Machine learning-driven prediction of Visual Range under changing climate conditions over complex terrain using AOD and CMIP6 climate simulations

Sadaf Javed [a], Muhammad Imran Shahzad [a,*], Muhammad Zeeshaan Shahid [b], Jun Wang [c], Imran Shahid [d]

[a] Earth & Atmospheric Remote Sensing Lab (EARL), Department of Meteorology, COMSATS University Islamabad, Islamabad, 45550, Pakistan
[b] College of Earth and Environmental Sciences, University of the Punjab, Lahore, Pakistan
[c] Department of Chemical and Environmental Engineering, University of Iowa, Iowa City, 52242-1503, United States
[d] Environmental Science Centre, Qatar University, PO BOX 2713, Doha, Qatar

A R T I C L E   I N F O

A B S T R A C T

Visibility through the atmosphere, or Visual Range (VR), is a key indicator of ambient air quality, especially in areas with complex topography and vulnerability to climate change. The specific aims of this study were to (1) evaluate the ability of the Coupled Model Intercomparison Project Phase 6 (CMIP6) climate model outputs and satellite Aerosol Optical Depth (AOD) to predict VR across diverse topography; (2) select important meteorological parameters for VR; and (3) design an ensemble Machine Learning (ML) model with high accuracy using Bagged Extreme Gradient Boosting (BG-XG) for long-term VR trends under future climate scenarios. This study contributes to the significant gap in regional visibility prediction by combining climate model projections, remotely sensed AOD, and ML to project future VR through 2100 across Pakistan. The BG-XG model was trained using in situ meteorological data, AOD, and six CMIP6 models (Euro-Mediterranean Centre on Climate Change Climate Model 2 High Resolution – version SR5 (CMCCCM2-SR5 (Italy))was the most consistently accurate model across all the topography). For the results computed at Lahore (LHR), the BG-XG model achieved the highest correlation coefficient of R = 0.98 and Root Mean Square Error (RMSE) = 0.24 km for the validation dataset. It is expected that the region will observe an average VR of 5.88 km with a standard deviation of 1.66 km by the end of 2100. The predictive strength of climate model parameters for VR was high (>90 %), with significant dependencies on sea-level pressure (SLP), relative humidity (RH), eastward wind (EW), and AOD. The region is expected to witness a significant decrease in average VR at a rate of −281.3 m/year due to an increase in AOD at a rate of 0.14/year from 2003 to 2100. Among the regions, Karachi (KHI) is anticipated to experience the most substantial reduction in VR by 2100, followed by Sindh and the northwestern areas. This study provides the first long-term, region-specific VR forecasts for Pakistan by integrating ML with CMIP6 climate projections. These findings can guide climate adaptation strategies, particularly for regions at considerable risk of declining air quality due to reduced visibility.

---

* Corresponding author.
E-mail address: imran.shahzad@comsats.edu.pk (M.I. Shahzad).

1. Introduction

Visibility, or Visual Range (VR), can function as a surrogate to assess the clarity, pollutant levels, and environmental quality (Zhang et al., 2020). Reduced VR impacts tourism, transportation, and public health—contributing to eye strain, mental health effects (e.g., seasonal affective disorder), and physical hazards such as accidents due to fog or glare (Van Pelt et al., 2020; Wu et al., 2021). VR is dependent on weather conditions, and this complex relationship requires further research (Aman et al., 2019; Xi, 2021). A wide range of ground-based and satellite-based devices is available for visibility measurements, each with its own set of benefits and limitations (Shahzad et al., 2018; Bai et al., 2023; Pang et al., 2024; Verbeek and Hincks, 2022). A typical application of ground-based sensors is localized real-time monitoring of visibility. A nephelometer measures the amount of light scattered by airborne particles to provide information about visibility conditions (Shahzad et al., 2018). Transmissometers are another standard ground-based instrument for measuring visibility over short to medium distances; they work by directly measuring the attenuation of a light beam over a defined distance (Javed et al., 2023). In addition, forward scatter meters and other similar sensors evaluate visibility by analyzing the light scattered by pollutants (Javed et al., 2023; Shahzad et al., 2018). Installing visibility sensors at ground weather stations, roadways, and airports provide significant support to the transportation sector, meteorology, and aviation (Javed et al., 2023). In recent times, only ground-based equipment, such as the Automated Surface Observing System (ASOS), has provided VR measurements, but with limited spatial coverage (Javed et al., 2023). Ground-based observations are unable to capture spatio-temporal variations across entire regions. Therefore, a transmissometer is unable to calculate VR at distances greater than 5000 m, and the nephelometer cannot measure scattering beyond 170° due to its structural design. Lastly, digital cameras cover only small areas and are unsuitable for regional studies (Javed et al., 2023; Shahzad et al., 2018).

Satellite remote sensing has addressed the issue of spatial coverage with high accuracy, but it still does not provide extensive temporal coverage to estimate VR or air pollution, despite recent advancements in this field. Satellite data have limitations such as infrequent overpasses, cloud cover interference, low spatial resolution, and delayed data processing, all of which affect real-time accuracy and coverage. Climate models simulate meteorological data over acceptable spatial and temporal domains, but with reduced accuracy in regions with diverse topography. The latest Model Intercomparison Project (CMIP6) scenarios incorporate Representative Concentration Pathways (RCPs) and Shared Socioeconomic Pathways (SSPs), allowing for standardized socioeconomic assumptions (Anil et al., 2021; Pimonsree et al., 2023). Studies have used the SSP-RCP scenarios to investigate potential future approaches to reducing air pollution and their associated impacts on health and climate (Kamworapan et al., 2021; Navarro-Racines et al., 2020; Thi et al., 2023).

Aerosol Optical Depth (AOD) is an essential atmospheric property that represents the amount of aerosols that inhibit the transmission of sunlight through absorption or scattering. AOD is a valuable proxy for the average concentrations of PM2.5 and PM10 in the atmosphere (Chen et al., 2020; Zhang et al., 2020). Higher AOD values usually indicate poorer air quality and visibility; thus, AOD is another crucial parameter for estimating atmospheric clarity and pollution levels (Shahid et al., 2025). AOD provides spatially extensive coverage and can detect aerosol loading almost anywhere, especially in regions where continuous ground-based monitoring is limited or unavailable (Chen et al., 2020; Zhang et al., 2020). For this study, AOD is utilized to estimate visibility conditions, as it represents the aerosol burden directly related to VR. Understanding AOD concerning future climate scenarios is relevant for forecasting visibility in climate-sensitive dust-prone regions such as Pakistan.

Rahman et al. (2024) use the relationship between AOD and various meteorological parameters to enhance the monitoring of air quality. Using satellite data (e.g., Moderate Resolution Imaging Spectroradiometer (MODIS)) along with ground-based observations, the study identifies how Temperature (T), Wind Speed (WS), precipitation, and humidity impact AOD. Generally, AOD increases with high T and high WS, with a negative correlation observed between AOD and precipitation and other factors (increasing AOD due to dust mobilization and less rainfall, leading to the accumulation of aerosols). Such findings are imperative for developing warning systems against dust storms, supporting pollution control policies, and mitigating health risks related to air quality. The study emphasized such integration to improve air quality management strategies, particularly in regions like Saudi Arabia, where dust storms coupled with climate variability are common.

Machine Learning (ML) has shown significant potential in air pollution research, uncovering complex trends hidden from conventional statistical approaches (Ismanto et al., 2019; Maleki et al., 2022). ML algorithms can reliably assess and anticipate air pollution levels by evaluating large volumes of data collected from satellite imagery, meteorological stations, and air quality sensors (Chen et al., 2020).

This study addresses the limitation of long-term, spatially comprehensive visibility projections under a changing climate, particularly in data-scarce and *climate-vulnerable regions like Pakistan*. Traditional monitoring of visibility relies on spatial information, and while AOD, as a satellite-derived climate variable, can be measured over larger geographic areas, these data do not enable long-term forecasts. CMIP6 climate models provide useful projections of meteorological drivers that affect AOD and visibility. However, they were designed to project climate change, and the information available on the relationships between visibility and AOD is often underutilized in direct forecasting within visibility research. The specific aims of this study were to (1) evaluate the ability of CMIP6 climate model outputs and satellite AOD to predict VR across diverse topography; (2) select important meteorological parameters for VR; and (3) design an ensemble ML model with high accuracy using Bagged Extreme Gradient Boosting (BG-XG) for long-term VR trends under future climate scenarios through 2100. By bridging climate modeling, remote sensing, and ML, this work informs adaptation strategies for transportation safety and environmental sustainability in climate-vulnerable regions. This work provides the first visibility projections at a national scale under future climate scenarios and may support environmental planning, transportation safety, and air quality policy.

2. Materials and methods

2.1. Study sites

Pakistan shares a southern border with the Arabian Sea, an eastern border with India, a western border with Iran and Afghanistan, and a northern border with China, which is defined by the 8600-m-tall Himalayan and Karakoram mountain ranges. Fig. 1 shows the locations of eight international airports, including Karachi (KHI), Nawab Shah (NWS), Sukkar (SKR), Multan (MLN), Lahore (LHR), Sialkot (SKT), Islamabad (ISB), and Peshawar (PSH), which were used for the collection of VR and meteorological data. Except for the sub-mountain region and the southern slopes of the Himalayas, most of Pakistan has an arid to semi-arid climate, with hot summers, cold winters, and highly variable rainfall (760 mm–2000 mm per year) (Ali, 2018). Regional concerns over air pollution and environmental degradation arise from deteriorating air quality in megacities, driven by rapid urbanization and economic development (Tabinda et al., 2020).

2.2. Data sources

For this research, VR measurements (2003–2020) were collected from Pakistan's Automated Surface Observing System (ASOS) network managed by Pakistan Meteorological Department (PMD) and the Pakistan Civil Aviation Authority (CAA). Regions of interest from ASOS were primarily located at international airports, which used forward-scatter sensors or transmissometers to provide visibility data continuously and in a standard way (in meters or kilometers). Table 1 presents climatic variables retrieved from in situ and modeled data.

Since VR is measured on a scale that can take any value within a given range (e.g., in meters or kilometers), it should be regarded as a continuous variable, and minor fluctuations can be quite meaningful (Burrows, 2020; Sulistya et al., 2019). Using Forward Scatter Visibility Sensors (FSVS), which offer consistent and objective measurements based on light scattering by aerosols at a 550 nm wavelength (Landolt et al., 2020), the ASOS network provides high-quality data through minute-by-minute automated recordings. By reducing noise and capturing long-term variability trends, aggregating these data into monthly averages from 2003 to 2020 further strengthens the datasets. This renders the dataset suitable for statistical and ML analysis as well as for highly-resolution modeling (Burrows, 2020; Sulistya et al., 2019). In this study, monthly averaged aerosol optical depth (AOD) data were extracted from both the MODIS Level 3 products (Aqua MYD08 and Terra MOD08) at a 10 km spatial resolution (linear distance based on geospatial standards) (Carmona et al., 2021). The data (specifically in South Asia) were filtered for retrievals that had a permitted error of ≤20 %, matching the anticipated performance of MODIS Collection 6.1 AOD (Carmona et al., 2021; Su et al., 2022). The products apply utilization improvements of the cloud masking, cloud standard deviation, surface reflectance corrections, and limitations of observable derivatives in Collection 6.1, and therefore reduce bias over high-topographic and urban sites (Burrows, 2020; Sulistya et al., 2019). The data build a representation of aerosol aggregates on daytime use by taking data from both the Terra (morning's sample) and Aqua (afternoon sample) satellites, which should reduce diurnal sampling limitations on aerosol representativeness over time (Shahzad
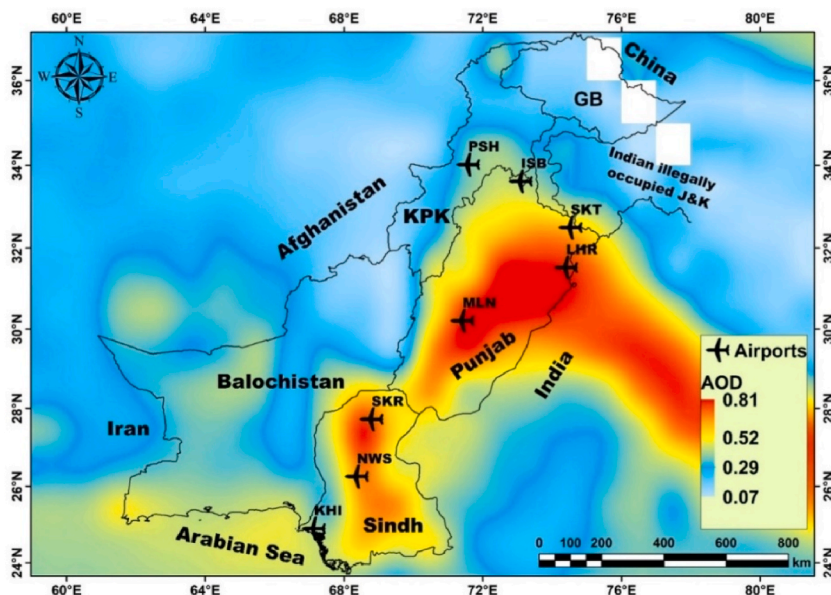
Fig. 1. Locations of eight international airports– Karachi (KHI), Nawab Shah (NWS), Sukkar (SKR), Multan (MLN), Lahore (LHR), Sialkot (SKT), Islamabad (ISB), and Peshawar (PSH) –used for the collection of VR and meteorological data. Colours on the map represent time-average AOD (MYD08 & MOD08) at 550 nm from 2003 to 2020 over Pakistan. White areas indicate missing data.

Table 1
Climatic variables retrieved from in situ and modeled data.

Sr #	Data	Unit	Source	Duration
1	*Visual Range (VR)*	Km	*ASOS*	2003–2020
2	Aerosol Optical Depth (AOD)	–	MODIS	2003–2020
3	*Air Temperature (T)*	C°	SSP5/RCP 8.5	2003–2100
4	*Relative Humidity (RH)*	%	SSP5/RCP 8.5	2003–2100
5	*Wind Speed (WS)*	m s^{-1}	SSP5/RCP 8.5	2003–2100
6	*Northward Wind (NW)*	m s^{-1}	SSP5/RCP 8.5	2003–2100
7	*Eastward Wind (EW)*	m s^{-1}	SSP5/RCP 8.5	2003–2100

et al., 2018).

Meteorological data (SLP, WS, NW, EW, RH, and T) were *collected from historical simulations (2003–2014) and future climate projections (2015–2100) under Shared Socioeconomic Pathway 5 (SSP5) and Representative Concentration Pathway 8.5 (RCP 8.5) forcing scenarios,* which are generally accepted as valid long-term climate projections and are widely endorsed by climate researchers across the globe. We did not create additional future projections; instead, we relied on CMIP6 scenarios that are published datasets relying on a long-established modeling framework recognized by peers. We opted for this approach to ensure compatibility with global climate scenarios for comparison with studies, while avoiding doubling of computation time.

CMIP6 SSP5/RCP8.5 future scenarios include six climate models: Alfred Wegener Institute Climate Model 1.1 Medium Resolution (AWI-CM-1-1-MR (Germany)), Canadian Earth System Model 5 with Canadian Ocean Ecosystem model (CANESM5-CANOE (Canada)), Euro-Mediterranean Centre on Climate Change Climate Model 2 High Resolution – version SR5 (CMCCCM2-SR5 (Italy)), Euro-Mediterranean Centre on Climate Change Earth System Model 2 (CMCC-ESM2 (Italy)), European Consortium Earth System Model 3 with dynamic Vegetation – Low Resolution (EC-EARTH3-VEG-LR (Europe)), and First Institute of Oceanography Earth System Model version 2.0 (FIO-ESM-2-0 (China)), all from the Coupled Model Intercomparison Project Phase 6 (CMIP6). The CMIP6 models selected in Table 2 are labeled "r1i1p1f1", which refers to a specific model, where "r1" represents the first realization (ensemble member), "i1" refers to the first initialization procedure, "p1" is the first set of parameterizations of the physics, and f1 refers to the first forcing. These models were examined at the 1000 hPa (hectopascal) level, a standard reference level more commonly used in the evaluation of tropospheric variables in climate change. The r1i1p1f1 variant label describes the model in such a way that outputs can be directly compared across many CMIP6 models, following the standardized experimental design framework given to the climate models by the Coupled Model Intercomparison Project (Tebaldi et al., 2021; Merrifield et al., 2023). The interest in the 1000 hPa was to specifically analyze climate processes near the surface that would be evaluated in the models. This will follow the CMIP6 approach to model intercomparisons and allow for a solid evaluation against observations and reanalysis.

The RCPs are hypothetical future scenarios for greenhouse gas concentrations used in climate models. Under RCP 8.5, the global population, GDP, and industrialization continue to rise (Meinshausen et al., 2020; Miller et al., 2021). The German Alfred Wegener Institute developed the AWI-CM-1-1-MR climate model, which comprises the atmosphere, ocean, and sea ice as unified systems (Ashfaq et al., 2022). The model's horizontal resolution is 0.94° × 0.93° in both latitude and longitude, and it features extensive atmospheric and oceanic detail (Abbas et al., 2022). The model was developed to make long-term projections of the Earth's climate system under various scenarios, i.e., RCP 8.5 high greenhouse gas emissions scenario (Semmler et al., 2020). CANESM5-CANOE climate model considers the atmosphere, ocean, sea ice, and land surface. It was created in Canada at the Canadian Centre for Climate Modelling and Analysis (CCCma). Horizontally, this model resolves to around 2.80° × 2.80° in both latitude and longitude, allowing the researchers to observe all the details of the atmosphere and water (Aylmer et al., 2022). The Italian research organization Centro Euro-Mediterraneo per i Cambiamenti Climatici (CMCC) modeled the CMCCCM2-SR5 climate model. The model's horizontal resolution is about 1.25° latitude and 0.94° longitude (Mbigi et al., 2022; Wang et al., 2021). It includes components from the atmosphere, ocean, and sea ice (Deepthi and Sivakumar, 2022; Nicolì et al., 2023). Italy's CMCC also modeled the Earth System Model known as CMCC-ESM2. Multiple atmosphere and ocean tiers are depicted with a horizontal resolution of 1.92° latitude and 2.80° longitude (Abbas et al., 2022; Deepthi and Sivakumar, 2022; Lovato et al., 2022). Continuous use of fossil fuels is expected to have several negative consequences for Earth's climate, including changes to severe events, ocean circulation patterns, and the cryosphere (Kim et al., 2023; Pimonsree et al., 2023). The ECMWF and other European organizations created the EC-EARTH3-VEG-LR Earth System Model in collaboration. The atmosphere and ocean have different tiers, and the horizontal resolution is about 1.10° by 1.10° (Abbas et al., 2022; Chen et al., 2021), to examine regional and global climate variability and change under several emission scenarios,

Table 2
Selected CMIP6 models with variant label of "**r1i1p1f1**″ at 1000 hPa.

Sr #	Climate Model	Country	Resolution (Lon. & Lat.)
1	AWI-CM-1-1-MR	Germany	0.94° × 0.93°
2	CanESM5-CanOE	Canada	2.80° × 2.80°
3	CMCC-CM2-SR5	Italy	0.94° × 1.25°
4	CMCC-ESM2	Italy	2.80° × 1.92°
5	EC-Earth3-Veg-LR	Europe	1.10° × 1.10°
6	FIO-ESM-2-0	China	1.30° × 0.90°

including RCP 8.5 (Halder et al., 2021; Liu and Lu, 2022). Parts of the carbon cycle, as well as the atmosphere, ocean, sea ice, land, and general environment, are included in the First Institute of Oceanography (FIO), Ministry of Natural Resources, China; FIO-ESM-2-0 (Abbas et al., 2022; Kamworapan et al., 2021). The atmosphere and ocean have different tiers, and the horizontal resolution is about 0.90° latitude and 1.30° longitude (Abbas et al., 2022; Kamworapan et al., 2021).

2.3. Methodology

The methodology for this research is divided into 3 phases. In phase 1, data were collected and subjected to a quality control process, including outlier removal, missing data management, and satellite data downscaling. In phase 2, statistically significant variables were selected using the Boruta algorithm and the BG-XG model was developed for prediction and forecasting. In phase 3, future trends of AOD and visibility were projected using the BG-XG model. A schematic diagram of the VR modeling and prediction framework is presented in Fig. 2.

2.3.1. Data Pre-processing

Data for this study underwent a comprehensive quality control process, where outliers were detected and removed using the Interquartile Range (IQR) method, and missing data were managed through linear interpolation and mean imputation. Results were validated through cross-verification against ground-based independent observations. To ensure compatibility with the spatial resolution of climate models and MODIS satellite observations, the data were downscaled to a 1° spatial resolution using cubic interpolation, which redistributed the data onto a uniform grid and was subsequently aggregated into monthly averages, matching the temporal resolution of the climate simulation. Cubic interpolation was preferred over bilinear interpolation due to its smoother and more precise surface fitting, particularly for meteorological fields such as visibility, which exhibit gradual spatial variation. While bilinear interpolation considers only the four nearest points and can create sharp discontinuities or artefacts, cubic interpolation uses additional surrounding data points, producing a more realistic and continuous surface. This is especially relevant when interpolating high-resolution ASOS data onto coarser satellite and climate model grids to preserve spatial integrity and consistency of regional trends. Quantile mapping was then applied for bias correction to minimize discrepancies between ASOS data and MODIS observations. The downscaled and average data were combined with output from six climate models and MODIS observations across space and time, and cross-validated using statistical metrics such as Root Mean Square Error (RMSE) and correlation coefficients. Trend and analysis of validated data under different climate scenarios were conducted to estimate the future VR dynamics. This effort enhanced data usability and compatibility for subsequent forecasting of climate and VR responses.

2.3.2. Variable selection for VR estimation

The Boruta algorithm was used to uncover the relationship between climatic variables and VR, due to its effective performance in an earlier regional study (Javed et al., 2024). The primary goal of the Boruta algorithm is to rank the importance of each independent
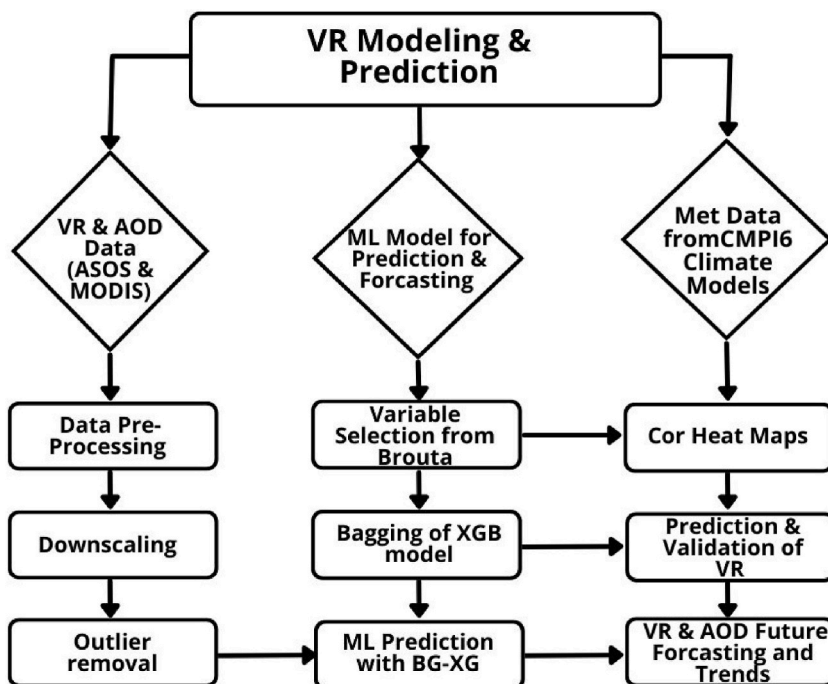

Fig. 2. A schematic diagram of the ML model working and methodology.

variable (Subbiah et al., 2022). The method generates a set of "shadow features" based on the original features, which are jumbled or permuted versions of the original features (Zhou et al., 2023). The Boruta method compares each feature's Z-score to the highest Z-score derived from its shadow features and uses the difference to determine the feature's significance. If a feature's Z-score is greater than the highest Z-score of its shadow features, it is considered a "confirmed" relevant feature (Zhou et al., 2023). In addition, the algorithm identifies "tentative" features with Z-scores lower than the highest Z-score of their shadow features. Once all features have been validated or rejected, the algorithm terminates (Zhou et al., 2023). After collecting data from multiple sources, the dataset was divided into two parts for training and validation (70:30), for each station as well as for the entire study area (Pakistan). The selection of climatic variables for model simulation was based on statistical significance at a 99 % significance level (P-value <0.01).

2.3.3. Correlation heatmaps

A correlation matrix illustrates mutual dependencies among different variables (Akoglu, 2018). Correlation matrix heatmaps were computed to understand the relationships among the variables at each station.

2.3.4. Bagging XGBoost (BG-XG)

Bootstrap aggregating, or bagging, is an ensemble learning method that combines data from several individual models to improve performance while reducing the likelihood of overfitting (Javed et al., 2024; Képeš et al., 2023; Kumar and Jain, 2020). Since the bagging process is an ensemble learning approach, we used this method to refine the Extreme Gradient Boosting (XGBoost or XGB) algorithm, increasing its generalization ability and reducing output variance. The core of bagging is a random subsample drawn from a larger data pool. A separate base model is trained using the subset data, and the final output is determined by averaging the models' predictions. Bagging helps reduce variance and overfitting (Képeš et al., 2023; Kumar and Jain, 2020; Nguyen et al., 2021).

It is essential to select hyperparameters that limit the risk of overfitting and maximize R^2 on the validation set when employing a bagging ensemble of XGB models (BG-XG) for VR estimation (Izanloo et al., 2022; Mastropietro and Moya, 2021). To evaluate the model's efficacy, values of num_models ranging from 5 to 10 were assessed, and num_models = 7 was selected. To begin with, a small learning rate, such as 0.01, was used. The learning rate (eta) was assessed over a range from 0.01 to 0.30, with a final selected value of eta = 0.08. Values from 3 to 6 were evaluated to avoid overfitting, and the final value was set to max_depth = 4. Various subsample values ranging from 0.5 to 0.8 were evaluated before settling on 0.6. Values of sample_bytree ranging from 0.5 to 0.8 were experimented with before finalizing the value at 0.7. The initial number of boosting rounds (nrounds = 1000) was reduced to 500. The value of early_stopping_rounds was set to 20. To maximize R2, the "maximize" parameter was set to TRUE. Five-fold cross-validation was also used for hyperparameter tuning and early stopping.

2.3.5. Performance assessment

Statistical measures, such as the correlation coefficient (R), coefficient of determination (R^2), mean absolute error (MAE), and RMSE as defined in equations (1)–(4), are used to evaluate estimated VR concentrations using the respective modeling methods (Nguyen et al., 2022).

$$R = \frac{\Sigma(xy) - (\Sigma x)(\Sigma y)/n}{\sqrt{\left(\Sigma x^2 - (\Sigma x)^2/n\right)\left(\Sigma y^2 - (\Sigma y)^2/n\right)}} \tag{1}$$

$$R^2 = 1 - \left\{ \frac{\sum_{i=1}^{n}(c_i - m_i)^2}{\sum_{i=1}^{n}\left(c_i - \bar{m}_i\right)^2} \right\} \tag{2}$$

$$MAE = \sum_{i=1}^{n} \frac{(c_i - m_i)}{n} \tag{3}$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(c_i - m_i)^2}{n}} \tag{4}$$

Here, "c_i" signifies the estimated value from the model, "m_i" represents the observed value, and "n" denotes the number of observed data pairs. For monthly VR model estimations, the aforementioned metrics were used to evaluate the performance of the BG-XG model at each station and across all of Pakistan.

2.3.6. Trend analysis

For assessing future VR forecasting at each station, it was necessary to obtain adequate data from each climatic model to 2100, excluding VR data. To get them, MODIS AOD and climatic variables were first modeled into VR using the BG-XG model, based on the availability period of the MODIS AOD data. After establishing the VR model, VR trend predictions were projected through the year 2100, to provide an overview of the expected development of VR based on available data and climatic parameters (Javed et al., 2024). We applied the Mann–Kendall (MK) test (equations (5)–(8)) and Sen's Slope (SS) estimator (equation (9)) to observe trends in AOD and

VR data (2003–2100) (Javed et al., 2023). The MK test is a non-parametric method for identifying statistically significant trends in time-series data without assuming any specific data distribution, whereas the SS estimator measures the magnitude of such a trend (Javed et al., 2023).

$$S = \sum_{i=1}^{n} \sum_{j=1}^{i-1} \text{sign}(X_i - X_j) \tag{5}$$

where:

X_i, X_j: indicates data values of visibility and AOD at time i and j
sign(X_j–X_i): For visibility, +1 indicates clearer conditions at time j compared to I. For AOD, +1 indicates increased aerosol loading at time j

sign(xj - xi): indicates the sign function defined as:

$$\text{sign}(X_i - X_j) = \begin{cases} if(X_i - X_j) > 0 \text{ then } +1 \\ if(X_i - X_j) < 0 \text{ then } -1 \\ if(X_i - X_j) = 0 \text{ then } 0 \end{cases} \tag{6}$$

$$\text{Var}(s) = \frac{n(n-1)(2n+5)}{18} \tag{7}$$

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{\text{Var}(S)}}, \text{if } S > 0 \\ 0, \text{if } S = 0 \\ \dfrac{S+1}{\sqrt{\text{Var}(S)}}, \text{if } S < 0 \end{cases} \tag{8}$$

where:

Z > 0: Indicates increasing trend
Z < 0: Indicates decreasing trend

For VR, positive Z presents good visibility, while negative Z shows deteriorating conditions. For AOD, Positive Z denotes high AOD, and negative Z shows low AOD and good air quality.

$$Q_i = \frac{x_j - x_k}{j - k}, \text{for all } j > k \tag{9}$$

where:

x_j, x_k: indicates data values at times j and k,
$j > k$ indicates all combinations are used.

The Sen's slope estimate is the median of all Q_i values:

$$Q = \text{median}(Q_i)$$

where:

Q: rate of change per unit time (km/year)
Positive Q: Indicates upward trend
Negative Q: Indicates a downward trend

For VR, Q (km/year) > 0 indicates improved visibility, while for AOD, Q > 0 denotes high aerosol loading.
The MK test examines the presence of a monotonic trend using test statistic S, which sums the signed pairwise differences between sequential measurements (equations (5) and (6)). To clarify, a positive S indicates that clarity is improving over time, whereas a negative S signifies an increase in haze. For AOD, a positive S denotes an increase in aerosol concentrations, and a negative S signifies cleaner conditions. The variance of S (equation (7)) accounts for tied observations, and the statistical significance of the trend can be assessed by taking the standardized Z-score (equation (8)). SS (equation (9)) does the additional work of quantifying the trend magnitude - the median rate of change (km/year for VR, unitless/year for AOD) from all paired slopes.

3. Results and discussion

3.1. Variable selection

The Boruta algorithm was applied to the monthly dataset (VR, AOD, WS, NW, EW, RH, and T) to select statistically significant (p < 0.01) variables for ML modeling at each station and to identify the key features affecting monthly VR. The dataset of six features was subjected to the Boruta algorithm in the first stage of development, using data from previously listed climatic models along with MODIS AOD and VR. The selected parameters for each station were recorded to evaluate the performance of training and prediction VR models (Fig. S3). Out of the six features, nearly all were confirmed at all stations except for the WS, EW, and NW variables. These three were not selected simultaneously; instead, the chosen algorithm was only one out of three for each station. WS was selected for MLN, PSH, and SKT, while NW was selected for ISB, LHR, NWS, and Pakistan. Additionally, EW was chosen for KHI and SKR.

3.2. Correlation matrix heatmaps

This study also evaluated the strength of relationships between the variables listed in Table 2 and the monthly VR dataset. Fig. 3 illustrates correlation matrix heatmaps of significant variables with monthly VR data, where VR is the dependent variable at each station. None of the coefficients exceeded 0.74 (T) or −0.73 (SLP) at the NWS station, indicating that these parameters play a complex role rather than exhibiting a simple, direct linear correlation with VR (Zhou et al., 2023). Moreover, the association between VR and the independent variables at all stations was less than 0.74. An inverse relationship (−0.73) exists between VR and SLP, where increasing SLP results in reduced visibility.

Atmospheric dynamics and the behavior of aerosols under varying pressure conditions can partly explain this. When SLP is high,
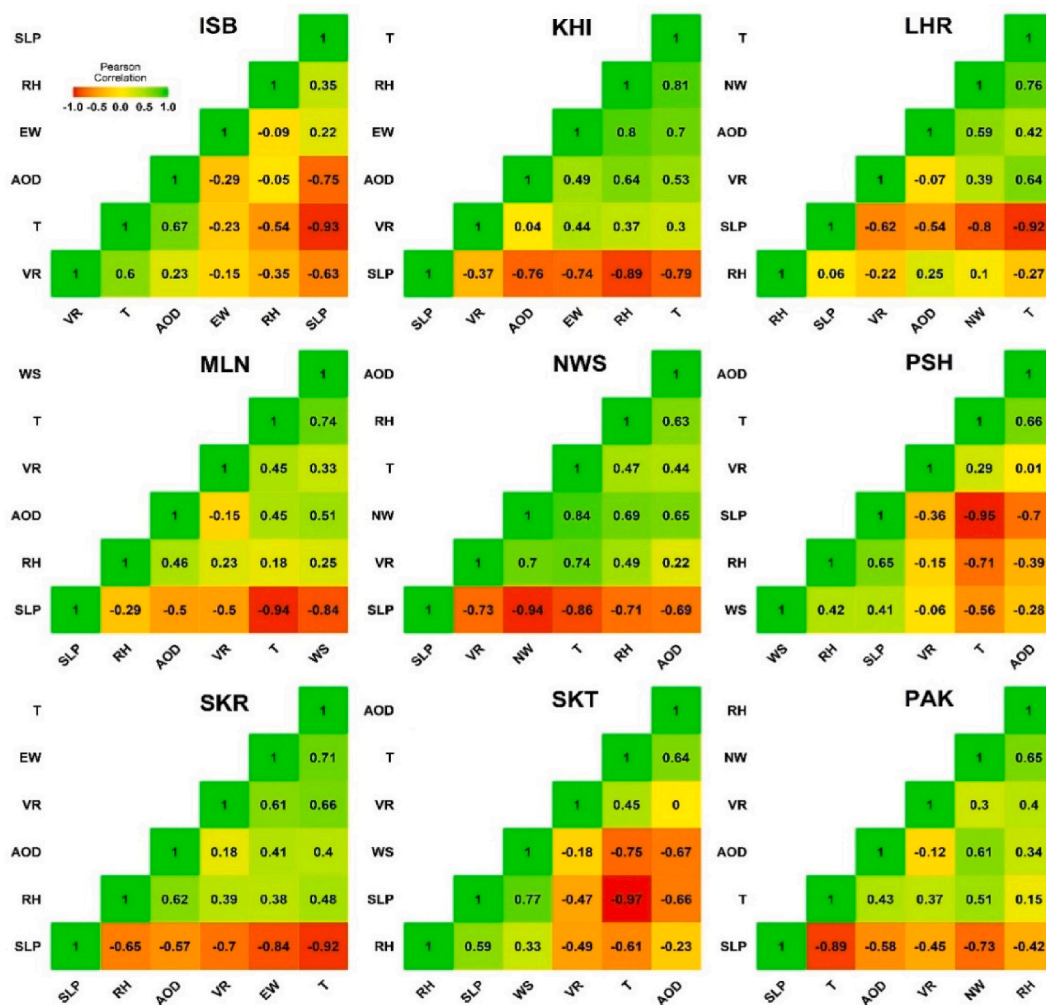
Fig. 3. Correlation matrix heatmaps of significant variables with monthly VR data, where VR is the dependent variable at each station.

the atmosphere stabilizes and exhibits reduced vertical mixing (Caputi et al., 2019; Grassi et al., 2020). Atmospheric stability inhibits the dispersion of aerosols–such as dust, pollutants, and water vapor, allowing them to remain trapped near the surface (Rahman et al., 2022). Thus, higher SLP typically corresponds to lower VR due to increased aerosol concentrations in the lower atmosphere. In contrast, lower SLP indicates dynamic weather systems with stronger winds and improved vertical mixing, which facilitate aerosol dispersal and enhance visibility (Maleki et al., 2022). In addition, at high altitudes– where naturally low ambient pressures prevail– air molecules are less dense, and weaker vertical mixing prolongs aerosol residence times. This is expected to reduce VR, as aerosols remain suspended in the atmosphere for prolonged periods. The inverse relationship between SLP and visibility indicates the key atmospheric control over air quality and visibility in regions with high aerosol concentrations, including both arid and semi-arid areas (Rahman et al., 2022). The strong positive correlation (0.74) between T and VR indicates that VR tends to improve as T increases and may decline as it decreases.

Warm air can hold more moisture, which may explain the correlation between T and VR. Reduced VR in the form of fog or haze may result from moisture condensation caused by cooler air (Lan et al., 2020; Li et al., 2022; Yang et al., 2023). Wind is associated with aerosol dispersion; however, the wind direction determines whether convection or accumulation occurs over a given area (Javed et al., 2024; Sun et al., 2020). Hence, the relationship between wind and VR varied across stations. AOD had the strongest effect on VR, followed by northerly and easterly winds (Anwar et al., 2021). Other meteorological variables (WS, NW, EW, and RH) with AOD showed lower correlations with VR, suggesting that their linear associations with VR are weak.

3.3. VR estimation with the BG-XG model

For monthly VR estimation in Pakistan, a dataset of 1309 observations was used, including four independent variables: T, RH, NW, and AOD. 70 % of the data was used to train the BG-XG model, and the results were then compared with the validation data (30 %). The BG-XG model for Pakistan achieved 88 % accuracy on the validation dataset. The level of accuracy indicates that the model avoided overfitting and generalized well to new data, as its performance on the tested dataset is only marginally lower than on the training data. The RMSE for the Pakistan monthly VR validation set was estimated at 0.98 km based on the model output.

3.3.1. Variable importance

Variation in VR depends on the selected parameters. Fig. 4 shows variable importance ranking based on the selected climate model for each station and Pakistan. The importance of each selected parameter at each station was ranked to gauge the sensitivity of VR to
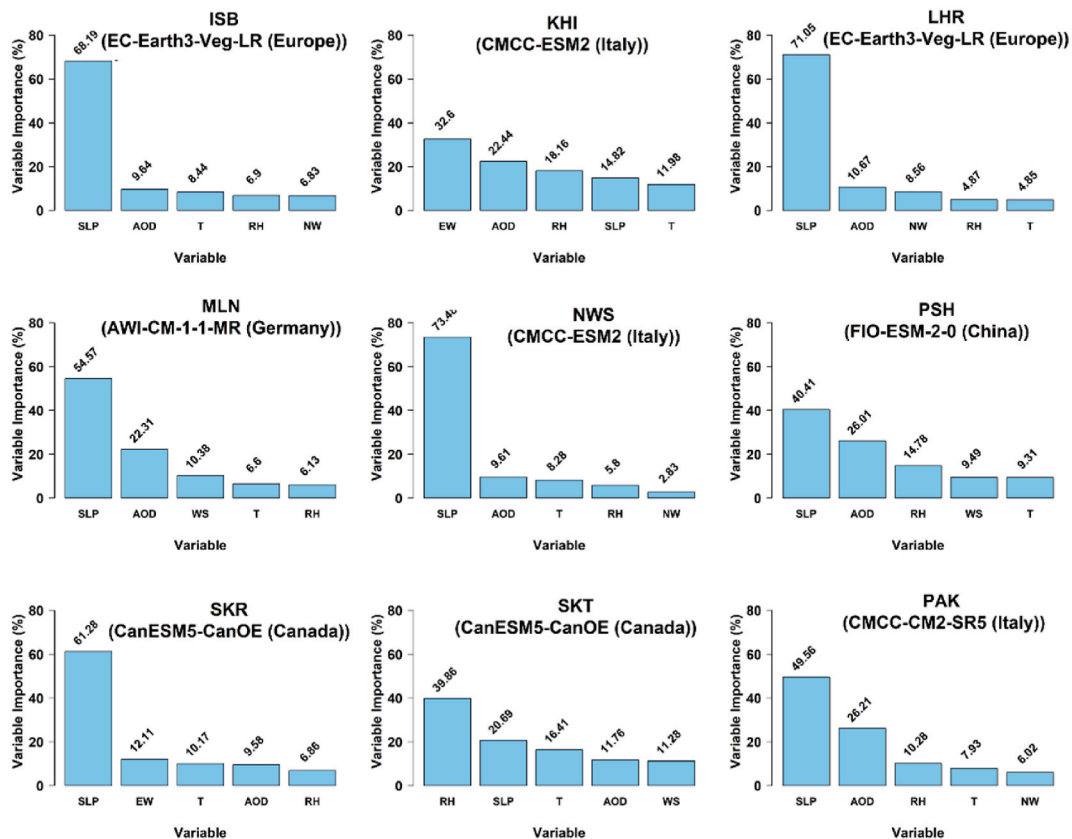

Fig. 4. Variable importance ranking based on the selected climate model for each station and Pakistan.

individual variables, where rank 1 indicates the most important and rank 5 the least significant variable for estimating VR at a given location.

The BG-XG model ranked SLP 1st for LHR, ISB, MLN, NWS, PSH, and SKR, whereas EW and RH were selected only once for KHI and SKT, respectively. The model ranked AOD in 2nd position after SLP at each station, including the national model for Pakistan–except at SKR and SKT, where EW and SLP were ranked 2nd position, respectively, followed by T and *Relative Humidity (RH)*. VR can be reduced by low-pressure system-related weather disturbances, such as storms, heavy rain, and fog, due to cloud cover, precipitation, and atmospheric instability (Zhang et al., 2019). These conditions alter the availability of aerosols at a location, depending on the source region, which affects light scattering and changes VR. RH may rise or fall depending on the prevailing atmospheric pressure. The formation of fog due to high humidity can significantly reduce VR. Clear skies and improved VR may result from lower humidity levels (Liu and Lu, 2022). Low-pressure systems trapped in valleys can lead to persistently low VR due to fog and pollution (Zhao et al., 2019).

3.3.2. Variable's partial dependency

To better explain the importance of parameters and the behavior of the BG-XG model for monthly VR prediction, partial dependency plots (PDPs) were generated (Fig. S4). All stations show steep slopes from top to bottom, indicating a strong effect of the given variable on VR (Nguyen et al., 2022). The non-linear patterns in the PDPs indicate complex interdependencies or interactions between the features and the target variable (VR) (Ben Jabeur et al., 2023; Hu et al., 2022).

The PDP analysis highlights contrasting regional influences towards visibility. For ISB, factors such as altitude and proximity to the Margalla Hills could play a role in affecting VR because, if the weather is settled (high SLP), aerosols and other pollutants will accumulate, resulting in reduced visibility. In the coastal station of KHI, degradation in VR was aligned with high humidity/sea breezes
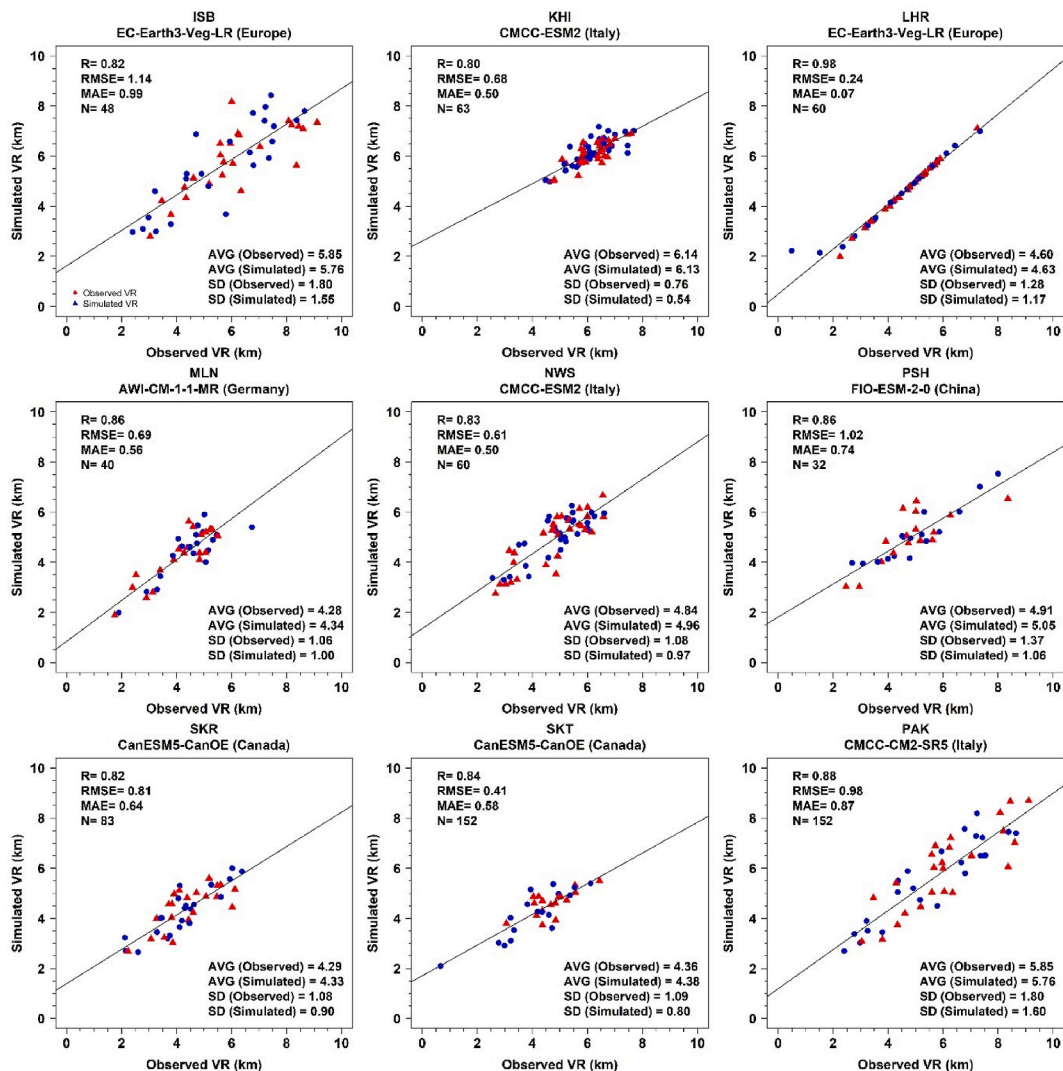

Fig. 5. BG-XG validation set of monthly VR scatter plots for all stations and Pakistan.

during pollution episodes. Conversely, LHR is more urbanized, and it is surprising to see a decline in VR, which is primarily dependent on vehicular emissions, industrial pollution, and winter smog. For MLN and SKR located in an arid region, they mainly exhibit declines in visibility due to dust storms, high T and low rainfall during summer. Similar spatial patterns of visibility are observed in NWS due to topography and wind patterns. In PSH, low visibility is due to dust storms or agricultural burning, and as a smaller station, SKT shows seasonal variations in visibility due to local farming and industrial activities. The national model for Pakistan demonstrates expected visibility trends: first, visibility is lower in arid regions and urban centers due to the presence of dust and pollution generated from anthropogenic activities.

3.4. Accuracy assessment

RMSE, MAE, and R metrics were employed to calculate the accuracy of the BG-XG model for the monthly dataset. BG-XG validation is presented in Fig. 5, showing VR scatter plots for all the stations, including Pakistan, with values of R, RMSE, number of observations (N), mean, and standard deviation (SD) for both observed and simulated VR.

At Lahore (LHR), the BG-XG model achieved the highest correlation coefficient, R = 0.98, with an RMSE of 0.24 km for the validation dataset. The model estimated a mean VR of 4.63 km, compared to the observed VR (4.60), with a SD of 1.17 for the simulated VR. In contrast, the lowest correlation coefficient was recorded at Karachi (KHI), where the model achieved R = 0.80 and an RMSE of 0.68 km for the validation set. The model estimated a mean VR of 6.13 km, compared to the observed VR (6.14), with a SD of 0.54 for the simulated VR. The scatter plots show that the correlation achieved by the model was more substantial than that reported by Shahani et al. (2021) and Shehadeh et al. (2021), who examined the relationship between VR and the meteorological variables in metropolitan regions, as well as Won et al. (2020), who assessed the exponential relationship between VR and $PM_{2.5}$ (R = 0.87).

Among all locations, the best correlation was observed for LHR with the EC-Earth3-Veg-LR (Europe) model, reflecting excellent agreement between simulated and observed VR values. The mean VR values of both observed and simulated data are in near-perfect agreement, reflecting minimum bias. The general fit of values on the 1:1 line indicates that the model performs very well in reproducing both spatial and temporal variability in visibility for this site. Following Lahore, high correlations were also observed for Pakistan, Multan (MLN), and Peshawar (PSH), with R values of 0.88, 0.86, and 0.86, respectively. These models not only identify trends accurately but also show strong agreement in mean VR and variability (standard deviation), indicating their reliability for inland urban and semi-urban areas. The poorest model performance was observed for Karachi (KHI) using the CMCC-ESM2 model, where the correlation dropped to 0.80, meaning that the model also performs well in simulating VR trends in this coastal setting. The plot demonstrates considerable variation in model performance by location, with interior locations such as Lahore showing strong agreement and coastal urban centers such as Karachi proving more challenging to simulate accurately, highlighting the importance of regional model validation and calibration.

In addition, the correlation was more substantial than that reported by Javed et al. (2024), who recently estimated daily VR in Pakistan (R = 0.97). These earlier studies often estimate VR on sunny days or under low RH (<80 %) conditions. This study found that the VR estimated using the BG-XR algorithm, along with meteorological and pollution data, accounted for more than 95 % of the observed VR at each station in the training set, and 70–90 % in the validation set.

3.5. VR trend dynamics (2003–2100)

Past, present, and future trend dynamics (2003–2100) of AOD and VR are presented in this section. Future simulations for VR and AOD were not available from any climate models. Therefore, AOD was first forecasted up to the year 2100 using the BG-XG model, which was then used to forecast VR through 2100. AOD and VR forecasting using the BG-XG model based on climate simulations from 2003 to 2100 are presented in Figs. 6 and 7, respectively. Overall, VR trends indicate that from 2003 to 2100, VR in Pakistan will decrease at an average rate of −281.3 m/year under the influence of an increase in AOD at an average rate of 0.1393/year.

The lowest VR is projected for KHI by 2100 with a decreasing rate of −396.8 m/year due to an increase in AOD at an average rate of 0.1242/year. SKR and NWS are expected to show improved VR than KHI with an average VR decrease rate of −136.0 m/year and
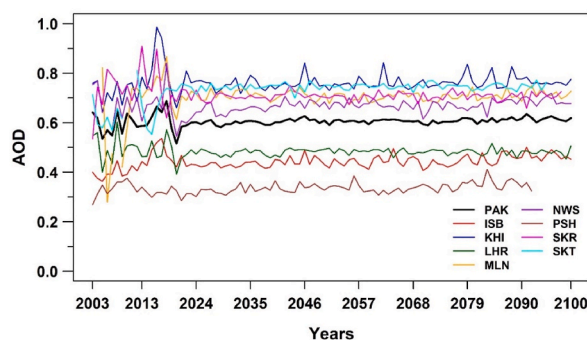

Fig. 6. AOD forecasting using the BG-XG model based on climate simulations from 2003 to 2100.
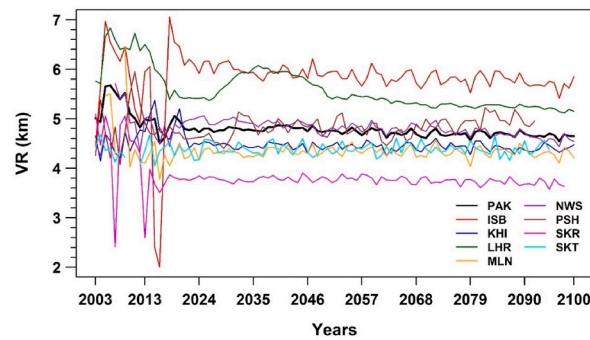
Fig. 7. VR forecasting using the BG-XG model, based on climate simulations and AOD data from 2003 to 2100.

−96.6 m/year, respectively. AOD in SKR and NWS is projected to increase at average rates of 0.0374/year and 0.1698/year, respectively. In Punjab, SKT is projected to experience the smallest decline in VR, decreasing at an average rate of −5.6 m/year alongside an increasing AOD trend at an average rate of 0.1527/year. Among the eight stations, only PSH is projected to show a slight increase in VR. According to Javed et al. (2023), plain areas such as SKT, KHI, NWS and MLN have experienced the highest number of poor VR days (>90 %). These lowland areas in central, southern, eastern Punjab, as well as northern, central, and eastern Sindh, exhibited decreasing VR trends. In contrast, VR in the south and southwestern coastal areas of Sindh, the western highlands, and parts of KPK province indicated positive trends (Javed et al., 2023).

We also applied the MK test on meteorological parameters to assess the cascading effects of ongoing climate change in the region on VR. MK and SS analysis of monthly time series data for eight stations is presented in Table 3, showing variations in meteorological trends dynamics (2003–2100). The MK values for T, RH, and SLP are positive, indicating increasing trends per year, while the value for WS is statistically insignificant, indicating no discernible trend. Specifically, the MK test shows trends of 0.1292 °C/year for T, 0.0762 %/year for RH, 0.0223 hPa/year for SLP, and −0.0011 mps/year for WS. The values from the SS estimator indicate that T is increasing by 0.0005 °C/year, RH by 0.0006 %/year, SLP by 0.0001 hpa/year, and WS is decreasing by 0.0002 mps/year. The analysis reveals that over the period from 2003 to 2100, Pakistan will experience slight increasing trends in T, RH, SLP, and AOD, indicating both direct and indirect impacts of future climate change. To examine the spatial variability of VR, AOD and other meteorological parameters as well as future dynamics, detailed results are provided in supplementary material as Figs. S5–S6.

It is expected that AOD and meteorological parameters will significantly impact visibility in the future under prevailing climate change scenarios. Average spatial MK Trends are presented in Fig. 8, showing the variability of VR, AOD, and other meteorological parameters from 2003 to 2100. Increasing T and RH are likely to result in more frequent and dense fog, as warm air can hold more moisture, which may condense into fine water droplets, reducing visibility (Anwar et al., 2021; Subhanullah et al., 2022). Conversely, a slight decrease in WS would further exacerbate poor visibility conditions by allowing pollutants and particulate matter to remain in the atmosphere for more extended periods, reducing both air quality and visibility. Additionally, changes in SLP could affect atmospheric stability, potentially leading to more stagnant air conditions that degrade visibility. These combined trends suggest that, over time, Pakistan could experience more frequent poor visibility events, impacting transportation, public health, and overall quality of

Table 3
MK and SS analysis of monthly time series data for eight stations, showing variations in meteorological trends dynamics (2003–2100). Italicized values represent statistically insignificant decreasing trends (no trend). T = Air Temperature; RH = Relative Humidity; WS = Wind Speed; SLP = Sea Level Pressure; AOD = Aerosol Optical Depth; VR = Visual Range.

Station	Test	T	RH	WS	SLP	AOD	VR
ISB	MK	0.1250	0.1002	*−0.0026*	*0.0283*	0.0995	*−0.0611*
	SS	0.0045	0.0052	*−0.0002*	*0.0009*	0.0001	*−0.0002*
KHI	MK	0.3386	0.0631	*0.0040*	*0.0081*	0.1242	*−0.3968*
	SS	0.0040	0.0037	*0.0001*	*0.0001*	0.0000	*−0.0002*
LHR	MK	0.1335	0.0959	*0.0001*	*0.0286*	0.0601	*−0.0816*
	SS	0.0042	0.0060	*0.0002*	*0.0009*	0.0000	*−0.0002*
MLN	MK	0.1340	0.1169	*−0.0046*	*0.0235*	0.0009	*−0.0222*
	SS	0.0044	0.0063	*−0.0001*	*0.0007*	0.0001	*−0.0000*
NWS	MK	0.2544	0.0686	*0.0058*	*0.0109*	0.1698	*−0.0966*
	SS	0.0041	0.0049	*0.0001*	*0.0002*	0.0000	*−0.0003*
PSH	MK	0.1328	0.0934	*−0.0032*	*0.0545*	0.0596	0.0685
	SS	0.0055	0.0045	*−0.0001*	*0.0016*	0.0000	0.0002
SKT	MK	0.1365	0.0889	*−0.0068*	*0.0480*	0.1527	*−0.0056*
	SS	0.0042	0.0053	*−0.0001*	*0.0015*	0.0000	*−0.0000*
SKR	MK	0.1915	0.0923	*0.0031*	*0.0149*	0.0374	*−0.1360*
	SS	0.0044	0.0041	*0.0001*	*0.0004*	0.0000	*−0.0002*
Pakistan	MK	0.1292	0.0762	*−0.0011*	0.0223	0.1393	*−0.2813*
	SS	0.0005	0.0006	*−0.0002*	0.0002	0.0001	*−0.0001*
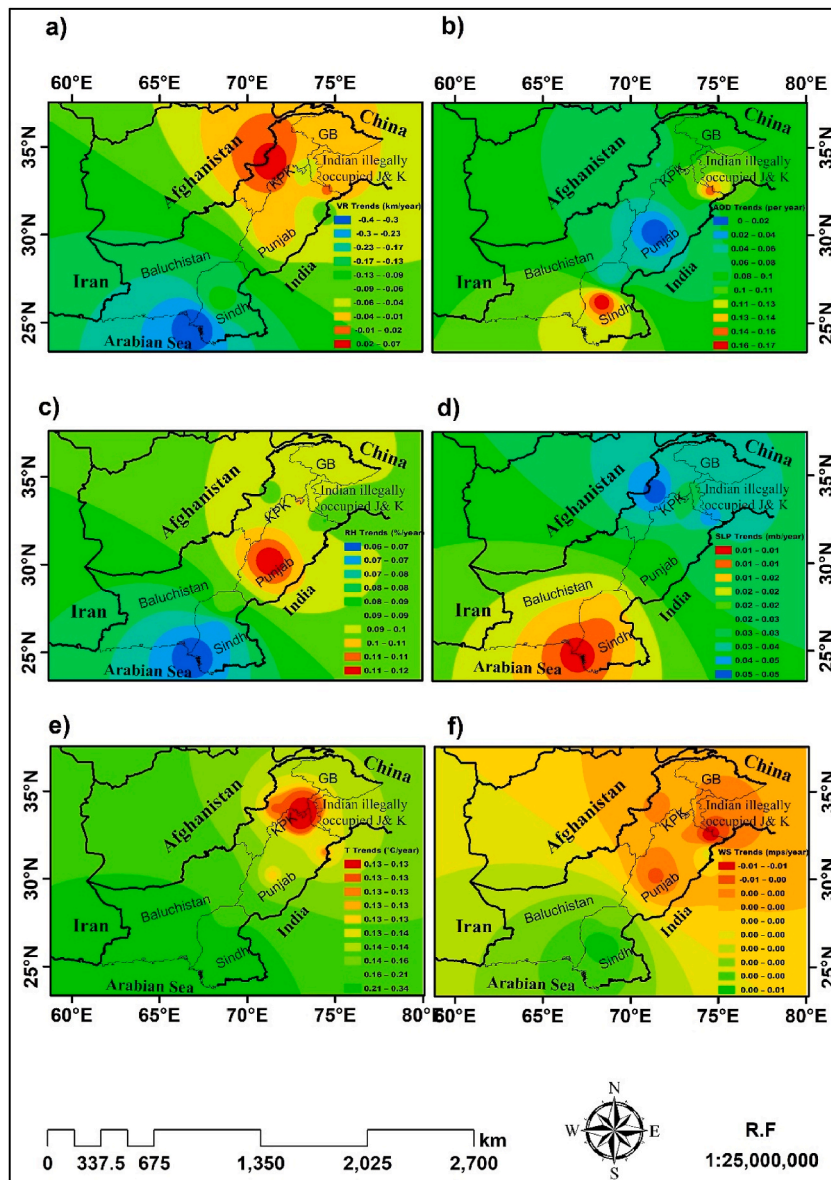
Fig. 8. Average spatial MK Trends showing the variability of VR, AOD and other meteorological parameters from 2003 to 2100: (a) VR; (b) AOD; (c) RH; (d) SLP; (e) T; (f) WS.

life–particularly in urban areas where pollution is already a significant concern, with KHI being the most affected (Fig. 8).

The projected decrease in VR across much of Pakistan also signals a long-term deterioration in atmospheric clarity, especially over urban and low-lying regions. This decreasing trend is consistent with observed increases in AOD, T, and RH—variables that individually and collectively reduce visibility by increasing aerosol loading, causing fog formation, and leading to stagnant air masses. Urban agglomerates such as KHI, MLN, and LHR—already infamous for their pollution levels—are especially vulnerable, with model simulations projecting extensive VR losses due to the combined effects of anthropogenic emissions and climate-induced atmospheric changes. Conversely, stable or improving VR trends in urban centers such as PSH and regions of KPK suggest a complex interplay of topographical, meteorological, and emission-related processes. These regional contrasts underscore the need for location-specific mitigation measures and localized policy interventions. Furthermore, the ongoing trend of low VR may be inflicting heavy socio-economic burdens on key sectors such as aviation, transportation, and public health. Given that Pakistan already experiences frequent flight delays and road accidents due to poor visibility, further reductions in VR will add to socioeconomic burdens. Finally, the persistent and spatially consistent reduction in VR underscores the increasing influence of climate change and air pollution in regulating atmospheric transparency and reinforces the need to integrate VR forecasting into more comprehensive environmental planning strategies.

To provide a broader context, the performance of CMIP6 models was compared with different datasets and modeling frameworks used globally. Satellite products such as MODIS AOD and Sentinel-5P have provided hydrologically valid, near-real-time datasets with high spatial and temporal resolution over the past 3–20 years (Wang et al., 2021; Reshi et al., 2024). These datasets allow for snapshot assessments of aerosol loading, aiding air quality monitoring and short-term aerosol distribution analysis across arbitrary periods. However, satellite-derived observations have limitations due to multiple factors, such as cloud cover, missing data, and limited temporal span. CMIP6 models and datasets are helpful because they provide long-term, physically consistent climate simulations aligned with various socioeconomic scenarios spanning centuries into the future (Schroeter et al., 2024; Nazrul et al., 2025).

The selection of CMIP6 models in this research was based on a need to evaluate the long-term climate projections and their relevance to future extreme events. In contrast to satellite-based datasets such as MODIS AOD and Sentinel-5P, CMIP6 are physically based, climate system simulations described well into the 21st century, making them a better selection for evaluating potential shifts under varying climate scenarios, which is a primary objective of the present research. Additionally, CMIP6 models include some improvements, including new aerosol schemes in some models and improved spatial resolution compared to CMIP5, further justifying their inclusion for the analysis (Wang et al., 2021; Reshi et al., 2024; Schroeter et al., 2024; Nazrul et al., 2025).

This study presents several methodological and contextual strengths. First, it employs a unique methodology by integrating diverse datasets–observed meteorological data, satellite-derived AOD, and CMIP6 climate projections– into a single ML framework that can deliver long-term visibility forecasts. Secondly, it applies the bagged XG-Boost (BG-XG) modeling approach to enhance the robustness of predictions across Pakistan's diverse landscape, while still maintaining high accuracy at both national and local levels. The further strength lies in the use of the Boruta algorithm, which enables a rigorous feature selection process, making the model more interpretable. Finally, the study is the first to provide long-term, spatially explicit visibility projections for Pakistan, filling a significant regional gap in the literature while providing relevant climate application scenarios for adaptation planning in transport, health and environment sectors.

4. Conclusions

Using CMIP6 simulations, satellite-derived AOD, and in situ meteorological variables, this study proposes a novel BG-XG ensemble ML model to predict long-term visibility trends for Pakistan. An integrated BG-XG model was developed to address the key challenge of predicting future climate scenarios and regional visibility in an environmentally sensitive region with limited observational data. The validated model achieved accuracy rates of 95 % and 90 %on training and validation datasets, respectively. The proposed highly adaptable BG-XG model is designed for forecasting VR at local scales and is regionally applicable across diverse and complex topographies. Using the Boruta feature selection method, the variables AOD, SLP, RH, WS, and wind components (NW and EW) were identified as significant predictors of visibility, demonstrating the complex and interactive roles of aerosols and their ambient meteorological conditions.

Statistically, we found a significant national decline in average visibility at a rate of -281.3 m/year through 2100. The most significant decreases in visibility are projected near Karachi, Sindh and in the most northwestern regions of Pakistan. In contrast, some areas, such as Peshawar and certain coastal and high-altitude localities, may exhibit greater resilience and be less sensitive to climate shifts, indicating significant regional differences in climate sensitivity to visibility. These results suggest that visibility and air quality management in terms of adaptation to change must account for distinct development vulnerabilities of each locality.

The integration of long-term climate scenarios with AOD represents a scalable and actionable decision-support tool for policy-makers with a focus on transportation, public health, and environmental planning. The use of visibility projections could also provide value in achieving climate finance by providing performance measures of emission-reduction efforts. Although the model's strength lies in its inclusion of emissions and climate projections, its limitations, including the spatial resolution of input datasets and the reliance on a single CMIP6 GCM (CMCC-CM2-SR5), may limit generalizability. Future work should be expanded to include an ensemble of GCM outputs, separately examine other predictors, such as emission inventories or land-use change, and even explore hybrid ML approaches to enhance model performance and policy relevance.

CRediT authorship contribution statement

Sadaf Javed: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis. **Muhammad Imran Shahzad:** Writing – review & editing, Supervision, Project administration, Methodology, Data curation, Conceptualization. **Muhammad Zeeshaan Shahid:** Writing – review & editing, Validation, Resources. **Jun Wang:** Writing – review & editing, Supervision. **Imran Shahid:** Writing – review & editing, Validation, Supervision, Resources.

Declaration of competing interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rsase.2025.101712.

Data availability

Data will be made available on request.

References

Abbas, A., Ullah, S., Ullah, W., Waseem, M., Dou, X., Zhao, C., Karim, A., Zhu, J., Hagan, D.F.T., Bhatti, A.S., Ali, G., 2022. Evaluation and projection of precipitation in Pakistan using the coupled model intercomparison project phase 6 model simulations. Int. J. Climatol. 42, 6665–6684. https://doi.org/10.1002/joc.7602.

Akoglu, H., 2018. User's guide to correlation coefficients. Turkish J. Emergency Me. 18, 91–93. https://doi.org/10.1016/j.tjem.2018.08.001.

Ali, G., 2018. Climate change and associated spatial heterogeneity of Pakistan: empirical evidence using multidisciplinary approach. Sci. Total Environ. 634, 95–108. https://doi.org/10.1016/j.scitotenv.2018.03.170.

Aman, N., Manomaiphiboon, K., Pengchai, P., Suwanathada, P., Srichawana, J., Assareh, N., 2019. Long-term observed visibility in eastern Thailand: temporal variation, association with air pollutants and meteorological factors, and trends. Atmosphere 10, 122. https://doi.org/10.3390/atmos10030122.

Anil, S., Manikanta, V., Pallakury, A.R., 2021. Unravelling the influence of subjectivity on ranking of CMIP6 based climate models: a case study. Int. J. Climatol. 41, 5998–6016. https://doi.org/10.1002/joc.7164.

Anwar, M.N., Shabbir, M., Tahir, E., Iftikhar, M., Saif, H., Tahir, A., Murtaza, M.A., Khokhar, M.F., Rehan, M., Aghbashlo, M., Tabatabaei, M., Nizami, A.S., 2021. Emerging challenges of air pollution and particulate matter in China, India, and Pakistan and mitigating solutions. J. Hazard Mater. 416, 125851. https://doi.org/10.1016/j.jhazmat.2021.125851.

Ashfaq, M., Rastogi, D., Kitson, J., Abid, M.A., Kao, S.C., 2022. Evaluation of CMIP6 GCMs over the CONUS for downscaling studies. J. Geophys. Res. Atmos. 127, e2022JD036659. https://doi.org/10.1029/2022JD036659.

Aylmer, J., Ferreira, D., Feltham, D., 2022. Different mechanisms of arctic and antarctic sea ice response to ocean heat transport. Clim. Dyn. 59, 315–329. https://doi.org/10.1007/s00382-021-06131-x.

Bai, H., Gao, W., Seong, M., Yan, R., Wei, J., Liu, C., 2023. Evaluating and optimizing PM2.5 stations in yangtze river Delta from a spatial representativeness perspective. Appl. Geogr. 154, 102949. https://doi.org/10.1016/j.apgeog.2023.102949.

Ben Jabeur, S., Stef, N., Carmona, P., 2023. Bankruptcy prediction using the XGBoost algorithm and variable importance feature engineering. Comput. Econ. 61, 715–741. https://doi.org/10.1007/s10614-021-10227-1.

Burrows, E., 2020. Evaluating the Impact of Land Surface Properties on Simulated Dust Emissions and Air Quality. The University of Alabama in Huntsville.

Caputi, D.J., Faloona, I., Trousdell, J., Smoot, J., Falk, N., Conley, S., 2019. Residual layer ozone, mixing, and the nocturnal jet in California's san joaquin valley. Atmos. Chem. Phys. 19, 4721–4740. https://doi.org/10.5194/acp-19-4721-2019.

Carmona, J.M., Gupta, P., Lozano-García, D.F., Vanoye, A.Y., Hernández-Paniagua, I.Y., Mendoza, A., 2021. Evaluation of MODIS aerosol optical depth and surface data using an ensemble modeling approach to assess PM2.5 temporal and spatial distributions. Remote Sens. 13, 3102. https://doi.org/10.3390/rs13163102.

Chen, J., Zhang, Q., Huang, W., Lu, Z., Zhang, Z., Chen, F., 2021. Northwestward shift of the northern boundary of the East Asian summer monsoon during the mid-Holocene caused by orbital forcing and vegetation feedback. Quat. Sci. Rev. 268, 107136. https://doi.org/10.1016/j.quascirev.2021.107136.

Chen, T., Li, M., Luo, L., Deng, S., Zhou, R., Chen, D., 2020. Simulating the effects of land urbanization on regional meteorology and air quality in yangtze river Delta, China. Appl. Geogr. 120, 102228. https://doi.org/10.1016/j.apgeog.2020.102228.

Deepthi, B., Sivakumar, B., 2022. General circulation models for rainfall simulations: performance assessment using complex networks. Atmos. Res. 278, 106333. https://doi.org/10.1016/j.atmosres.2022.106333.

Grassi, D., Adriani, A., Mura, A., Atreya, S.K., Fletcher, L.N., Lunine, J.I., Orton, G.S., Bolton, S., Plainaki, C., Sindoni, G., Altieri, F., Cicchetti, A., Dinelli, B.M., Filacchione, G., Migliorini, A., Moriconi, M.L., Noschese, R., Olivieri, A., Piccioni, G., Sordini, R., Stefani, S., Tosi, F., Turrini, D., 2020. On the spatial distribution of minor species in jupiter's troposphere as inferred from juno JIRAM data. J. Geophys. Res.: Planets 125, e2019JE006206. https://doi.org/10.1029/2019JE006206.

Halder, S., Parekh, A., Chowdary, J.S., Gnanaseelan, C., Kulkarni, A., 2021. Assessment of CMIP6 models' skill for tropical Indian Ocean sea surface temperature variability. Int. J. Climatol. 41, 2568–2588. https://doi.org/10.1002/joc.6975.

Hu, M., Zhang, H., Wu, B., Li, G., Zhou, L., 2022. Interpretable predictive model for shield attitude control performance based on XGboost and SHAP. Sci. Rep. 12, 18226. https://doi.org/10.1038/s41598-022-22948-w.

Ismanto, H., Hartono, Marfai, M.A., 2019. Classification tree analysis (Gini-Index) smoke detection using Himawari-8 satellite data over sumatera-borneo maritime continent sout east Asia. IOP Conf. Ser. Earth Environ. Sci. 256, 012043. https://doi.org/10.1088/1755-1315/256/1/012043. NA.

Izanloo, M., Aslani, A., Zahedi, R., 2022. Development of a machine learning assessment method for renewable energy investment decision making. Appl. Energy 327, 120096. https://doi.org/10.1016/j.apenergy.2022.120096.

Javed, S., Imran Shahzad, M., Shahid, I., 2024. Unveiling the nexus between atmospheric visibility, remotely sensed pollutants, and climatic variables across diverse topographies: a data-driven exploration empowered by artificial intelligence. Atmos. Pollut. Res. 15, 102200. https://doi.org/10.1016/j.apr.2024.102200.

Javed, S., Shahzad, M.I., Abbas, S., Nazeer, M., 2023. Long-term variability of atmospheric visual range (1980–2020) over diverse topography of Pakistan. Remote Sens. 15, 46. https://doi.org/10.3390/rs15010046.

Kamworapan, S., Thao, P.T.B., Gheewala, S.H., Pimonsree, S., Prueksakorn, K., 2021. Evaluation of CMIP6 GCMs for simulations of temperature over Thailand and nearby areas in the early 21st century. Heliyon 7, e08263. https://doi.org/10.1016/j.heliyon.2021.e08263.

Képeš, E., Vrábel, J., Haddad, J. El, Harhira, A., Pořízka, P., Kaiser, J., 2023. Machine learning in the context of laser-induced breakdown spectroscopy. Laser Induced Breakdown Spectroscopy (LIBS) 1, 305–330. https://doi.org/10.1002/9781119758396.ch15.

Kim, H.H., Laufkötter, C., Lovato, T., Doney, S.C., Ducklow, H.W., 2023. Projected 21st-century changes in marine heterotrophic bacteria under climate change. Front. Microbiol. 14, 1049579. https://doi.org/10.3389/fmicb.2023.1049579.

Kumar, A., Jain, M., 2020. Using ensemble learning libraries. Ensemble learning for AI developers. https://doi.org/10.1007/978-1-4842-5940-5_5.

Lan, L., Xia, L., Hejjo, R., Wyon, D.P., Wargocki, P., 2020. Perceived air quality and cognitive performance decrease at moderately raised indoor temperatures Even when clothed for comfort. Indoor Air 30, 841–859. https://doi.org/10.1111/ina.12685.

Landolt, S.D., Gaydos, A., Porter, D., Divito, S., Jacobson, D., Schwartz, A.J., Thompson, G., Lave, J., 2020. Inferring the presence of freezing drizzle using archived data from the automated surface observing system (ASOS). J. Atmos. Ocean. Technol. 37, 2239–2250. https://doi.org/10.1175/JTECH-D-20-0098.1.

Li, Y., Li, S., Gao, W., Xu, W., Xu, Y., Wang, J., 2022. Exploring the effects of indoor temperature on college students' physiological responses, cognitive performance and a concentration index derived from EEG signals. Develop. Built Environ. 12, 100095. https://doi.org/10.1016/j.dibe.2022.100095.

Liu, J., Lu, Y., 2022. How well do CMIP6 models simulate the greening of the Tibetan Plateau? Remote Sens. 14, 4633. https://doi.org/10.3390/rs14184633.

Lovato, T., Peano, D., Butenschön, M., Materia, S., Iovino, D., Scoccimarro, E., Fogli, P.G., Cherchi, A., Bellucci, A., Gualdi, S., Masina, S., Navarra, A., 2022. CMIP6 simulations with the CMCC Earth system model (CMCC-ESM2). J. Adv. Model. Earth Syst. 14, e2021MS002814. https://doi.org/10.1029/2021MS002814.

Maleki, H., Sorooshian, A., Alam, K., Fathi, A., Weckwerth, T., Moazed, H., Jamshidi, A., Babaei, A.A., Hamid, V., Soltani, F., Goudarzi, G., 2022. The impact of meteorological parameters on PM10 and visibility during the Middle Eastern dust storms. J. Environ. Health Sci. Eng. 20, 495–507. https://doi.org/10.1007/s40201-022-00795-1.

Mastropietro, D.G., Moya, J.A., 2021. Design of Fe-based bulk metallic glasses for maximum amorphous diameter (dmax) using machine learning models. Comput. Mater. Sci. 188, 110230. https://doi.org/10.1016/j.commatsci.2020.110230.

Mbigi, D., Onyango, A.O., Mtewele, Z.F., Kiprotich, P., Xiao, Z., 2022. Coupled model intercomparison project phase 6 simulations of the spatial structure of rainfall variability over East Africa: evaluation and projection. Int. J. Climatol. 42, 9865–9885. https://doi.org/10.1002/joc.7868.

Meinshausen, M., Nicholls, Z.R.J., Lewis, J., Gidden, M.J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J.G., Daniel, J.S., John, A., Krummel, P.B., Luderer, G., Meinshausen, N., Montzka, S.A., Rayner, P.J., Reimann, S., Smith, S.J., Van Den Berg, M., Velders, G.J.M., Vollmer, M.K., Wang, R.H.

J., 2020. The shared socioeconomic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. Geosci. Model Dev. (GMD) 13, 3571–3605. https://doi.org/10.5194/gmd-13-3571-2020.

Merrifield, A.L., et al., 2023. Climate model selection by Independence, performance, and spread (ClimSIPS v1.0.1) for regional applications. Geosci. Model Dev. (GMD) 16, 4715–4747. https://doi.org/10.5194/gmd-16-4715-2023.

Miller, O.L., Putman, A.L., Alder, J., Miller, M., Jones, D.K., Wise, D.R., 2021. Changing climate drives future streamflow declines and challenges in meeting water demand across the Southwestern United States. J. Hydrol. X 11, 100074. https://doi.org/10.1016/j.hydroa.2021.100074.

Navarro-Racines, C., Tarapues, J., Thornton, P., Jarvis, A., Ramirez-Villegas, J., 2020. High-resolution and bias-corrected CMIP5 projections for climate change impact assessments. Sci. Data 7, 1–14. https://doi.org/10.1038/s41597-019-0343-8.

Nguyen, D., Sadeghnejad Barkousaraie, A., Bohara, G., Balagopal, A., McBeth, R., Lin, M.H., Jiang, S., 2021. A comparison of monte carlo dropout and bootstrap aggregation on the performance and uncertainty estimation in radiation therapy dose prediction with deep learning neural networks. Phys. Med. Biol. 66, 54002. https://doi.org/10.1088/1361-6560/abe04f.

Nguyen, N.H., Abellán-García, J., Lee, S., Garcia-Castano, E., Vo, T.P., 2022. Efficient estimating compressive strength of ultra-high-performance concrete using XGBoost model. J. Build. Eng. 52, 104302. https://doi.org/10.1016/j.jobe.2022.104302.

Nicolì, D., Bellucci, A., Ruggieri, P., Athanasiadis, P.J., Materia, S., Peano, D., Fedele, G., Hénin, R., Gualdi, S., 2023. The euro-mediterranean center on climate change (CMCC) decadal prediction system. Geosci. Model Dev. (GMD) 16, 179–197. https://doi.org/10.5194/gmd-16-179-2023.

Nazrul, F.B., et al., 2025. Future climate extremes in Bangladesh: insights from a multi-model ensemble of CMIP6 simulations. Earth Syst. Environ. https://doi.org/10.1007/s41748-024-00420-1 early online.

Pang, Z., Xie, B., An, Z., Wang, L., 2024. Spatial and moderating effects of greenspace on the association between air pollution and lung cancer incidence. Appl. Geogr. 164, 103207. https://doi.org/10.1016/j.apgeog.2024.103207.

Pimonsree, S., Kamworapan, S., Gheewala, S.H., Thongbhakdi, A., Prueksakorn, K., 2023. Evaluation of CMIP6 GCMs performance to simulate precipitation over southeast Asia. Atmos. Res. 282, 106522. https://doi.org/10.1016/j.atmosres.2022.106522.

Rahman, M.M., Shults, M., Hasan, M.G., Arshad, A., Alsubhi, Y.H., Alsubhi, A.S., 2024. Exploring the trends of aerosol optical depth and its relationship with climate variables over Saudi Arabia. Earth Syst. Environ. 8 (4), 1247–1265. https://doi.org/10.1007/s41748-024-00452-7.

Rahman, M.M., Shuo, W., Zhao, W., Xu, X., Zhang, W., Arshad, A., 2022. Investigating the relationship between air pollutants and meteorological parameters using satellite data over Bangladesh. Remote Sens. 14 (12). https://doi.org/10.3390/rs14122757.

Reshi, A.R., Pichuka, S., Tripathi, A., 2024. Applications of Sentinel-5P TROPOMI satellite sensor: a review. IEEE Sens. J. 24, 20312–20321. https://doi.org/10.1109/JSEN.2024.3401234.

Schroeter, S., et al., 2024. Global-scale future climate projections from ACCESS model contributions to CMIP6. J. Southern Hemisphere Earth Syst. Sci. 74. https://doi.org/10.1071/ES24012.

Semmler, T., Danilov, S., Gierz, P., Goessling, H.F., Hegewald, J., Hinrichs, C., Koldunov, N., Khosravi, N., Mu, L., Rackow, T., Sein, D.V., Sidorenko, D., Wang, Q., Jung, T., 2020. Simulations for CMIP6 with the AWI climate model AWI-CM-1-1. J. Adv. Model. Earth Syst. 12. https://doi.org/10.1029/2019MS002009.

Shahani, N.M., Zheng, X., Liu, C., Hassan, F.U., Li, P., 2021. Developing an XGBoost regression model for predicting young's modulus of intact sedimentary rocks for the stability of surface and subsurface structures. Front. Earth Sci. 9, 1–13. https://doi.org/10.3389/feart.2021.761990.

Shahid, M.Z., Shahzad, M.I., Jaweria, S., et al., 2025. Characteristics of aerosols and planetary boundary layer dynamics during biomass burning season. Nat. Hazards. https://doi.org/10.1007/s11069-025-07295-z.

Shahzad, M.I., Nichol, J.E., Campbell, J.R., Wong, M.S., 2018. Assessment of MODIS, OMI, MISR and CALIOP aerosol products for estimating surface visual range: a mathematical model for Hong Kong. Remote Sens. 10, 1333. https://doi.org/10.3390/rs10091333.

Shehadeh, A., Alshboul, O., Al Mamlook, R.E., Hamedat, O., 2021. Machine learning models for predicting the residual value of heavy construction equipment: an evaluation of modified decision tree, LightGBM, and XGBoost regression. Autom. ConStruct. 129, 103827. https://doi.org/10.1016/j.autcon.2021.103827.

Su, X., et al., 2022. Accuracy, stability, and continuity of AVHRR, SeaWiFS, MODIS, and VIIRS deep blue long-term land aerosol retrieval in Asia. Sci. Total Environ. 832, 155048. https://doi.org/10.1016/j.scitotenv.2022.155048.

Subbiah, S., Muthu Anbananthen, K.S., Thangaraj, S., Kannan, S., Chelliah, D., 2022. Intrusion detection technique in wireless sensor network using grid search random forest with boruta feature selection algorithm. J. Commun. Network. 24, 264–273. https://doi.org/10.23919/JCN.2022.000002.

Subhanullah, M., Ullah, U., Javed, M.F., Ullah, R., Akbar, T.A., Ullah, W., Baig, S.A., Aziz, M., Mohamed, A., Sajjad, R.U., 2022. Assessment and impacts of air pollution from brick kilns on public health in Northern Pakistan. Atmosphere 13, 1231. https://doi.org/10.3390/atmos13081231.

Sulistya, W., Nugraha, H.A., Budhi Dharmawan, G.S., Putra, M., Furqon, A., Sugiarto, S., Pramagusta, A.P., 2019. Development of automated weather observing system based on real-time web display. In: IES 2019 - International Electronics Symposium: the Role of Techno-Intelligence in Creating an Open Energy System Towards Energy Democracy. Proceedings. IEEE, pp. 577–581. https://doi.org/10.1109/ELECSYM.2019.8901542.

Sun, X., Zhao, T., Liu, D., Gong, S., Xu, J., Ma, X., 2020. Quantifying the influences of PM2.5 and relative humidity on change of atmospheric visibility over recent winters in an urban area of East China. Atmosphere 11, 461. https://doi.org/10.3390/ATMOS11050461.

Tabinda, A.B., Ali, H., Yasar, A., Rasheed, R., Mahmood, A., Iqbal, A., 2020. Comparative assessment of ambient air quality of major cities of Pakistan. MAPAN-J. Metrol. Soc. India 35, 25–32. https://doi.org/10.1007/s12647-019-00335-7.

Tebaldi, C., et al., 2021. Climate model projections from the scenario model intercomparison project (ScenarioMIP) of CMIP6. Earth System Dynam. 12, 253–293. https://doi.org/10.5194/esd-12-253-2021.

Thi, N., Kim, Y., Kwon, H., 2023. Uncorrected proof evaluation and selection of CMIP6 GCMs for long-term hydrological projections based on spatial performance assessment metrics across South Korea uncorrected proof 00, 1–17. https://doi.org/10.2166/wcc.2023.021.

Van Pelt, R.S., Tatarko, J., Gill, T.E., Chang, C., Li, J., Eibedingil, I.G., Mendez, M., 2020. Dust emission source characterization for visibility hazard assessment on lordsburg playa in Southwestern New Mexico, USA. Geoenvironmental Disasters 7, 34. https://doi.org/10.1186/s40677-020-00171-x.

Verbeek, T., Hincks, S., 2022. The 'just' management of urban air pollution? A geospatial analysis of low emission zones in brussels and London. Appl. Geogr. 140, 102642. https://doi.org/10.1016/j.apgeog.2022.102642.

Wang, Y., et al., 2021. Full-coverage spatiotemporal mapping of ambient PM2.5 and PM10 over China from Sentinel-5P and assimilated datasets: considering the precursors and chemical compositions. Sci. Total Environ. 793, 148535. https://doi.org/10.1016/j.scitotenv.2021.148535.

Wang, Y.C., Hsu, H.H., Chen, C.A., Tseng, W.L., Hsu, P.C., Lin, C.W., Chen, Y.L., Jiang, L.C., Lee, Y.C., Liang, H.C., Chang, W.M., Lee, W.L., Shiu, C.J., 2021. Performance of the Taiwan Earth system model in simulating climate variability compared with observations and CMIP6 model simulations. J. Adv. Model. Earth Syst. 13, e2020MS002353. https://doi.org/10.1029/2020MS002353.

Won, W.S., Oh, R., Lee, W., Kim, K.Y., Ku, S., Su, P.C., Yoon, Y.J., 2020. Impact of fine particulate matter on visibility at Incheon international airport, south korea. Aerosol Air Qual. Res. 20, 1048–1061. https://doi.org/10.4209/aaqr.2019.03.0106.

Wu, J., Zhang, S., Yang, Q., Zhao, D., Fan, W., Zhao, J., Shen, C., 2021. Using particle swarm optimization to improve visibility-aerosol optical depth retrieval method. npj Clim. Atmospheric Sci. 4, 49. https://doi.org/10.1038/s41612-021-00207-5.

Xi, X., 2021. Revisiting the recent dust trends and climate drivers using horizontal visibility and present weather observations. J. Geophys. Res. Atmos. 126. https://doi.org/10.1029/2021JD034687.

Yang, M., Xi, C., Wang, J., Feng, Z., Cao, S. jie, 2023. An interactive design framework for large-scale public buildings based on comfort and carbon abatement. Energy Build. 279, 112679. https://doi.org/10.1016/j.enbuild.2022.112679.

Zhang, Chuang, Wu, M., Chen, J., Chen, K., Zhang, Chi, Xie, C., Huang, B., He, Z., 2019. Weather visibility prediction based on multimodal fusion. IEEE Access 7, 74776–74786. https://doi.org/10.1109/ACCESS.2019.2920865.

Zhang, S., Wu, J., Fan, W., Yang, Q., Zhao, D., 2020. Review of aerosol optical depth retrieval using visibility data. Earth Sci. Rev. 200, 102986. https://doi.org/10.1016/j.earscirev.2019.102986.

Zhao, D., Chen, H., Yu, E., Luo, T., 2019. PM 2.5 /PM 10 ratios in eight economic regions and their relationship with meteorology in China. Adv. Meteorol. 2019, 1–15. https://doi.org/10.1155/2019/5295726.

Zhou, H., Xin, Y., Li, S., 2023. A diabetes prediction model based on boruta feature selection and ensemble learning. BMC Bioinf. 24, 224. https://doi.org/10.1186/s12859-023-05300-5.