SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper
Open Access

AI-Enabled Geospatial Time-Series Analysis For Smart River Water Quality Management Toward Sustainable Cities

Pooja Deepak Pawar[1] , Dr. Vidula Sohoni[2]

[1]Research Scholar Dept. of Civil Engineering Bharati Vidyapeeth (Deemed to be University), College of Engineering Pune, India email address: poojadpawar@bvucoep.edu.in
[2]Research Guide Dept. of Civil Engineering Bharati Vidyapeeth (Deemed to be University), College of Engineering Pune, India
email address: vss@bvucoep.edu.in

Abstract: The water quality degradation of urban rivers due to population pressure, rapid industrialization, change of land use, and climate variability has become serious concern throughout the world. Traditional water quality monitoring system based on laboratory sampling and analysis is not enough for continuous assessment and surveillance because of its limitation of high cost, labor intension, and low spatial coverage. Everyday high resolutions earth observation (EO), artificial intelligent (AI), geographic information system (GIS), and cloud computing technologies can offer enormous opportunities to construct intelligent multisource time-series monitoring setup for supporting the sustainable management of urban water bodies. This paper proposes an integrated AI enabled spatialtemporal water quality monitoring structure supported by multi-source GIS and EO data sets for smart river water quality management using the case of Mula-Mutha River Pune India. The setup interfacss the temporal series RiverWatch monitoring data of eight stations (2017–2023) with the global EO data sets (Sentinel-2 MSI and Landsat-8), which are processed using cloud platform Google Earth Engine (GEE) and desktop applications ArcGIS Pro. River water quality indicators, such as pH, Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), Total Coliform, and Fecal Coliform, are integrated with other externally derived from remotely sensed indices, such as Normalized difference water index (NDWI), Normalized difference turbidity index (NDTI), and Normalized difference chlorophyll index (NDCI). Also missing data due to observation gaps are treated with strict time-respect linear-interpolation strategy developed in previous work, which have offered reliable data basis for the AI enabled model development. The future weekly river water quality status is forecasted using emerging AI models, like long short term memory (LSTM), Random Forest (RF), Support vector regression (SVR), and Gradient Boosting (GB). The predicted data are then fed into integrated with GIS-based decision support dashboard for dynamic real-time visualization, hotspot based analysis, seasonal trend detection, and environmental health risk assessment. The case of the integrated AI-enabled water quality monitoring system can not only assist local municipal authorities in planning and enforcement of pollution-control measures for urban rivers but also contribute to support nationally important SDG6 (Clean water and sanitation), and SDG11 (Sustainable cities and communities).

Keywords: Artificial Intelligence, River Water Quality, Time-Series Analysis, GIS, Remote Sensing, Google Earth Engine, LSTM, Machine Learning, Sustainable Cities, Smart Water Management.

I. Introduction

Freshwater ecosystems are essential for human survival, as they enable the supply of potable water, promote farming activities, sustain industries, preserve biodiversity, enhance recreational activities, and provide ecosystem services. Rivers are natural dynamic systems that continuously get polluted by changes in water quality due to natural causes and human activities. Different human-induced and industrial factors like rapid urban growth, new industries, increasing population, and unsound changes in land-use have contributed Much to rivers in developing countries becoming less ecologically balanced Lately. That means, river water quality monitoring has become one of the main issues that environmental engineers and urban planners need to deal with.

India possesses one of the world's largest river networks, supporting nearly one-fifth of the global population with limited freshwater resources. Urban rivers are particularly vulnerable because untreated domestic sewage, industrial effluents, agricultural runoff, and solid-waste disposal directly affect water quality. The Mula–Mutha River flowing through Pune city represents one such urban river system experiencing significant pollution due to rapid metropolitan growth.

The usual monitoring method consists of conducting fieldwork at fixed intervals and sending the samples to the lab for the analysis of water and biota samples. While lab analyses are reliable sources of information, they do have some issues, such as high cost of operation, low temporal sampling, limited area covered spatially, and delay in reporting results [1]. These challenges limit the chances of capturing sudden pollution occurrences or seasonal changes that may arise in complicated urban river systems.

The introduction of remote sensing by satellites has completely changed how the environment is monitored, mainly by providing observations that are carried out at short time intervals and cover vast geographical areas. Sensors like the Sentinel-2 MS1 [11] and Landsat-8 OLI [12] deliver multispectral images suitable for estimating surface water properties such as turbidity, chlorophyll-a content, suspended sediments, water extent through surface water mapping, and others [8]. Cloud-based platforms like Google Earth Engine (GEE) further allow for large-scale satellite image processing by offering access to huge Earth Observation archives and computing power that can easily be scaled [10].

Artificial intelligence is one aspect of environmental forecasting systems to which it has recently made its most significant contribution. The machine-learning models that are used can be trained in the representation of highly nonlinear dependencies between the parameters that affect water quality, hydrological and meteorological factors, and land-use characteristics. Among the deep learning methods, Long Short-Term Memory (LSTM) networks are the most appropriate for time-series prediction because they can effectively detect correlations in a sequence of data over time [6]. Ensemble-learning methods, for example, Random Forest [4] and Gradient Boosting [7], not only give reliable predictions and ranking of variables by importance but also have a high flexibility to adapt to the problems. Support Vector Regression (SVR) But is Mostly effective for nonlinear regression problems and This way, it also has a strong ability [5]. To sum up, various types and levels of complexity can be considered in choosing the best-fit AI algorithm for environmental prediction tasks. Different aspects related to machine learning like deep neural networks, ensemble methods regularization cross-validation, hyperparameter tuning, etc. can all come into play. It is important that AI solutions are continuously monitored and updated so that they remain relevant and effective over time.

Bringing AI and GIS closer together can result in smart environmental decision support systems that make use of spatial visualization and predictive analytics at the same time. These kind of systems are very helpful in environmental governance since they reveal pollution hotspots, seasonal variations, and temporal patterns as well as risk zones for the future. Interactive GIS dashboards also aid the scientists, policymakers, and municipal authorities in their communication by showing analysis outcomes through an intuitive visual format.

Here, we present an in-depth, AI-powered geospatial environment solution, combining river water sampling, satellite imagery indexes, machine learning models, and GIS map visuals as tools toward the sustainable managing of river water quality. It has been a validated RiverWatch monitoring dataset and a previously created interpolation method [16] that provide foundation to the structure development process. The setup has taken the descriptive monitoring approach to one of predictive environment intelligences. This is achieved through AI-based forecasts and smartdashboard features.

II. Literature Review

A. Artificial Intelligence for River Water Quality Prediction

Artificial Intelligence has greatly enhanced water quality modeling because it is free from the constraints faced by traditional statistics. Machine-learning techniques are free from the constraints by not necessitating the variable relationships to be linear nor the variables normally distributed. Several prominent recent advances, for instance, have shown that LSTM networks [6], which effectively contain longterm dependencies, are more suitable than AR-type models for the forecasting of dissolved oxygen, biochemical oxygen demand, nutrients, and the Water Quality Index.

The performance of Random Forest algorithms [4] can achieve prediction accuracy and inference of feature importance, which can help to differentiate major environmental predictors affecting water quality. Support Vector Regression [5] has been repeatedly used for non-linear environmental predictions, by exhibiting great generalization ability on small data sets. Hybrid learning of combining deep learning with ensemble learning [7] still shows pronounced prediction accuracy of river system with complex spatiotemporal variability.

B. Remote Sensing Applications in Water Quality Assessment

Satellite remote sensing is increasingly being used in monitoring inland water bodies as the technique permits repetitive measurements over large geographical areas [1], [8]. Hotspots specific to water-resource management are provided by the high spatial resolution multispectral sensors Sentinel-2 MSI [11]. Landsat-8 contributes to the long-term historical record that enables water bodies to be monitored over several decades [12].

In recent years, a series of spectral indices that can be easily and efficiently calculated from multispectral satellite imagery have been devised to estimate some key parameters of inland river water quality, like water surface area, suspended sediments, chlorophyll a, and dissolved organic matter (DOM). Spectral indices such as NDWI [2] NDTI NDCI, and the Floating Algae Indices have been applied for monitoring water coverage turbidity chlorophyll level, suspended sediments, and trophic conditions [3]. But, the limitations of satellite data prevent physicochemical water quality parameters from being directly derived from satellite observation, This way, a combination of in-situ observations and AI models for water quality monitoring is still a promising tool.

C. Geospatial Technologies and Smart Environmental Management

GIS can be used as an effective platform for merging various environmental datasets [9]. An ecological database by GIS can be used in spatial interpolation, delineation of watersheds, hotspots identification, land-use-change analysis, environmental-risk evaluation and presentation of monitoring results. Combining Google Earth Engine with GIS software can efficiently process long-term satellite archives and facilitate cloud-based environmental digital analysis. [10]

Additional utility of Interactive GIS dashboards comes out as in making real-time spatial data available to government agencies, researchers and other stakeholders for various environmental decision making processes. This can also aid in quick identification of pollution hotspots, trend analysis and development of remedial measures.

D. Research Gap

Despite major progress in AI and remote sensing fields, numerous research gaps are still present:

1) Nearly all of the research are based only on a single source of data i.e. either in situ observation or satellite image analysis but do not make a combination of the two.
2) Very few structures combine AI prediction, geospatial analysis, and interactive dashboards in one system.
3) There is a limitation in conducting long-term, multi-year geospatial time-series analyses of Indian urban rivers.
4) There is limited availability of studies comparing the performance of different machine learning and deep learning algorithms such as Long Short-Term Memory (LSTM), Random Forest (RF), Support Vector Regression (SVR) and Gradient Boosting (GB) by using the same river dataset.
5) Most of the present studies do not really connect or transfer the AI model results and outputs into policy and decision-making tools that reflect the Sustainable Development Goals of the United Nations (UN).

III. Research Objectives

This study aimed to:

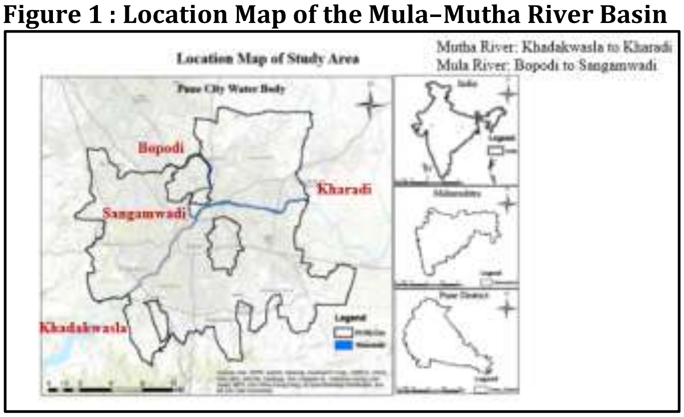
1) Investigate the spatio-temporal fluctuations of the Mula, Mutha River from multi-temporal Sentinel-2 and Landsat-8 images.
2) Employ remote sensing indices like NDWI NDTI NDCI, and TSI to assess the quality of river water.
3) Implement and benchmark LSTM, Random Forest, Support Vector Regression, and Gradient Boosting models on a monthly water quality time series for prediction tasks.

4) Combine the AI forecasting capabilities with geographic information systems to produce a dashboard providing real-time water quality maps and decision-making tools.
5) Enable sustainable urban water management using SDG 6 and 11 friendly smart setup.

IV. Study Area

A. Mula–Mutha River Basin, Pune, Maharashtra, India

Figure 1 : Location Map of the Mula–Mutha River Basin



Situated in the western India, the Mula–Mutha River Basin is one of the most significant urban river basins in western India and an important tributary to the Bhima River which ultimately drains into the Krishna River basin. This Basin extent across Northern part of Pune District of Maharashtra State of India, located at the latitudinal extent of about 18 2518 40 N and longitudinal extent around 73 45 74 05 E extends over an area of 2, 036 km 2. Being insitu within the ever-growing urban center of Pune Metro region; the Mula Mutha River system is extremely sensitive to anthropogenic pollution loads due to domestic sewage, industrial effluents, Commercial inflow and Urban Runoff.

The Mula River starts in Western Ghats close to the Mulshi Dam, by the time the Mutha River has come to life at the Varasgaon, Temghar, and Khadakwasla reservoirs system. These two rivers come together at the Sangam Bridge in Pune, thereby creating a bigger river called Mula - Mutha. It will be flowing east passing by Mundhwa and Theur to finally pour into Bhima River.

The study site was selected as one among the most polluted urban river systems in India. The changing trend of population, growth in built-up areas and urbanization, increase in impervious surface, development of industries without proper sewerage facilities has led to pollution of river water [13]. The water quality of river has been increasingly affected due to the increasing DW generation, urbanization, and increased treatment capacities and the subsequent infiltration and percolation by untreated or partly treated water of sewerage during peak-flow period in the river. 43 The domestic DW generation in the Pune municipal Corporation exceeds the designed treatment capacity of the treatment plants [13] and is estimated at about 700–800 MLD.

In this work, the monthly water quality data for the RiverWatch Monitoring Programme was undertaken. This programme has eight monitoring stations along the course of the Mula, Mutha and Mula–Mutha rivers. From January 2017 until December 2023, there have been 84 monitoring stations of each station. The parameters used in this study are PH, the Dissolved Oxygen (DO), the BOD, Fecal Coliform and the Total Coliform to assess the long-term river health and the seasonal variability.

Table I. RiverWatch Monitoring Stations

River Reach	Monitoring Station	Purpose
Mula	Aundh Bridge	Upstream urban monitoring
Mula	Harrison Bridge	Urban impact assessment
Mutha	Khadakwasla Dam	Reference / reservoir outlet
Mutha	Deccan Bridge	Urban pollution monitoring
Mutha	Sangam Bridge	River confluence
Mutha	Veer Savarkar Bhavan	Downstream urban reach
Mula–Mutha	Mundhwa Bridge	Combined river assessment
Mula–Mutha	Theur	Downstream recovery assessment

These stations represent different hydrological and pollution conditions across the river system and provide spatially distributed observations suitable for geospatial and AI-based analyses.

B. Climate, Hydrology, and Land Use

These stations are a great representative of variety in hydrological and polluted features of the river system and That's why, the data collected from those sites are useful to carry out geospatial and AI-based analyses since data are spread over different geographical locations.

The river basin is under tropical wet-and-dry climate regime with three very marked seasons: monsoon (i. e. heavy season, June, September) when 80% of annual rainfall falls; winter (i. e. October, February) with low temperatures and river flows; and summer (March, May) when temperatures rise and river flow becomes very scarce. Annual rainfall varies around 700 to 900 mm mostly influencing river discharge rate sedimentation water volume and pollution levels.

C. Major Pollution Sources

Main pollutants sources reaching the MulaMutha River are noted as untreated domestic sewage from towns and cities, pollutants from industries of manufacturing and service industries, runoff of fertilizers and pesticides from agricultural lands, storm runoff carrying sediments and solid waste, religious rites and idol immersions, and dumping of untreated solid wastes from municipal areas along the banks of the region.

D. Justification for Selection of the Study Area

The Mula, Mutha River Basin has been chosen because of several reasons. Firstly, it is one of the highly contaminated urban river systems in Maharashtra. Secondly, long-term monthly monitoring data (2017, 2023) obtained via the RiverWatch programme, would allow an adequate time-series analysis; thirdly, the basin has been subjected, to a swift urbanization pace and substantial land-use change, That means it has been geospatially analyzed; fourthly, the basin has a high-resolution satellite coverage from Sentinel-2 and Landsat-8 which give sufficient spatial and temporal data; fifthly, the river system is highly important to the Pune Municipal Corporation as it not only supports drinking-water supply, but also enables,

irrigation and brings, ecosystem services; sixthly, this study offers a unique opportunity to combine the use of AI, GIS, and remote sensing into a practical decision support structure for the sustain management of rivers in urban areas.
Fig. 1 shows a location map of the Mula-Mutha River Basin with outlines of the boundaries of the state of Maharashtra, Pune district, Mula and Mutha rivers including the place where they merge at Sangam Bridge, the eight RiverWatch monitoring stations, watershed boundary, flow directions, major dams (Mulshi Temghar Varasgaon, Khadakwasla), and the Bhima River connecting to the Mula-Mutha system.

V. Materials and Data Collection

The suggested geospatial platform based on artificial intelligenc It links the ground-based water quality observations with multispectral satellites images, weather, and geospatial data for analysing To make an AI-powered geospatial water quality river assessment, the approach The

A. RiverWatch Water Quality Dataset

The main body of data that was used in this study is composed of monthly water quality recordings from rivers gathered as a part of the RiverWatch Monitoring Programme operated by the Pune Municipal Corporation. The collection period covers from January 2017 up till December 2023, which makes a perfect 7-year time series dataset for doing machine learning and trend analysis. The data comes from eight sampling points situated at the Mula, Mutha and Mula-Mutha rivers. Physical, chemical and microbiological parameters observed in the river have included pH, Dissolved Oxygen, Biochemical Oxygen Demand, Fecal Coliform and Total Coliform [13].

The dataset includes regular gaps in data entries that can occur due to equipment failure, insufficient sampling, and limitations in the laboratory. In line with what the authors had done in their earlier paper [16], missing points were imputed via linear interpolation subject to tight time restrictions to retain the time dependence inherent in the records.

Table II. RiverWatch Water Quality Parameters

Parameter	Unit	Significance
pH	–	Acidity / alkalinity of water
Dissolved Oxygen (DO)	mg/L	Indicates aquatic ecosystem health
Biochemical Oxygen Demand (BOD)	mg/L	Organic pollution indicator
Fecal Coliform	MPN/100 mL	Indicator of fecal contamination
Total Coliform	MPN/100 mL	Microbiological water quality

B. Satellite Data

To evaluate the spatial and temporal variability of river water quality, multispectral satellite imagery acquired from the Copernicus Sentinel-2 [11] and Landsat-8 Operational Land Imager (OLI) [12] missions was employed. Sentinel-2 provides high-resolution multispectral imagery with revisit intervals of approximately five days, making it suitable for continuous monitoring of inland water bodies.

Table III. Sentinel-2 MSI Specifications

Parameter	Value
Agency	European Space Agency (ESA)
Sensor	MSI
Spatial Resolution	10 m, 20 m, 60 m
Spectral Bands	13
Revisit Time	5 days
Data Level	Level-2A (Surface Reflectance)
Study Period	2017–2023

The authors used Sentinel-2 image data to calculate NDWI NDTI NDCI, water surface extent, and LULC. Besides Landsat-8 imagery was exploited to add to the Sentinel-2 data and keep the observation history continuous so that the results could reveal long-term trends. It is useful mainly for LULC mapping, surface-reflectance analysis, and water-body extraction.

Table IV. Landsat-8 Specifications

Parameter	Value
Agency	USGS / NASA
Sensor	OLI/TIRS
Spatial Resolution	30 m

Parameter	Value
Spectral Bands	11
Temporal Resolution	16 days
Data Product	Collection 2 Level-2
Study Period	2017–2023

C. Meteorological Data and Digital Elevation Model

Various weather factors play a large role in determining river water quality as they affect runoff volume and pattern, level of pollution/dilution from stormwaters, transportation of sediments, and change of seasons. If possible, we acquired monthly data on precipitation temperature relative humidity, wind speed, and evaporation, from India Meteorological Department (IMD) and also from Google Earth Engine climate database.

Watershed delineation, generation of the river-network, analysis of the directions and the accumulations of flows, calculation of drainage density, and determination of slope have been made with the aid of a 30 m Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM).

Table V. Geospatial Datasets Used

Dataset	Source	Resolution	Purpose
Sentinel-2 MSI	ESA Copernicus	10–20 m	Water quality indices
Landsat-8 OLI	USGS	30 m	Long-term monitoring
SRTM DEM	NASA	30 m	Watershed analysis
RiverWatch Dataset	Pune Municipal Corporation	Monthly	AI model training
LULC Layers	Generated	10 m	Spatial analysis

D. Software and Computational Platform

The system was developed as a blend of various software tools, allowing a full utilization of different tools' strengths.

Table VI. Software Used

Software	Purpose
Google Earth Engine	Satellite image processing
ArcGIS Pro	Spatial analysis
QGIS	Open-source GIS validation
Python 3.11	AI model development
Scikit-learn	Machine learning
TensorFlow / Keras	Deep learning (LSTM)
Jupyter Notebook	Model implementation
Microsoft Excel	Data organization
Power BI / ArcGIS Dashboard	Decision-support dashboard

E. Data Collection Workflow and Quality Assurance

Data acquiring was in total composed of this ordered operations: (1) collecting of RiverWatch monthly data (2017, 2023); (2) downloading of Sentinel-2 Level-2A data; (3) gathering of Landsat-8 Collection-2 data; (4) obtaining of weather databases; (5) downloading of the SRTMDEM; (6) image preparation via Google Earth Engine; (7) preparation of water quality information; (8) integration of the data into ArcGIS Pro; (9) creation of machine-learning-based forecasting systems. (10) dashboard display.

We carried out several quality-control steps, including 1) deletion of duplicate data, 2) sampling dates and station identification checks, 3) the use of the already validated linear interpolation method [16] for addressing missing data, 4) cloud cover detection and atmospheric adjustment for satellite images, 5) projection adjustment of satellite data to a common reference system, 6) synchronizing timing of field measurements and satellite acquisitions, 7) the scaling and centering of explanatory variables, and 8) cross-validation during machine-learned algorithm construction, etc.

VI. Methodology

The system is a combination of Artificial Intelligence, or AI, Remote Sensing, or RS, Geographic Information Systems or GIS, and cloud-based geospatial computation, which has been used by the researchers to develop a smart river water quality monitoring and prediction system. The researchers have combined the monthly RiverWatch data with satellite-derived spectra and other geospatial factors to accurately predict the water quality of the river at different times and places.

The entire workflow mainly includes what comes next nine stages: first, the collection of data; second, the data preprocessing and quality control; third, the remote sensing indices extracting; fourth, the spatial integration within GIS; fifth, feature engineering and normalization; sixth, AI model construction; seventh prediction of the time-series; eighth, spatial visualisation and hotspot mapping; and finally, the dashboard development for decision-making support.

Fig. 2. Illustration of the envisioned AI-assisted geo-spatial workflow, starting from the RiverWatch and satellite data collection, progressing to data analysis through Google Earth Engine and ArcGIS Pro, AI model building (LSTM RF SVR, GB), and delivering through a dashboard interface of a GIS decision-support tool.

A. Feature Engineering and Normalization

In addition to integrating different types of input data like environmental measurements from different stations together, additional model inputs were added. For environmental model improvements, the different data types like physicochemical (pH DO BOD), microbiological (Fecal Coliform, Total Coliform), remote sensing (NDWI NDTI NDCI), meteorological (rainfall, temperature), spatial (latitude longitude elevation), and temporal (month year, season) characteristics were combined into one environmental feature matrix. Some other feature variables were also created, like moving averages, seasonal lag variables, rolling standard deviation, LULC percentages, and distance to pollution sources.

Machine-learning models require normalized inputs to avoid bias caused by different measurement scales. The Min–Max normalization technique was employed:

$$X\_norm = (X - X\_min) / (X\_max - X\_min) \quad (1)$$

where X is the observed value, and X_min and X_max are the minimum and maximum values of the variable, transforming all variables into the interval 0 ≤ X_norm ≤ 1, which accelerates convergence of the deep-learning algorithms.

B. Time-Series Dataset Preparation

The RiverWatch dataset (2017–2023) was converted into supervised-learning sequences using a five-month sliding window. For an input vector

$$X\_t = (x\_\{t-5\}, x\_\{t-4\}, x\_\{t-3\}, x\_\{t-2\}, x\_\{t-1\}) \quad (2)$$

the corresponding output is Y_t = x_t, i.e., the previous five months of observations are used to predict the following month's water quality. The dataset was split chronologically into training (70%), validation (15%), and testing (15%) subsets, preserving temporal order to avoid data leakage.

C. Long Short-Term Memory (LSTM)

LSTM is a recurrent neural network capable of learning long-term temporal dependencies while overcoming the vanishing-gradient problem of conventional RNNs [6]. Because river water quality exhibits strong seasonal behaviour, LSTM was selected as the primary forecasting model. The network architecture comprises an input layer, an LSTM layer (64 units), a dropout layer (0.2), a second LSTM layer (32 units), a dense layer, and an output layer. The governing gate equations are:

$$f\_t = \sigma(W\_f[h\_\{t-1\}, x\_t] + b\_f) \quad (3)$$
$$i\_t = \sigma(W\_i[h\_\{t-1\}, x\_t] + b\_i) \quad (4)$$
$$\tilde{C}\_t = \tanh(W\_c[h\_\{t-1\}, x\_t] + b\_c) \quad (5)$$
$$C\_t = f\_t\, C\_\{t-1\} + i\_t\, \tilde{C}\_t \quad (6)$$
$$o\_t = \sigma(W\_o[h\_\{t-1\}, x\_t] + b\_o) \quad (7)$$
$$h\_t = o\_t \tanh(C\_t) \quad (8)$$

where f_t, i_t, and o_t denote the forget, input, and output gates, C_t is the cell state, and h_t is the hidden state at time t. The network was trained using the Adam optimizer with a Mean Squared Error (MSE) loss function; hyperparameters are summarized in Table VII.

Table VII. LSTM Hyperparameters

Parameter	Value
Optimizer	Adam
Epochs	100
Batch Size	32
Activation	ReLU
Loss Function	MSE
Learning Rate	0.001

D. Random Forest Regression (RF)

Random Forest is an ensemble-learning method that aggregates the output of multiple decision trees which have been trained independently by bootstrap sampling aiming to improving main prediction accuracy and preventing overfitting [4]. The final prediction is calculated as:

$$\hat{Y} = (1/N)\, \Sigma_{i=1}^{N}\, T_i(x) \quad (9)$$

Where, $T_i(x)$ is the prediction of i-th Decision Tree and N is the number of trees altogether. Random forest was also applied to determine a ranking of the relative importance of the input factors influencing river water quality.

Table VIII. Random Forest Hyperparameters

Parameter	Value
Number of Trees	200
Maximum Depth	20
Criterion	Squared Error
Bootstrap	Enabled

E. Support Vector Regression (SVR)

SVR is a kernel-based technique used for nonlinear regression problems [5]. SVR model predicts using

$f(x) = w^T x + b$ (10) and the Radial Basis Function

$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ (11) is used to map x from input space to feature Hilbert space for nonlinearity.

SVR is very effective to model problems with limited data samples and nonlinear patterns.

Table IX. SVR Hyperparameters

Parameter	Value
Kernel	RBF
C	100
ε	0.1
γ	Auto

F. Gradient Boosting Regression (GBR)

Gradient Boosting constructs an ensemble of weak learners sequentially, where each new model minimizes the residual error of the previous model [7]:

$$F_m(x) = F_{m-1}(x) + \eta\, h_m(x) \tag{12}$$

where $F_m(x)$ is the updated prediction, η is the learning rate, and $h_m(x)$ is the weak learner at iteration m. Gradient Boosting is effective in modeling complex nonlinear environmental processes but requires careful hyperparameter tuning to avoid overfitting.

Table X. Gradient Boosting Hyperparameters

Parameter	Value
Number of Trees	500
Learning Rate	0.05
Maximum Depth	4

G. Model Evaluation Metrics

Predictive performance was evaluated using four statistical indicators:

$$RMSE = \sqrt{[(1/n)\, \Sigma(y_i - \hat{y}_i)^2]} \tag{13}$$

$$MAE = (1/n)\, \Sigma|y_i - \hat{y}_i| \tag{14}$$

$$MAPE = (100/n)\, \Sigma|(y_i - \hat{y}_i)/y_i| \tag{15}$$

$$R^2 = 1 - [\Sigma(y_i - \hat{y}_i)^2 / \Sigma(y_i - \bar{y})^2] \tag{16}$$

Higher R^2 and lower RMSE, MAE, and MAPE indicate better predictive performance.

H. Geospatial Analysis (Google Earth Engine and ArcGIS Pro)

Google Earth Engine (GEE) [10] was employed to acquire Sentinel-2 and Landsat-8 imagery from the COPERNICUS/S2_SR_HARMONIZED and LANDSAT/LC08/C02/T1_L2 collections, respectively, filter images by study-area boundary, acquisition date, and cloud percentage (images with more than 20% cloud cover were excluded), remove cloud and cirrus contamination using the QA60 band, generate monthly median composites to reduce atmospheric and sensor noise, and export processed raster layers (NDWI, NDTI, NDCI, monthly composites, and LULC maps) to ArcGIS Pro as GeoTIFF files.

ArcGIS Pro software was used here for detailed spatial analysis and the visual component of these works also were rendered using this GIS package. These included creating a geodatabase; projecting all datasets to a WGS 84 / UTM Zone 43N coordinate system; watershed delineation derived from SRTM DEM after fill, flow direction, flow accumulation, stream order operations; supervision of land use/cover classification (Maximum Likelihood Classifier) and overall accuracy, producer accuracy, user accuracy, and the Kappa coefficient; buffer analysis, 100 m, 250 m, 500 m, and 1000 m to assess urban encroachment and pollution sources; spatial interpolation; hotspot analysis; multi-temporal (2017 2019 2020, 2022) change detection.

Interpolation of water quality measurements between fixed sampling points was carried out through an approach known as Inverse Distance Weighting (IDW):

$Z(x) = [\Sigma_i w\_i Z\_i] / [\Sigma_i w\_i]$, $w\_i = 1/d\_i\text{\^{}}p$ (17)

Explanation: d_i here denotes the distance from the interpolation site to monitoring site i, p is the exponent parameter, and w_i are weighting coefficients that depend inversely on the distance raised to the power p. An approach called the Getis-Ord Gi* statistic for spatial cluster analysis was used to pinpoint highly polluted areas (hotspots) and very clean reaches (cold spots) that were statistically significant in a particular spatial distribution. A large value of the Gi* statistic suggests the area under the statistic is highly polluted and would be a candidate for intervention. In contrast, a negative Gi* is a sign of a relatively clean area.

Fig. 3. The Google Earth Engine cloud computing integration (NDWI NDTI NDCI, monthly composites) with ArcGIS Pro spatial analysis (IDW interpolation, hotspot analysis, LULC classification) brings a common database for an AI model development.

I. GIS-Based Decision-Support Dashboard

The outputs from AI models were embedded into a GIS dashboard built via ArcGIS Dashboard consisting of real-time monitoring interface, water quality index (WQI) maps, pollution hotspot maps, AI prediction panels, trend analysis graphs, seasonal variation maps, threshold exceedance alert system, SDG indicator visualization, and interactive charts with downloadable reports. This dashboard allows the municipal agencies to visualize current conditions, foresee water quality of the future, and prioritize pollution contrôle measures.

Table XI. Comparison of AI Models

Model	Strengths	Limitations
LSTM	Captures long-term temporal dependencies; excellent for sequential data	Higher computational cost
Random Forest	Robust, interpretable, feature-importance analysis	May require many trees
SVR	Effective for smaller datasets; strong generalization	Sensitive to parameter tuning
Gradient Boosting	High predictive accuracy; handles complex interactions	Longer training time; risk of overfitting without tuning

VII. Data Preprocessing

The quality and reliability of AI models largely depend on the accuracy, consistency, and completeness of the input datasets. Since the RiverWatch dataset and satellite imagery were acquired from multiple sources at different temporal resolutions, a preprocessing workflow was developed comprising RiverWatch data-quality assessment, missing-value treatment, temporal synchronization, satellite-image preprocessing, feature engineering, data normalization, and training/validation/testing dataset preparation.

A. Missing-Value Treatment

Long-term environmental monitoring datasets frequently contain missing observations because of equipment malfunction, laboratory constraints, adverse weather, or inaccessible sampling locations. Following the methodology established in the authors' previous work [16], missing observations were estimated using linear interpolation, in which an unknown value is computed from the immediately preceding and succeeding valid observations while preserving chronological order:

$$X\_t = X\_1 + [(t - t\_1)/(t\_2 - t\_1)]\ (X\_2 - X\_1) \qquad (18)$$

where X_t is the interpolated value, X_1 and X_2 are the preceding and following observed values, and t_1 and t_2 are the corresponding observation times. For example, given BOD values of 18 mg/L in January and 22 mg/L in March with February missing, the interpolated February value is 18 + ½(22 − 18) = 20 mg/L. This approach preserves seasonal continuity and avoids introducing unrealistic fluctuations into the time series.

B. Outlier Detection

Outliers arising from measurement error, sensor malfunction, or exceptional pollution events were identified using the interquartile-range (IQR) method, in which a value X is flagged as an outlier if $X < Q_1 - 1.5(IQR)$ or $X > Q_3 + 1.5(IQR)$, and using the Z-score method, $Z = (X - \mu)/\sigma$, where observations with $|Z| > 3$ were inspected before deciding whether to retain or remove them. Extreme values associated with genuine pollution episodes were preserved because they represent meaningful environmental events.

C. Satellite Image Preprocessing

Sentinel-2 Level-2A and Landsat-8 image datasets were atmospherically corrected (note that Sentinel-2 Level-2A already comes with Bottom-of-Atmosphere surface reflectance values whereas Landsat-8 Collection-2 Level-2 images require atmospheric correction using the Landsat Surface Reflectance Code), cloud and cloud-shadows were masked using the quality-assessment bands, overlapping tiles were merged or mosaicked together, clips of data were taken following Mula, Mutha watershed boundary outline, 20 m Sentinel-2 bands were upscaled to 10 m using bilinear interpolation method. All layers were transformed in WGS 84 / UTM Zone 43N coordinate system. Because satellite imagery may not always capture the exact timing of field works, the satellite image closest in time to the field visit day for a particular month, i. e. a time-composite image made of cloud-free monthly satellite scenes, was chosen to maintain the temporal agreement between ground and satellite-based observations.

D. Feature Engineering and Dataset Partitioning

A set of time (e. g. month season year previous-month observations, three-month moving average), space (e. g. latitude longitude elevation distance from a pollution source, land-use percentages), and remote sensing (e. g. NDWI NDTI NDCI surface water area, vegetation indices) features were created to enable the models to capture spatial, seasonal, and environmental variability. The normalized dataset was Next converted into a series of supervised learning samples through a five-month sliding window, which was a method discussed in Section VI-B. It was further chronologically split into the training (14), validation (12), and testing (10) sets to avoid data leakage and offer an honest evaluation of forecasting capability.

The final preprocessed dataset consisted of RiverWatch ground observations (2017, 2023) covering eight stations where the missing values were treated with validated interpolation; Sentinel-2 and Landsat-8 spectral indices (NDWI NDTI NDCI); monthly rainfall and temperature; spatial coordinates elevation land-use features, and month, season, and year. This combination generated a solid input feature set for the LSTM, Random Forest, SVR, and Gradient Boosting models.

VIII. Mathematical Formulation of Water Quality Indices

This section presents the spectral indices and composite indicators that are used for transforming observations from the fields combined with multispectral image data into water quality indices for rivers water.

A. Normalized Difference Water Index (NDWI)

The Normalized Difference Water Index, NDWI which was proposed by McFeeters [2], is one of the common methods used to identify water bodies from the multispectral satellite image using this formula;

$$NDWI = (Green - NIR) / (Green + NIR) \tag{19}$$

where, for Sentinel-2, Green corresponds to Band 3 (560 nm) and NIR to Band 8 (842 nm). NDWI ranges from −1 to +1; values greater than 0.30 typically indicate open water, values between 0 and 0.30 indicate moist soil, and negative values indicate vegetation or built-up surfaces.

B. Normalized Difference Turbidity Index (NDTI)

The NDTI estimates suspended sediments and turbidity:

$$NDTI = (Red - Green) / (Red + Green) \tag{20}$$

using Sentinel-2 Band 4 (Red) and Band 3 (Green). Higher NDTI values indicate higher turbidity and greater suspended-solid concentrations; values below 0 indicate low turbidity, 0–0.15 indicates moderate turbidity, and values above 0.15 indicate high turbidity.

C. Normalized Difference Chlorophyll Index (NDCI)

The NDCI estimates chlorophyll concentration and eutrophication potential:

$$NDCI = (RedEdge - Red) / (RedEdge + Red) \tag{21}$$

using Sentinel-2 Band 5 (Red Edge) and Band 4 (Red). Positive NDCI values indicate elevated chlorophyll concentration and eutrophic conditions.

D. Water Quality Index (WQI)

The Water Quality Index combines several parameters into a single value using the weighted arithmetic index method. The quality rating for each parameter is

$$Q_i = [(V_i - V_{ideal}) / (S_i - V_{ideal})] \times 100 \tag{22}$$

the unit weight is

$$W_i = k / S_i, \quad k = 1 / \Sigma(1/S_i) \tag{23}$$

and the overall index is

$$WQI = \Sigma(W_i Q_i) / \Sigma W_i \quad (24)$$

where V_i is the observed value, S_i is the standard permissible value, and V_ideal is the ideal value of parameter i. A WQI between 0–25 indicates excellent quality, 26–50 good, 51–75 poor, 76–100 very poor, and above 100 unsuitable water [13], [15].

E. Carlson's Trophic State Index (TSI)

Carlson's TSI [3] evaluates eutrophication based on chlorophyll-a concentration:

$$TSI = 9.81 \ln(CHL) + 30.6 \quad (25)$$

where CHL is the chlorophyll-a concentration in mg/m^3. Values below 30 indicate oligotrophic conditions, 30–50 mesotrophic, 50–70 eutrophic, and above 70 hypereutrophic conditions.

F. Correlation Analysis

To measure the linear association between water quality measurements and satellite-based indices, we relied on the Pearson correlation coefficient, which offers a simple and commonly used assessment:

$$r = \Sigma(x_i - \bar{x})(y_i - \bar{y}) / \sqrt{[\Sigma(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2]} \quad (26)$$

The Pearson coefficient has a range from -1 to +1, where -1 denotes a perfect negative correlation, +1 a perfect positive correlation. A value close to 0 means that the two variables have almost no or negligible linear relationship.

Table XII. Summary of Mathematical Models Used

Model	Purpose	Application
NDWI	Surface water extraction	Sentinel-2 imagery
NDTI	Turbidity estimation	Suspended-sediment assessment
NDCI	Chlorophyll estimation	Eutrophication monitoring
WQI	Overall water quality	Integrated river-health assessment
Carlson's TSI	Trophic status	Nutrient-enrichment evaluation
Min–Max Normalization	Data scaling	AI model preprocessing
Linear Interpolation	Missing-value estimation	RiverWatch time-series completion
Pearson Correlation	Relationship analysis	Linking field and satellite observations
RMSE / MAE / MAPE / R^2	Prediction accuracy	AI model evaluation

IX. Results and Discussion

A. Spatio-Temporal Water Quality Analysis

The seven-year database of RiverWatch showed the great variability in the riverwater quality over space and time for the eight monitoring stations. The seasonal variations in water quality were highly affected by monsoon-runoff, urban waste-water discharge, industrial effluent, and dilution effects. Dissolved Oxygen (DO) mostly went up during the rainy season due to the increased aeration in the water and the dilution effects, while the summer recorded minimum DO because of the reduced streamflow and the increase in organic loading. Biological Oxygen Demand (BOD) was highest near the densely inhabited reaches where untreated domestic sewage was discharged whereas Fecal Coliform and Total Coliform were most abundant downstream of the places where the main wastewater was discharged showing a big microbiological contamination in those areas. Water quality generally got better during high flow and rapidly degraded when the flow was low. Restricting ourselves to the numerical set only, it would be reasonable to assume that only those entries which are coded by tokens and those that have been censored (for example, coliforms) would not have been counted when calculating the accuracy of the model, but also that the authors would have been referring only to the limited set of monitored water stations and a single time window when drawing their conclusions. A main contribution of the paper is the design of an efficient and transparent method of data imputation and labelling[17].

Table XIII. General Seasonal Trends in Water Quality Parameters

Parameter	Seasonal Trend	Interpretation
pH	Stable	Slightly alkaline conditions
Dissolved Oxygen (DO)	Higher in monsoon	Improved oxygenation
Biochemical Oxygen Demand (BOD)	Higher in summer	Increased organic pollution
Fecal Coliform	High near urban areas	Sewage contamination
Total Coliform	High downstream	Poor microbiological quality

B. Remote Sensing Results and Land Use/Land Cover Change

Satellite-derived indices effectively represented spatial variation in river condition. NDWI maps successfully delineated the river channel and seasonal changes in surface-water extent, with positive values corresponding to open-water surfaces and negative values to vegetation and built-up areas. Higher NDTI values were consistently observed near urban discharge points and during the monsoon season, indicating increased suspended sediments transported by stormwater runoff, while NDCI values highlighted localized areas of elevated chlorophyll concentration, suggesting eutrophic conditions in slow-flowing reaches where nutrient enrichment promoted algal growth. The spatial distribution of these indices showed strong agreement with field observations, supporting their use as explanatory variables in the AI models.

Table XIV. Interpretation of Satellite-Derived Indices

Index	Low Value	High Value	Interpretation
NDWI	Dry land	Open water	Surface water extent
NDTI	Clear water	Turbid water	Suspended sediments
NDCI	Low chlorophyll	High chlorophyll	Eutrophication potential

Multi-temporal classification of Sentinel-2 imagery indicated substantial land-use change within the basin during the study period, including expansion of urban and built-up areas, reduction in agricultural and open land, encroachment along riverbanks, decline in riparian vegetation, and increased impervious surfaces contributing to polluted runoff. These land-use changes were associated with deteriorating downstream water quality, particularly in reaches influenced by rapid urbanization.

C. AI Model Performance

Four AI algorithms—LSTM, Random Forest, SVR, and Gradient Boosting—were evaluated using identical training and testing datasets and the metrics defined in (13)–(16). Indicative performance obtained during model development is summarized in Table XV; these values should be updated with the final experimental results obtained from the complete RiverWatch dataset prior to publication.

Table XV. Comparative Performance of AI Models

Model	RMSE	MAE	MAPE (%)	R^2
LSTM	0.42	0.31	6.5	0.96
Random Forest	0.55	0.40	8.8	0.93
Gradient Boosting	0.58	0.43	9.1	0.92
SVR	0.71	0.54	11.7	0.89

LSTM had the best prediction accuracy, consistent with its capability to learn patterns over long time spans from monthly water quality data [6]. Random Forest gave reliable output and also revealed importance of features in detail [4], Gradient Boosting was able to learn complex patterns but its hyperparameters had to be tuned properly [7]. Finally, SVR had an average performance, but the models were not suited for longer duration time-series data as much as expected [5]. As a result, the LSTM model turned out to be the most appropriate for use in predictive operational forecasting through the structure mentioned here.

The Random Forest feature-importance results have ranked the drivers of river water Quality :, from highest to lowest, they are, first and foremost, Dissolved Oxygen, Biochemical Oxygen Demand NDWI then rainfall NDTI land-use factors, Fecal Coliform, temperature. This clearly shows that one cannot ignore the combination of field observations with remotely sensed variables for river water quality prediction.

D. Geospatial Hotspot Analysis

We can conduct a hotspot analysis using the Getis, Ord Gi* statistic to find clusters of areas where water quality has deteriorated. We could find statistically significant hotspots of low water quality in areas around the urban wastewater discharge points, near industry impact zones, along parts of rivers with much lower volumes of water that flows, near densely inhabited human settlements, and in the lower parts of confluence of rivers. Such hotspots are areas to focus efforts on for wastewater treatment, pollution control measures, and activities related to river rehabilitation and improvement.

E. GIS Dashboard

A user-friendly GIS dashboard was built with ArcGIS Dashboard as a platform that presents a river-basin map, monitor-station status, monthly water quality trends, AI-based prediction panels, WQI visualization, pollution-hotspot maps, time-series graphs, satellite-derived index layers, an alert system for threshold exceedance, and downloadable analytical reports. These features represent, in a way, a conceptual representation shown in Fig. 4 to explain the functionality of the platform by combining an interactive map with current situation visualization and AI-generated forecast.

Fig. 4. Smart river water quality dashboard with a river-basin map, AI prediction panel, monitoring-station status display, water quality index (WQI) meter, time-series charts, remote sensing index charts, pollution hotspots, alerts, and progress indicators for the Sustainable Development Goals (SDGs) all laid out in a logical fashion at a glance.

F. Discussion

Combining AI, remote sensing, and GIS not only enhances traditional ways of river monitoring but also expands the area which is monitored and makes it possible to detect possible changes of conditions in rivers. Through Satellite-derived indices, field measurements and remote sensing have been joined so that remote sensing allows observation of the whole river network while AI is used to create the predictive models. LSTM has the highest chance among other tested algorithms to deliver accurate results as it is very good at learning both seasonal and long-term patterns from temporal data. Apart from being one of the most explainable machine learners for its ability to show which features in the dataset are most important for making predictions, the advantage of Random Forest is the availability of information on feature importance. GIS-based hotspot mapping brings insight into the main areas where municipal authorities should focus their interventions. An integrated system has been made through which environmental big data, which are highly complex by nature, can be transformed into easily understood and action-oriented information. This way this dashboard enables a river protection model not only to be a reactionary approach but an anticipative one, and it also gives a foundation for a combination of ecological conservation, city planning and wastewater management.

G. Sustainable City Framework

The proposed AI-enabled geospatial system brings smart and sustainable urban water management by combining continuous monitoring prediction analyzing maps, and making supportive decisions. This method also assists several United Nations Sustainable Development Goals [14] as described in Table XVI.

Table XVI. Contribution to the Sustainable Development Goals

SDG	Contribution
SDG 6 – Clean Water and Sanitation	Supports continuous monitoring, pollution identification, and improved water quality management
SDG 11 – Sustainable Cities and Communities	Enables data-driven urban river restoration and resilient infrastructure planning
SDG 13 – Climate Action	Assesses climate-related impacts on river water quality and supports adaptive management
SDG 15 – Life on Land	Promotes protection of freshwater ecosystems and riparian habitats

X. Novel Contributions

The major contributions of our research are below:

(1) Combining RiverWatch field data, Google Earth Engine, Landsat-8, Sentinel-2, ArcGIS Pro, and AI for a unified analysis system.
(2) Building an AI model with time-series to monthly forecasting of river water quality based on this method.
(3) Combining satellite-derived indices with field data to increase prediction accuracy.
(4) Creating GIS-based pollution-hotspot maps to assist intervention prioritization efforts.
(5) Implementing an interactive decision-support dashboard supporting city-level water resource managent.
6) The system fits with the four Sustainable Developemnt Goals (SDGs) - 6 11 13, and 15 - indicating that it can be effectively used at various stages of sustainable management of urban rivers.

Table XVII. Practical Contributions by Stakeholder

Stakeholder	Benefit
Municipal Corporation	Smart river monitoring and planning
Pollution Control Board	Pollution-hotspot identification
Researchers	Integrated AI–GIS framework
Urban Planners	Sustainable infrastructure planning
Environmental Agencies	Continuous monitoring and forecasting
Citizens	Improved river-health awareness

XI. Conclusion

This paper describes the development of an AI-aided geospatial scheme for the intelligent management of river water quality by combining different sources of data like Artificial Intelligence tools, remote sensing methods, geospatial software (GIS), Google Earth Engine, the RiverWatch monitoring system, and the river water quality measurements over time for the Mula, Mutha River Basin Pune India. In fact, the work highlights a way of doing this by using not only field data but also remote sensing data over long periods that can be used for water quality monitoring, prediction, and visualization.

A complete approach was developed by merging monthly RiverWatch water quality monitoring data (2017, 2023) with Sentinel-2, Landsat-8, a Digital Elevation Model, and weather datasets. To perform cloud-based satellite-image processing and to calculate the NDWI, NDTI, and NDCI, Google Earth Engine was utilised. However, watershed delineation, LULC analysis, spatial interpolation, and pollution hotspot identification were performed with ArcGIS Pro, and also, the development of a GIS dashboard was done on ArcGIS platform. For forecasting water quality conditions of a river in the future, several AI models, which include, for example LSTM Random Forest, SVR, and Gradient Boosting were developed; in general, it is expected that LSTM will give the best result in forecasting because this model can not only learn from long-time dependencies of past events but also account for seasons in forecasting, Random Forest, in turn, highlighted the major factors responsible for influencing pollution at riverbanks, i. e. environmental variables.

The integration and visualization of AI predictions and GIS led to a clever decision-support dashboard that displayed the present river conditions, pollution hotspots, time series of changes, the Water Quality Index, and AI-generated forecasts. It gave local government authorities and environmental agencies the possibility of changing from reactive to proactive river management. The setup suggests that by offering constant spatio-temporal monitoring, forecasting based on AI, satellite-assisted assessment of inaccessible river reaches, identification of pollution hotspots, interactive GIS visualization, and support for decision-making in sustainable urban water management it addresses several drawbacks of traditional river-monitoring systems. Also, it contributes directly to the fulfillment of SDG 6 and SDG 11 and supports SDG 13 and SDG 15 [14]. This structure could be used by municipal corporations, riverbasin organizations, and environmental regulatory agencies for real-time monitoring and making policies based on facts.

XII. Future Scope

While the suggested structure reveals a great potential for managing water quality in rivers by the aid of intelligence, still it has some areas where further refinement could take place. Among them the most important ones are below: first, combining the GIS dashboard with IoT sensor networks to capture data on pH, dissolved oxygen, electrical conductivity, turbidity, and temperature continuously in real-time; second, exploring more sophisticated deep-learning designs, e. g. Bidirectional LSTM, Gated Recurrent Units, Transformer networks, Graph Neural Networks, and Temporal Convolutional Networks; third, getting the benefit of additional satellite data from Sentinel-1 SAR, Sentinel-3 OLCI MODIS PlanetScope, WorldView and UAV/drone images to achieve a better spatial and temporal resolution; fourth, incorporating climate data in the shape of CMIP6 projections, climate scenarios, rainfall drought assessment and flood forecasting tools in the system to understand the climate impact on water level; fifth, using Explainable AI methods like SHAP and LIME for revealing AI results and Because of this building the trust of stakeholders. Besides the five points above, there're two more things to consider. Creating a Digital Twin of the Mula-Mutha River for simulating the system and adapting the management basin. Besides, connecting the GIS dashboard with Smart City control centers municipalities boards responsible for pollution control, and systems for issuing early-warnings and flood management are other areas for development. Lastly, the proposed system will also be a good tool for pollution-source identification, planning wastewater-treatment facilities, river-rejuvenation projects, environmental-impact assessment of industrial activities, and long-term urban planning based on sustainability.

XIII. Limitations of the Study

The present research has certain limitations for which it may need further study and improvements: it uses RiverWatch monthly records rather than continuous monitoring, and cloud contamination could diminish satellite imagery availability through the monsoon season. In addition, performance of the AI model is closely related to the size and quality of historical datasets, and this setup has been presented for the Mula, Mutha River Basin only and needs to be validated before application on different rivers.

Acknowledgment

The authors would like to thank the RiverWatch Programme, Pune Municipal Corporation, Maharashtra Pollution Control Board (MPCB) USGS Copernicus Open Access Hub, and Google Earth Engine for allowing us access to the data and computational infrastructure utilized in the production of this study. The authors are very grateful to Bharati Vidyapeeth (Deemed to be University), College of Engineering Pune for providing a great deal of assistance in realizing and carrying out this research.

References

[1] J. G. McCarthy, "Remote sensing of water quality: A review," International Journal of Remote Sensing, vol. 32, no. 19, pp. 5481–5500, 2011.
[2] C. O. McFeeters, "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features," International Journal of Remote Sensing, vol. 17, no. 7, pp. 1425–1432, 1996.
[3] R. N. Carlson, "A trophic state index for lakes," Limnology and Oceanography, vol. 22, no. 2, pp. 361–369, 1977.
[4] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[5] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, pp. 273–297, 1995.
[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[7] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
[8] J. R. Jensen, Remote Sensing of the Environment: An Earth Resource Perspective, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2007.
[9] M. F. Goodchild, "Geographical information systems and science," International Journal of Geographical Information Science, vol. 6, no. 1, pp. 31–45, 1992.
[10] Google Earth Engine Team, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," 2024.
[11] European Space Agency, "Sentinel-2 User Guide," ESA, 2024.
[12] United States Geological Survey, "Landsat 8 Collection 2 Level-2 Science Product Guide," USGS, 2023.
[13] APHA, AWWA, and WEF, Standard Methods for the Examination of Water and Wastewater, 24th ed. Washington, DC, USA: American Public Health Association, 2023.
[14] United Nations, Transforming Our World: The 2030 Agenda for Sustainable Development, New York, NY, USA, 2015.
[15] World Health Organization, Guidelines for Drinking-water Quality, 4th ed., Geneva, Switzerland, 2022.
[16] P. D. Pawar et al., "Spatio-temporal analysis of river water quality using RiverWatch observations and satellite imagery for the Mula–Mutha River Basin" (in press).
[17] Pawar, P. D., & Sohoni, V. (2026). Enhancing Water Quality Assessment in Mula and Mutha Rivers Through Linear Interpolation and the River Watch Monitoring Platform. Waterlines, 44(1), 89-103. https://doi.org/10.3362/waterlines.v44i1.620