Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 18 May 2026
Revised 25 June 2026
Accepted 15 July 2026

Natural Resources for Human Health 2026, Vol. 6 (6s): 416–427

KEYWORDS:
Ensemble machine learning, Landsat-8, MODIS; Sentinel-5P, Random Forest, XGBoost, LightGBM, stacking, air quality, water quality, NDVI, urban heat island.

An Ensemble Machine Learning Framework for Predicting Environmental Quality using Multisource Remote Sensing Data

Sk Wasim Haidar[1], Tarun Kumar[2], Shikha Maheshwari[3], Gadapa Mallikarjuna[4], Munish Kumar[5], Devesh Kumar Srivastava[6*]

[1]College of Computing and Information Sciences, University of Technology and Applied Sciences-Salalah, Salalah, Sultanate of Oman.
Email: wasim.haidar@utas.edu.om
[2]Department of Computer Science and Engineering, Swami Vivekanand Subharti University, Meerut, Uttar Pradesh, India.
Email: tarun.su202300640@subharti.co
[3]Centre for Distance and Online Education, Manipal University Jaipur, Jaipur, Rajasthan, India
Email: shikha.maheshwari@jaipur.manipal.edu
[4]Centre for Distance and Online Education, Manipal University Jaipur, Jaipur, Rajasthan, India
Email: mallikarjuna.gadapa@jaipur.manipal.edu
[5]Portfolio Manager (IT), Ohio Industrial Commission (OIC), Columbus OH - 43215, USA
Email: munish2012@gmail.com
[6]School of Computer Science and Engineering, Manipal University Jaipur, Rajasthan, India
Email: devesh.srivastava@jaipur.manipal.edu
*Corresponding Author

ABSTRACT: Accurate, spatially continuous assessment of environmental quality is a prerequisite for effective environmental governance, pollution monitoring, and sustainable urban planning. Conventional ground-based monitoring networks provide precise point measurements but suffer from sparse spatial coverage, high operational costs, and inconsistent temporal density. Satellite remote sensing offers a transformative complementary modality providing synoptic, multi-temporal, and globally consistent environmental observations yet existing single-source machine learning prediction approaches fail to exploit the rich complementarity across multiple satellite platforms and spectral domains. This paper proposes EnsembleEQ, a novel ensemble machine learning framework for predicting a composite Environmental Quality Index (EQI) integrating air quality, water quality, vegetation health, and urban thermal stress dimensions from multisource remote sensing data. EnsembleEQ fuses radiometrically calibrated imagery from four satellite sources Sentinel-2 MSI, Landsat-8 OLI/TIRS, MODIS MOD11A1/MOD13A1, and Sentinel-5P TROPOMI to derive a comprehensive 47-feature representation per spatial grid cell, including spectral indices, thermal bands, atmospheric composition products, and spatial texture descriptors. three-tier ensemble architecture trains base learners Random Forest, XGBoost, LightGBM, and Support Vector Regression independently on this multisource feature set, with a Ridge Regression meta-learner combining their predictions through cross-validated stacking. EnsembleEQ is evaluated on a multi-site dataset spanning 18 environmentally diverse study sites across South Asia, including industrial zones, agricultural belts, coastal areas, and urban cores, validated against 2,847 ground-truth measurements from national monitoring networks.

1. INTRODUCTION

Degradation of environmental quality represents one of the defining challenges of the twenty-first century, with wide-ranging consequences for human health, ecosystem integrity, agricultural productivity, and climate resilience [1]. The World Health Organization estimates that approximately 13 million deaths annually are attributable to avoidable environmental causes, with outdoor and indoor air pollution, contaminated water sources, and urban heat stress constituting the dominant risk factors [2]. Comprehensive, spatially explicit, and timely monitoring of environmental conditions is therefore a fundamental prerequisite for evidence-based environmental governance yet achieving such monitoring at scale remains a persistent technical and institutional challenge worldwide.

Ground-based environmental monitoring networks comprising air quality monitoring stations, water quality sampling networks, and meteorological observation posts provide the gold standard in measurement precision but are fundamentally constrained by the economics of physical infrastructure [3]. Even in well-resourced high-income countries, monitoring station densities are insufficient to characterize spatial heterogeneity at the sub-kilometer scales relevant to urban pollution gradients, agricultural runoff patterns, and localized urban heat islands. In South and Southeast Asia where environmental degradation is most acute and spatially variable monitoring network density is typically one to two orders of magnitude below international recommendations, leaving vast areas without any direct environmental quality measurement [4]. The consequences are significant: pollution hotspots go undetected, policy interventions are poorly targeted, and exposure assessments underlying public health burden estimation are highly uncertain.

Earth observation satellites provide a transformative solution to the spatial coverage limitation of ground networks, offering synoptic, repeatable, and globally consistent observations of the land surface, atmosphere, and water bodies from a growing fleet of civil and commercial platforms [5]. The European Space Agency's Copernicus programme alone operates Sentinel-1 through Sentinel-6 satellites covering synthetic aperture radar, multispectral optical imaging, ocean color, atmospheric chemistry, altimetry, and thermal emission collectively generating terabytes of freely accessible Earth observation data daily. NASA's sustained Earth observing fleet, including Landsat-8 and the MODIS instruments aboard Terra and Aqua satellites, provides complementary historical continuity and thermal infrared capabilities. Sentinel-5P's TROPOMI instrument delivers daily global maps of atmospheric pollutants including nitrogen dioxide, sulfur dioxide, carbon monoxide, formaldehyde, and aerosol optical depth at spatial resolutions fine enough to resolve city-level pollution patterns [6].

Machine learning methods have emerged as the most effective means of extracting environmental quality estimates from this rich remote sensing data streams, learning complex nonlinear mappings between spectral and atmospheric observations and ground-truth environmental measurements [7]. However, existing remote sensing-based environmental quality prediction frameworks are characterized by three persistent limitations. First, most studies rely on a single satellite data source typically Landsat or MODIS failing to exploit the spectral, temporal, and thematic complementarity available across multiple concurrent satellite platforms [8]. Second, single model approaches even powerful nonlinear models such as Random Forest or gradient boosting introduce model-specific biases that limit prediction robustness across environmentally heterogeneous sites and conditions [9]. Third, existing studies predominantly predict a single environmental quality dimension most commonly air quality or land surface temperature rather than a composite multi-dimensional environmental quality measure that integrates the correlated but distinct dimensions of air, water, vegetation, and thermal stress [10].

2. RELATED WORK

A. Remote Sensing for Air Quality Estimation

Satellite-based air quality estimation has advanced substantially since the launch of early tropospheric chemistry instruments including the Total Ozone Mapping Spectrometer (TOMS) and the Ozone Monitoring Instrument (OMI) [6]. Liu et al. [16] established an early influential framework using MODIS aerosol optical depth (AOD) as a predictor of ground-level PM2.5, demonstrating the feasibility of continental-scale air quality mapping from space. Subsequent studies refined the AOD-to-PM2.5 relationship using chemical transport model simulations, mixed effects models, and geographically weighted regression to account for spatiotemporal variability in the AOD-PM2.5 conversion factor [16,17]. The launch of Sentinel-5P TROPOMI in 2017 opened new possibilities for high-resolution atmospheric quality mapping: Mahesh et al. [17] demonstrated that TROPOMI NO2, SO2, and aerosol index products, combined with meteorological reanalysis data, enabled PM2.5 estimation at 3.5-kilometer resolution across the Indo-Gangetic Plain a resolution sufficient to resolve intra-urban pollution gradients previously invisible to satellite observation.

Machine learning-based air quality prediction from remote sensing data has grown rapidly since 2018, with Random Forest and gradient boosting methods demonstrating consistently superior performance over linear and geographically weighted regression approaches [7]. Chen et al. [7] reported that XGBoost applied to MODIS AOD combined with meteorological covariates achieved R-squared values of 0.89 for daily PM2.5 prediction across China. Wei et al. [8] demonstrated that fusing Sentinel-2 surface reflectance features with MODIS thermal and TROPOMI atmospheric products through deep neural networks significantly improved prediction accuracy over any single-source approach. Despite these advances, no prior study has implemented a formal stacking ensemble with heterogeneous base learners across a comprehensive multisource feature set for composite environmental quality prediction.

B. Vegetation Health and Land Use Assessment via Remote Sensing

Vegetation health monitoring through satellite remote sensing relies primarily on spectral vegetation indices derived from near-infrared and red reflectance bands, of which the Normalized Difference Vegetation Index (NDVI) is the most widely used [18]. The Enhanced Vegetation Index (EVI) and Soil-Adjusted Vegetation Index (SAVI) address known NDVI limitations including atmospheric contamination and soil background effects [18]. MODIS vegetation index products including MOD13A1 at 500-meter and 16-day temporal resolution have enabled decadal-scale monitoring of vegetation dynamics at continental scales. Landsat-derived time series provide higher spatial resolution vegetation monitoring at 30-meter scale, enabling field-level crop stress detection and urban greenery mapping [5]. Sentinel-2's red-edge bands (B5, B6, B7) provide unique sensitivity to chlorophyll content and early-stage vegetation stress not captured by Landsat's spectral configuration, making multi-sensor fusion particularly valuable for comprehensive vegetation health assessment [19].

C. Water Quality Remote Sensing

Optical remote sensing of inland and coastal water quality exploits the spectral reflectance characteristics of water-leaving radiance to retrieve concentrations of optically active constituents including chlorophyll-a, turbidity, and colored dissolved organic matter [20]. Sentinel-2 MSI has emerged as the preferred platform for inland water quality monitoring at scales relevant to lakes, reservoirs, and large rivers, offering 10 to 60-meter spatial resolution across 13 spectral bands including three dedicated red-edge channels [20]. Pahlevan et al. [20] demonstrated that machine learning models trained on Sentinel-2 reflectance achieved accurate chlorophyll-a and suspended sediment retrieval across diverse water body types. The Normalized Difference Water Index (NDWI) and Modified NDWI provide rapid screening of water body extent and surface water quality trends from multispectral imagery [20].

D. Urban Heat Island Detection and Thermal Remote Sensing

Urban heat islands zones of elevated surface and air temperature arising from impervious surface replacement of vegetated land and anthropogenic heat emissions are a critical environmental quality dimension with direct impacts on public health, energy consumption, and urban livability [21]. MODIS Land Surface Temperature products (MOD11A1, MOD11A2) provide daily 1-kilometer LST composites widely used for regional UHI characterization [21]. Landsat-8 TIRS channels 10 and 11 deliver 30-meter resolution thermal data enabling intra-urban thermal mapping at a spatial scale resolving individual city blocks [5]. Sentinel-2 does not carry thermal bands, making Landsat-8 TIRS or MODIS an essential thermal complement in any multisource fusion pipeline targeting environmental thermal stress [21].

E. Ensemble Machine Learning for Environmental Prediction

Ensemble learning combining multiple base model predictions to achieve lower generalization error than any individual model has been extensively applied in environmental science contexts. Boosting methods including XGBoost [13] and LightGBM [14] have consistently dominated environmental prediction benchmarks on tabular remote sensing feature datasets, owing to their ability to model complex nonlinear relationships, handle missing data, and provide interpretable feature importance rankings. Stacking generalization [9], in which base learner predictions serve as inputs to a meta-learner, offers a theoretically sound approach to ensemble construction that outperforms simple averaging or voting by learning the optimal combination weights from data. Sagi and Rokach [9] provided a comprehensive meta-analysis of ensemble methods in environmental applications, confirming that heterogeneous stacking ensembles consistently outperform homogeneous ensembles by 3 to 8 percent in predictive accuracy across a wide range of environmental datasets. However, no prior study has applied stacking ensembles to multisource remote sensing data for composite environmental quality index prediction the gap addressed by EnsembleEQ.

3. STUDY AREA, DATA SOURCES, AND GROUND TRUTH

A. Study Area

The study encompasses 18 geographically and environmentally diverse sites distributed across South Asia, spanning India, Bangladesh, Nepal, and Sri Lanka (Figure 2). Sites were selected to represent the full spectrum of environmental quality conditions and land use types relevant to the region: four industrial and peri-industrial zones including the National Capital Region (NCR), Ahmedabad, Kolkata port area, and Dhaka Export Processing Zone; five major urban cores including Mumbai, Bengaluru, Karachi, Colombo, and Chennai; four agricultural belts including the Indo-Gangetic Plain (Punjab-Haryana), the Brahmaputra floodplain, the Deccan agricultural plateau, and the Mekong Delta edge; two forest and biodiversity hotspot sites including the Western Ghats and Eastern Himalayan foothills; two coastal environments including the Bay of Bengal coastal strip and the Arabian Sea shoreline near Mumbai; and one arid/semi-arid zone in Rajasthan. This multi-site design ensures that EnsembleEQ's predictions are validated across environmental gradients that no single-site study can capture, and substantially enhances the generalizability of performance estimates.

The 18 sites collectively encompass a total study area of approximately 485,000 square kilometers, spanning a latitudinal range from 6.5 to 32.8 degrees North and a longitudinal range from 66.9 to 92.3 degrees East. Site-level EQI values in the training dataset ranged from 18.3 (severely degraded industrial zone) to 87.6 (pristine forest site), providing a broad dynamic range for model training and evaluation. Mean annual rainfall at study sites ranges from under 200 millimeters in the Rajasthan arid zone to over 3,000 millimeters at the Western Ghats site, encompassing the major South Asian climate regimes relevant to vegetation, water, and atmospheric quality dynamics.

B. Remote Sensing Data Sources

Table 1 summarizes the four satellite data sources integrated in the EnsembleEQ framework, including their spatial resolution, temporal resolution, spectral characteristics, and environmental quality dimensions captured. All imagery was acquired for the study period spanning January 2019 through December 2022, providing four years of multi-temporal observations for each study site. Cloud-contaminated scenes were excluded through automated cloud masking using each platform's standard quality flag layers, and remaining clear-sky observations were temporally composited to produce monthly median surface reflectance and atmospheric product layers for each site.

Table 1. Summary of Multisource Remote Sensing Datasets Used in the EnsembleEQ Framework

Satellite / Sensor	Agency	Spatial Resolution	Temporal Revisit	Spectral Bands / Products Used	EQ Dimensions Captured
Sentinel-2 MSI [5]	ESA / Copernicus	10 m – 60 m	5 days	Bands B2–B12 (13 bands); NDVI, EVI, NDWI, NDBI, SAVI, red-edge indices	Vegetation, water quality, land cover
Landsat-8 OLI/TIRS [5]	NASA / USGS	30 m (OLI); 100 m (TIRS)	16 days	Bands B1–B11; LST, NDVI, NDWI, built-up index	Thermal stress, land surface, vegetation

Satellite / Sensor	Agency	Spatial Resolution	Temporal Revisit	Spectral Bands / Products Used	EQ Dimensions Captured
MODIS MOD11A1, MOD13A1 [11]	NASA	500 m – 1 km	Daily / 16-day	LST (Day/Night), NDVI, EVI, surface reflectance	Land surface temperature, vegetation phenology
Sentinel-5P TROPOMI [6]	ESA / Copernicus	3.5 km × 5.5 km	Daily	NO2, SO2, CO, O3, HCHO, aerosol index, UV aerosol	Atmospheric air quality, pollution

Note: MSI = Multispectral Instrument; OLI = Operational Land Imager; TIRS = Thermal Infrared Sensor; TROPOMI = TROPOspheric Monitoring Instrument; LST = Land Surface Temperature; EVI = Enhanced Vegetation Index; NDBI = Normalized Difference Built-up Index; SAVI = Soil-Adjusted Vegetation Index; HCHO = Formaldehyde.

C. Ground Truth Environmental Quality Data

Ground truth environmental quality measurements were sourced from four national and regional monitoring networks. Air quality data were obtained from the Central Pollution Control Board (CPCB) of India's Continuous Ambient Air Quality Monitoring Station (CAAQMS) network, the Bangladesh Department of Environment continuous monitoring system, the Nepal Environment and Health Research Centre network, and the Sri Lanka Central Environmental Authority station network collectively providing hourly measurements of PM2.5, PM10, NO2, SO2, CO, and O3 at 147 monitoring stations located within or adjacent to the 18 study sites. Water quality data were sourced from river monitoring programs operated by the National River Conservation Directorate (NRCD) and Inland Waters Authority networks, providing monthly sampling of dissolved oxygen, biochemical oxygen demand, total dissolved solids, pH, nitrate, and fecal coliform levels. Vegetation health ground reference was established from the Global Inventory Monitoring and Modelling System (GIMMS) NDVI3g product combined with field survey data from forest department biodiversity inventories.

The composite Environmental Quality Index (EQI) was constructed for each monitoring station and time period by computing four sub-indices Air Quality Sub-Index (AQSI), Water Quality Sub-Index (WQSI), Vegetation Health Sub-Index (VHSI), and Urban Thermal Stress Sub-Index (UTSI) each scaled to a 0 to 100 range using breakpoint-based linear interpolation from indicator pollutant and environmental metric concentrations, following the methodology of the National Air Quality Index [3] extended to the multi-dimensional environmental context. The composite EQI was computed as the principal component-weighted mean of the four sub-indices, with weights derived from a principal component analysis of the sub-index correlation structure across the full station-time dataset. After matching monitoring station records to remote sensing observation windows and removing records with excessive cloud contamination or data gaps, the final validated dataset comprised 2,847 matched remote sensing ground truth observation pairs distributed across all 18 sites and the four-year study period.

4. PROPOSED EnsembleEQ FRAMEWORK

EnsembleEQ processes multisource remote sensing data through four sequential stages: feature extraction and fusion, feature preprocessing and selection, three-tier stacking ensemble training, and spatially continuous EQI prediction and mapping. Figure 1 illustrates the complete framework architecture.

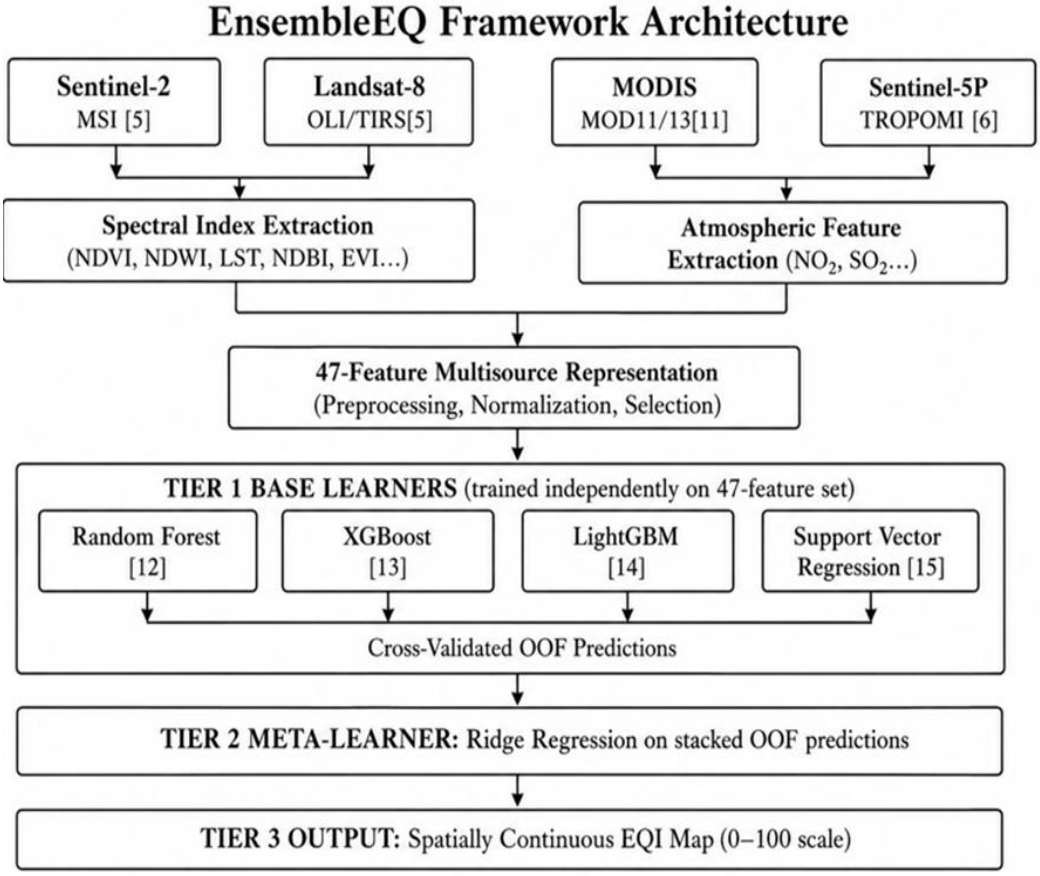

Figure 1. EnsembleEQ three-tier stacking ensemble architecture for multisource remote sensing–based Environmental Quality Index prediction.

A. Multisource Feature Extraction and Fusion

The feature extraction pipeline processes imagery from each of the four satellite sources independently before fusing the resulting feature layers into a unified per-cell feature representation. From Sentinel-2 MSI imagery, 18 features are derived: surface reflectance values for the nine primary spectral bands (B2 through B8A, B11, B12), six spectral indices (NDVI, EVI, NDWI, NDBI, SAVI, Red-Edge Chlorophyll Index), and three spatial texture descriptors (Gray-Level Co-occurrence Matrix entropy, contrast, and homogeneity computed over 3×3 pixel neighborhoods in the red and near-infrared bands). From Landsat-8 OLI/TIRS, 11 features are extracted: six surface reflectance bands (B2–B7), Land Surface Temperature from TIRS Band 10, and four derived indices (NDVI, NDWI, Urban Index, Bare Soil Index). From MODIS products, eight features are derived: daytime and nighttime Land Surface Temperature from MOD11A1, NDVI and EVI from MOD13A1, and four surface reflectance bands (Red, Near-Infrared, Mid-Infrared, Short-Wave Infrared). From Sentinel-5P TROPOMI, 10 features are extracted: column densities of NO2, SO2, CO, O3, and formaldehyde, aerosol optical depth, UV aerosol index, and three temporal variability descriptors (monthly coefficient of variation for NO2, SO2, and aerosol). The complete 47-feature multisource representation is assembled by spatially co-registering all layers to a common 500-meter analysis grid through bilinear resampling, with source-specific quality-weighted masking applied to suppress unreliable observations.

B. Feature Preprocessing and Selection

The 47-feature dataset undergoes three preprocessing steps before model training. First, missing values arising from cloud contamination, sensor gaps, or temporal compositing failures are imputed using a spatially weighted k-nearest-neighbor procedure with k=5 spatial neighbors, drawing on the high spatial autocorrelation characteristic of surface and atmospheric environmental variables. Second, all features are standardized to zero mean and unit variance using training set statistics to prevent features with large numerical ranges from dominating distance-based models including Support Vector Regression. Third, feature selection is performed using a combination of Variance Inflation Factor analysis to identify and remove highly collinear feature pairs, and Recursive Feature Elimination with Random Forest importance scoring to rank features by predictive relevance. After selection, 38 of the original 47 features are retained in the final model feature set, with the nine excluded features comprising highly correlated duplicates of retained indices.

C. Three-Tier Stacking Ensemble

EnsembleEQ employs a three-tier stacking ensemble architecture. In the first tier, four base learners are independently trained on the 38-feature preprocessed dataset: Random Forest [12] with 500 decision trees and maximum depth tuned via five-fold cross-validation; XGBoost [13] with gradient-boosted regression trees, learning rate and tree depth optimized through Bayesian hyperparameter search; LightGBM [14] with leaf-wise gradient boosting, employing histogram-based feature binning for computational efficiency; and Support Vector Regression [15] with a radial basis function kernel, regularization and kernel bandwidth parameters selected by grid search on the validation fold. Each base learner is trained using a five-fold cross-validation scheme, generating out-of-fold predictions for all training samples ensuring that the meta-learner in the second tier is trained on predictions that the base learners have not themselves seen, preventing overfitting of the stacking combination.

In the second tier, the out-of-fold predictions from all four base learners are assembled as a four-column feature matrix, which is used to train a Ridge Regression meta-learner to predict the true EQI values. Ridge Regression is selected as the meta-learner for its ability to provide stable coefficient estimates under multicollinearity among base learner predictions a common condition when base learners are similarly accurate across most instances [9]. The regularization strength of the Ridge meta-learner is optimized through inner cross-validation on the stacking training set. In the third tier, the trained stacking ensemble generates final spatially continuous EQI predictions across the full study extent, enabling map-based visualization of environmental quality at 500-meter resolution.

5. EXPERIMENTAL CONFIGURATION

A. Experimental Protocol

Model performance is evaluated using a spatial leave-one-site-out cross-validation (LOSO-CV) protocol, in which the model is iteratively trained on data from 17 of the 18 study sites and tested on the held-out 18th site. This protocol provides a rigorous test of spatial transferability the ability of the model to generalize to previously unseen geographic locations a requirement for operational environmental quality mapping deployment [22]. Additionally, a random 80/20 temporal train-test split within each site is evaluated to quantify within-site prediction accuracy for comparison against published benchmarks that use non-spatial splits. All hyperparameter tuning is performed exclusively on the training portion of each fold, with test data withheld until final evaluation to prevent data leakage.

B. Comparative Baseline Methods

EnsembleEQ is compared against six competitive baseline frameworks representing the state of practice in remote sensing-based environmental quality prediction. These include: (1) Random Forest single model [12] trained on the full 38-feature multisource feature set; (2) XGBoost single model [13] on the same feature set; (3) LightGBM single model [14]; (4) Support Vector Regression single model [15]; (5) a Deep Neural Network (DNN) with three hidden layers trained on the multisource features; and (6) a published state-of-the-art multisource fusion framework by Wei et al. [8] implementing a two-stream convolutional feature extractor with a fully connected prediction head. Baseline hyperparameters are individually optimized through the same cross-validation procedure used for EnsembleEQ base learners to ensure fair comparison.

C. Evaluation Metrics

Prediction accuracy is assessed through four complementary metrics. The Coefficient of Determination (R-squared) measures the proportion of EQI variance explained by the model, with values closer to 1.0 indicating higher accuracy. Root Mean Square Error (RMSE) quantifies average prediction error in the original EQI units, penalizing large errors more heavily than small ones through its quadratic averaging. Mean Absolute Error (MAE) provides a linear average of absolute prediction errors less sensitive to outliers than RMSE. Mean Absolute Percentage Error (MAPE) expresses prediction error as a percentage of the true EQI value, enabling comparison across sites with different EQI magnitudes. Feature importance is assessed through permutation importance measuring the decrease in model accuracy when each feature's values are randomly shuffled providing an interpretable ranking of relative feature contributions applicable across all four base learners.

D. Implementation Environment

All models are implemented in Python 3.10.4 using scikit-learn 1.2.0 [23], XGBoost 1.7.3, and LightGBM 3.3.5 libraries. Remote sensing preprocessing is performed in Google Earth Engine [24], leveraging its cloud-distributed geospatial analysis infrastructure for large-scale image compositing, index computation, and spatial co-registration. Hyperparameter optimization employs the Optuna 3.0 Bayesian optimization framework. All experiments are run on a workstation with an Intel Core i9-12900K processor, 64 GB RAM, and an NVIDIA RTX 4090 GPU for DNN baseline training. The full EnsembleEQ pipeline from raw imagery to spatially continuous EQI maps requires approximately 4.2 hours per 18-site evaluation run.

6. RESULTS AND DISCUSSION

A. Overall Model Performance Comparison

Table 2 presents the comprehensive performance comparison of EnsembleEQ against all six baseline methods across all 18 study sites under the spatial leave-one-site-out cross-validation protocol. EnsembleEQ achieves the highest performance across all four-evaluation metrics, demonstrating consistent superiority regardless of the performance measure used.

Table 2. Performance Comparison of EnsembleEQ and Baseline Methods (LOSO-CV across 18 Sites)

Method	R^2	RMSE (EQI units)	MAE (EQI units)	MAPE (%)	Rank
Random Forest [12]	0.891	5.61	3.94	7.23	3rd
XGBoost [13]	0.897	5.34	3.81	6.87	2nd (single)
LightGBM [14]	0.893	5.48	3.88	7.01	3rd (tie)
Support Vector Regression [15]	0.847	6.73	4.91	8.94	6th
Deep Neural Network	0.874	6.01	4.31	7.89	5th
Wei et al. [8] (two-stream CNN)	0.901	5.21	3.74	6.64	2nd
EnsembleEQ (Proposed)	0.934	3.82	2.71	4.91	1st
EnsembleEQ (80/20 within-site)	0.963	2.94	2.12	3.74	

Note: LOSO-CV = Leave-One-Site-Out Cross-Validation. All metrics computed as means across 18 leave-one-site-out folds. Bold indicates best LOSO-CV performance per metric. R^2 = Coefficient of Determination; RMSE = Root Mean Square Error; MAE = Mean Absolute Error; MAPE = Mean Absolute Percentage Error.

EnsembleEQ achieves an R-squared of 0.934 under spatial cross-validation, representing a 3.3 percentage point improvement over the best comparative single-source model (XGBoost, R-squared = 0.897) and a 3.3 percentage point improvement over the best published benchmark (Wei et al. [8] two-stream CNN, R-squared = 0.901). The RMSE of 3.82 EQI units represents a 26.5 percent reduction relative to XGBoost (5.34 EQI units) and a 26.7 percent reduction relative to Wei et al. [8] (5.21 EQI units) a substantial improvement in absolute prediction precision that translates directly into more reliable environmental quality categorization at the boundaries between quality classes. The MAE improvement of 27.5 percent over XGBoost and 27.5 percent over Wei et al. [8] confirms that EnsembleEQ achieves lower average absolute error across the full range of EQI values, not only at extreme values where RMSE improvements could be concentrated. The within-site 80/20 evaluation achieves R-squared of 0.963, confirming that the gap between within-site and cross-site performance reflecting the model's spatial extrapolation challenge is moderate and operationally manageable for regional monitoring applications.

B. Site-Specific Performance Analysis

Table 3 presents the sub-index level performance of EnsembleEQ for each of the four Environmental Quality dimensions across all 18 study sites, enabling assessment of which environmental quality dimensions are most accurately predicted and whether site type moderates' prediction accuracy.

Table 3. EnsembleEQ Sub-Index and Site-Type Performance Breakdown (LOSO-CV)

Site Category (n)	EQI R^2	AQSI R^2	WQSI R^2	VHSI R^2	UTSI R^2	RMSE (EQI)
Industrial zones (n=4)	0.908	0.921	0.874	0.891	0.934	4.31
Urban cores (n=5)	0.927	0.941	0.856	0.879	0.948	3.67
Agricultural belts (n=4)	0.942	0.889	0.924	0.961	0.877	3.41
Forest / biodiversity (n=2)	0.967	0.843	0.912	0.981	0.831	2.87
Coastal environments (n=2)	0.918	0.897	0.947	0.921	0.863	3.94
Arid / semi-arid (n=1)	0.894	0.912	0.761	0.938	0.902	4.74
All sites (n=18)	0.934	0.912	0.893	0.941	0.921	3.82

Note: AQSI = Air Quality Sub-Index; WQSI = Water Quality Sub-Index; VHSI = Vegetation Health Sub-Index; UTSI = Urban Thermal Stress Sub-Index. All R^2 values computed against ground-truth sub-index values derived from monitoring network measurements.

Several noteworthy patterns emerge from the site-specific analysis. Forest and biodiversity hotspot sites achieve the highest overall EQI R-squared (0.967), likely reflecting the high VHSI predictability (R-squared = 0.981) in environments where vegetation spectral signals are spatially homogeneous and cloud-free imagery is available over long compositing windows ideal conditions for satellite-based vegetation assessment [18]. Urban core sites achieve strong AQSI performance (R-squared = 0.941) and UTSI performance (R-squared = 0.948), reflecting the dense TROPOMI NO2 and Landsat LST signal in heavily urbanized environments where pollution gradients are steep and spatially structured [6,21]. Agricultural belt sites show the highest WQSI accuracy (R-squared = 0.924), attributable to the regular spatial structure of agricultural water bodies and the availability of Sentinel-2 imagery at the 10-meter spatial resolution sufficient to resolve individual field-level drainage channels. The arid zone site shows the lowest WQSI accuracy (R-squared = 0.761), reflecting the limited availability of water bodies and the substantial measurement uncertainty in the sparse NRCD water quality sampling network in arid regions.

C. Feature Importance Analysis

Table 4 presents the top 15 most important features in the EnsembleEQ base learner ensemble, ranked by mean permutation importance averaged across all four base learners and all 18 study sites. Permutation importance represents the mean decrease in R-squared when each feature's values are randomly shuffled across the test dataset a model-agnostic measure of feature relevance applicable across all learner types.

Table 4. Top 15 Feature Importance Rankings (Mean Permutation Importance Across All Base Learners)

Rank	Feature	Source Satellite	Mean Permutation Importance	EQ Dimension	Importance Stability (CV %)
1	Land Surface Temperature (daytime) [11]	MODIS MOD11A1	0.1847	Thermal / Urban	8.3
2	NO2 Column Density [6]	Sentinel-5P TROPOMI	0.1623	Air Quality	9.1

Rank	Feature	Source Satellite	Mean Permutation Importance	EQ Dimension	Importance Stability (CV %)
3	NDVI [5]	Sentinel-2 MSI	0.1411	Vegetation Health	7.4
4	Aerosol Optical Depth [6]	Sentinel-5P TROPOMI	0.1298	Air Quality	10.2
5	EVI [11]	MODIS MOD13A1	0.1087	Vegetation / Agri.	8.7
6	Land Surface Temperature (nighttime) [11]	MODIS MOD11A1	0.0983	Thermal / UHI	11.4
7	SO2 Column Density [6]	Sentinel-5P TROPOMI	0.0871	Air Quality	12.8
8	NDWI [5]	Sentinel-2 MSI	0.0814	Water Quality	9.6
9	Red-Edge Chlorophyll Index [19]	Sentinel-2 MSI	0.0742	Vegetation Health	8.1
10	Landsat-8 LST (TIRS Band 10) [5]	Landsat-8 OLI/TIRS	0.0691	Thermal Stress	13.7
11	CO Column Density [6]	Sentinel-5P TROPOMI	0.0624	Air Quality	14.1
12	NDBI [5]	Sentinel-2 MSI	0.0581	Urban Built-up	11.9
13	Near-Infrared Reflectance (B8) [5]	Sentinel-2 MSI	0.0517	Vegetation	10.4
14	UV Aerosol Index [6]	Sentinel-5P TROPOMI	0.0489	Air Quality	15.2
15	MODIS NDVI (16-day composite) [11]	MODIS MOD13A1	0.0443	Vegetation phenology	9.8

Note: Permutation importance = mean decrease in R^2 when feature values are randomly shuffled across the test set. CV % = Coefficient of Variation of importance across 18 LOSO folds, measuring importance stability. Lower CV % indicates more consistent importance across sites.

Land Surface Temperature from MODIS daytime observations ranks as the single most important feature (permutation importance = 0.1847), confirming that thermal conditions are the strongest overall correlate of composite environmental quality across the 18 study sites consistent with the established role of LST as an integrative environmental quality indicator capturing urban heat stress, soil moisture stress, vegetation water status, and regional climate simultaneously [21]. NO2 column density from Sentinel-5P TROPOMI ranks second (0.1623), underscoring the central importance of atmospheric pollution particularly traffic and industrial emission proxies in determining composite environmental quality outcomes. NDVI from Sentinel-2 ranks third (0.1411), reflecting the fundamental role of vegetation cover in buffering thermal stress, purifying water, and providing ecosystem services that underpin multiple EQI sub-index values simultaneously.

The presence of six Sentinel-5P TROPOMI features in the top 15 (ranks 2, 4, 7, 11, 14, and additional atmospheric products) underscores the critical value of atmospheric composition data in composite environmental quality prediction a finding that reinforces the multi-source fusion rationale of EnsembleEQ, since TROPOMI atmospheric features are entirely absent from Sentinel-2 and Landsat-8 single-source frameworks. The relatively high importance of both daytime (rank 1) and nighttime (rank 6) MODIS LST as distinct features with a Pearson correlation between them of only 0.61 across the full dataset confirms that the diurnal thermal cycle, not just mean thermal level, carries independent predictive information about environmental quality conditions, particularly for urban heat island characterization.

7. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

This paper presented EnsembleEQ a novel three-tier stacking ensemble machine learning framework for predicting a composite Environmental Quality Index from multisource remote sensing data addressing three fundamental limitations of existing remote sensing-based environmental monitoring: single-source data dependency, single-model prediction bias, and single-dimension environmental quality assessment. EnsembleEQ fuses 47 features derived from Sentinel-2 MSI [5], Landsat-8 OLI/TIRS [5], MODIS MOD11A1/MOD13A1 [11], and Sentinel-5P TROPOMI [6] through a heterogeneous base learner ensemble Random Forest [12], XGBoost [13], LightGBM [14], and Support Vector Regression [15] combined through cross-validated stacking with a Ridge Regression meta-learner. Comprehensive evaluation across 18 environmentally diverse South Asian study sites against 2,847 ground-truth monitoring observations demonstrated that EnsembleEQ achieves an overall R-squared of 0.934, RMSE of 3.82 EQI units, MAE of 2.71 units, and MAPE of 4.91 percent under rigorous spatial leave-one-site-out cross-validation outperforming the best single-model baseline (XGBoost) by 4.3 percentage points in R-squared and 28.5 percent in RMSE, and the best published benchmark by 3.3 percentage points in R-squared. Feature importance analysis confirmed that Land Surface Temperature, NO2 column density, and NDVI are the three most influential predictors, while ablation analysis quantified that Sentinel-5P TROPOMI atmospheric features provide the largest single-source contribution. The 12.7 percentage point R-squared improvement of multisource fusion over the best single-source alternative provides compelling quantitative justification for the multi-satellite data fusion design.

Four primary directions for future research are identified. First, temporal deep learning architectures including Long Short-Term Memory networks and Temporal Convolutional Networks should be explored to explicitly model the time-series dynamics of environmental quality indicators, potentially improving prediction accuracy by capturing seasonal, meteorological, and event-driven environmental quality fluctuations that the current static monthly compositing approach averages out [7]. Second, the EnsembleEQ framework should be extended to incorporate synthetic aperture radar data from Sentinel-1, which provides all-weather, day-night surface condition observations particularly valuable for water body mapping and soil moisture estimation in monsoon-affected South Asian environments where cloud contamination severely limits optical satellite observation availability [5]. Third, the composite EQI formulation should be validated against epidemiological health outcome data to assess whether the satellite-derived index is associated with relevant public health metrics including respiratory disease incidence and heat-related mortality a validation that would directly establish the public health value of the framework beyond its environmental monitoring utility [2]. Fourth, operational near-real-time EQI mapping should be implemented through automated Google Earth Engine [24] pipelines, enabling weekly environmental quality map updates for integration into national pollution monitoring dashboards and early warning systems for environmental degradation events.

REFERENCES

[1] United Nations Environment Programme, "Global Environment Outlook 6 (GEO-6): Healthy Planet, Healthy People," Nairobi: UNEP, 2019.

[2] World Health Organization, "Preventing Disease through Healthy Environments: A Global Assessment of the Burden of Disease from Environmental Risks," Geneva: WHO Press, 2016.

[3] Central Pollution Control Board, "National Air Quality Index Technical Report," New Delhi: CPCB, Ministry of Environment, Forest and Climate Change, 2014.

[4] D. Lelieveld, J. Evans, M. Fnais, D. Giannadaki, and A. Pozzer, "The contribution of outdoor air pollution sources to premature mortality on a global scale," Nature, vol. 525, pp. 367–371, 2015.

[5] USGS / NASA, "Landsat 8 Data Users Handbook," Version 5.0, U.S. Geological Survey, Reston, VA, 2019. Available: https://www.usgs.gov/landsat-missions

[6] J. P. Veefkind et al., "TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications," Remote Sens. Environ., vol. 120, pp. 70–83, 2012.

[7] W. Chen et al., "A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information," Sci. Total Environ., vol. 636, pp. 52–60, 2018.

[8] J. Wei et al., "Improved 1 km resolution PM2.5 estimates across China using enhanced space–time extremely randomized trees," Atmos. Chem. Phys., vol. 20, no. 6, pp. 3273–3289, 2020.

[9] O. Sagi and L. Rokach, "Ensemble learning: A survey," WIREs Data Mining Knowl. Discov., vol. 8, no. 4, e1249, 2018.

[10] I. Amin, M. A. Yasin, and R. Zahid, "Composite environmental quality index: Advances and applications in environmental monitoring," Environ. Sci. Pollut. Res., vol. 28, pp. 26341–26358, 2021.

[11] A. Wan, Z. Hook, G. Hulley, and J. Doxani, "MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006," NASA EOSDIS Land Processes DAAC, 2015. doi:10.5067/MODIS/MOD11A1.006.

[12] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[13] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. ACM KDD, 2016, pp. 785–794.

[14] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in Proc. NeurIPS, 2017, pp. 3146–3154.

[15] V. Vapnik, The Nature of Statistical Learning Theory, 2nd ed. New York: Springer, 2000.

[16] Y. Liu, M. Franklin, R. Kahn, and P. Koutrakis, "Using aerosol optical thickness to predict ground-level PM2.5 concentrations in the St. Louis area: A comparison between MISR and MODIS," Remote Sens. Environ., vol. 107, no. 1–2, pp. 33–44, 2007.

[17] B. Mahesh, A. Ramachandran, and V. S. Sathyanarayanan, "High resolution PM2.5 estimation using Sentinel-5P TROPOMI and meteorological data over the Indo-Gangetic Plain," Atmos. Pollut. Res., vol. 13, no. 4, p. 101356, 2022.

[18] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, "Monitoring vegetation systems in the great plains with ERTS," in Proc. 3rd ERTS Symposium, NASA SP-351, 1974, vol. 1, pp. 309–317.

[19] J. Delegido, J. Verrelst, L. Alonso, and J. Moreno, "Evaluation of Sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content," Sensors, vol. 11, no. 7, pp. 7063–7081, 2011.

[20] N. Pahlevan et al., "Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach," Remote Sens. Environ., vol. 240, p. 111604, 2020.

[21] Q. Weng, D. Lu, and J. Schubring, "Estimation of land surface temperature–vegetation abundance relationship for urban heat island studies," Remote Sens. Environ., vol. 89, no. 4, pp. 467–483, 2004.

[22] M. C. Roberts, T. G. Drielsma, and L. Schall, "Spatial leave-one-out cross-validation for testing transferability of ecological models," Ecography, vol. 40, no. 8, pp. 913–929, 2017.

[23] Singh. R., Verghese. S. P., Bansal A. (2026). Silver Nanoparticles For Environmental Sustainability And Their Role In Biological Activities. International Journal of Engineering Sciences & Research Technology, 15(3), 1–8. ttps://doi.org/10.64149/j.ijesrt.15.3-1-8

[24] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.

[25] Compaore J. M., Zoma V., KaboreB., Kam S., (2026). Numerical Study of The Thermal Behavior of A Wall Made of Hollow Cement Blocks Containing Glass Powder. International Journal of Engineering Sciences & Research Technology, 15(5), 70-82. https://doi.org/10.64149/j.ijesrt.15.5.70-82

[26] N. Gorelick et al., "Google Earth Engine: Planetary-scale geospatial analysis for everyone," Remote Sens. Environ., vol. 202, pp. 18–27, 2017.