parameters	Enhanced prediction accuracy	Improves model tuning	Risk of overfitting
Hybrid deep learning	CNN-LSTM, decomposition-LSTM	Spatial and temporal inputs	Fine-scale pollution forecasts	Captures spatiotemporal patterns	Computationally demanding

5.5 Strengths and Limitations of Hybrid ML

Hybrid machine learning methods offer multiple benefits to air pollution risk-zone mapping. They enhance the accuracy of prediction, combine various sources of data, model nonlinear relationships and facilitate spatiotemporal modelling. They can also be used to combine concentration of pollutants with meteorology, satellite data, land use and exposure indicators. Nevertheless, hybrid models also have drawbacks. They can be overfit, have poor explainability, be computationally expensive, have poor cross-city transferability, and can be uncertain due to incomplete or inconsistent input data. Interpretability is particularly significant in the case of risk-zone mapping since policymakers require to know not only the high-risk areas but also the reasons why a particular area is considered a high-risk area. Spatiotemporal air-quality prediction systematic reviews highlight the importance of improved validation, explainability, and trustworthy combination of machine learning and deep learning approaches (Agbehadji and Obagbuwa, 2024). The overall hybrid machine learning process of air pollution prediction and risk-zone mapping is shown in Figure 1.

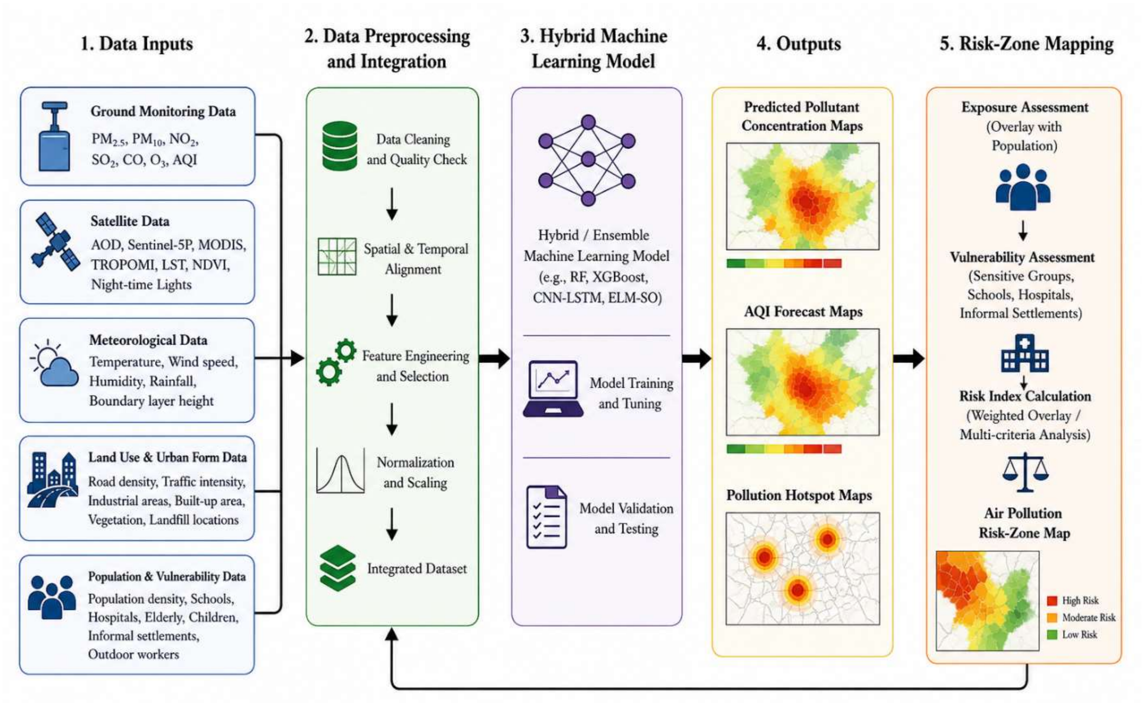

Figure 1. Hybrid machine learning workflow for air pollution prediction and risk-zone mapping

6. Evidence from Indian and Global Literature

6.1 Indian Evidence on Air-Quality Prediction and Modelling

Machine learning and deep learning are becoming more popular in Indian studies to predict AQI and PM 2.5. These studies reveal that air pollution differs among cities because of geographic, meteorological, traffic, land use, and seasonal variations. The spatiotemporal clustering of the Indian cities proves that the behaviour of air-quality in urban areas varies regionally and needs to be modelled in place-specific ways (Dar & Shaik, 2025).

AQI prediction in major Indian cities has also been done using optimized ML models. The models can be applied in forecasting, early warning, and pollution-management planning since they are able to model nonlinear relationships among pollutants and environmental variables (Natarajan et al., 2024).

New Delhi has received major attention because of its severe $PM_{2.5}$ pollution. Hybrid models such as extreme learning machine with snake optimization show the potential of optimization-based approaches for improving $PM_{2.5}$ prediction in highly polluted urban settings (Masood et al., 2023).

6.2 International Evidence on Geospatial and Hybrid ML Models

Indian cities can get a good methodological advice on international studies. Comparative evidence in China indicates that both ML and DL models can forecast air pollution, with model performance varying based on the type of pollutant, quality of data, and the scale of prediction (Ayus et al., 2023).

In mobile monitoring studies, it has been demonstrated that the variation in pollutants at the street level can be better captured than just using fixed stations. This is significant in determining small-scale urban exposure and risk zones at the neighbourhood level (Xu et al., 2022).

Spatial autocorrelation can be used to enhance prediction using hybrid deep-learning models that take into account the effect of nearby locations. Likewise, CNN-LSTM models can be used to make fine-grained predictions of air-pollution in cities by integrating both spatial and temporal learning (Zhao et al., 2023).

6.3 Critical Synthesis of Existing Studies

Current literature reveals that there is a good advancement in the prediction of air quality, although most of the studies are limited to the concentration or AQI forecasting of the pollutant, as opposed to the complete risk-zone classification. According to spatial modelling reviews, GIS, remote sensing, and ML are becoming more and more utilized, yet exposure and vulnerability indicators remain poorly interconnected (Badawy, 2025).

The overreliance on statistical accuracy measures (R 2, RMSE, MAE and MAPE) is another weakness. These measures are handy, yet do not entirely show whether a model can detect vulnerable communities or policy-relevant risk areas. Spatial and temporal resolution also have a significant impact on model reliability, particularly in mapping at the neighbourhood level (Zhalehdoost & Taleai, 2025). In general, the literature indicates that hybrid ML has a potential, yet the research in urban India in the future should integrate the prediction of pollutants with population exposure, vulnerability, proximity to sources, and seasonal persistence. Table 5 summarizes briefly some of the air-pollution modelling studies.

Table 5. Summary and critical synthesis of selected air-pollution modelling studies

Author/Year	Study Area	Method	Key Contribution	Limitation
Dar & Shaik (2025)	Indian cities	ML/DL clustering and forecasting	Shows spatiotemporal variation across Indian cities	Limited risk-zone classification
Natarajan et al. (2024)	Major Indian cities	Optimized ML	Improves AQI prediction	Mainly prediction-focused
Masood et al. (2023)	New Delhi	Hybrid ELM-SO	Improves $PM_{2.5}$ prediction	Limited exposure integration
Ayus et al. (2023)	China	ML/DL comparison	Shows model performance varies by data and pollutant	Non-Indian context
Xu et al. (2022)	Metropolitan area	Mobile monitoring and spatial modelling	Captures fine-scale pollutant variation	Requires intensive monitoring
Zhao et al. (2023)	Chinese metropolitan context	Hybrid DL with spatial autocorrelation	Includes spatial dependence	Context-specific
Zhang et al. (2022)	Metropolitan cities	CNN-LSTM	Supports fine-grained estimation and forecasting	High data demand

7. Conceptual Framework for Air Pollution Risk-Zone Mapping in Urban India

7.1 Framework Components

A geospatial air pollution risk-zone model ought to integrate pollution intensity with exposure and vulnerability measures. The key elements are pollution hazard, population exposure, socioeconomic vulnerability, health vulnerability, land-use sensitivity, proximity to sources, seasonal persistence, and model uncertainty. Pollution hazard is the concentration of the pollutants and exposure is the number and type of people exposed. The factors of vulnerability are used to determine sensitive populations like children, the elderly, roadside communities, and individuals with underlying health conditions. This integration can be enabled by a machine learning-based health-risk mapping framework, which would connect the prediction of pollution with population and health-risk indicators (Rajesh et al., 2025).

7.2 Stepwise Risk-Zone Mapping Process

The risk-zone mapping process proposed must start with the data gathering of the ground monitoring, satellite, meteorology, land use, traffic, and population data. These data are then to be cleaned, standardized and incorporated into a spatial database in a GIS format. Once the predictors have been selected, a hybrid machine learning model can be created and tested.

Pollutant concentration surfaces should be generated using the model output, and then hotspots should be identified and exposure-vulnerability overlay. The last process is categorization of cities into various risk groups to interpret policy. Geospatial PM 0 hotspot modelling: This illustrates the application of ML-based spatial outputs to inform specific pollution interventions (Yasin et al., 2025). Figure 2 demonstrates the conceptual framework that is proposed in the identification of air pollution risk zones in the urban India.

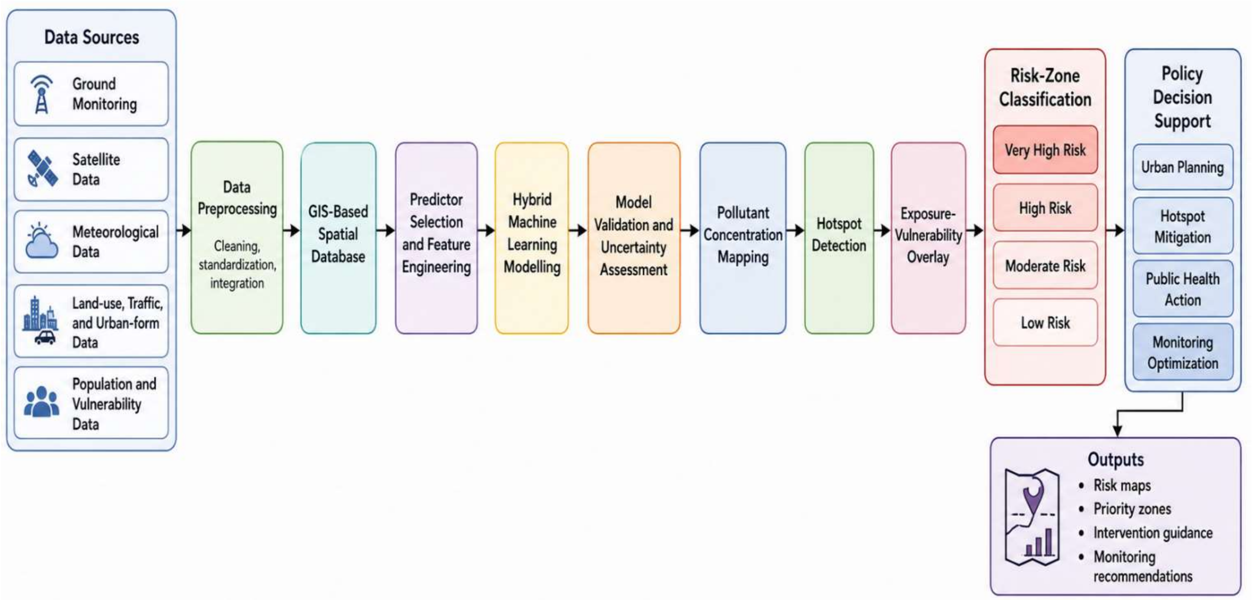

Figure 2. Proposed conceptual framework for geospatial identification of air pollution risk zones in urban India

7.3 Risk-Zone Classification and Index Development

The urban areas should be classified into very high-risk zones, high-risk zones, moderate-risk zones, and low-risk zones. The classification cannot be based on the concentration of the pollutants. It must encompass exposure load, vulnerability score, proximity of the source, seasonal persistence, and model uncertainty. A conceptual risk index can be represented as:

Air Pollution Risk-Zone Index = Pollution Hazard + Exposure Load + Vulnerability Score + Persistence Factor + Source-Proximity Score + Model-Uncertainty Adjustment

Normalization of this index can be based on weighted overlay, multi-criteria decision analysis or feature importance based on ML. Multi-pollution modelling based on GeoAI can be used to facilitate such an integrated urban environmental-risk assessment through the integration of various environmental indicators in spatial decision-making (Razavi-Termeh et al., 2024). Table 6 shows the key indicators of air pollution risk-zone classification.

Table 6. Indicators for air pollution risk-zone classification

Risk Component	Indicator Examples	Possible Data Source	Role in Risk Mapping
Pollution hazard	$PM_{2.5}$, PM_{10}, NO_2, AQI	Monitoring, satellite, ML outputs	Measures pollution intensity
Exposure load	Population density, commuter density	Census, gridded population	Identifies exposed population
Socioeconomic vulnerability	Poverty, informal settlements, housing quality	Census, municipal data	Captures social sensitivity
Health vulnerability	Schools, hospitals, elderly population	Health and municipal data	Identifies sensitive receptors
Source proximity	Roads, industries, landfills, construction zones	GIS and land-use data	Links risk with emission sources
Persistence	Exceedance days, seasonal recurrence	Time-series pollution data	Identifies chronic risk zones
Model uncertainty	Prediction error, confidence interval	Model validation outputs	Supports reliable classification

8. Policy Implications

8.1 Urban Planning and Land-Use Management

Evidence-based urban planning can be supported through geospatial air pollution risk-zone maps that help determine regions in which pollution, exposure, and vulnerability intersect. Such maps may assist planners to control land use in areas around industrial areas, high traffic routes, construction belts, and settlements near landfills. Risk-zone mapping can be used in Indian cities, where residential, commercial, transport, and industrial activities are frequently co-located, to inform zoning decisions, green-buffer development, school and hospital locations, and restrictions on new sensitive land uses in high-risk areas. The management of air-pollution using AI can also enhance the urban decision-making process by connecting the monitoring, prediction, and planning systems more closely (Rautela and Goyal, 2024).

8.2 Pollution-Control and Hotspot Mitigation

Risk-zone mapping can assist pollution-control agencies to shift their general citywide interventions to hotspots mitigation. High-risk areas may be targeted to control traffic-emission, road-dust, construction-site, industrial inspection, waste-burning prevention and local emission-source. As an example, traffic-prone areas might need to have congestion control, improvement of public transport, low-emission mobility, and enforcement of vehicle-emission. Industrial areas might need stack-emission control, the use of cleaner fuel, and more rigorous compliance inspections. The benefit of geospatial hotspot modelling is that it helps to identify the priority areas of intervention instead of considering the whole city as a homogenous surface of pollution (Yasin et al., 2025).

8.3 Public Health and Exposure Reduction

Public-health planning can also be supported using air pollution risk-zone maps. Health advisories and exposure-reduction programmes should focus on high-risk areas with high population density, schools, hospitals, elderly populations, outdoor workers, or informal settlements. Such maps can assist the authorities to develop school-level protection measures, hospital preparedness plans, occupational safety and community-level warning systems during extreme pollution events. Risk-zone maps can also be used by the public-health agencies to identify populations that need special awareness, respiratory-health screening, or preventive measures. The predictive environmental health-risk mapping is particularly applicable since it connects the prediction of pollution with the vulnerable population and health sensitive areas (Rajesh et al., 2025).

8.4 Monitoring, Governance, and Decision Support

Risk-zone mapping may also be used to enhance the design of monitoring networks by determining where new fixed stations, mobile monitoring routes, or inexpensive sensors are required. The current monitoring networks in most Indian cities are too coarse to measure neighbourhood level variation. Thus, geospatial risk analysis can be used to optimize the location of the stations in traffic routes, industrial belts, populated areas, and new urban growth areas. It is also capable of supporting real-time dashboards, early-warning systems, public communication platforms, and ward-level environmental reporting. Mobile monitoring is especially effective in identifying fine-scale spatial variation which can be overlooked by fixed monitoring stations (Xu et al., 2022). The policy translation pathway of air pollution risk-zone mapping is described in Figure 3.

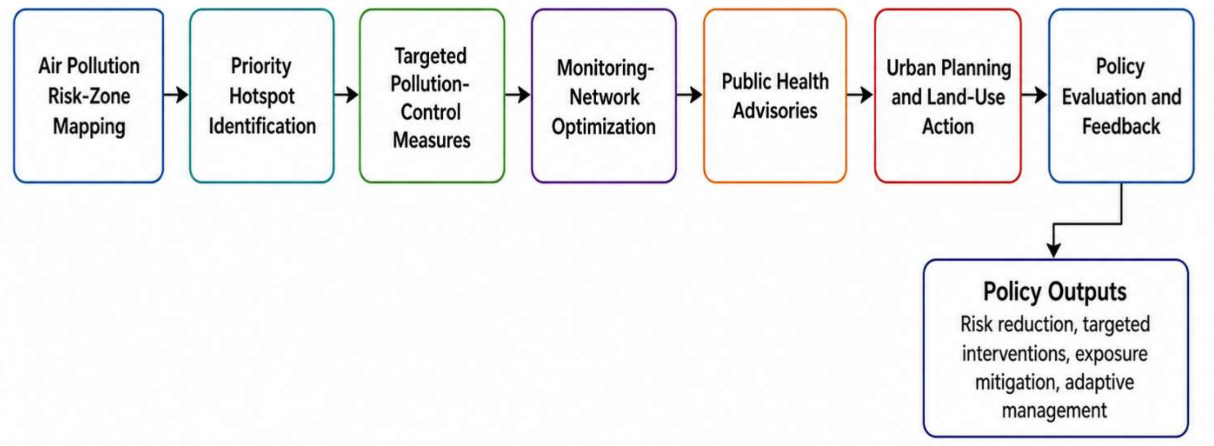

Figure 3. Policy translation pathway for air pollution risk-zone mapping

9. Challenges, Research Gaps, and Future Directions

9.1 Data and Monitoring Challenges

A significant issue in geospatial risk-zone mapping of air pollution is the disproportional distribution of quality monitoring data. Limited coverage of monitoring, absence of observations, poor temporal coverage and poor neighbourhood level exposure data are still present in many Indian cities. Such constraints decrease the accuracy of model calibration and validation. Air-quality monitoring can be aided by statistical learning techniques, yet their effectiveness is highly reliant on the quality, completeness, and representativeness of the input data (Imam et al., 2024).

The second problem is the lack of real-time monitoring, satellite data, meteorology, traffic indicators, and population exposure layers integration. Past machine-learning regression research indicates that pollutant prediction can be conducted with the available data, but these methods might be constrained in case of the absence of spatial and contextual variables (Harishkumar and Yogesh, 2020).

9.2 Methodological Challenges

Hybrid machine learning models have the potential to enhance prediction accuracy, but also present methodological issues, including overfitting, poor explainability, poor transferability and high computational requirements. Complex urban prediction tasks can be solved with deep learning models, which need large datasets and can be hard to interpret to use in policies (Bekkar et al., 2021).

Deep learning that decomposes time-series can enhance PM 2.5 forecasting, by isolating complex patterns of time-series prior to prediction, although these methods can raise the computational cost and need to be thoroughly validated before applied to operational risk-zone mapping (Zaini et al., 2022).

Another issue is model evaluation. A lot of literature uses primarily statistical measures like RMSE, MAE, R 2, or accuracy, yet these measures do not necessarily indicate that a model is able to identify vulnerable populations or policy-relevant risk areas. Systematic reviews of supervised ML methods demonstrate that model choice, type of pollutant, data structure, and validation strategy strongly influence air-quality prediction outcomes (Essamlali et al., 2024).

9.3 Research Gaps

Existing studies are still more focused on pollutant prediction than on complete risk-zone classification. Notable gaps are lack of multi-pollutant evaluation, inadequate incorporation of health vulnerability, inadequate ward-level exposure mapping, and inadequate focus on medium and small cities in India. The shift towards deep learning as opposed to neural-network-based prediction has enhanced pollution modelling in the Indian context, and further integration with geospatial exposure and vulnerability indicators is required (Nandi et al., 2024).

9.4 Future Research Directions

Future research should develop explainable, validated, and policy-oriented hybrid ML frameworks for urban India. The emphasis should be placed on low-cost sensors, mobile surveillance, multi-pollutant risk index, health-risk-integrated mapping, real-time risk-zone forecasting, and decision-support systems. The data on human mobility can also enhance exposure assessment since the mobility patterns determine the location and timing of exposure to pollution (Rahman et al., 2021). Table 7 shows the key research gaps and future directions of geospatial air pollution risk-zone mapping.

Table 7. Research gaps and future directions in geospatial air pollution risk-zone mapping

Research Gap	Current Limitation	Future Direction
Sparse monitoring	Poor intra-urban representation	Expand fixed, mobile, and low-cost sensor networks
Missing data	Incomplete temporal records	Use robust imputation and data-fusion methods
$PM_{2.5}$/AQI dominance	Limited multi-pollutant assessment	Develop composite multi-pollutant risk indices
Weak vulnerability integration	Pollution maps not linked with social risk	Integrate population, health, and socioeconomic indicators
Black-box ML models	Limited interpretability	Use explainable AI and uncertainty mapping
Poor transferability	City-specific model performance	Develop region-specific and transferable hybrid models
Limited policy linkage	Research outputs not operationalized	Build decision-support systems for urban governance
Limited health linkage	Weak disease-risk integration	Connect exposure maps with health-outcome data

Conclusion

Air pollution in Indian cities is spatially non-uniform and cannot be sufficiently explained by city-level averages or sparsely distributed monitoring stations. The review shows that geospatial methods, remote sensing, mobile monitoring, GIS-based analysis, and hybrid machine learning models are significant in detecting the hotspots of pollution and transforming them into meaningful risk areas. Risk-zone mapping must combine pollutant concentration, population exposure, socioeconomic vulnerability, health sensitivity, source proximity, seasonal persistence and model uncertainty. The value of hybrid ML methods is that they have the potential to integrate heterogeneous data and nonlinear spatiotemporal relationships, but their use needs to be thoroughly validated, explained, and transferred to different Indian urban settings. Another point that the review makes is that the future research must go beyond prediction accuracy and aim at policy-relevant classification of high-risk regions. An integrated risk-zone framework has the potential to underpin specific pollution-reduction strategies, enhanced monitoring-network design, community-health guidance, urban land-use planning, and evidence-based environmental governance. Thus, geospatial and hybrid ML-based risk mapping presents a viable avenue to enhance air pollution control and mitigate health risks associated with exposure in urban India.

References

1. Agbehadji, I. E., & Obagbuwa, I. C. (2024). Systematic review of machine learning and deep learning techniques for spatiotemporal air quality prediction. *Atmosphere*, *15*(11), 1352.
2. Ayus, I., Natarajan, N., & Gupta, D. (2023). Comparison of machine learning and deep learning techniques for the prediction of air pollution: a case study from China. *Asian Journal of Atmospheric Environment*, *17*(1), 4.
3. Badawy, H. (2025). Recent trends in spatial modeling studies of air quality (2010–2023). *Earth Science Informatics*, *18*(2), 193.
4. Bekkar, A., Hssina, B., Douzi, S., & Douzi, K. (2021). Air-pollution prediction in smart city, deep learning approach. *Journal of big Data*, *8*(1), 161.
5. Dar, A. A., & Shaik, A. J. (2025). Spatiotemporal clustering and multivariate forecasting of air quality index across Indian cities using machine learning and deep learning models. *Franklin Open*, 100435.
6. Dheekwal, A., Singh, K., & Sharma, A. (2025). GIS-based modeling of traffic-related air pollution and public health risks: a systematic review with an emphasis on developing urban contexts. *Computational Urban Science*, *5*(1), 65.
7. Essamlali, I., Nhaila, H., & El Khaili, M. (2024). Supervised machine learning approaches for predicting key pollutants and for the sustainable enhancement of urban air quality: A systematic review. *Sustainability*, *16*(3), 976.
8. Halder, B., Ahmadianfar, I., Heddam, S., Mussa, Z. H., Goliatt, L., Tan, M. L., ... & Yaseen, Z. M. (2023). Machine learning-based country-level annual air pollutants exploration using Sentinel-5P and Google Earth Engine. *Scientific reports*, *13*(1), 7968.
9. Harishkumar, K. S., & Yogesh, K. M. (2020). Forecasting air pollution particulate matter (PM2. 5) using machine learning regression models. *Procedia Computer Science*, *171*, 2057-2066.
10. Imam, M., Adam, S., Dev, S., & Nesa, N. (2024). Air quality monitoring using statistical learning models for sustainable environment. *Intelligent Systems with Applications*, *22*, 200333.
11. Karroum, K., Lin, Y., Chiang, Y. Y., Ben Maissa, Y., El Haziti, M., Sokolov, A., & Delbarre, H. (2020). A Review of Air Quality Modeling: K. Karroum et al. *Mapan*, *35*(2), 287-300.
12. Masood, A., Hameed, M. M., Srivastava, A., Pham, Q. B., Ahmad, K., Razali, S. F. M., & Baowidan, S. A. (2023). Improving PM2. 5 prediction in New Delhi using a hybrid extreme learning machine coupled with snake optimization algorithm. *Scientific Reports*, *13*(1), 21057.
13. Méndez, M., Merayo, M. G., & Núñez, M. (2023). Machine learning algorithms to forecast air quality: a survey. *Artificial intelligence review*, *56*(9), 10031-10066.
14. Nandi, B. P., Singh, G., Jain, A., & Tayal, D. K. (2024). Evolution of neural network to deep learning in prediction of air, water pollution and its Indian context. *International Journal of Environmental Science and Technology*, *21*(1), 1021-1036.
15. Natarajan, S. K., Shanmurthy, P., Arockiam, D., Balusamy, B., & Selvarajan, S. (2024). Optimized machine learning model for air quality index prediction in major cities in India. *Scientific reports*, *14*(1), 6795.
16. Rahman, M. M., Paul, K. C., Hossain, M. A., Ali, G. M. N., Rahman, M. S., & Thill, J. C. (2021). Machine learning on the COVID-19 pandemic, human mobility and air quality: A review. *Ieee Access*, *9*, 72420-72450.
17. Rajesh, M., Babu, R. G., Moorthy, U., & Easwaramoorthy, S. V. (2025). Machine learning-driven framework for realtime air quality assessment and predictive environmental health risk mapping. *Scientific Reports*, *15*(1), 28801.
18. Rautela, K. S., & Goyal, M. K. (2024). Transforming air pollution management in India with AI and machine learning technologies. *Scientific Reports*, *14*(1), 20412.
19. Razavi-Termeh, S. V., Sadeghi-Niaraki, A., Jelokhani-Niaraki, M., & Choi, S. M. (2024). Exploring multi-pollution variability in the urban environment: geospatial AI-driven modeling of air and noise. *International Journal of Digital Earth*, *17*(1), 2378819.
20. Thapa, I., Devkota, B., Lamichhane, B. R., Devkota, B. P., Dhakal, R., & Horanont, T. (2025). Applying Deep Learning and Machine Learning Algorithms to Estimate PM $ _ {2.5} $ Concentration Using Satellite Data and Meteorological Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
21. Xu, J., Yang, W., Bai, Z., Zhang, R., Zheng, J., Wang, M., & Zhu, T. (2022). Modeling spatial variation of gaseous air pollutants and particulate matters in a Metropolitan area using mobile monitoring data. *Environmental research*, *210*, 112858.
22. Yasin, K. H., Yasin, M. I., Iguala, A. D., Gelete, T. B., & Kebede, E. (2025). Methodological integration of machine learning and Geospatial analysis for PM10 pollution mapping. *MethodsX*, *14*, 103322.
23. Yasin, K. H., Yasin, M. I., Iguala, A. D., Gelete, T. B., Tulu, D., & Kebede, E. (2025). Predictive machine learning and geospatial modeling reveal PM10 hotspots and guide targeted air pollution interventions in Addis Ababa, Ethiopia. *Discover Applied Sciences*, *7*(4), 263.
24. Zaini, N. A., Ean, L. W., Ahmed, A. N., Abdul Malek, M., & Chow, M. F. (2022). PM2. 5 forecasting for an urban area based on deep learning and decomposition method. *Scientific reports*, *12*(1), 17565.
25. Zhalehdoost, A., & Taleai, M. (2025). Unravelling the importance of spatial and temporal resolutions in

modeling urban air pollution using a machine learning approach. *Scientific Reports*, *15*(1), 27708.

26. Zhang, Q., Han, Y., Li, V. O., & Lam, J. C. (2022). Deep-AIR: A hybrid CNN-LSTM framework for fine-grained air pollution estimation and forecast in metropolitan cities. *IEEE access*, *10*, 55818-55841.
27. Zhao, Z., Wu, J., Cai, F., Zhang, S., & Wang, Y. G. (2023). A hybrid deep learning framework for air quality prediction with spatial autocorrelation during the COVID-19 pandemic. *Scientific Reports*, *13*(1), 1015.