**Discover** Applied Sciences

Research

# Predictive machine learning and geospatial modeling reveal $PM_{10}$ hotspots and guide targeted air pollution interventions in Addis Ababa, Ethiopia

**Kalid Hassen Yasin[1] · Muaz Ismael Yasin[2] · Anteneh Derribew Iguala[3] · Tadele Bedo Gelete[1] · Diriba Tulu[4] · Erana Kebede[5]**

Received: 23 December 2024 / Accepted: 13 March 2025
Published online: 26 March 2025
© The Author(s) 2025 OPEN

## Abstract

Air pollution is a critical twenty-first century environmental and public health challenge that is linked to millions of deaths and ecological harm. Accurate prediction of pollutants such as $PM_{10}$ is vital for mitigation and urban sustainability. This study combines geospatial modeling with three machine learning algorithms (MLAs), Random Forest (RF), K-Nearest Neighbor (KNN), and Naïve Bayes (NB), to identify $PM_{10}$ hotspots in Addis Ababa, Ethiopia. $PM_{10}$ data from 11 stations (August 2021–August 2023) were analyzed alongside 25 atmospheric, climatic, anthropogenic, and pollution source predictors. A concentric zonal approach was used to assess spatial variability across radial distances and directional sectors and was supported by 30 m-resolution satellite imagery, climate data, and local geospatial repositories. The model accuracies were 0.95 (KNN), 0.93 (RF), and 0.88 (NB), with distinct performance trade-offs: RF predicted the largest "Good" $PM_{10}$ zones (78.98 $km^2$), KNN highlighted the most "UnHealSen" areas (279 $km^2$), and NB predict "Moderate" coverage (311 $km^2$). High $PM_{10}$ concentrations clustered in eastern and northwestern sectors, aligning with industrial zones and traffic density. The results demonstrate the efficacy of MLAs and geospatial integration in producing high-resolution pollution maps. We advocate for targeted emission controls in hotspots, expanding public transit to reduce vehicular emissions, and incorporating air quality metrics into urban planning. This study advances air quality assessment methods for rapidly urbanizing regions, providing data-driven strategies to combat pollution and enhance ecological resilience in African cities.

## Article Highlights

- Geospatial modeling combined with machine learning algorithms offers promising tools for $PM_{10}$ prediction.
- Compared with the other models, the random forest (RF) algorithm achieved better accuracy in $PM_{10}$ hotspot detection.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s42452-025-06723-w.

✉ Kalid Hassen Yasin, kalidh84@gmail.com | [1]Geo-Information Science Program, School of Geography and Environmental Studies, Haramaya University, PO Box 138, 3220 Dire Dawa, Ethiopia. [2]School of Medicine, College of Health and Medical Sciences, Haramaya University, PO Box 235, Harar, Ethiopia. [3]Center for Rural Development, Oromia State University, P.O. Box 209, Batu, Ethiopia. [4]School of Animal and Range Sciences, College of Agriculture and Environmental Sciences, Haramaya University, P.O. Box 138, Dire Dawa, Ethiopia. [5]School of Plant Sciences, College of Agriculture and Environmental Sciences, Haramaya University, P.O. Box 138, Dire Dawa, Ethiopia.

- The spatial prediction reveals hotspots 8–16.5 km in E, ENE, ESE and NW from the city center.
- Addis Ababa $PM_{10}$ reduction via emission controls and monitoring needs policy reforms, stakeholder engagement to tackle fiscal/regulatory barriers.

**Keywords** Air quality · Atmospheric pollution · Particulate matter · Public health · Remote sensing

## 1 Introduction

Air pollution is a significant risk factor for human health, contributing to increased morbidity and mortality worldwide [1]. Long-term exposure to hazardous urban air pollutants, particularly particulate matter (PM), has been associated with increased hospital admissions and deaths related to respiratory and cardiovascular diseases [2]. According to the World Health Organization [3], PM with an aerodynamic diameter ≤ 10 µm ($PM_{10}$) is more hazardous and poses severe health risks than larger particles due to its complex composition, which includes nitrates, sulfates, and heavy metals capable of penetrating deeply into the lungs and bloodstream [4, 5]. Both short and long-term exposure to respirable fractions that can enter the body through the mouth and nose can exacerbate respiratory and cardiovascular diseases, impair lung function, and increase cancer risk [6, 7]. Air pollution caused 8.1 million deaths globally in 2021, ranking as the second-leading risk factor for mortality worldwide, including among children under five, with over 700,000 deaths (15% of all global under-five mortalities). Noncommunicable diseases (heart disease, stroke, diabetes, lung cancer, and COPD) account for nearly 90% of the disease burden of air pollution [8]. In Africa, air pollution now represents the foremost environmental cause of premature death, surpassing malaria and HIV [9]. Urban ambient air pollution alone drives stark disparities; the number of annual premature deaths per 100,000 urban population reached 73 in Cairo, 46 in Lagos, and 35 in Abidjan, underscoring the urgency of region-specific interventions [9–12] Moreover, $PM_{10}$ affects social welfare and economic development by impacting public health, increasing medical costs and reducing productivity [13].

Major global sources contributing to $PM_{10}$ loads include residential fuel combustion, vehicle/engine exhaust emissions, industrial activities, biomass burning, agricultural practices, and natural dust production [14]. However, the extent of the anthropogenically induced risk of $PM_{10}$ varies spatially worldwide. For example, as cities in developing nations grow rapidly and industrialize, increased $PM_{10}$ levels pose an acute threat to environmental health, highlighting the need for proper monitoring and prediction of $PM_{10}$ levels as crucial steps in air quality management and public health protection [15]. Traditional ground-based air quality monitoring stations provide accurate but spatially limited data, often overlooking localized variations crucial for intra-urban exposure assessment [16]. To overcome these limitations, remote sensing data from satellite sensors provide frequent, low-cost observations over vast areas with high spatial resolution, making them ideal for intra-urban exposure assessment [17].

Additionally, geographic information systems (GIS) facilitate the understanding of $PM_{10}$ distribution by integrating spatial data and analyzing patterns [18]. Recent advances linking remotely sensed aerosol optical depth (AOD) to particulate concentrations have enabled indirect monitoring of near-surface $PM_{10}$ levels inaccessible to ground stations. While physical AOD—$PM_{10}$ models using classical regression have shown reasonable predictions [19, 20], the intricate, nonlinear nature of this relationship allows the exploration of more sophisticated machine learning algorithms (MLAs). These advanced algorithms can predict results more precisely than traditional statistical methods. These algorithms adapt to regional conditions, integrate different data sets and model spatiotemporal emission patterns in urban landscapes more effectively [20, 21].

Machine learning algorithms have proven to be powerful tools for analyzing the complex spatiotemporal dynamics of urban air pollutants such as $PM_{10}$ by capturing intricate relationships between $PM_{10}$ levels and predictor variables such as meteorology, terrain, and land surface characteristics, MLAs offer more comprehensive modeling than classical regression techniques [22, 23]. Previous studies have shown superior MLA predictive performance for $PM_{10}$ mapping [18, 21, 22, 24–27]. However, most studies have focused on Europe, China, Malaysia, and India, with few investigations conducted in parts of Africa [28–30]. This highlights the need for context-specific evaluation across diverse world regions to develop broadly generalizable models. Furthermore, different MLAs' distinct mathematical bases and assumptions translate into context-specific strengths and limitations, necessitating the testing of a range of algorithms to confirm the best model under various real-world conditions [31].

Despite the increased adoption of MLAs in air quality modeling, most research has been conducted in well-monitored cities in Asia and Europe, where dense ground-based sensor networks provide continuous, high-quality $PM_{10}$ measurements

[32, 33]. In contrast, African cities, including Addis Ababa, suffer from severe data scarcity, with limited air quality monitoring stations and irregular sensor coverage [34–36]. This lack of observational data poses significant challenges for ML applications, which require predictive models to compensate for missing or inconsistent $PM_{10}$ data by integrating multisource geospatial datasets, such as AOD, meteorological parameters, and urban land-use information [37, 38]. In addition to data limitations, socioeconomic constraints further hinder air quality management efforts [39]. Unlike developed regions with substantial financial support for environmental monitoring, many African cities lack the necessary investments and technological resources to establish comprehensive air quality networks [35]. Additionally, weak enforcement of environmental regulations and the absence of standardized emission control policies further exacerbate pollution challenges [40]. Many global air quality modeling studies overlook under monitored cities, reinforcing the urgent need for MLAs capable of operating in data-limited environments.

Urban expansion and industrial growth are major contributors to worsening air quality in African cities [41]. Addis Ababa, the capital of Ethiopia, is experiencing a rapid population growth rate of 4.63% annually, surpassing five million residents (estimated as of December 2023), making it one of the fastest-growing urban centers in Africa [42]. This unprecedented urban expansion is accompanied by increased vehicular emissions, industrial activities, and uncontrolled biomass burning, all of which contribute to rising $PM_{10}$ levels [43–45]. According to a recent report, Addis Ababa ranks sixth among African cities with the highest air pollution levels. The city's air pollution scale is 76.1 μg/$m^3$, surpassing the acceptable range of 0–50 μg/$m^3$, indicating poor air quality [46]. With increasing rural-to-urban migration and expanding informal settlements [47], air quality is expected to deteriorate without effective mitigation strategies. A study conducted in 2017 estimated that 2,700 lives were lost prematurely in Addis Ababa due to $PM_{10}$ exposure, highlighting the urgency of addressing air pollution in the city, which exceeds the WHO annual guideline value of 15 μg/$m^3$ by 2–5 times [48]. These trends emphasize the urgent need for data-driven pollution monitoring frameworks, such as MLAs, to support urban planning and air quality management.

The diverse geographic and climatic conditions across Africa also present a significant challenge for accurate $PM_{10}$ modeling, which often exceeds the capabilities of classical statistical models. This is particularly evident in Addis Ababa, where the city's diverse topography and high altitude (2047–3015 m above sea level) create microclimatic variations that significantly influence pollutant dispersion, leading to localized pollution hotspots with fluctuating $PM_{10}$ concentrations [49]. Seasonal variations further complicate $PM_{10}$ patterns, with dust storms and biomass burning increasing $PM_{10}$ levels during the dry season (October–January and May), whereas rainfall-driven pollutant washout affects concentrations in the wet season (February–April and June–September) [50–52]. Hence, many classical statistical models based air quality models fail to incorporate these crucial topographical and climatic factors, limiting their effectiveness in predicting air pollution trends in African urban environments [53]. This underscores the pressing need for scalable, cost-effective, and adaptive MLAs for $PM_{10}$ prediction in African cities.

Several studies in Addis Ababa have explored various air quality aspects, including TSP and $PM_{10}$ compositions in urban and peri-urban areas [54], $PM_{10}$ exposure during cooking [55], carcinogenic and non-carcinogenic health risks from roadside $PM_{10}$ [56], and the relationship between $PM_{2.5}$ toxicity and $PM_{10}$ distribution [57]. However, comprehensive source apportionment and accurate exposure assessment are hampered by the limited $PM_{10}$ monitoring infrastructure, which relies on a few sparsely and unevenly distributed operational, regulatory-grade stations. Therefore, explicit $PM_{10}$ prediction and mapping, which integrate spatial data and MLAs, are crucial for environmental planning and public health protection.

This study presents an integrated framework for MLAs and geospatial data to improve the accuracy of $PM_{10}$ prediction while enabling data-driven policy interventions in Addis Ababa, Ethiopia. We evaluate three MLAs, RF, KNN and NB, to (1) predict pollution hotspots using geospatial variables, (2) quantify environmental/anthropogenic drivers, and (3) prioritize mitigation zones via high-resolution mapping. Spatial prediction of $PM_{10}$ reveal actionable pathways for localized air quality strategies, demonstrating how computational advances bridge environmental monitoring gaps and inform equitable policy interventions.

## 2 Materials and methods

### 2.1 Description of the study area

Addis Ababa, the capital of Ethiopia with an estimated population of five million [58], is situated at an average elevation of 2355 m above sea level. Surrounding mountain ranges to the northeast through the west and opening onto plains in the east and south, the city's topography channels windblown emissions predominantly toward the west and northwest sectors. Temperature inversions commonly develop during colder seasons, especially in the early morning and evening

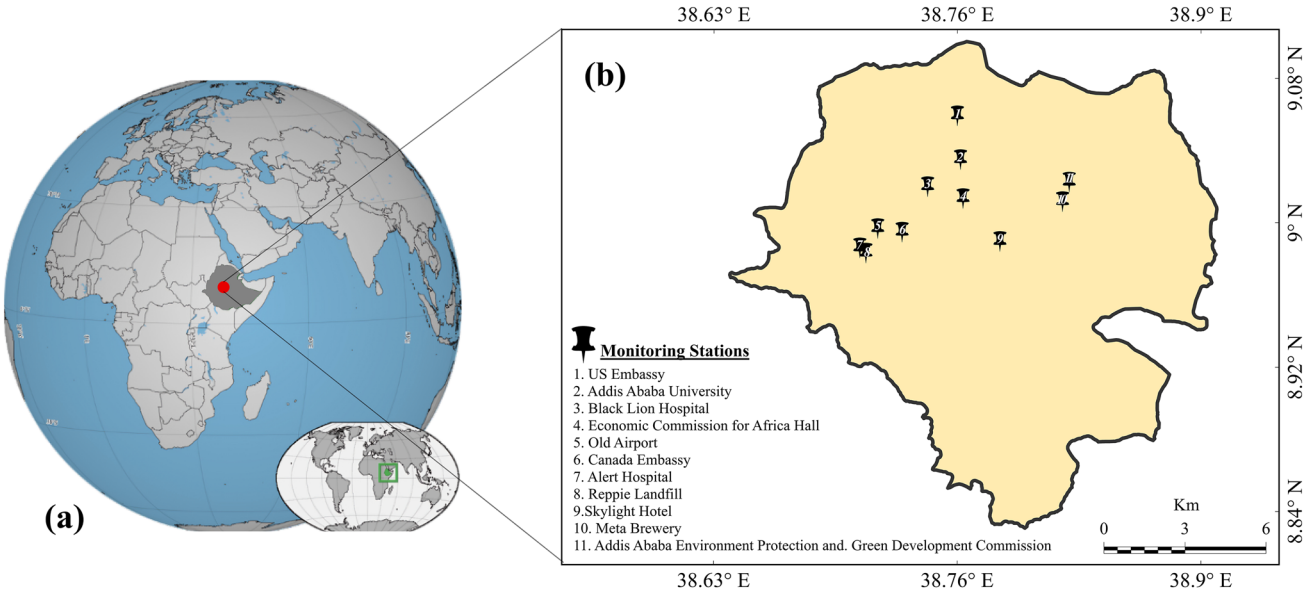



**Fig. 1** Geospatial Context of the Study Area and Monitoring Locations in Addis Ababa, Ethiopia. (a) A global map highlighting Ethiopia's geographic position within Africa, with a red marker indicating the capital city, Addis Ababa. (b) A detailed map of Addis Ababa delineating the city boundary and pinpointing the locations of 11 $PM_{10}$ monitoring stations

[59]. Addis Ababa experiences three primary climatic seasons based on rainfall patterns: the dry season from October to January and May; the light to moderate wet season (Wet 1) from February to April; and the major wet season (Wet 2) from June to September [52]. According to the National Meteorological Agency (NMA), monthly average temperatures range from a minimum of 8 °C to a maximum of 25 °C. The monthly rainfall averages range from 7 mm (November/December) to 280–290 mm (July/August), whereas the relative humidity varies monthly from 45.5% (December) to 79.5% (July) [60]. The monthly temperatures and precipitation patterns from 1990–2023 at the study location are presented in the supplementary file (Fig. S1).

## 2.2 Data acquisition

### 2.2.1 $PM_{10}$ ground measurement data

The data used in this study to detect $PM_{10}$ levels were obtained from the World Air Quality Index Project (WAQI) database (https://aqicn.org/). This platform provides daily ambient air quality monitoring data from ground stations worldwide. The locations of the air quality monitoring stations are shown in Fig. 1. Because the stations were not established concurrently, inconsistent data series lengths exist across the stations. Therefore, the study was based on annual $PM_{10}$ averages (µg/m3) from August 2021 to August 2023 for 11 stations due to data availability constraints. A stringent signal-to-noise ratio (SNR) threshold, specifically the coefficient of variation derived from station readings, was used to ensure optimal data quality. Readings showing above 10% of the SNR levels were considered susceptible to bias from erroneous sensor readings, potentially due to malfunctioning fans or dust accumulation.

Additionally, several measures were implemented to maintain data validity. The WAQI employs rigorous quality control measures, such as outlier detection and sensor calibration, to ensure data reliability. Moreover, the data undergo hourly validation against neighboring stations, whereby deviations from the median are assessed, and a weighted value is calculated. If the weighted value exceeds a predetermined threshold, the station is automatically disabled to maintain data accuracy. Moreover, the confidence zone around the median indicates data reliability, further enhancing the validity of the dataset. Thus, per the established protocol [61, 62], only readings with an SNR below 33% were incorporated to maintain data integrity. All available stations (N = 11) in Addis Ababa were included on the basis of stringent data quality and continuity criteria. Owing to the limited number of sites posing challenges, performing geostatistical analyses was statistically unfeasible given the small sample size, violating assumptions of normality and reliable variance estimation [63, 64]. However, we contend that these stations sufficiently capture urban variability in $PM_{10}$ levels across industrial, residential, and high-traffic areas despite known limitations in detecting spatial autocorrelation with $N < 30$ [65–67].

The $PM_{10}$ concentrations were categorized into three index classes (Table 1) defined by the WHO and EPA with flexible ranges [22].

### 2.2.2 PM 10 predictors (independent variables)

Studies have not established universal guidelines for selecting predictors for modeling $PM_{10}$ concentrations [68]. Therefore, this study identified relevant variables associated with $PM_{10}$ concentrations on the basis of different literatures. $PM_{10}$ concentrations are influenced by various factors, including climatic variables, transport, urban morphology, terrain, and potential sources [24, 68, 69]. These factors impact $PM_{10}$ levels to varying degrees and have been commonly incorporated into air pollution models. Therefore, in this study, 25 factors contributing to $PM_{10}$ concentrations were identified on the basis of a thorough literature review and analysis of publicly available data (Fig. 2).

**2.2.2.1 Climatic factors** Five climatic factors were selected based on their impacts on $PM_{10}$ concentrations: atmospheric density (AD), rainfall (RF), temperature (Ta), wind exposure (WEX), and wind speed (WS). The spatial data for AD and WS were sourced from the Global Wind Atlas (https://globalwindatlas.info/) at a 3 km resolution, whereas the Ta and RF datasets were acquired from the WorldClim platform (https://www.worldclim.org/) at a 1 km spatial resolution. These climatic variables, which typically appear in air pollution models, affect the generation, formation, propagation, diffusion, transformation, and sedimentation of atmospheric pollutants [70]. For example, Ta influences the formation, physical properties, and chemical characteristics of $PM_{10}$, and a relatively high Ta can increase wildfires [71]. Moreover, wind is crucial in dispersing harmful gases, such as wildfire emissions, and transporting air pollutants to neighboring areas. WS further influences the movement and dispersion of pollutants [72]. WEX, which calculates the Euclidean direction from the nearest source and uses angular steps to demonstrate the relationship between monitoring sites and average wind direction (Eq. 1), is another essential factor [73, 74]. Additionally, RF contributes to the wet deposition of atmospheric pollutants, particularly $PM_{10}$ [75], while AD affects $PM_{10}$ during dry deposition [76, 77].

The WEX was calculated using the following equation:

$$\mathbf{WEX} = \sum \frac{(\cos(\alpha - \beta))}{n} \tag{1}$$

where $\alpha$ is the angle between the wind direction and the normal to the terrain surface, $\beta$ is the angle between the wind direction and the x-axis, and $n$ is the number of wind directions considered.

**2.2.2.2 Transportation factors** Vehicular emissions significantly contribute to $PM_{10}$ concentrations [78], a primary source of air pollution in Addis Ababa [79]. Road traffic produces substantial emissions due to the combustion of fossil fuels, with PM in the form of exhaust being a direct product released into the atmosphere during this process [80]. In addition to major roads, railways generate airborne contaminants and are a major source of $PM_{10}$ [81]. Furthermore, bus stations represent substantial sources of $PM_{10}$ due to various factors, including vehicle component wear, road dust resuspension, smoking at shelters, and traffic flow [82]. Accordingly, the transportation-related factors included in this study were road density (RD), distance from roads (DFRO), distance from railways (DFR), and distance from bus stations (DFBS). The road and railway vector datasets were sourced from OpenStreetMap (https://www.openstreetmap.org/), whereas the bus station location data were derived from the SSGI Geo-Portal (http://www.ethionsdi.gov.et/) at a 1:25,000 scale. For the proximity variables, a buffer distance was established around the features, and the road density (RD) was then calculated as the total length of roads per unit area using the line density tool in ArcGIS10.8.

**2.2.2.3 Urban morphology factors** Urban morphology factors, including night light (NL), land surface temperature (LST), the urban index (UI), building density (BD), and the soil-adjusted vegetation index (SAVI), play crucial roles in influencing $PM_{10}$ concentrations. Vegetation reduces urban air pollution by absorbing gaseous pollutants through plant stomata [83, 84], and regions with high vegetation coverage typically have a lower air quality index. Conversely, increased urbanization reduces green infrastructure and contributes to higher emissions from vehicles and industries, thus impacting $PM_{10}$ [85]. Furthermore, LST is another factor affecting air quality, as extreme LST can increase atmospheric particulate levels and influence drought, leading to dry soil and forests, which may cause wildfires and release harmful gases into the atmosphere [68, 86]. Additionally, bright city NL can aggravate air pollution due to their glare interferes with nighttime chemical reactions that typically remove pollutants [87].

**Table 1** $PM_{10}$ Levels, Cautionary Statements, and Measurements from Monitoring Stations in Addis Ababa

| Level Index | Cautionary statements | $PM_{10}$ values ($\mu g/m^3$) | Monitoring stations in Addia Ababa and their $PM_{10}$ values ($\mu g/m^3$) |
|---|---|---|---|
| Good | None | 0–50 | Old Airport (38.96), AAEPGDC (47.32) and AAU (49.06) |
| Moderate | Unusually sensitive people should consider reducing prolonged or heavy exertion | 51–100 | Alert Hosp (65.98), Canada Embassy (55.77), Skyline (68.28), Meta Brewery (84.17), ECA (61.68), and USA Embassy (82.63) |
| Unhealthy for Sensitive Groups (UHealSesn) | People with heart or lung disease, older adults, children, and people of lower socioeconomic status should reduce prolonged or heavy exertion | 101–150 | Black Lion Hosp (101.43), and REPPI WTE (149) |

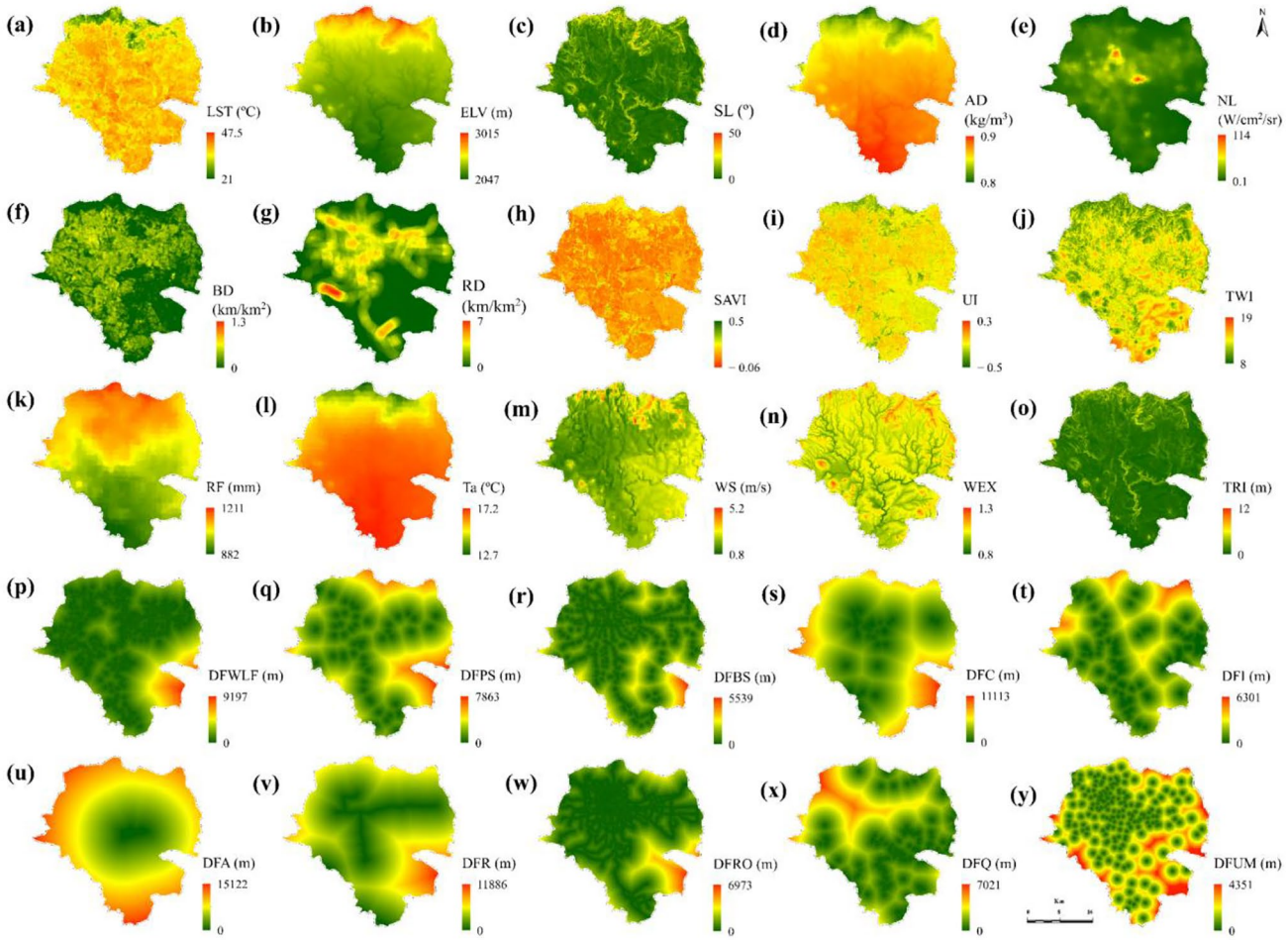



**Fig. 2** Spatial distributions of 25 predictor variables influencing $PM_{10}$ in Addis Ababa, Ethiopia. Each panel illustrates the spatial distribution of a specific predictor variable, with value ranges displayed in the legends: **a** LST, **b** ELV, **c** SL, **d** AD, **e** NL, **f** BD, **g** RD, **h** SAVI, **i** UI, **j** TWI, **k** RF, **l** Ta, **m** WS, **n** WEX, **o** TRI, **p** DFWLF, **q** DFPS, **r** DFBS, **s** DFC, **t** DFI, **u** DFA, **v** DFR, **w** DFRO, **x** DFQ, and **y** DFUM. For detailed descriptions and definitions of these predictor variables and their abbreviations, refer to Sect. 2.2.2

Building footprint data were obtained from Open Buildings (https://sites.research.google/open-buildings) and used to generate the BD layer using the kernel density tool in ArcGIS 10.8, while proximity distances to relevant features were calculated using the Euclidean distance tool. Moreover, NL datasets were sourced from the Visible Infrared Imaging Radiometer Suite (VIIRS) database (https://eogdata.mines.edu/) at 500 m spatial resolution, whereas LST, UI, and SAVI layers were generated from Sentinel-2A satellite imagery (https://scihub.copernicus.eu/) at 10 m resolution, processed via the Google Earth Engine (GEE) platform (Eqs. 2–4).

The LST was calculated as follows:

$$\mathbf{LST} = BT/(1 + \lambda * BT/\rho) * ln\varepsilon) \tag{2}$$

where BT is the at-satellite brightness temperature, λ is the wavelength of the emitted radiance, ρ = h * c/σ (1.438 * $10^{-2}$ m K), σ is the Boltzmann constant (1.38 * $10^{-23}$ J/K), h is Planck's constant (6.626 * $10^{-34}$ J s), and c is the velocity of light (3 * $10^{8}$ m/s).

The SAVI was calculated as follows:

$$\mathbf{SAVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED} + L} * (1 + L) \tag{3}$$

where NIR is the near-infrared band 8, RED is the red band 4, and L is the soil brightness correction factor (0.5).

The UI was derived the following equation:

$$\mathbf{UI} = \frac{\mathrm{SWIR1} - \mathrm{NIR}}{\mathrm{SWIR1} + \mathrm{NIR}} \tag{4}$$

where SWIR1 is the shortwave infrared band and NIR is the near-infrared band.

**2.2.2.4 Terrain factors** The terrain factors analyzed in this study are elevation (ELV), slope (SL), the topographic wetness index (TWI), and the terrain roughness index (TRI), which were selected based on their documented influence on air pollutant dispersion and deposition dynamics in urban settings. Although the predictive contributions of these variables may vary across analytical models, their inclusion is justified by mechanistic relationships with the physical processes governing $PM_{10}$ concentrations, as evidenced by prior geospatial air quality studies [18, 27, 88]. These parameters were generated using a 30 m spatial resolution Copernicus digital surface model (DSM) (https://spacedata.copernicus.eu/), processed through the terrain analysis module in SAGA-GIS 7.2.

Topography substantially affects pollution levels, with strong winds at high elevations dispersing air pollutants, whereas pollutants become trapped in low-lying areas, increasing their concentration [89, 90]. Furthermore, changes in altitude influence air pollutants, with concentrations rising near ground level and decreasing with increasing altitude [89]. The relationships between these factors and air pollution are significant, as ELV and SL directly influence atmospheric mixing and pollutant trapping. Lower elevations and gentler slopes are prone to thermal inversions and reduced ventilation, exacerbating $PM_{10}$ accumulation [89]. Conversely, steeper slopes at higher elevations enhance turbulent airflow, promoting dispersion, while TWI and TRI can affect the deposition and resuspension of $PM_{10}$. Higher TWI values indicate areas with greater potential for moisture accumulation, which can lead to increased wet deposition of pollutants. High TRI values correlate with impeded pollutant dispersion due to increased surface friction, particularly in Addis Ababa's peri-urban highlands [91]. By considering these terrain factors and their relationships with air pollution, a better understanding of the spatial distribution of $PM_{10}$ concentrations can be achieved.

The TWI was calculated using Eq. 5:

$$\mathbf{TWI} = ln\left(\frac{\alpha}{tan\beta}\right) \tag{5}$$

where α is the cumulative upslope area draining through a point per unit contour length, and β is the slope gradient (in radians).

The TRI was computed using Eq. 6:

$$\mathbf{TRI}(\mathbf{C_X}) = \sqrt{\sum_{k\epsilon N8}\left(C_k - C_x\right)^2} \tag{6}$$

where Cx is the cell under analysis and N8 is the set of 8 neighbors of the cell.

**2.2.2.5 Potential $PM_{10}$ source factors** The factors related to potential $PM_{10}$ sources included distance from waste plate and landfill sites (DFWLF), distance from petrol stations (DFPS), distance from construction sites (DFC), distance from industries (DFI), distance from the airport (DFA), distance from unions and markets (DFUM), and distance from quarries (DFQ). Location data for waste plate and landfill sites, petrol stations, airport, unions and markets and quarries were obtained from the SSGI Geo-Portal (http://www.ethionsdi.gov.et/) and used to calculate proximity distances via the Euclidean distance tool in ArcGIS 10.8. Aircraft operations produce PM emissions at varying rates during different phases, such as idling, taxiing, takeoff, and landing [92, 93]. In addition, petrol stations are potential sources of $PM_{10}$ due to high pollutant concentrations from fueling activities and leaks [94, 95]. Various industrial activities also release PM into the atmosphere, with quarries being a primary source of air pollution due to rock blasting and crushing [96, 97]. Additionally, construction site dust generated during demolition and construction and bacterial fragments from landfills and waste piles constitute other primary $PM_{10}$ sources [98]. Moreover, $PM_{10}$ levels are high in market areas due to human activities, while industrial processes, such as bulk material handling, combustion, and manufacturing, also generate $PM_{10}$. Burning fossil fuels in factories emits significant amounts of PM [99].

All thematic layers created in raster format for model integration were standardized to a uniform spatial resolution of 30 m × 30 m and reprojected to a standard reference of World Geodetic System 1984 (WGS84) spheroid with the Universal Transverse Mercator (UTM) Zone 37 N.

## 2.3 Spatial data modeling

This study analyzed annual $PM_{10}$ data collected from 11 air quality monitoring stations. A 2 km circular buffer around each monitoring site was generated to capture the localized conditions near each station. A 30 m resolution grid was subsequently created within these 11 buffers, resulting in 39,885 grid cells or observations. This methodological approach enabled us to reliably represent the $PM_{10}$ concentrations within a 2 km radius of each monitoring station while avoiding the potential oversimplification that could arise from the use of larger buffer sizes, which may fail to discern fine-scale spatial variations [18, 22, 27, 100, 101]. The predictor variables were then extracted for each of the 39,885 grid cells and saved in CSV format using the Raster statistics to polygon tool in the SAGAGIS module in QGIS 3.18. These extracted observations served as the input data for the subsequent statistical analysis conducted in the R programming environment.

The sample size and response variable type were considered when choosing a suitable MLA. $PM_{10}$ concentrations were converted into categorical data for our study. R's classification and regression training (CARET) package implemented three MLAs: KNN, RF, and NB. This package offers crucial features for data preprocessing, sample selection, variable selection, data splitting, and model building [102]. The datasets were thoroughly cleaned to reduce noise; nonpredictive observations were eliminated, leaving the relevant labels and features necessary for the study, which decreased the number of observations from 39,885 to 38,408. Data normalization, which places the data under the 0–1 range (Eq. 7), is crucial for enabling ML estimation [103].

The dataset was partitioned via spatially stratified random sampling: 70% (26,886) for training and 30% (11,522) for testing. This spatial stratification protocol preserved autocorrelation structures and ensured a representative $PM_{10}$ distribution across Addis Ababa, thereby minimizing spatial bias in model evaluation. The partitioning methodology optimized the bias−variance trade-off while maintaining spatial integrity, facilitating rigorous validation of the MLAs. The detailed methodological framework is presented in Fig. 3.

$$\boldsymbol{X_n} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{7}$$

where $X_{min}$ is the lowest value of $X$, $X_{max}$ is the greatest of $X$, $X$ is the original value, and $X_n$ is the normalized value.

Early stopping was implemented during RF training to optimize generalization, dynamically terminating tree growth when validation accuracy plateaued, thereby tailoring the ensemble size (*ntree*) to data complexity. This protocol balanced

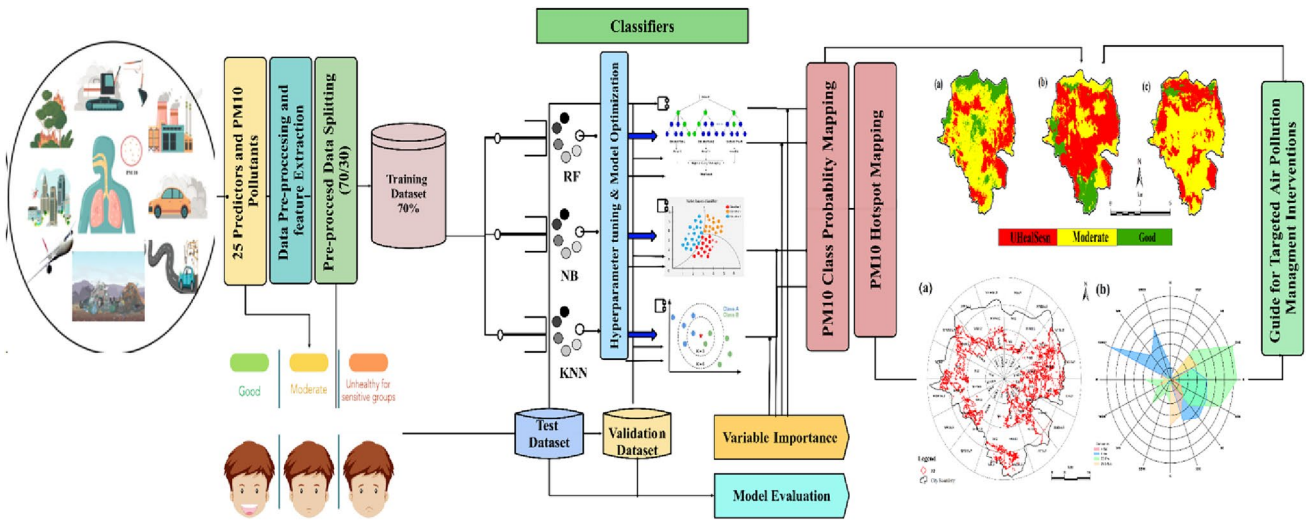



**Fig. 3** Methodological framework for identifying and mapping $PM_{10}$ hotspots and guiding targeted interventions for air pollution management in Addis Ababa, Ethiopia

computational efficiency and performance, avoiding redundant iterations. For NB and KNN, early stopping was structurally irrelevant: NB's single-pass probabilistic design and KNN's training-free prediction mechanism inherently preclude iterative optimization [22]. This methodological rigor prevented overfitting by establishing precise termination criteria at the model's optimal performance threshold. The model architectural framework underwent systematic refinement through hyperparameter optimization, employing a grid search methodology to traverse the parameter space. Cross-validation (CV) metrics guided the selection of optimal hyperparameter configurations, thereby minimizing prediction error margins. The resulting model architecture demonstrated robust performance metrics that satisfied predetermined validation criteria, aligning with established methodological standards [18, 22]. The models' descriptions are presented as follows:

#### 2.3.1 Random forest (RF)

The RF algorithm, a supervised ensemble learning method proposed by Breiman [104], excels in CART, particularly in predicting air pollution levels [22, 105]. The algorithm develops multiple decision trees during training and aggregates their predictions, utilizing bootstrap aggregating (*bagging*) to train each tree on a random subsample of instances. RF was used in this study due to its robust prediction accuracy and resilience to overfitting when handling multidimensional problems [106, 107]. The RF can also calculate variable importance metrics on the basis of variable selection frequency during sampling and limit the pool of candidate-splitting variables at each decision tree node to a random subset of all variables [22, 68, 108].

In this study, the RF R package [109] was used, which uses observations in the bag for training and observations outside the bag for the CV. The RF hyperparameters were optimized through a structured, iterative process. The *mtry* parameter, which defines the number of variables sampled at each split, was evaluated using a grid search combined with a tenfold CV (repeated three times) over a range of 1–20, resulting in an optimal value of 4. The *ntree* parameter, which controls the number of trees, remained unchanged at the default value of 500 as preliminary tests confirmed stable performance above this threshold. *Maxnodes* (maximum number of end nodes per tree) was examined (5–30).

#### 2.3.2 K-nearest neighbor (KNN)

KNN, a supervised learning technique suitable for classification and regression, operates on the principle of lazy learning without assuming underlying data distribution [24, 110]. This characteristic makes KNN suitable for air pollution prediction tasks lacking universal predictors [18, 27]. The algorithm classifies new data points by finding their $k$ closest neighbors in the training set and allocating them to the most frequently represented class in that neighborhood. The nearest neighbor computation relies on distance metrics and $k$ selection [111, 112]. The algorithm uses the given distance function to classify new points on the basis of a majority vote among the $k$ closest training instances [27, 113]. The model's accuracy increases when the relative weights assigned to neighbors in determining class membership match their significance [114]. The Euclidean distance is frequently employed for continuous features, whereas the Hamming distance is favored for distinct features [24, 115]. The distance between points $a$ and $b$ with $n$ features is computed using Eq. 2:

$$d_{euclidean} = \sqrt{\sum_{i=1}^{n} (ai - bi)^2} \tag{8}$$

where $n$ is the number of dimensions, $a$ is the data point from the data, and $b$ is the new data to be predicted.

The ideal $k$ (number of neighbors) was systematically tested across odd integers 1–31 to avoid classification ties [110]. This study employed a tenfold CV to evaluate model performance and prevent overfitting [115]. After several $k$ values were compared, $k = 5$ was chosen for optimal performance.

#### 2.3.3 Naive Bayes (NB)

NB classifiers, which are grounded in Bayes' theorem, assume feature independence within a given class [116, 117]. Despite this simplifying assumption, NB performance can rival decision trees and neural networks in specific contexts [118]. The algorithm's advantages include noise insensitivity, ease of implementation, and insensitivity to dataset size [27, 118]. However, air pollution forecasting rarely adopts NB despite its speed, simplicity, and efficacy on large datasets [117]. The NB theorem is defined below in Eq. 3:

$$P(c/d) = \frac{P\left(\frac{d}{c}\right)P(c)}{P(d)} \tag{9}$$

where P(c/d) is the posterior probability, P(d/c) is the likelihood, P(c) is the class prior probability, and P(d) is the predictor prior probability.

The R software's "*nb*" function executes the model, employing a kernel density estimate to represent the conditional probability density function nonparametrically. Hyperparameter tuning for the NB classifier focused on refining probabilistic calibration and handling zero-frequency edge cases. Three critical parameters were examined: Laplace correction (*fL*), which was tested at 0, 0.5, and 1 to modulate class-related probability smoothing; the kernel density bandwidth adjustment (*adjust*), rated 0, 0.5, and 1 to optimize the nonparametric density estimate; and the distribution type (*usekernel*), which was set to core-based estimation to provide flexibility for continuous predetermination to increase actors. The CV result identified an optimal configuration ($fL = 0$, $adjust = 0.5$) that balanced trade-offs between distortion and variance while preventing excessive smoothing.

### 2.4 Model validation and variable importance

Model evaluation and validation are critical for assessing reliability [119]. This study used several vital metrics for model evaluation. A confusion matrix (truth table) was constructed to facilitate the model assessment. It is a standard technique for evaluating MLAs. For both binary and multiclass classification tasks, the confusion matrix reliably reports model performance [88]. From the confusion matrix, various statistical measures were calculated to assess the predictive abilities of the MLAs. These included overall accuracy, 95% confidence interval (95% CI), no information rate (NIR), *p*-value, kappa index (*κi*), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), detection prevalence (DP), balanced accuracy, detection rate (DR), and F1 score. The area under the receiver operating characteristic (ROC) curve (AUC) was also calculated. The equations for these metrics are listed in the supplementary file (Table 1).

These metrics are widely accepted for assessing the reliability of MLAs in classification problems [120]. Sensitivity, also known as the recall or true positive rate, measures the proportion of actual positives correctly classified, whereas specificity (true negative rate) measures the ratio of actual negatives correctly classified. The precision (PPV) indicates the accuracy of the classifier in identifying positive classes. Overall accuracy provides an aggregate performance measure by determining the proportion of correct predictions among all cases. The 95% CI gives a range that can be 95% confident and contains true accuracy. Multiple other metrics provide nuanced and balanced performance assessments across classes, including balanced accuracy, *κi*, negative predictive value (NPV), and the F1 score [121]. The ROC curve visualizes the performance of a classifier across different thresholds, whereas the AUC provides a scalar summary. An AUC of 1 denotes perfect classification. The statistical significance of the model's accuracy over random chance is evaluated using the NIR and *p*-value [122]. We calculate the AUC using the '*auc*' function from the *'PROC'* package.

The feature selection methodology utilized within the CARET environment integrates predictor evaluation concurrently with model optimization protocols. This analytical approach generates variable importance hierarchies through the implementation of R's *'VarImp'* function. While conventional spatial autocorrelation metrics traditionally favor continuous data structures, the analytical framework incorporates systematic visual diagnostics to identify potential spatial clustering patterns within model residuals. To quantify the linear associations among the $PM_{10}$ conditioning factors, we computed a Pearson correlation matrix, which was selected for its capacity to capture both the magnitude and directionality of pairwise relationships.

## 3 Results and discussion

### 3.1 Intervariable correlation analysis

The correlation heatmap of the $PM_{10}$ conditioning factors (Fig. 4) illustrates the interrelationships among the 25 variables for MLA integration and reveals moderate to low correlations (−0.36 to +0.57) that prevent multicollinearity risk (< 0.7). Significant positive correlations, such as TWI-ELV (+0.51), RD-DFC (+0.57), and DFA-TRI (+0.51), highlight the interdependencies between topography and anthropogenic activities, whereas WS-TWI (+0.42) and WS-NL (+0.39) suggest

interactions between the atmosphere and topography. Conversely, LST-DFPS (− 0.36) and LST-DFWLF (− 0.30) represent inverse relationships between the LST and anthropogenic factors.

Domain-specific analyses identified RD and DFC as positive influences (+ 0.57 for DFC), whereas SAVI had weak correlations ($|r| < 0.32$), suggesting a limited direct effect of vegetation. In ML integration, the lack of high collinearity ensures the differentiating power of the variables. This matrix provides an empirical basis for including variables and balances interpretability and model efficiency.

### 3.2 Variable importance analysis

The variable importance analysis for the MLAs revealed variations in the significance of individual models for predicting $PM_{10}$ concentrations, as shown in Fig. 5a–c. The models were categorized into Good, Moderate, and UHealSesn models on the basis of the WHO and EPA classifications. In the KNN model, proximity-based factors such as DFI, DFBS, and DFUM consistently ranked as the most influential variables across all the $PM_{10}$ levels, with importance scores ranging from 99.4% to 32.7% (Fig. 5b). These findings are consistent with those of previous studies that highlighted the impacts of industrial emissions, transportation networks, and urban activities on $PM_{10}$ concentrations [21, 123, 124]. DFA exhibited high importance for both Good and Moderate $PM_{10}$ levels (Fig. 5a, b). The RF model identified different top-ranking variables, such as DFBS, DFA, DFQ, and DFRO, emphasizing the significance of transportation-related variables and proximity to potential emission sources. The prominence of transportation proximity factors aligns with Addis Ababa's reliance on aging diesel-powered buses and minibuses, which are known to emit high PM during frequent stops and idle at congested terminals

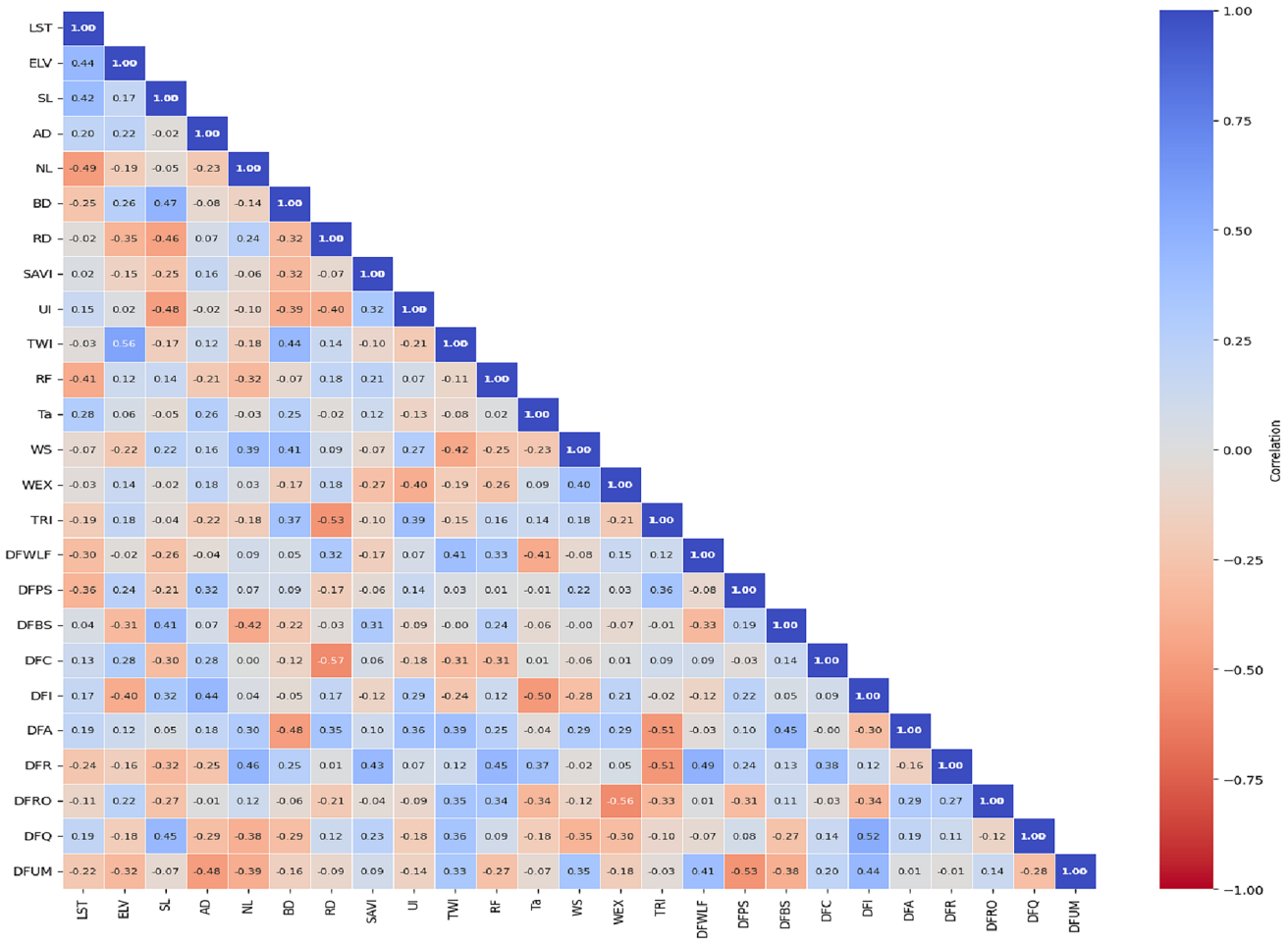



**Fig. 4** Correlation heatmap of 25 $PM_{10}$ predictor variables, illustrating pairwise Pearson correlation coefficients (*r*). The color map indicates the strength, direction, and magnitude of the correlations: red denotes negative associations, blue denotes positive associations, and the intensity of the colors corresponds to correlation values ranging from -1.00 (strong negative correlation) to 1.00 (strong positive correlation)

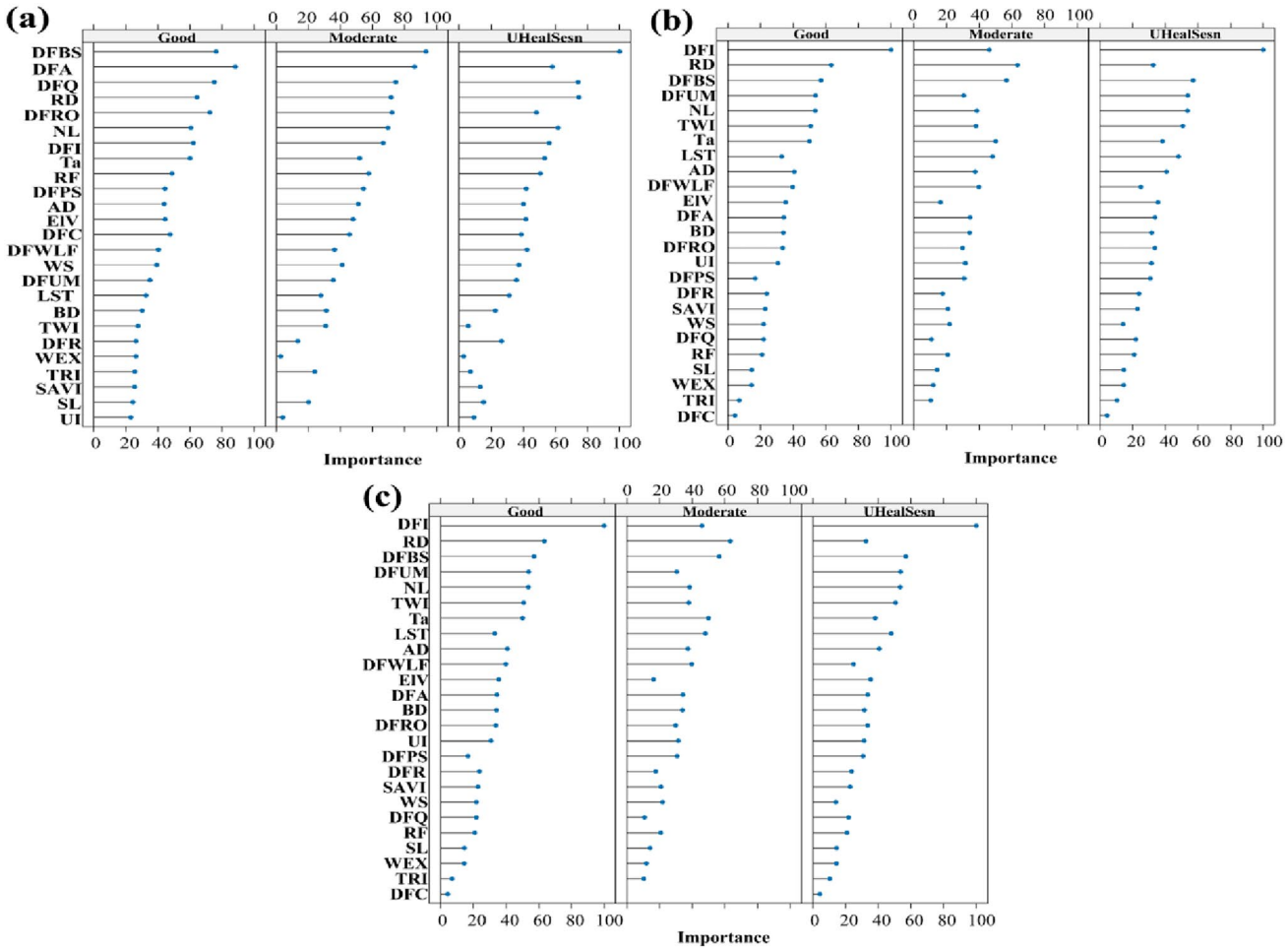



**Fig. 5** Percentage importance of predictive variables in determining $PM_{10}$ levels for the Good, Moderate, and UHealSesn groups across three MLAs: **a** RF, **b** KNN, and **c** NB. The height of each bar indicates the percentage contribution of the corresponding variable to the predictions for each MLA and group

[125]. These findings are consistent with pollution source distribution studies in rapidly industrializing African cities such as Rabat and Mpumalanga, where vehicular emissions and informal industries dominate $PM_{10}$ profiles [126, 127].

The UI appeared to be the least significant factor among all the predictive variables, with importance scores ranging from 3.7% to 23% (Fig. 5a). Similarly, Arowosegbe et al. [127] reported that urban representative parameters have low importance in predicting $PM_{10}$ concentrations across four South African provinces. The NB model showed a similar pattern to that of the KNN model, with DFI, RD, DFBS, and DFUM rankings among the most influential predictors. However, slight variations were observed in the importance rankings of certain variables, such as NL and Ta, compared with those of the KNN model, possibly reflecting the unique assumptions and characteristics of the NB algorithm.

Our study highlights the varying significance of input variables across different MLAs for predicting $PM_{10}$ concentrations in the study area. Proximity-based factors, such as transportation networks and potential emission sources, have consistently emerged as influential predictors, aligning with the well-established influence of these variables on $PM_{10}$ levels in urban environments. However, terrain characteristics, such as TRI and SL, consistently exhibited low importance across all the models and $PM_{10}$ levels, indicating their limited usefulness for $PM_{10}$ prediction in the study area. These results differ from those of Tella et al. [27] and Tella & Balogun [18], who reported that terrain factors were most influential in Asia–Pacific cities. This discrepancy may be attributed to geographic differences in pollution sources and urban forms between the African study area and the Asian cities examined in previous studies. The divergence underscores the importance of testing multiple modeling approaches, as the significance of variables can vary substantially depending on the algorithms used [128].

### 3.3 Classifier performance evaluation

This study evaluated the performance of three MLAs in predicting $PM_{10}$ levels. The results revealed that the ensemble-based RF model outperformed the KNN and NB models, which aligns with recent studies that have demonstrated the superior performance of these algorithms in air pollution modeling tasks [18, 22, 105]. The performance of RF can be attributed to its ability to handle nonlinear relationships effectively, capture complex interactions between predictor variables, and its robustness to overfitting and noisy data [18]. The findings of this study were supported by the high overall accuracy of RF (95%) and exceptional performance across various evaluation metrics (Table 2).

The RF models exhibited perfect AUC scores of 1.00 for the Good $PM_{10}$ category, indicating excellent classification performance. For the Moderate category, RF and KNN demonstrated outstanding AUC values of 0.99, whereas NB had a slightly lower but still notable AUC of 0.96 (Fig. 6). In the UnHealSen category, RF maintained its superior performance, with an AUC value of 0.99, closely followed by KNN and NB, with an AUC of 0.98.

The KNN model achieved an overall accuracy of 93%, although it was slightly behind the RF model (Table 3). This observation is consistent with the work of Kujawska et al. [129], who employed KNN, along with other MLAs, for $PM_{10}$ prediction in Lublin, Poland, and reported that KNN exhibited good predictive performance, albeit slightly lower than that of ensemble-based methods. The NB model performed reasonably well in specific categories, such as the UnHealSen category, but showed the lowest overall accuracy (88%) among the evaluated models. This finding aligns with those of previous studies that highlighted the limitations of NB in handling complex, nonlinear relationships and high-dimensional data, which are common in air pollution modeling tasks [27, 130]. Rigorous evaluation techniques, including CV and performance metrics, were employed to ensure the robustness and reliability of the findings. We implemented a *k*-fold CV with varying fold sizes (e.g., twofold, 13-fold, 15-fold, and 17-fold), and the results were consistently high with minimal variance across folds, as detailed in Fig. 7.

The RF model achieved overall accuracies of 95.16% and 95.25% for twofold and 13-fold CVs, respectively, whereas the KNN model peaked accuracy of 97.22% with 15- and 17-fold CVs. To mitigate overfitting, an inherent risk given the high accuracy scores, our approach leverages the intrinsic ability of the RF model to generalize by aggregating predictions from multiple decision trees trained on random subsets of data and predictors. These CV outcomes and consistent performance across independent folds underscore the model's generalizability and confirm that high accuracy reflects true predictive power rather than overfitting.

The results demonstrate that RF achieves superior accuracy compared with other models, which aligns with prior research validating RF's effectiveness. For example, studies such as Son and Kim [131] reported that RF outperforms linear regression and support vector machines in forecasting $PM_{10}$ levels in Seoul. Similarly, Shree et al. [132] reported that RF outperformed mixture discriminant analysis (MDA) and bagged CART with 93% accuracy. A study by Choubin et al. [68] and Abu El-Magd et al. [105] further reinforced RF reliability and robustness. Comparative analyses conducted by Tella et al. [27] and Tella & Balogun [18] demonstrate RF superiority over KNN and NB MLAs in predictive performance metrics. However, KNN remains a viable option for $PM_{10}$ prediction, achieving accuracies exceeding 0.9. For instance, Aini and Mustafa [114] applied KNN to predict $PM_{10}$ in Makassar, attaining 96% accuracy, whereas

**Table 2** Performance Metrics of MLAs (RF, KNN, and NB) for $PM_{10}$ prediction

| MLAs | Class | Classifier performance measures | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Sensitivity | Specificity | PPV | NPV | Detection Rate | Detection Prevalence | Balanced Accuracy | F1 score |
| RF | Good | 1.00 | 1.00 | 1.00 | 1.00 | 0.17 | 0.17 | 1.00 | 1.00 |
| | Moderate | 0.95 | 0.95 | 0.96 | 0.94 | 0.52 | 0.55 | 0.95 | 0.95 |
| | UHealSesn | 0.92 | 0.96 | 0.90 | 0.97 | 0.25 | 0.28 | 0.94 | 0.91 |
| KNN | Good | 1.00 | 1.00 | 1.00 | 1.00 | 0.17 | 0.17 | 1.00 | 1.00 |
| | Moderate | 0.94 | 0.93 | 0.95 | 0.93 | 0.52 | 0.55 | 0.94 | 0.94 |
| | UHealSesn | 0.89 | 0.95 | 0.88 | 0.96 | 0.25 | 0.28 | 0.92 | 0.89 |
| NB | Good | 0.94 | 0.93 | 0.75 | 0.99 | 0.17 | 0.23 | 0.93 | 0.83 |
| | Moderate | 0.83 | 0.94 | 0.94 | 0.83 | 0.44 | 0.47 | 0.89 | 0.88 |
| | UHealSesn | 0.91 | 0.94 | 0.86 | 0.96 | 0.26 | 0.30 | 0.93 | 0.89 |

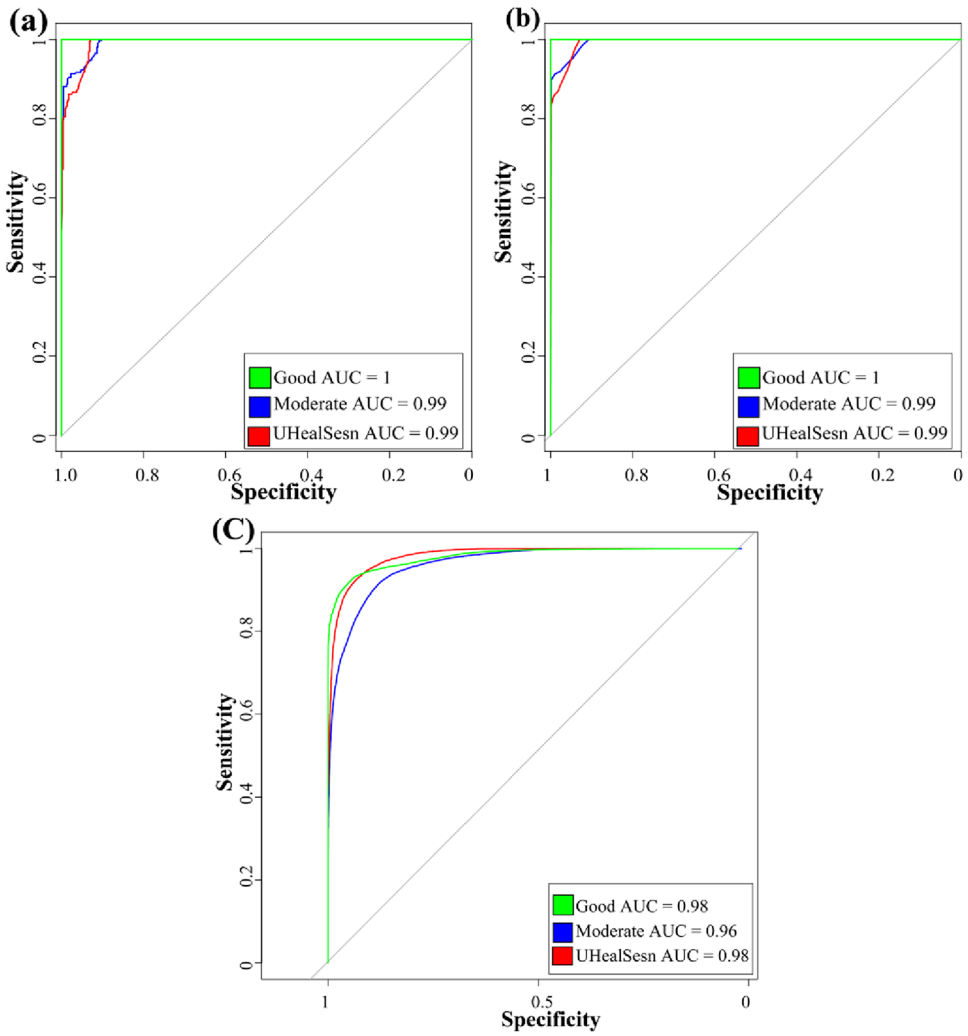



**Fig. 6** ROC curve analysis of three MLAs: **a** RF, **b** KNN, and **c** NB applied to the testing dataset. The AUC values indicate the model's classification performance for the Good, Moderate, and UHealSesn categories, with higher AUC values signifying better predictive accuracy

**Table 3** Overall performance of the three MLAs (RF, KNN, and NB)

| Performance | RF | KNN | NB |
|---|---|---|---|
| Overall Accuracy | 0.95 | 0.93 | 0.88 |
| 95% CI | 0.93, 0.96 | 0.92, 0.95 | 0.87, 0.90 |
| SE | 0.00765 | 0.00765 | 0.0102 |
| NIR | 0.55 | 0.56 | 0.53 |
| *P*-Value [Acc > NIR] | $< 2.2 \times 10^{-16}$ | $< 2.2 \times 10^{-16}$ | $< 2.2 \times 10^{-16}$ |
| *κi* | 0.91 | 0.89 | 0.81 |

Maheshwari and Lamba [133] reported that KNN outperforms linear regression and stochastic gradient descent in Beijing's air quality modeling. Although NB is relatively new to air pollution prediction, its competitive performance in this study mirrors findings by Tella et al. [27], suggesting its growing potential in the field.

### 3.4 Spatial extent of $PM_{10}$ pollution in Addis Ababa

This study quantitatively assessed the spatial extent of $PM_{10}$ pollution levels across Addis Ababa via three MLAs. The results revealed varying predictions among the MLAs, with the RF model estimating 78.98 $km^2$ (15.82%) under the Good class, whereas the KNN model predicts 70.69 $km^2$ (14.27%). The model's spatial prediction (Fig. 8b) reveals limited areas in southern residential neighborhoods with Good air quality, owing to green cover and residential planning [45].

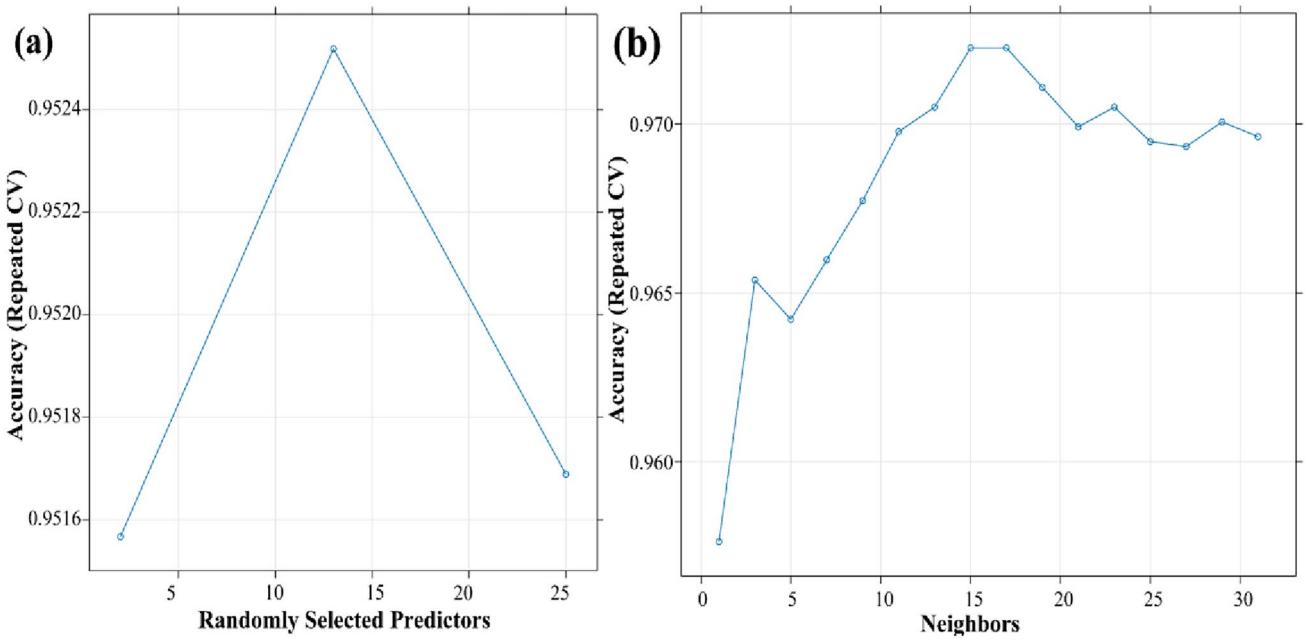



**Fig. 7** K-fold CV outcomes for the **a** RF model across varying numbers of randomly selected predictors and **b** KNN model as a function of the number of neighbors

The KNN model predicts 279.37 km$^2$ (56.41%). This observation resonates with previous research linking elevated $PM_{10}$ concentrations to industrial activities, transportation networks, and landfill effects [134–137]. The NB model predicts the smallest area of 15.48 km$^2$ (3.13%) with Good air quality.

All the models consistently predict Moderate $PM_{10}$ pollution levels as the most widespread across the city. The NB model estimates the largest area under Moderate pollution at 311.01 km$^2$ (62.8%), followed by the RF model at 283.9 km$^2$ (57.33%) and the KNN model at 145.12 km$^2$ (29.3%). The RF model indicates that these Moderate $PM_{10}$ pollution levels are distributed across the study area, with relatively high concentrations in the western and central corridors. In the model's spatial prediction (Fig. 8a), the commercial district in the west-central region experiences Moderate pollution, likely due to traffic-related emissions [44, 138, 139]. The model's spatial prediction (Fig. 8c) shows that the central business district and Addis Ababa University campus experienced Moderate pollution levels.

Alarmingly, the KNN model predicts the largest extent of UHealSesn $PM_{10}$ levels, covering 279.37 km$^2$ (56.41%) of the study area. The NB model also showed a significant portion of 168.67 km$^2$ (34.06%) under UHealSesn conditions, followed by the RF model at 132.98 km$^2$ (26.86%) (Table 4).

The spatial extent of the $PM_{10}$ pollution predictions varied among the three MLAs employed in this study. To quantify the uncertainty associated with each model's predictions, we calculated the relative percentage differences in the spatial extent of the $PM_{10}$ pollution levels between each pair of MLAs (Table 5). This approach allows us to assess the degree of agreement or divergence among the models, providing insight into the reliability of the predictions. Several factors can be attributed to discrepancies in spatial patterns, including the inherent variability among MLAs stemming from their underlying assumptions, modeling techniques, input variables, tuning parameters, and sensitivity to training data [140]. The relative percentage difference for the RF and KNN models is relatively

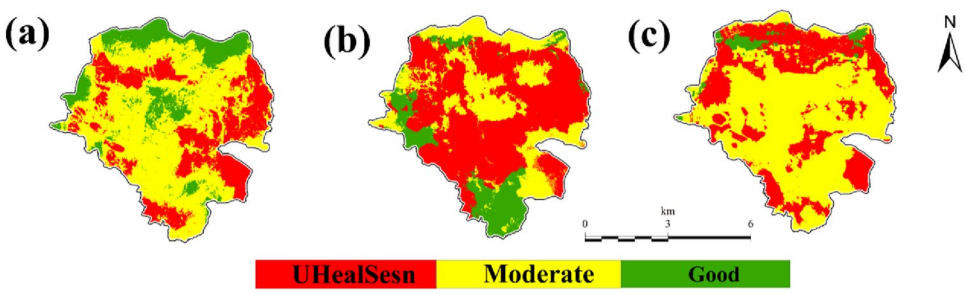



**Fig. 8** Map for spatial predictions of $PM_{10}$ pollution levels generated by three MLAs: **a** RF, **b** KNN, and **c** NB

**Table 4** Spatial extent predictions of $PM_{10}$ pollution levels (Good, Moderate, and UHealSesn) using three MLAs: RF, KNN, and NB

| $PM_{10}$ Level | Area | RF | KNN | NB |
|---|---|---|---|---|
| Good | $Km^2$ | 78.98 | 70.69 | 15.48 |
| | % | 15.82 | 14.27 | 3.13 |
| Moderate | $Km^2$ | 283.9 | 145.12 | 311.01 |
| | % | 57.33 | 29.3 | 62.8 |
| UHealSesn | $Km^2$ | 132.98 | 279.37 | 168.67 |
| | % | 26.86 | 56.41 | 34.06 |

**Table 5** Relative percentage differences in the spatial extent of $PM_{10}$ pollution predictions between MLAs

| MLAs | Relative difference | Good | Moderate | UHealSesn |
|---|---|---|---|---|
| RF vs. KNN | $km^2$ | 8.29 | 138.8 | 146.39 |
| | % | 11.08 | 64.71 | 71.2 |
| RF vs. NB | $km^2$ | 63.5 | 27.1 | 35.69 |
| | % | 115.35 | 11.1 | 18.43 |
| KNN vs. NB | $km^2$ | 55.21 | 165.9 | 110.7 |
| | % | 128.14 | 72.74 | 49.41 |

small (11.08%) in predicting areas with good air quality but diverges significantly for the Moderate (64.71%) and UHealSesn (71.2%) levels. This suggests that RF and KNN perform similarly for low pollution levels but differ in their ability to predict areas with relatively high pollution. The comparison between the RF and NB models revealed a significant difference in the Good $PM_{10}$ level (115.35%), indicating that the NB model may be less reliable than the RF model in predicting areas with low pollution levels. However, the differences are smaller for Moderate (11.1%) and UHealSesn (18.43%) levels, suggesting some agreement in high-pollution areas. The KNN and NB models show the greatest discrepancies across all categories, with differences of 128.14% for Good, 72.74% for Moderate, and 49.41% for UHealSesn, highlighting substantial discrepancies between these models, particularly in predicting areas with Good to Moderate pollution levels.

Despite these variations, the RF model consistently outperformed the other models in our study, and other researchers have conducted various comparisons [18, 22, 27]. The superiority of the RF model can be attributed to its ability to implicitly detect complex data interactions, even in the presence of noise variables. Moreover, RF helps avoid overfitting issues commonly seen in other methods through ensemble averaging [105]. These characteristics make the RF model's classification more scientific, realistic, and acceptable for real-world applications. However, decision-makers should not ignore the regions predicted as polluted areas by other models. The discrepancies in the spatial extent of $PM_{10}$ pollution predictions among MLAs underscore the importance of considering multiple modeling approaches to capture the full range of potential $PM_{10}$ hotspots.

Furthermore, the relative absolute differences in $PM_{10}$ predictions between MLAs can serve as a measure of uncertainty in the spatial mapping of $PM_{10}$ levels. Areas with high discrepancies among MLAs may indicate regions where additional data collection or model refinement is needed to improve the accuracy and reliability of pollution estimates. This information can guide future research efforts and help prioritize air quality monitoring and management resources, ensuring that the most vulnerable areas receive the necessary attention and interventions.

### 3.5 Spatial distribution of $PM_{10}$ hotspot areas

To systematically investigate the spatial distribution of $PM_{10}$ pollution in Addis Ababa, Ethiopia, this study employed concentric zonal analysis to identify pollution hotspots across radial distances and directional sectors. The methodological framework divided the study area into concentric rings at radii of 4, 8, 12, and 16.5 km, which were subdivided into 16 directional sectors originating from the city center (Fig. 9a, b). Given its superior accuracy and real-world applicability, the RF model's UHealSesn level results were overlaid onto these spatial units. This facilitated a dual examination of pollution gradients along both radial and angular dimensions. This approach delineates hotspot areas and provides critical insights into their driving factors, informing spatially targeted mitigation policies.

Our $PM_{10}$ hotspot analysis (Fig. 9a) revealed pronounced UHealSesn concentrations in the eastern (E), east–northeastern (ENE), and east-southeast (ESE) sectors, which were located mainly between 8–16.5 km from the city center. Quantitatively, the 12 km ring in the E direction accounts for 7.38% of the study area, whereas the 8 km and 12 km rings in the ENE direction account for 2.54% and 7.42%, respectively. Similarly, the 8 km and 12 km rings of the ESE sector encompass 4.31% and 5.9%, respectively, of the total area. These spatial trends correlate strongly with documented $PM_{10}$ drivers, including transportation infrastructure expansion and unregulated urban sprawl in these zones [138, 141, 142].

In contrast, the northern and western sectors present markedly lower hotspot intensities. Notably, the northwest (NW) and west—northwest (WNW) sectors within a 12 km radius constitute 0.46% and 7.09% of the study area, respectively. The southern (S) and southeastern (SE) regions also demonstrated reduced pollution, with the 16.5 km ring constituting 3.83% and 1.87% of the area, respectively. This spatial heterogeneity underscores the multifactorial nature of $PM_{10}$ dynamics, which are shaped by topography, urban morphology, meteorological conditions, and emission source distribution [22, 143]. Notably, the cold spot in the 16.5 km north—northwest (NNW) sector, a pollution-free zone, stems from particulate scavenging by high altitude and rainfall in the Entoto forest, which dissolves $PM_{10}$ from vehicular and domestic emissions [144]. Conversely, localized hotspots in WNW regions align with industrial zones and transportation corridors (Fig. 9b) [145–147].

The spatially explicit analysis identified $PM_{10}$ hotspots concentrated in eastern (E, ENE) and western (NW, WNW) corridors 8–16.5 km from the city center, where industrial clusters, traffic density, and urbanization intersect. The eastern corridor exhibited peak $PM_{10}$ concentrations clustered around the Lemi Kura and Akaki–Kaliti sub-city, a spatial legacy of long-standing land-use policies prioritizing export-driven industrialization. This zone's development has been heavily influenced by its strategic location along major freight transport corridors, which, since the early 2000s, have anchored cement factories, metallurgical industries, and textile manufacturing hubs [49]. Weak regulatory oversight of zoning amplified risks by permitting informal settlements to proliferate near industrial developments, heightening community exposure to particulate emissions, a pattern corroborated by recent spatial autocorrelation analyses identifying Akaki–Kaliti as a pollution hotspot [49]. These systemic challenges are compounded by transportation frameworks prioritizing industrial logistics over environmental safeguards, transforming critical supply chain routes into localized $PM_{10}$ epicenters. Particulate spikes exceeding background levels within 500 m of these corridors are linked to unregulated emissions from aging diesel freight vehicles, exacerbated by the influx of outdated, high-emission trucks diverted from toll roads [41, 56]. Studies further attributed 12–18% of the $PM_{10}$ variability in this corridor to proximity to railways and

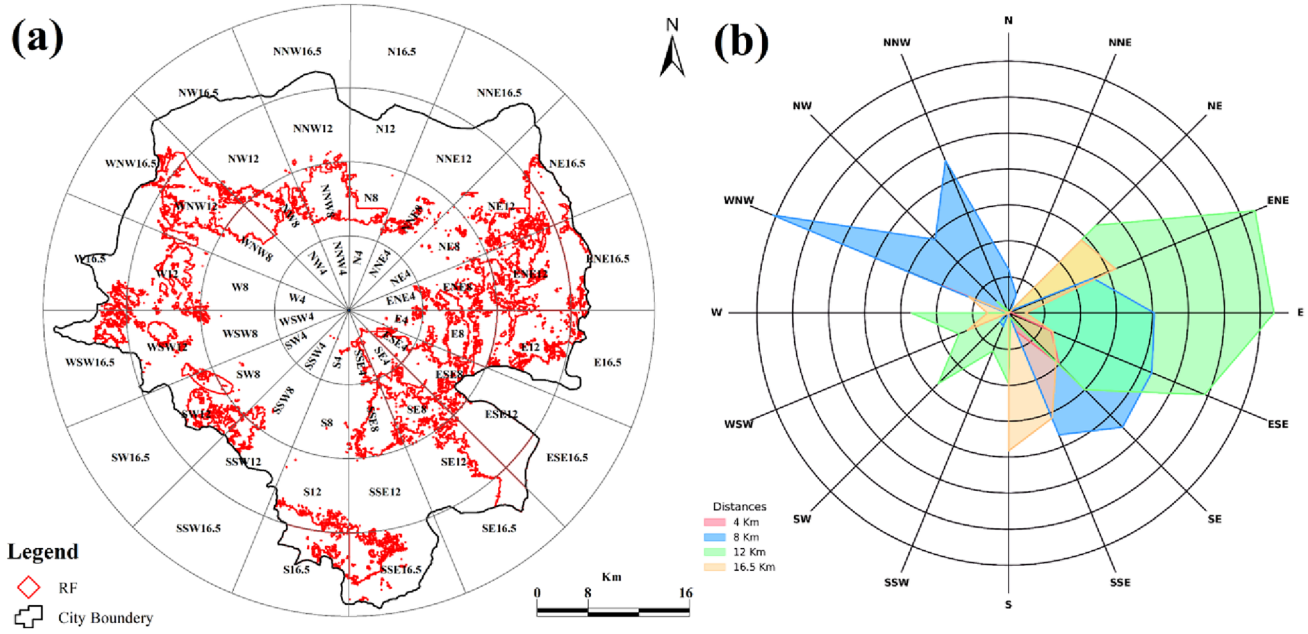



**Fig. 9** Hotspot analysis of $PM_{10}$ pollution from Addis Ababa city center: **a** Spatial distribution of $PM_{10}$ pollution hotspots categorized by concentric distance rings at 4 km, 8 km, 12 km, and 16.5 km, **b** Windrose diagram depicting the directional distribution of $PM_{10}$ pollution across 16 sectors

industrial clusters, with elevated trace metal concentrations aligning with vehicle brake wear and resuspended road dust [56].

The western corridors within the boundaries of the Nifas Silk-Lafto and Kolfe Keranio sub-cities face analogous pressures from construction-driven particulate emissions and pollution from the Koshe landfill (Fig. 10a, b), where decades of open waste burning and windborne debris persistently degrade air quality [148, 149]. These spatial disparities reflect entrenched socioeconomic inequities, as evidenced by recent studies linking $PM_{10}$ hotspots to informal settlements and industrial-residential overlaps [40, 49]. In Akaki-Kaliti, 62% of households rely on biomass fuels due to energy poverty, creating indoor—outdoor $PM_{10}$ feedback loops where residential cooking emissions intersect with industrial particulate dispersion [150]. Concurrently, informal enterprises, auto workshops, brick kilns, and small-scale recycling units operating without emission controls introduce granular pollution sources that are often omitted from conventional monitoring frameworks [148, 149, 151]. Such dynamics underscore how $PM_{10}$ hotspots emerge not solely from industrial activity but also from systemic policy gaps, including inadequate bypass roads funneling freight traffic through residential zones and unplanned settlements lacking green buffers [152, 153].

To address these challenges, interventions must align with geospatial and socioeconomic realities. In the eastern industrial corridor, retrofitting electrostatic precipitators (ESPs) in cement plants and enforcing real-time emission monitoring standards should be prioritized, as these measures optimize performance through upgraded components and compliance with dust resistivity requirements [154, 155]. Concurrently, land-use reforms such as mandating 300-m green buffers between industrial and residential zones and relocating nonstrategic industries to zones equipped with advanced filtration systems are critical, as evidenced by spatial analyses of the environmental impacts of industrial relocation [156]. These reforms align with China's 14th Five-Year Plan (2021–2025), which emphasizes emission reductions and green manufacturing systems [157].

In the western corridor, scaling the reppie waste-to-energy plant's success, which converts 1,400 tons/day of waste into 25 MW of electricity, requires expanded waste segregation protocols and geotextile covers to suppress fugitive dust. Furthermore, stringent regulation of enclosed debris transport and mandatory water spraying at construction sites is essential, mirroring strategies to maintain ESP efficiency in high-resistivity dust conditions [158–160]

Additionally, community-level strategies are critical to bridging the gap between policy implementation and equity in air quality initiatives. For example, in regions with high biomass dependency, such as sub-Saharan Africa, microfinance programs supporting the adoption of clean cookstoves have gained traction, aligning with global efforts to reduce household air pollution linked to 3 million premature deaths annually [161]. These initiatives are bolstered by industrial policies such as China's 2023–2025 air quality action plan, which prioritizes green innovation, energy efficiency, and the




**Fig. 10** Snapshots of key sources of $PM_{10}$ pollution in Addis Ababa: **a** Side and **b** top views of dry waste landfills (Reppie waste-to-energy plant and Koshe landfill located in southwestern Addis Ababa), **c** industrial emissions from a glass factory emitting smog in western Addis Ababa, and **d** transportation-related air pollution from vehicle emissions on the city's roads in eastern Addis Ababa

phaseout of outdated industrial capacities to curb pollutants such as $PM_{10}$ [157]. Concurrently, the deployment of low-cost air quality sensor networks, highlighted in recent studies as vital for addressing global monitoring gaps, enables real-time data sharing via public dashboards. This empowers residents to engage in participatory mitigation efforts, particularly in underserved communities where countries lack government-level air quality monitoring infrastructure [162]. Integrating these approaches with localized vulnerability assessments can ensure that interventions are both equitable and responsive to the unique challenges of high-exposure areas.

While this study provides actionable insights via annual aggregated $PM_{10}$ data, temporal variations, particularly seasonal fluctuations in rainfall, temperature, and emission patterns, merit further investigation. Although current data limitations preclude detailed seasonal analysis, future research incorporating temporal resolution could refine mitigation strategies. Additionally, the transferability of the RF model requires validation in diverse urban contexts beyond Addis Ababa, necessitating region-specific predictors to ensure broader applicability. Although comprehensive cost–benefit analyses remain outstanding, preliminary evidence suggests that interventions such as emission control retrofits and real-time air quality monitoring systems hold promise for long-term economic and public health gains through reduced $PM_{10}$-related morbidity. These benefits, however, must contend with persistent implementation challenges, including fiscal limitations, regulatory stagnation, and community resistance barriers, as corroborated by recent institutional analyses of Addis Ababa's environmental governance frameworks. Addressing these obstacles will require coordinated policy reforms, innovative financing mechanisms, and stakeholder engagement to align short-term economic priorities with long-term environmental health imperatives [45, 48, 163–166].

## 4 Conclusion

This study used a robust array of geospatial variables to assess the efficacy of three MLAs—RF, KNN, and NB, in forecasting $PM_{10}$ concentrations within Addis Ababa, Ethiopia. The study incorporated 25 variables, with a keen focus on their significance. Variable importance analysis across the models shed light on the principal factors influencing $PM_{10}$ levels. Factors related to proximity, such as distance from roads, industrial zones, bus stations, and significant emission sources, were identified as pivotal determinants, highlighting the profound influence of urban infrastructure and activities on air quality. Conversely, terrain features were found to have minimal impact. The comparative analysis underscored the superior predictive accuracy of RF MLA, with an accuracy of 0.95. Its ability to navigate nonlinear dynamics, intricate variable interplay, and resilience to overfitting and data irregularities accounts for exceptional performance. Spatial prediction indicated that Moderate $PM_{10}$ levels predominated across Addis Ababa, with the NB model estimating 62.8% of the area falling within this range.

Conversely, the KNN model's projections are particularly alarming, indicating that 56.41% of the area has $PM_{10}$ concentrations that are UHealSesn, highlighting an urgent call for intervention. The pinpointed hotspots, primarily within industrial areas, transport hubs, and urban development zones, are located mainly in the eastern (E, ENE) and western (NW, WNW) sectors and extend 8–16.5 km from the city center. The findings underscore the significant promise of MLAs combined with geospatial data in crafting cutting-edge $PM_{10}$ mapping and assessment tools. Using the predictive capabilities of the RF model and the rich tapestry of geospatial insights, this approach offers substantial contributions to formulating effective strategies to address the critical issue of air pollution in fast-growing African metropolises such as Addis Ababa. Pinpointing areas with the highest $PM_{10}$ levels and potential hotspots can facilitate targeted measures, including stricter emission regulations, sustainable urban development, and comprehensive public awareness initiatives, all of which are vital in reducing the detrimental health effects of air pollution on vulnerable populations.

**Author contributions** K.H.Y: Writing –original draft, Writing –review & editing, Visualization, Validation, Software, Resources, Methodology, Formal analysis, Conceptualization; M.I.Y: Conceptualization; Writing—review, and editing, Resources; A.D.I: Resources, Conceptualization; Writing-review and editing; T.B.G: Methodology, writing-review and editing, D.T: Writing-review and editing, Resources, E.K: Writing-review and editing, Resources.

**Funding** This study received no external funding.

**Data availability** Data will be made available on request.

## Declarations

**Ethics approval and consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/.

## References

1. Vilcins D, Christofferson RC, Yoon J-H, Nazli SN, Sly PD, Cormier SA, Shen G. Updates in air pollution: current research and future challenges. Ann Glob Heal. 2024. https://doi.org/10.5334/aogh.4363.
2. Wang H, Lv S, Diao Z, Wang B, Zhang H, Yu C. Study on sandstorm PM10 exposure assessment in the large-scale region: a case study in inner Mongolia. Environ Sci Pollut Res. 2018;25:17144–55. https://doi.org/10.1007/s11356-018-1841-5.
3. WHO, world health organization WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide; Geneva, 2021; ISBN 9789240034228.
4. Orellano P, Kasdagli M-I, Pérez Velasco R, Samoli E. Long-term exposure to particulate matter and mortality: an update of the WHO global air quality guidelines systematic review and meta-analysis. Int J Public Health. 2024. https://doi.org/10.3389/ijph.2024.1607683.
5. Losacco C, Perillo A. Particulate matter air pollution and respiratory impact on humans and animals. Environ Sci Pollut Res. 2018;25:33901–10. https://doi.org/10.1007/s11356-018-3344-9.
6. Kumar N, Hamzah FM, Diantoro M, Muhd Zailani NA. Suman physiochemical characterization of ambient PM10 and PM2.5 in an urban environment. Curr Appl Phys. 2025;71:57–69. https://doi.org/10.1016/j.cap.2024.12.006.
7. Tornevi A, Olstrup H, Forsberg B. Short-term associations between PM10 and respiratory health effects in Visby, Sweden. Toxics. 2022;10:333. https://doi.org/10.3390/toxics10060333.
8. HEI; IHME Health Effects Institute. 2024. State of Global Air 2024. Special Report.; 2024;
9. HEI; IHME The state of air quality and health impacts in Africa. 2022, 23.
10. Croitoru L, Miranda JJ, Khattabi A, Lee JJ. The cost of coastal zone degradation in Nigeria. Cost Coast Zo Degrad Niger. 2020. https://doi.org/10.1596/34758.
11. Croitoru L, Chang JC, Kelly A. The cost of air pollution in Lagos; 2020;
12. HEI, H.E.I. *Trusted Science for a Changing World*; Boston, 2021;
13. World Bank Group; IHME The cost of air pollution: strengthening the economic case for action. World Bank Inst. Heal. Metrics Eval Univ Washington, Seattle 2016, 1–102.
14. Hoffer A, Jancsek-Turóczi B, Tóth Á, Kiss G, Naghiu A, Levei EA, Marmureanu L, Machon A, Gelencsér A. Emission factors for PM10 and polycyclic aromatic hydrocarbons (PAHs) from illegal burning of different types of municipal waste in households. Atmos Chem Phys. 2020;20:16135–44. https://doi.org/10.5194/acp-20-16135-2020.
15. Alyousifi Y, Othman M, Husin A, Rathnayake U. A New hybrid fuzzy time series model with an application to predict PM10 concentration. Ecotoxicol Environ Saf. 2021;227: 112875. https://doi.org/10.1016/j.ecoenv.2021.112875.
16. Saraswat I, Mishra RK, Kumar A. Estimation of PM10 concentration from landsat 8 OLI satellite imagery over Delhi, India. Remote Sens Appl Soc Environ. 2017;8:251–7. https://doi.org/10.1016/j.rsase.2017.10.006.
17. Imani M. Particulate matter (PM2.5 and PM10) generation map using MODIS level-1 satellite images and deep neural network. J Environ Manage. 2021;281:111888. https://doi.org/10.1016/j.jenvman.2020.111888.
18. Tella A, Balogun A-L. GIS-based air quality modelling: spatial prediction of PM10 for selangor state, malaysia using machine learning algorithms. Environ Sci Pollut Res. 2022;29:86109–25. https://doi.org/10.1007/s11356-021-16150-0.
19. Pedde M, Kloog I, Szpiro A, Dorman M, Larson TV, Adar SD. Estimating long-term PM10–25 concentrations in six US cities using satellite-based aerosol optical depth data. Atmos Environ. 2022;272:118945. https://doi.org/10.1016/j.atmosenv.2022.118945.
20. Yao L, Lu N. Particulate matter pollution and population exposure assessment over Mainland China in 2010 with remote sensing. Int J Environ Res Public Health. 2014;11:5241–50. https://doi.org/10.3390/ijerph110505241.
21. Shaziayani WN, Ul-Saufie AZ, Ahmat H, Al-Jumeily D. Coupling of quantile regression into boosted regression trees (BRT) technique in forecasting emission model of PM10 concentration. Air Qual Atmos Heal. 2021;14:1647–63. https://doi.org/10.1007/s11869-021-01045-3.
22. AlThuwaynee OF, Kim SW, Najemaden MA, Aydda A, Balogun AL, Fayyadh MM, Park HJ. Demystifying uncertainty in PM10 susceptibility mapping using variable drop-off in extreme-gradient boosting (XGB) and random forest (RF) algorithms. Environ Sci Pollut Res. 2021;28:43544–66. https://doi.org/10.1007/s11356-021-13255-4.

23. Ren X, Mi Z, Georgopoulos PG. Comparison of machine learning and land use regression for fine scale spatiotemporal estimation of ambient air pollution: modeling ozone concentrations across the contiguous United States. Environ Int. 2020;142: 105827. https://doi.org/10.1016/j.envint.2020.105827.
24. Bozdağ A, Dokuz Y, Gökçek ÖB. Spatial prediction of PM10 concentration using machine learning algorithms in Ankara, Turkey. Environ Pollut. 2020;263: 114635. https://doi.org/10.1016/j.envpol.2020.114635.
25. Lin L, Liang Y, Liu L, Zhang Y, Xie D, Yin F, Ashraf T. Estimating PM2.5 concentrations using the machine learning RF-XGBoost model in Guanzhong Urban Agglomeration, China. Remote Sens. 2022;14:5239. https://doi.org/10.3390/rs14205239.
26. Mampitiya L, Rathnayake N, Leon LP, Mandala V, Azamathulla HM, Shelton S, Hoshino Y, Rathnayake U. Machine learning techniques to predict the air quality using meteorological data in two urban areas in Sri Lanka. Environments. 2023;10:141. https://doi.org/10.3390/environments10080141.
27. Tella A, Balogun A-L, Adebisi N, Abdullah S. Spatial Assessment of PM10 Hotspots Using Random Forest, K-Nearest Neighbour and Naïve Bayes. Atmos Pollut Res. 2021;12:1–12. https://doi.org/10.1016/j.apr.2021.101202.
28. Houdou A, Khomsi K, Monache LD, Hu W, Boutayeb S, Belyamani L, Abdulla F, Al-Delaimy WK, Khalis M. Predicting particulate matter (PM10) levels in Morocco: a 5-day forecast using the analog ensemble method. 2024, 46, 151
29. Morapedi TD, Obagbuwa IC. Air pollution particulate matter (PM2.5) prediction in South African cities using machine learning techniques. Front Artif Intell. 2023. https://doi.org/10.3389/frai.2023.1230087.
30. Govindasamy TR, Chetty N. machine learning models to quantify the influence of PM10 aerosol concentration on global solar radiation prediction in South Africa. Clean Eng Technol. 2021;2: 100042. https://doi.org/10.1016/j.clet.2021.100042.
31. Yang R, Yin L, Hao X, Liu L, Wang C, Li X, Liu Q. Identifying a suitable model for predicting hourly pollutant concentrations by using low-cost microstation data and machine learning. Sci Rep. 2022;12:19949. https://doi.org/10.1038/s41598-022-24470-5.
32. Okorn K, Iraci LT. An overview of outdoor low-cost gas-phase air quality sensor deployments: current efforts, trends, and limitations. Atmos Meas Tech. 2024;17:6425–57. https://doi.org/10.5194/amt-17-6425-2024.
33. Carotenuto F, Bisignano A, Brilli L, Gualtieri G, Giovannini L. Low-cost air quality monitoring networks for long-term field campaigns: a review. Meteorol Appl. 2023. https://doi.org/10.1002/met.2161.
34. Raheja G, Nimo J, Appoh EK-E, Essien B, Sunu M, Nyante J, Amegah M, Quansah R, Arku RE, Penn SL, et al. Low-cost sensor performance intercomparison, correction factor development, and 2+ years of ambient PM 2.5 monitoring in Accra, Ghana. Environ Sci Technol. 2023;57:10708–20. https://doi.org/10.1021/acs.est.2c09264.
35. Shiferaw AB, Kumie A, Tefera W. The spatial and temporal variation of fine particulate matter pollution in ethiopia: data from the atmospheric composition analysis group (1998–2019). PLoS ONE. 2023;18: e0283457. https://doi.org/10.1371/journal.pone.0283457.
36. Amare AN, Sorsa S, Gebremariam Z. Levels and health risk assessments of particulate matter and inorganic gaseous pollutants in urban and industrial areas of Hawassa City. Ethiopia Heliyon. 2024;10: e33286. https://doi.org/10.1016/j.heliyon.2024.e33286.
37. Baruah A, Bousiotis D, Damayanti S, Bigi A, Ghermandi G, Ghaffarpasand O, Harrison RM, Pope FD. A novel spatiotemporal prediction approach to fill air pollution data gaps using mobile sensors, machine learning and citizen science techniques. NPJ Clim Atmos Sci. 2024;7:310. https://doi.org/10.1038/s41612-024-00859-z.
38. Hua V, Nguyen T, Dao M-S, Nguyen HD, Nguyen BT. The impact of data imputation on air quality prediction problem. PLoS ONE. 2024;19: e0306303. https://doi.org/10.1371/journal.pone.0306303.
39. Imam M, Adam S, Dev S, Nesa N. Air quality monitoring using statistical learning models for sustainable environment. Intell Syst with Appl. 2024;22: 200333. https://doi.org/10.1016/j.iswa.2024.200333.
40. Sanbata H, Asfaw A, Kumie A. Indoor air pollution in slum neighbourhoods of Addis Ababa. Ethiopia Atmos Environ. 2014;89:230–4. https://doi.org/10.1016/j.atmosenv.2014.01.003.
41. Flanagan E, Mattisson K, Walles J, Abera A, Eriksson A, Balidemaj F, Oudin A, Isaxon C, Malmqvist E. Air pollution and urban green space: evidence of environmental injustice in Adama, Ethiopia. Front Sustain Cities. 2021. https://doi.org/10.3389/frsc.2021.728384.
42. Yasin KH, Iguala AD, Gelete TB. Spatiotemporal analysis of urban expansion and its impact on farmlands in the Central Ethiopia Metropolitan Area. Discov Sustain. 2025;6:36. https://doi.org/10.1007/s43621-024-00749-7.
43. Kumie A. Air pollution in Ethiopia: in-door air pollution in rural butajira and traffic air pollution in Addis Ababa, Addis Ababa, 2009.
44. Roychowdhury A, Chandola P, Vivek C. Urban air quality management in Ethiopia: a guidance framework; New Delhi, 2016;
45. Tsegaye D. Ambient air pollution status of Addis Ababa City; the case of selected roadside. Am J Environ Prot. 2019;8:39. https://doi.org/10.11648/j.ajep.20190802.11.
46. Amare, M. Addis Ababa City Is Ranked 6th among the Cities with Air Pollution. Available online: https://hagerie.tv/home/detail/1750/ethiopia#:~:text=Addis Ababa city is ranked 6th among,Ababa city is ranked 6th among the. Accessed 7 Feb 2024.
47. Weldegebriel AT, Tekalign M, Van Rompaey A. Socio-spatial analysis of regime shifts in Addis Ababa's urbanisation. Appl Geogr. 2023;154: 102918. https://doi.org/10.1016/j.apgeog.2023.102918.
48. AAEPGDC, (Addis Ababa Environmental Protection Green Development Commission) ADDIS ABABA CITY AIR QUALITY MANAGEMENT PLAN (2021 to 2025); Addis Ababa, Ethiopia, 2021
49. Mulgeta D, Gotu B, Temesgen S, Belina M, Likassa HT, Tsegaye D. Statistical analysis of spatial distribution of ambient air pollution in Addis Ababa, Ethiopia. Stoch Environ Res Risk Assess. 2024. https://doi.org/10.1007/s00477-024-02748-6.
50. Mekonnen EN, Fetene A, Gebremariam E. Grid-based climate variability analysis of Addis Ababa, Ethiopia. Heliyon. 2024;10: e27116. https://doi.org/10.1016/j.heliyon.2024.e27116.
51. Mekonnen EN, Gebremariam E, Fetene A, Damene S. Remote sensing-based spatio-temporal rainfall variability analysis: the case of Addis Ababa City. Ethiopia Appl Geomatics. 2024;16:365–85. https://doi.org/10.1007/s12518-024-00554-x.
52. NMA, N.M.A. of E. Climate of City: Addis Ababa.; Addis Ababa, Ethiopia, 2016;
53. Jida SN, Jean-François H, Chesse P. Artificial neural network modelling to predict PM2.5 and PM10 exhaust emissions from on-road vehicles in Addis Ababa, Ethiopia. Environ Pollut Clim Chang. 2022;6:279.
54. Gebre G, Feleke Z, Sahle-Demissie E. Mass concentrations and elemental composition of urban atmospheric aerosols in Addis Ababa. Ethiopia Bull Chem Soc Ethiop. 2010. https://doi.org/10.4314/bcse.v24i3.60787.

55. Kumar P, Hama S, Abbass RA, Nogueira T, Brand VS, Wu H-W, Abulude FO, Adelodun AA, Anand P, Andrade MF, et al. In-kitchen aerosol exposure in twelve cities across the globe. Environ Int. 2022;162:107155. https://doi.org/10.1016/j.envint.2022.107155.
56. Embiale A, Zewge F, Chandravanshi BS, Sahle-Demessie E. Levels of trace elements in PM10 collected at roadsides of Addis Ababa, Ethiopia, and exposure risk assessment. Environ Monit Assess. 2019;191:397. https://doi.org/10.1007/s10661-019-7503-3.
57. Sricharoenvech P, Lai A, Tefera W, Kumie A, Berhane K, Gilliland F, Samet J, Patz J, Schauer JJ. Reactive oxygen species (ROS) activity of fine particulate matter health impacts in Addis Ababa. Ethiopia Atmos Pollut Res. 2021;12: 101149. https://doi.org/10.1016/j.apr.2021.101149.
58. World Bank Group, W. Population Growth (Annual %) https://data.worldbank.org/indicator/SP.POP.GROW?locations=ET, https://worldpopulationreview.com/world-cities/addis-ababa-population,.
59. Etyemezian V, Tesfaye M, Yimer A, Chow JC, Mesfin D, Nega T, Nikolich G, Watson JG, Wondmagegn M. Results from a pilot-scale air quality study in Addis Ababa, Ethiopia. Atmos Environ. 2005;39:7849–60. https://doi.org/10.1016/j.atmosenv.2005.08.033.
60. NMA, N.M.A. of E. Long-Term Weather Forecast Addis Ababa, Ethiopia. Ethiopia National Meteorological Agency (Weather Atlas) https://www.weather-atlas.com/en/ethiopia/addis-ababa-long-term-weather-forecast.
61. Tkachenko NV. Optical Spectroscopy; Elsevier, 2006; ISBN 9780444521262.
62. VanCuren RA, Cliff SS, Perry KD, Jimenez-Cruz M. Asian continental aerosol persistence above the marine boundary layer over the Eastern North Pacific: continuous aerosol measurements from intercontinental transport and chemical transformation 2002 (ITCT 2K2). J Geophys Res Atmos. 2005. https://doi.org/10.1029/2004JD004973.
63. Bivand RS, Pebesma E, Gómez-Rubio V. Applied Spatial Data Analysis with R; Springer New York: New York, NY, 2013; ISBN 978–1–4614–7617–7.
64. Anselin L. Local indicators of spatial association—LISA. Geogr Anal. 1995;27:93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x.
65. Griffith DA, Plant RE. Statistical analysis in the presence of spatial autocorrelation: selected sampling strategy effects. Stats. 2022;5:1334–53. https://doi.org/10.3390/stats5040081.
66. Anselin L, Li X. Tobler's law in a multivariate world. Geogr Anal. 2020;52:494–510. https://doi.org/10.1111/gean.12237.
67. ESRI How Spatial Autocorrelation (Global Moran's I) Works https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm.
68. Choubin B, Abdolshahnejad M, Moradi E, Querol X, Mosavi A, Shamshirband S, Ghamisi P. Spatial Hazard assessment of the PM10 using machine learning models in Barcelona. Spain Sci Total Environ. 2020;701:2–11. https://doi.org/10.1016/j.scitotenv.2019.134474.
69. Adams MD, Massey F, Chastko K, Cupini C. Spatial modelling of particulate matter air pollution sensor measurements collected by community scientists while cycling, land use regression with spatial cross-validation, and applications of machine learning for data correction. Atmos Environ. 2020;230: 117479. https://doi.org/10.1016/j.atmosenv.2020.117479.
70. Kleine DJ, Zalakeviciute R, Gonzalez M, Rybarczyk Y. Modeling PM 2.5 urban pollution using machine learning and selected meteorological parameters. J Elect Comput Eng. 2017;2017:1–14. https://doi.org/10.1155/2017/5106045.
71. Usmani RSA, Saeed A, Abdullahi AM, Pillai TR, Jhanjhi NZ, Hashem IAT. Air pollution and its health impacts in Malaysia: a review. Air Qual Atmos Heal. 2020;13:1093–118. https://doi.org/10.1007/s11869-020-00867-x.
72. Ganguly R, Sharma D, Kumar P. Trend analysis of observational PM10 concentrations in Shimla City. India Sustain Cities Soc. 2019;51:101719. https://doi.org/10.1016/j.scs.2019.101719.
73. Böhner J, Antonić O. Land-surface parameters specific to topo-climatology. Dev Soil Sci. 2009;33:195–226. https://doi.org/10.1016/S0166-2481(08)00008-1.
74. Li C, Du S, Bai Z, Shao-fei K, Yan Y, Bin H, Dao-wen H, Li Z. Application of land use regression for estimating concentrations of major outdoor air pollutants in Jinan, China. J Zhejiang Univ A. 2010;11:857–67. https://doi.org/10.1631/jzus.A1000092.
75. Wang Y, He L. Effect of rainfall intensity on PM 10 and PM 2.5 scavenging in Guangzhou at night. IOP Conf Ser Earth Environ Sci. 2020;450:012075. https://doi.org/10.1088/1755-1315/450/1/012075.
76. Birim NG, Turhan C, Atalay AS, Gokcen Akkurt G. The influence of meteorological parameters on PM10: a statistical analysis of an urban and rural environment in Izmir/Türkiye. Atmosphere (Basel). 2023;14:421. https://doi.org/10.3390/atmos14030421.
77. Zolfaghari Nikanjam S, Khalesifard HR, Abedini Y. Estimation of dust downfall time in dusty days using the correlation between PM10 and sunphotometer data. Atmos Meas Tech Discuss. 2017, 1–8.
78. Xu C, Zhao J, Liu P. A geographically weighted regression approach to investigate the effects of traffic conditions and road characteristics on air pollutant emissions. J Clean Prod. 2019. https://doi.org/10.1016/j.jclepro.2019.118084.
79. Singh A, Ng'ang'a D, Gatari MJ, Kidane AW, Alemu ZA, Derrick N, Webster MJ, Bartington SE, Thomas GN, Avis W, et al. Air quality assessment in three east african cities using calibrated low-cost sensors with a focus on road-based hotspots. Environ Res Commun. 2021;3:075007. https://doi.org/10.1088/2515-7620/ac0e0a.
80. Dewangan A, Mallick A, Yadav AK, Kumar R. Combustion-generated pollutions and strategy for its control in CI engines: a review. Mater Today Proc. 2020;21:1728–33. https://doi.org/10.1016/j.matpr.2019.12.155.
81. Tu M, Olofsson U. PM10 in underground stations from different types of trains. Transp Res Part D Transp Environ. 2021;95: 102867. https://doi.org/10.1016/j.trd.2021.102867.
82. Ngoc L, Kim M, Bui V, Park D, Lee Y-C. Particulate matter exposure of passengers at bus stations: a review. Int J Environ Res Public Health. 2018;15:2886. https://doi.org/10.3390/ijerph15122886.
83. Irga PJ, Burchett MD, Torpy FR. Does urban forestry have a quantitative effect on ambient air quality in an urban environment? Atmos Environ. 2015;120:173–81. https://doi.org/10.1016/j.atmosenv.2015.08.050.
84. Zhou M, Huang Y, Li G. Changes in the concentration of air pollutants before and after the COVID-19 blockade period and their correlation with vegetation coverage. Environ Sci Pollut Res. 2021;28:23405–19. https://doi.org/10.1007/s11356-020-12164-2.
85. Abdul-Halim ND, Latif MT, Mohamed AF, Maulud KNA, Idrus S, Azhari A, Othman M, Sofwan NM. Spatial assessment of land use impact on air quality in mega urban regions, Malaysia. Sustain Cities Soc. 2020;63: 102436. https://doi.org/10.1016/j.scs.2020.102436.
86. Kalisa E, Fadlallah S, Amani M, Nahayo L, Habiyaremye G. Temperature and air pollution relationship during heatwaves in Birmingham, UK. Sustain Cities Soc. 2018;43:111–20. https://doi.org/10.1016/j.scs.2018.08.033.

87. Misra P, Takeuchi W. Analysis of air quality and nighttime light for indian urban regions. IOP Conf Ser Earth Environ Sci. 2016;37: 012077. https://doi.org/10.1088/1755-1315/37/1/012077.
88. Balogun AL, Yekeen ST, Pradhan B, Althuwaynee OF. Spatio-temporal analysis of oil spill impact and recovery pattern of coastal vegetation and wetland using multispectral satellite landsat 8-oli imagery and machine learning models. Remote Sens. 2020. https://doi.org/10.3390/rs12071225.
89. Bisht DS, Tiwari S, Dumka UC, Srivastava AK, Safai PD, Ghude SD, Chate DM, Rao PSP, Ali K, Prabhakaran T, et al. Tethered balloon-born and ground-based measurements of black carbon and particulate profiles within the lower troposphere during the foggy period in Delhi, India. Sci Total Environ. 2016;573:894–905. https://doi.org/10.1016/j.scitotenv.2016.08.185.
90. Choi JE, Lee H, Song J. Forecasting daily PM10 concentrations in seoul using various data mining techniques. Commun Stat Appl Methods. 2018;25:199–215. https://doi.org/10.29220/CSAM.2018.25.2.199.
91. Wu L, Xie B, Wang W. Quantifying the impact of terrain–wind–governed close-effect on atmospheric polluted concentrations. J Clean Prod. 2022;367: 132995. https://doi.org/10.1016/j.jclepro.2022.132995.
92. Dodson RE, Andres Houseman E, Morin B, Levy JI. An analysis of continuous black carbon concentrations in proximity to an airport and major roadways. Atmos Environ. 2009;43:3764–73. https://doi.org/10.1016/j.atmosenv.2009.04.014.
93. Wolfe PJ, Yim SHL, Lee G, Ashok A, Barrett SRH, Waitz IA. Near-airport distribution of the environmental costs of aviation. Transp Policy. 2014;34:102–8. https://doi.org/10.1016/j.tranpol.2014.02.023.
94. Basahi JM, Ismail IM, Hussain HK, El Behaedi EH, Hassan IA. Air pollution assessment in fuel stations and its impact on workers ' health : a case study from Jeddah. KSA Adv Environ Biol. 2015;9:326–31.
95. Malavolti M, Malagoli C, Filippini T, Wise LA, Bellelli A, Palazzi G, Cellini M, Costanzini S, Teggi S, Vinceti M. Residential proximity to petrol stations and risk of childhood leukemia. Eur J Epidemiol. 2023;38:771–82. https://doi.org/10.1007/s10654-023-01009-0.
96. Bakamwesiga H, Mugisha W, Kisira Y, Muwanga A. An assessment of air and water pollution accrued from stone quarrying in Mukono District, Central Uganda. J Geosci Environ Prot. 2022;10:25–42. https://doi.org/10.4236/gep.2022.105003.
97. Nartey VK, Nanor JN, Klake RK. Effects of quarry activities on some selected communities in the lower Manya Krobo District of the Eastern Region of Ghana. Atmos Clim Sci. 2012;02:362–72. https://doi.org/10.4236/acs.2012.23032.
98. Yan H, Li Q, Feng K, Zhang L. The characteristics of pm emissions from construction sites during the earthwork and foundation stages: an empirical study evidence. Environ Sci Pollut Res. 2023;30:62716–32. https://doi.org/10.1007/s11356-023-26494-4.
99. Bruce K, Lam J, Lau S, Quinn H, Taylor K. PM10 Implementation maintenance plan and redesignation request for sacramento County; Sacramento, CA, 2010;
100. Son Y, Osornio-Vargas ÁR, O'Neill MS, Hystad P, Texcalac-Sangrador JL, Ohman-Strickland P, Meng Q, Schwander S. Land use regression models to assess air pollution exposure in Mexico City using finer spatial and temporal input parameters. Sci Total Environ. 2018;639:40–8. https://doi.org/10.1016/j.scitotenv.2018.05.144.
101. Xu G, Jiao L, Zhao S, Yuan M, Li X, Han Y, Zhang B, Dong T. Examining the impacts of land use on air quality from a spatio-temporal perspective in Wuhan, China. Atmosphere (Basel). 2016;7:62. https://doi.org/10.3390/atmos7050062.
102. Kuhn M, Johnson K. Applied Predictive Modeling; Springer New York: New York, NY, 2013; ISBN 978-1-4614-6848-6.
103. Davidson L, Kline K ,Klein S, Windisch K. The normalization process. In: Pro SQL Server 2008 relational database design and implementation; Apress: Berkeley, CA, 2008; pp. 117–175.
104. Breiman L. Random forests. Mach Learn. 2001;45:5–32. https://doi.org/10.1023/A:1010933404324.
105. Abu El-Magd S, Soliman G, Morsy M, Kharbish S. Environmental hazard assessment and monitoring for air pollution using machine learning and remote sensing. Int J Environ Sci Technol. 2023;20:6103–16. https://doi.org/10.1007/s13762-022-04367-6.
106. Mishra M, Guria R, Baraj B, Nanda AP, Santos CAGRM. Spatial analysis and machine learning prediction of forest fire susceptibility: a comprehensive approach for effective management and mitigation. Sci Total Environ. 2024;926: 171713. https://doi.org/10.1016/j.scitotenv.2024.171713.
107. Guria R, Mishra M, Silva RM. Predicting forest fire probability in similipal biosphere reserve (India) using sentinel-2 MSI data and machine learning. Remote Sens Appl Soc Environ. 2024;36: 101311. https://doi.org/10.1016/j.rsase.2024.101311.
108. Belgiu M, Drăguţ L. Random forest in remote sensing: a review of applications and future directions. ISPRS J Photogramm Remote Sens. 2016;114:24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.
109. Breiman L, Cutler A. Breiman and Cutler's random forests for classification and regression; 2022;
110. Yao Z, Ruzzo WL. A regression-based K nearest neighbor algorithm for gene function prediction from heterogeneous data. BMC Bioinformatics. 2006;7:S11. https://doi.org/10.1186/1471-2105-7-S1-S11.
111. Gou J, Qiu W, Yi Z, Shen X, Zhan Y, Ou W. Locality constrained representation-based K-nearest neighbor classification. Knowledge-Based Syst. 2019;167:38–52. https://doi.org/10.1016/j.knosys.2019.01.016.
112. Hu L-Y, Huang M-W, Ke S-W, Tsai C-F. The distance function effect on k-nearest neighbor classification for medical datasets. Springerplus. 2016;5:1304. https://doi.org/10.1186/s40064-016-2941-7.
113. Fix E, Hodges JL. Discriminatory analysis. Nonparametric discrimination: consistency properties. Int Stat Rev. 1989;57:238. https://doi.org/10.2307/1403797.
114. Aini N, Mustafa MS. Data mining approach to predict air pollution in makassar. In: Proceedings of the 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS); IEEE, October 27 2020; 1–5.
115. Shahabi H, Shirzadi A, Ghaderi K, Omidvar E, Al-Ansari N, Clague JJ, Geertsema M, Khosravi K, Amini A, Bahrami S, et al. Flood detection and susceptibility mapping using sentinel-1 remote sensing data and a machine learning approach: hybrid intelligence of bagging ensemble based on K-nearest neighbor classifier. Remote Sens. 2020;12:266. https://doi.org/10.3390/rs12020266.
116. Gore RW, Deshpande DS. An approach for classification of health risks based on air quality levels. In: Proceedings of the 2017 1st International Conference on Intelligent Systems and Information Management (ICISIM); IEEE, October 2017; 58–61.
117. Rish I. An empirical study of the naive bayes classifier. IJCAI 2001 Work. Empir. methods Artif. Intell. 2001, 41–46.
118. Wang H, Wang H, Wu Z, Zhou Y. Using multi-factor analysis to predict urban flood depth based on Naive Bayes. Water. 2021;13:432. https://doi.org/10.3390/w13040432.

119. Chen G, Li S, Knibbs LD, Hamm NAS, Cao W, Li T, Guo J, Ren H, Abramson MJ, Guo Y. A Machine learning method to estimate PM2.5 concentrations across china with remote sensing, meteorological and land use information. Sci Total Environ. 2018;636:52–60. https://doi.org/10.1016/j.scitotenv.2018.04.251.
120. Xu J, Zhang Y, Miao D. Three-way confusion matrix for classification: a measure driven view. Inf Sci. 2020;507:772–94. https://doi.org/10.1016/j.ins.2019.06.064.
121. Lasisi A, Ghazali R, Herawan T. Application of real-valued negative selection algorithm to improve medical diagnosis. In: Applied Computing in Medicine and Health; Elsevier, 2016; pp. 231–243.
122. Iacovacci J, Palorini F, Cicchetti A, Fiorino C, Rancati T. Dependence of the AUC of NTCP models on the observational dose-range highlights cautions in comparison of discriminative performance. Phys Medica. 2023;113: 102654. https://doi.org/10.1016/j.ejmp.2023.102654.
123. Azmi SZ, Latif MT, Ismail AS, Juneng L, Jemain AA. Trend and status of air quality at three different monitoring stations in the Klang Valley, Malaysia. Air Qual Atmos Heal. 2010;3:53–64. https://doi.org/10.1007/s11869-009-0051-1.
124. Zhu Y, Chen Q, Li G, She J, Zhu Y, Sun W, Liu X, Wang Q. Source and health risk apportionment of PM10 based on heavy metals in a city on the edge of the Tengger Desert. Air Qual Atmos Heal. 2023;16:391–9. https://doi.org/10.1007/s11869-022-01279-9.
125. Moges GT, Alemu GT. Contribution of congested traffic flow condition to air pollution at intersections in Addis Ababa, Ethiopia. Heliyon. 2024;10: e36477. https://doi.org/10.1016/j.heliyon.2024.e36477.
126. Salah Eddine S, Drissi LB, Mejjad N, Mabrouki J, Romanov AA. Machine learning models application for spatiotemporal patterns of particulate matter prediction and forecasting over Morocco in North of Africa. Atmos Pollut Res. 2024;15:102239. https://doi.org/10.1016/j.apr.2024.102239.
127. Arowosegbe OO, Röösli M, Adebayo-Ojo TC, Dalvie MA, de Hoogh K. Spatial and temporal variations in PM10 concentrations between 2010–2017 in South Africa. Int J Environ Res Public Health. 2021;18:13348. https://doi.org/10.3390/ijerph182413348.
128. Sivakumar N, Mura C, Peirce SM. Innovations in integrating machine learning and agent-based modeling of biomedical systems. Front Syst Biol. 2022. https://doi.org/10.3389/fsysb.2022.959665.
129. Kujawska J, Kulisz M, Oleszczuk P, Cel W. Machine learning methods to forecast the concentration of PM10 in Lublin, Poland. Energies. 2022;15:6428. https://doi.org/10.3390/en15176428.
130. Sethi JK, Mittal M. Efficient weighted naive bayes classifiers to predict air quality index. Earth Sci Informatics. 2022;15:541–52. https://doi.org/10.1007/s12145-021-00755-7.
131. Son S, Kim J. Evaluation and predicting PM10 concentration using multiple linear regression and machine learning. Korean J Remote Sens. 2020;36:1711–20. https://doi.org/10.7780/kjrs.2020.36.6.3.7.
132. Nithya Shree R, Santhiya G, Bhargavi R. Assessment of spatial hazard and impact of pm10 using machine learning. 4th Int Conf Comput Commun Signal Process. ICCCSP 2020 2020, https://doi.org/10.1109/ICCCSP49186.2020.9315283.
133. Maheshwari K, Lamba S. Air quality prediction using supervised regression model. In: Proceedings of the 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT); IEEE, September 2019; 1–7.
134. Diriba DB, Meng XZ. Rethinking of the solid waste management system of Addis Ababa, Ethiopia. J Adv Environ Heal Res. 2021;9:7–22. https://doi.org/10.32598/JAEHR.9.1.1198.
135. Fitaw Y, Zenebre B. Assessment of the landfill situation of Addis Ababa city administration. Ethiop J Health Sci. 2003;13:125–30.
136. Kebede TA, Hailu BT, Suryabhagavan KV. Evaluation of spectral built-up indices for impervious surface extraction using sentinel-2A MSI imageries: a case of Addis Ababa City, Ethiopia. Environ Challenges. 2022;8: 100568. https://doi.org/10.1016/j.envc.2022.100568.
137. USAID Clean air catalyst: sources of air pollution: Addis Ababa; 2023;
138. Bikis A. Urban air pollution and greenness in relation to public health. J Environ Public Health. 2023;2023:1–18. https://doi.org/10.1155/2023/8516622.
139. Worku H. Environmental and socioeconomic impacts of cobblestone quarries in Addis Ababa and implication for resource use efficiency, environmental quality, and sustainability of land after-use. Environ Qual Manag. 2017;27:41–61. https://doi.org/10.1002/tqem.21524.
140. Renard X, Laugel T, Detyniecki M. Understanding prediction discrepancies in machine learning classifiers. 2021, 1–20, https://doi.org/10.48550/arXiv.2104.05467.
141. Avis W, Weldetinsae A, Singh A, Getaneh Z. Vulnerability scoping study: air pollution exposure and public transport in Addis Ababa; 2018; Vol. 4;.
142. Kume A, Charles K, Berehane Y, Anders E, Ali A. Magnitude and variation of traffic air pollution as measured by CO in the city of Addis Ababa, Ethiopia. Ethiop J Heal Dev. 2010;24:156–66. https://doi.org/10.4314/ejhd.v24i3.68379.
143. Mathew A, Gokul PR, Raja Shekar P, Arunab KS, Ghassan Abdo H, Almohamad H, AbdullahAlDughairi A. Air quality analysis and PM 2.5 modelling using machine learning techniques: a study of Hyderabad City in India. Cogent Eng. 2023. https://doi.org/10.1080/23311916.2023.2243743.
144. Wallace JM, Hobbs PV. Cloud microphysics. In Atmospheric Science; Elsevier, 2006; pp. 209–269.
145. Alok T. Urban air pollution caused by automobiles in Addis Ababa (Ethiopia ) and its health consequences. IUP J Environ Sci. 2012;3:49–58.
146. Hirpe L, Yeom C. Municipal solid waste management policies, practices, and challenges in Ethiopia: a systematic review. Sustainability. 2021;13:11241. https://doi.org/10.3390/su132011241.
147. Soruma M, Woldeamanuel M. The level of air quality at public transport stations: the case of Torhailoch-Ayat main road in Addis Ababa. J Transp Heal. 2022;24: 101328. https://doi.org/10.1016/j.jth.2021.101328.
148. Abebe MA. Challenges of waste to energy facility in Reppi (Koshe). Int Res J Pharm Med Sci. 2018;1:9–16.
149. Yimam YM, Tesfaye N. Environmental and health impacts of and challenges on Rappie Koshe waste-to-energy facility, Addis Ababa. Int Res J Multidiscip Scope. 2020;1:9–17. https://doi.org/10.47857/irjms.2020.v01si02.027.
150. Taye AE, Chandravanshi BS, Beshah FZ, Sahle-Demessie E. Elemental composition and health risk assessment of PM10, PM2.5, at different microenvironments: Addis Ababa. Ethiopia PLoS One. 2024;19:e0309995. https://doi.org/10.1371/journal.pone.0309995.
151. Sisay SF, Gari SR, Ambelu A. Solid waste management service chain and sanitation safety: a case study of existing practice in Addis Ababa, Ethiopia. Front Environ Eng. 2024. https://doi.org/10.3389/fenve.2024.1414669.
152. Jacobsen K, Hollander JB, Sternlieb SR, Yimere A, Naegele AC, Schwalm C. Urban policy responses to climate hazards in Addis Ababa, Ethiopia. City Environ Interact. 2024;24: 100162. https://doi.org/10.1016/j.cacint.2024.100162.

153. Mekonnen GB, dos Muchangos LS, Ito L, Tokai A. Evaluation of evolving waste management strategies in Addis Ababa City, Ethiopia: a life cycle assessment approach. In: EcoDesign for Sustainable Products, Services and Social Systems II; Springer Nature Singapore: Singapore, 2024; pp. 171–186.
154. Bohm J. Electrostatic precipitators.; 1982; ISBN 0444997644.
155. Schmoch M, Steiner D. The electric field in the electrostatic precipitator. In: Electrostatic Precipitator Handbook; Springer Fachmedien Wiesbaden: Wiesbaden, 2024; pp. 7–40.
156. Assaye R, Suryabhagavan KV, Balakrishnan M, Hameed S. Geo-spatial approach for urban green space and environmental quality assessment: a case study in Addis Ababa City. J Geogr Inf Syst. 2017;09:191–206. https://doi.org/10.4236/jgis.2017.92012.
157. UN China's Policy Strategies for Green Low-Carbon Development: Perspective from South-South Cooperation. In Proceedings of the United Nations Conference on Trade and Development; 2023.
158. Constantinescu J, Rosu A, Ataman E, Popescu M. New solutions for the removal of the pollutants generated by electrical power plants in Romania. In: The modern problems of electrostatics with applications in environment protection; Springer Netherlands: Dordrecht, 1999; pp. 47–52.
159. Cheng Y, Zhou Z, Thiele L. ISpray: reducing urban air pollution with intelligent water spraying. Proc ACM Interact MobWearable Ubiquitous Technol. 2022;6:1–29. https://doi.org/10.1145/3517227.
160. Liu B, Wang P, Zhou J, Guo Y, Ma S, Chen W-Q, Li J, Chang VW-C. Refocusing on effectiveness over expansion in urban waste–energy–carbon development in China. Nat Energy. 2024. https://doi.org/10.1038/s41560-024-01683-8.
161. IEA, I.E.A. A vision for clean cooking access for all. World Energy Outlook Spec. Rep. 2023, 1–200.
162. Rahal F. Low-cost sensors, an interesting alternative for air quality monitoring in Africa. Clean Air J. 2020. https://doi.org/10.17159/caj/2020/30/2.9223.
163. Asmelash F. Green infrastructure for sustainability and climate resilience. Icccb 2012, 1–15.
164. Eticha TC, Emagnu YM. Evaluation of electric vehicle-dependent strategy in Addis Ababa Ethiopia transport system. Model Simul Eng. 2023;2023:1–20. https://doi.org/10.1155/2023/1874418.
165. UNEP Addis Ababa city air quality policy and regulatory situational analysis; Addis Ababa, Ethiopia, 2018;
166. Kumie A, Worku A, Tazu Z, Tefera W, Asfaw A, Boja G, Mekashu M, Siraw D, Teferra S, Zacharias K, et al. Fine particulate pollution concentration in Addis Ababa exceeds the WHO guideline value. Environ Epidemiol. 2021;5: e155. https://doi.org/10.1097/EE9.0000000000000155.

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