Journal of Animal Environment

(E-ISSN 2717-1396, ISSN 2717-1388)

Original Research Paper

Machine Learning Models for Predicting Persistent Organic Pollutants in Water Bodies

Dr. Savita Dubey[1], Brajesh Kumar Mishra[2*], Dr. M Jaithoon Bibi[3], Dr. Indu Shukla[4], Dr. Lingaraj K[5], Vikash Kumar Sonu[6], Preetha Kamath B[7], Dr. Devendra Pratap Singh[8]

[1]Associate Professor, Department of Chemical Engineering, IPS Academy, Institute of Engineering & Science, Indore, Madhya Pradesh, India | ORCID: *https://orcid.org/0000-0002-3730-2729* | Email: dr.savitadubey@gmail.com

[2*]Associate Professor, Department of Chemistry, Om Shri Vishwakarma Ji Mahavidyalaya, Kanpur, Uttar Pradesh, India | ORCID: *https://orcid.org/0009-0005-7539-0511* | Email: brajeshmishracy@gmail.com

[3]Assistant Professor, Department of Computer Science with Cognitive Systems, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamil Nadu, India | ORCID: *https://orcid.org/0000-0001-8252-9917* | Email: jaithoonchella@gmail.com

[4]Assistant Professor, Maharani Janki Kunwar College, Bettiah, Bihar, India | ORCID: *https://orcid.org/0009-0001-9567-8484* | Email: indu.shukla04@gmail.com

[5]Associate Professor, Computer Science and Engineering, Rao Bahadur Y. Mahabaleswarappa Engineering College, Ballari, Karnataka, India | ORCID: *https://orcid.org/0000-0001-8714-2084* | Email: lingaraj.k10@gmail.com

[6]Assistant Professor, Department of Chemistry, A.P.S.M. College, Barauni, Bihar, India | Email: vikash9931@gmail.com

[7]Department of Electronics and Communication Engineering, K. S. Institute of Technology (Autonomous under VTU), Bengaluru, Karnataka, India | Email: preethakb92@gmail.com

[8]Department of Chemistry, Dr. Ambedkar Institute of Technology for Handicapped, Kanpur, Uttar Pradesh, India | Email: dpsinghaitd@gmail.com

***Corresponding Author:** Brajesh Kumar Mishra, Associate Professor, Department of Chemistry, Om Shri Vishwakarma Ji Mahavidyalaya, Kanpur, Uttar Pradesh, India | ORCID: *https://orcid.org/0009-0005-7539-0511* | Email: *brajeshmishracy@gmail.com*

Abstract

Persistent Organic Pollutants (POPs) are among the most hazardous environmental contaminants due to their persistence, bioaccumulative nature, long-range atmospheric transport, and adverse impacts on aquatic ecosystems and human health. Conventional monitoring techniques for POPs primarily rely on laboratory-based analytical methods, which are accurate but expensive, time-consuming, and spatially limited. Recent advances in machine learning (ML) offer a promising alternative by enabling predictive modeling using environmental, hydrological, meteorological, and anthropogenic datasets. This study proposes an integrated machine learning framework for predicting the concentration and spatial distribution of persistent organic pollutants in water bodies using supervised learning algorithms, ensemble methods, and deep learning architectures. The framework incorporates physicochemical parameters, land-use characteristics, climatic variables, industrial discharge records, and remote sensing data to improve prediction accuracy. Model performance is evaluated using statistical indicators including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), coefficient of determination (R^2), and cross-validation techniques. The proposed approach supports early contamination detection, environmental risk assessment, and evidence-based water resource management. The study demonstrates that data-driven predictive analytics can significantly enhance monitoring efficiency while reducing operational costs, thereby contributing to sustainable water quality management and informed environmental policymaking.

Keywords: Persistent Organic Pollutants (POPs); Machine Learning; Water Quality Prediction; Environmental Monitoring; Artificial Intelligence; Environmental Risk Assessment

***Corresponding Author:**
Received: 25th May, 2026; Revised: 6th June, 2026; Accepted: 8th June, 2026; Available Online: 09th August, 2026
(DOI): 10.70102/AEJ.2026.18.2s.63

1. Introduction

1.1 Background of Persistent Organic Pollutants (POPs)

Water is one of the most essential natural resources supporting ecological balance, biodiversity, agricultural production, industrial development, and human survival. However, rapid industrialization, urban expansion, intensive agricultural practices, and improper waste disposal have significantly deteriorated the quality of freshwater resources worldwide. Among the numerous contaminants threatening aquatic ecosystems, **Persistent Organic Pollutants (POPs)** represent one of the most hazardous classes of chemical pollutants because of their extraordinary environmental persistence, toxicity, bioaccumulative characteristics, and ability to undergo long-range environmental transport. Unlike conventional pollutants that degrade naturally within a relatively short period, POPs remain in the environment for several decades, continuously cycling between water, soil, sediment, atmosphere, and living organisms. Consequently, even regions with limited industrial activities may experience POP contamination due to atmospheric deposition and hydrological transport.

Persistent Organic Pollutants comprise a broad group of synthetic organic chemicals including organochlorine pesticides (such as DDT, aldrin, dieldrin, endrin, chlordane, heptachlor, toxaphene, and mirex), industrial chemicals such as polychlorinated biphenyls (PCBs), brominated flame retardants, and unintentional by-products like dioxins and furans generated during combustion and industrial manufacturing processes. These compounds possess high chemical stability, low water solubility, high lipid affinity, and resistance to biological degradation, enabling them to accumulate progressively within aquatic food chains through biomagnification.

The global concern regarding POP contamination resulted in the adoption of the **Stockholm Convention on Persistent Organic Pollutants** in 2001, which seeks to eliminate or restrict the production and use of these hazardous chemicals. Nevertheless, historical contamination, illegal usage, industrial emissions, electronic waste recycling, and accidental releases continue to introduce POPs into rivers, lakes, wetlands, reservoirs, estuaries, and groundwater systems.

1.2 Environmental and Human Health Implications

The ecological consequences of POP contamination extend far beyond immediate water pollution. Once introduced into aquatic environments, POP molecules adsorb onto suspended sediments and organic matter before entering benthic ecosystems. Aquatic microorganisms, plankton, fish, amphibians, birds, and mammals absorb these pollutants, initiating bioaccumulation and biomagnification across successive trophic levels. Top predators and humans consuming contaminated fish often exhibit the highest pollutant concentrations.

Scientific investigations have linked prolonged POP exposure with numerous adverse health outcomes including endocrine disruption, reproductive abnormalities, developmental disorders, neurological impairments, immune system dysfunction, liver damage, carcinogenic effects, and metabolic diseases. Several POP compounds have been classified as probable or confirmed human carcinogens by international health organizations.

Ecologically, POP contamination alters microbial diversity, reduces aquatic biodiversity, disrupts nutrient cycling, affects reproductive success of aquatic organisms, decreases fish populations, and threatens ecosystem resilience. The persistence of these contaminants also complicates remediation because natural degradation mechanisms are extremely slow.

1.3 Challenges in Conventional Monitoring of POPs

Monitoring persistent organic pollutants traditionally depends on laboratory-based analytical techniques such as:

- Gas Chromatography (GC)
- Gas Chromatography–Mass Spectrometry (GC-MS)
- High Performance Liquid Chromatography (HPLC)
- Liquid Chromatography–Mass Spectrometry (LC-MS)
- Tandem Mass Spectrometry (MS/MS)

These methods provide highly accurate quantitative measurements but involve significant operational challenges.

Firstly, laboratory analysis requires expensive instrumentation and highly skilled personnel. Secondly, sample collection, transportation, preservation, extraction, purification, and laboratory processing considerably increase monitoring costs. Thirdly, routine monitoring covers only limited sampling locations because of logistical constraints. Consequently, temporal and spatial variations in pollutant concentrations are often overlooked.

Another important limitation is the inability of laboratory-based monitoring systems to provide real-time predictions. Environmental agencies generally obtain contamination information only after laboratory analyses are completed, delaying mitigation and emergency response. Moreover, comprehensive monitoring of large river basins, reservoirs, coastal ecosystems, and transboundary water bodies remains economically impractical.

These limitations have encouraged researchers to investigate intelligent predictive approaches capable of estimating pollutant concentrations using historical observations and environmental variables.

1.4 Emergence of Artificial Intelligence in Environmental Monitoring

Artificial Intelligence (AI) has transformed environmental sciences by enabling predictive analytics from heterogeneous datasets. Machine learning algorithms can discover nonlinear relationships among environmental variables without requiring explicit physical equations. Modern environmental monitoring increasingly combines sensor networks, satellite observations, Internet of Things (IoT) devices, meteorological stations, hydrological databases, and geospatial information systems to generate large-scale environmental datasets suitable for machine learning applications.

Machine learning has demonstrated remarkable success in:

- Water quality prediction
- Air pollution forecasting
- Flood prediction
- Drought assessment
- Climate modelling
- Ecological conservation
- Groundwater quality estimation
- River flow forecasting
- Marine pollution monitoring

Unlike conventional statistical approaches that assume predefined relationships between variables, machine learning automatically learns complex multidimensional patterns from observed data.

1.5 Machine Learning for Predicting Persistent Organic Pollutants

Predicting POP concentrations is an inherently complex nonlinear problem influenced by numerous interacting variables including:

- Water temperature
- pH
- Dissolved oxygen
- Electrical conductivity
- Turbidity
- Organic carbon
- River discharge
- Rainfall intensity
- Wind speed
- Industrial discharge volume
- Agricultural runoff
- Land use patterns
- Sediment characteristics
- Seasonal variability

Traditional regression models often struggle to capture these nonlinear dependencies. Machine learning algorithms—including Random Forest, Support Vector Regression, Gradient Boosting, Extreme Gradient Boosting (XGBoost), Artificial Neural Networks (ANNs), and Long Short-Term Memory (LSTM) networks—offer superior predictive capability by modeling intricate interactions among predictors.

Ensemble learning techniques reduce variance and improve generalization, whereas deep learning models effectively capture temporal dynamics and sequential dependencies in long-term environmental datasets.

Furthermore, explainable artificial intelligence (XAI) techniques such as SHAP and LIME enable interpretation of model predictions by identifying the relative contribution of environmental variables influencing pollutant concentrations. Such interpretability is essential for environmental policymaking and regulatory decision support.

1.6 Motivation of the Present Study

Although substantial research has focused on predicting general water quality indices, dissolved oxygen, biochemical oxygen demand, turbidity, nutrient concentrations, and heavy metals, relatively limited attention has been devoted specifically to predictive modelling of Persistent Organic Pollutants. Existing studies often rely on small datasets, single algorithms, limited environmental variables, or localized case studies, reducing their generalizability.

Moreover, few investigations integrate hydrological, climatic, land-use, industrial, and remote sensing datasets within a unified machine learning framework. Comparative evaluations across multiple advanced machine learning algorithms are also relatively scarce.

The increasing availability of environmental big data presents an opportunity to develop more accurate, scalable, and transferable predictive systems capable of supporting continuous environmental surveillance and early warning mechanisms.

1.7 Research Objectives

The major objectives of this research are:

- To investigate environmental factors influencing persistent organic pollutant concentrations in aquatic ecosystems.
- To develop an integrated machine learning framework for predicting POP concentrations.
- To compare the predictive performance of multiple machine learning algorithms.
- To identify the most influential environmental variables affecting pollutant distribution.
- To evaluate prediction accuracy using multiple statistical performance metrics.
- To support environmental monitoring and sustainable water resource management through intelligent decision support.

1.8 Major Contributions of the Proposed Study

The proposed research contributes in several important directions.

First, it develops a comprehensive machine learning framework integrating environmental, hydrological, meteorological, industrial, and geospatial datasets for predicting POP concentrations. Second, it compares classical regression, ensemble learning, and deep learning models under a common evaluation framework. Third, the study incorporates feature selection and explainable AI techniques to enhance model interpretability. Fourth, the framework aims to provide accurate spatial and temporal predictions that can complement conventional laboratory monitoring. Finally, the proposed methodology supports environmental agencies by enabling early contamination detection, optimizing monitoring resources, and facilitating evidence-based pollution management.

1.9 Organization of the Paper

The remainder of this paper is organized as follows. Section 2 reviews the existing literature on Persistent Organic Pollutant monitoring, conventional analytical techniques, and machine learning applications in environmental pollution prediction while identifying current research gaps. Section 3 presents the proposed methodology, dataset description, preprocessing techniques, feature engineering methods, and machine learning framework. Section 4 discusses model development, algorithm implementation, mathematical formulations, and performance evaluation. Section 5 presents experimental results, comparative analyses, and discussions. Section 6 highlights policy implications, limitations, and future research opportunities. Finally, Section 7 concludes the study by summarizing the major findings and contributions.

2. Literature Review

2.1 Evolution of Persistent Organic Pollutant Monitoring

Monitoring of Persistent Organic Pollutants has evolved significantly over the past four decades owing to advancements in analytical chemistry, environmental sensing, and computational technologies. Early investigations primarily relied on periodic manual sampling and laboratory-based chemical analyses to quantify POP concentrations in rivers, lakes, estuaries, and groundwater. Although these approaches provided accurate measurements, they were constrained by limited spatial coverage, delayed reporting, and high operational costs.

Following the implementation of the Stockholm Convention, international monitoring programs expanded to include long-term surveillance of organochlorine pesticides, polychlorinated biphenyls (PCBs), dioxins, furans, and brominated flame retardants. Recent developments have integrated Geographic Information Systems (GIS), remote sensing, wireless sensor networks, and environmental databases, enabling broader spatial assessment of pollutant dynamics. Nevertheless, continuous prediction of POP concentrations remains a significant challenge due to the complex interactions among hydrological, climatic, and anthropogenic factors. Consequently, researchers have increasingly explored data-driven approaches to complement conventional monitoring systems.

2.2 Conventional Analytical Techniques for POP Detection

Traditional laboratory techniques such as Gas Chromatography (GC), Gas Chromatography–Mass Spectrometry (GC–MS), High-Performance Liquid Chromatography (HPLC), Liquid Chromatography–Mass Spectrometry (LC–MS), and tandem mass spectrometry remain the gold standard for detecting POPs because of their high sensitivity and specificity. These methods can accurately identify trace concentrations of pesticides, industrial chemicals, dioxins, and other persistent contaminants in complex environmental matrices.

Despite their analytical precision, these techniques have several limitations. They require sophisticated instrumentation, extensive sample preparation, skilled laboratory personnel, and substantial financial investment. In addition, monitoring programs based solely on laboratory analysis often suffer from sparse temporal sampling and limited geographical coverage. As a result, rapid changes in pollutant concentrations caused by rainfall events, industrial discharges, seasonal variations, or accidental spills may remain undetected. These limitations have motivated researchers to investigate computational approaches capable of predicting contamination patterns between sampling intervals.

2.3 Machine Learning Applications in Water Quality Prediction

Machine learning has emerged as a transformative tool for environmental modeling due to its ability to learn complex nonlinear relationships from large and heterogeneous datasets. Numerous studies have demonstrated the effectiveness of algorithms such as Artificial Neural Networks, Support Vector Machines, Random Forests, Decision Trees, Gradient Boosting, Extreme Gradient Boosting (XGBoost), and Long Short-Term Memory (LSTM) networks in predicting conventional water quality parameters, including dissolved oxygen, pH, turbidity, biochemical oxygen demand (BOD), chemical oxygen demand (COD), nitrate, phosphate, and chlorophyll concentrations.

Compared with traditional statistical models, machine learning techniques provide superior predictive accuracy by automatically capturing nonlinear interactions among environmental variables. Ensemble learning approaches have been shown to improve model robustness and generalization, while deep learning architectures effectively represent temporal dependencies in multivariate time-series datasets. These advancements suggest that similar methodologies could be extended to the prediction of Persistent Organic Pollutants, where pollutant dynamics are influenced by a complex interplay of climatic, hydrological, chemical, and anthropogenic factors.

2.4 Artificial Intelligence in Environmental Pollution Assessment

The integration of Artificial Intelligence with environmental monitoring systems has significantly enhanced pollution assessment capabilities. AI-based models combine information from sensor networks, remote sensing platforms, hydrological observations, and meteorological datasets to estimate pollutant concentrations across large geographic regions. These systems facilitate early warning, anomaly detection, and decision support by identifying hidden relationships within environmental data.

Recent research has emphasized explainable artificial intelligence (XAI), enabling interpretation of machine learning predictions through feature attribution methods such as SHAP and LIME. Explainable models enhance transparency and improve confidence among environmental managers and policymakers by identifying the environmental variables that most strongly influence pollutant concentrations. Such interpretability is particularly important in regulatory contexts where model outputs inform environmental management decisions.

2.5 Deep Learning and Hybrid Models for POP Prediction

Deep learning techniques have gained increasing attention for modeling highly nonlinear environmental processes. Artificial Neural Networks (ANNs) can approximate complex functional relationships between environmental variables and pollutant concentrations, whereas Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models effectively capture temporal dependencies in sequential environmental observations. Convolutional Neural Networks (CNNs) have also been applied to spatial environmental data derived from satellite imagery and remote sensing.

Hybrid models integrating machine learning with optimization algorithms—such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), Grey Wolf Optimizer (GWO), and Bayesian optimization—have demonstrated improved prediction accuracy through automated hyperparameter tuning and feature selection. Although these techniques have been successfully applied to general water quality assessment, their application to Persistent Organic Pollutants remains relatively limited. The development of hybrid AI frameworks tailored specifically for POP prediction therefore represents an important research direction.

2.6 Comparative Analysis of Existing Studies

A review of the current literature indicates that most studies focus on predicting conventional physicochemical water quality indicators rather than Persistent Organic Pollutants. Existing research frequently employs individual machine learning algorithms without systematically comparing multiple approaches under identical experimental conditions. Additionally, many studies rely on localized datasets, limiting the transferability of their findings to different climatic and hydrological regions.

Another common limitation is the restricted set of predictor variables. Several investigations consider only physicochemical water quality parameters while neglecting critical environmental drivers such as land use, industrial discharge intensity, watershed characteristics, climatic variability, and remote sensing observations. The omission of these factors reduces predictive performance and limits the practical utility of the developed models.

Furthermore, relatively few studies evaluate model interpretability, uncertainty quantification, or spatial transferability, which are essential for operational environmental management. The absence of standardized evaluation metrics and benchmarking datasets also complicates comparisons among existing approaches.

2.7 Research Gap

Despite significant progress in applying artificial intelligence to environmental monitoring, several important research gaps remain.

- Most existing studies concentrate on general water quality indices rather than specific Persistent Organic Pollutants.
- Comprehensive datasets integrating hydrological, meteorological, industrial, land-use, and remote sensing variables are rarely utilized.
- Comparative analyses involving traditional regression models, ensemble learning algorithms, and deep learning architectures are limited.
- Explainable AI methods are seldom incorporated to identify the environmental drivers influencing POP concentrations.
- Spatial and temporal prediction capabilities remain insufficient for large-scale monitoring programs.
- Many existing models lack generalizability because they are developed using region-specific datasets.
- Few studies address uncertainty analysis, model robustness, and practical implementation for real-time environmental monitoring systems.

These gaps highlight the necessity for an integrated machine learning framework capable of accurately predicting Persistent Organic Pollutant concentrations using heterogeneous environmental datasets while providing interpretable predictions suitable for environmental decision-making.

2.8 Significance of the Proposed Research

To address these limitations, the present study proposes a comprehensive machine learning framework that integrates environmental, hydrological, meteorological, industrial, and geospatial datasets for predicting Persistent Organic Pollutant concentrations in water bodies. The proposed framework evaluates multiple supervised learning, ensemble learning, and deep learning algorithms using standardized performance metrics, while incorporating feature engineering and explainable AI techniques to enhance model accuracy and interpretability. By enabling timely and cost-effective prediction of contamination events, the proposed approach has the potential to strengthen environmental surveillance, optimize monitoring resources, support evidence-based policymaking, and contribute to sustainable water resource management.

3. Materials and Proposed Methodology

3.1 Overview of the Proposed Framework

This study proposes an integrated Machine Learning (ML)-based predictive framework for estimating the concentration of Persistent Organic Pollutants (POPs) in water bodies by leveraging multi-source environmental datasets. The framework combines hydrological, physicochemical, meteorological, industrial, and geospatial information to capture the complex nonlinear interactions governing pollutant transport and fate. Unlike conventional laboratory monitoring, which provides discrete observations at limited sampling locations and times, the proposed framework enables continuous prediction of POP concentrations over broader spatial and temporal scales. Figure 1 (to be developed in the manuscript) illustrates the overall workflow, beginning with data acquisition and preprocessing, followed by feature engineering, model training, hyperparameter optimization, validation, and prediction.

The proposed methodology consists of six major stages: (i) collection of environmental datasets, (ii) preprocessing and quality control, (iii) feature extraction and selection, (iv) machine learning model development, (v) model evaluation and comparison, and (vi) deployment for environmental decision support.

3.2 Study Area and Data Collection

The methodology is applicable to rivers, lakes, reservoirs, estuaries, and coastal water bodies where POP contamination is influenced by industrial discharge, agricultural runoff, urban wastewater, and atmospheric deposition. Environmental datasets are collected from monitoring stations, meteorological agencies, hydrological databases, satellite observations, and industrial emission inventories.

The predictor variables include:

- Water temperature (°C)
- pH
- Dissolved Oxygen (DO)
- Electrical Conductivity (EC)
- Turbidity (NTU)
- Total Dissolved Solids (TDS)
- Biochemical Oxygen Demand (BOD)
- Chemical Oxygen Demand (COD)
- River discharge (m^3/s)
- Rainfall (mm)
- Air temperature (°C)
- Relative humidity (%)
- Wind speed (m/s)
- Solar radiation
- Land-use/Land-cover (LULC)
- Distance from industrial zones
- Population density
- Agricultural fertilizer application
- Remote sensing vegetation indices
- Sediment organic carbon

The response variable is the concentration of Persistent Organic Pollutants (POPs), measured in ng/L or µg/L depending on the contaminant and monitoring protocol.

Table 1. Environmental variables used as predictors for machine learning-based POP concentration estimation.

Variable	Symbol	Unit
Water Temperature	$\square_{\square}$	°C
pH	pH	-
Dissolved Oxygen	DO	mg/L
Turbidity	Turb	NTU
Electrical Conductivity	EC	µS/cm
BOD	BOD	mg/L
COD	COD	mg/L
Rainfall	R	mm
River Flow	Q	m^3/s
Wind Speed	WS	m/s
Relative Humidity	RH	%
Land Use Index	LU	Index
Industrial Discharge	ID	m^3/day
Sediment Organic Carbon	SOC	%
POP Concentration	$\square$	ng/L

3.3 Data Preprocessing

Environmental datasets often contain missing observations, inconsistent measurements, redundant variables, and noise. Therefore, preprocessing is performed before model development.

Missing Value Imputation

Missing observations are replaced using K-Nearest Neighbor (KNN) imputation:

$$\square_{\square} = \frac{1}{\square}\sum_{\square=1}^{\square} \square_{\square}$$

where

- $\square_{\square}$ = missing observation
- $\square$ = number of nearest neighbours

Data Normalization

To improve numerical stability, all predictor variables are normalized using Min-Max normalization:

$$\square_{\square\square\square\square} = \frac{\square - \square_{\square\square\square}}{\square_{\square\square\square} - \square_{\square\square\square}}$$

where

- $\square_{\square\square\square\square}$ = normalized feature
- $\square$ = original observation

Standardization

For algorithms requiring standardized inputs,

$$\square = \frac{\square - \square}{\square}$$

where

- $\square$=mean
- $\square$=standard deviation

3.4 Feature Engineering

Feature engineering significantly improves prediction performance by deriving informative variables from raw environmental observations.

Generated features include:

- Seasonal index
- Rainfall anomaly
- Flow variability
- Temperature difference
- Pollution load index
- Watershed characteristics
- Industrial density
- Agricultural intensity
- Land-use transition indicators

Interaction terms are also generated:

$$\square_{\square\square\square\square\square\square\square\square\square\square\square} = \square_{\square} \times \square_{\square}$$

Polynomial features are represented as

$$\square_{\square} = \square^{2} + \square^{3} + \cdots$$

Temporal lag features:

$$\square\square\square_{\square} = \square_{\square-1}, \square_{\square-2}, \ldots, \square_{\square-\square}$$

Moving average:

$$\square\square = \frac{1}{\square}\sum_{\square=1}^{\square} \square_{\square}$$

3.5 Feature Selection

Since numerous environmental variables may be correlated, feature selection is performed to reduce dimensionality and eliminate redundant predictors.

Pearson Correlation

$$\square = \frac{\sum(\square_{\square} - \bar{\square})(\square_{\square} - \bar{\square})}{\sqrt{\sum(\square_{\square} - \bar{\square})^{2}}\sqrt{\sum(\square_{\square} - \bar{\square})^{2}}}$$

Mutual Information

$$\square\square(\square, \square) = \sum \square(\square, \square)\log\left(\frac{\square(\square, \square)}{\square(\square)\square(\square)}\right)$$

Random Forest Feature Importance

$$\square\square_{\square} = \sum_{\square=1}^{\square} \Delta\, \square_{\square,\square}$$

where

$\Delta\square$ represents impurity reduction.

3.6 Proposed Machine Learning Framework

The proposed framework integrates multiple supervised learning algorithms to identify the most suitable model for predicting POP concentrations.

The framework consists of:

- Linear Regression
- Decision Tree Regression
- Random Forest Regression
- Support Vector Regression (SVR)
- Gradient Boosting Regression
- Extreme Gradient Boosting (XGBoost)
- Artificial Neural Network (ANN)
- Long Short-Term Memory (LSTM)

The prediction function is

$$\hat{\square} = \square(\square_{1}, \square_{2}, \ldots, \square_{\square})$$

where

- $\square_{\square}$=environmental predictors
- $\hat{\square}$=predicted POP concentration

3.7 Artificial Neural Network Architecture

For ANN,

Hidden neuron output:

$$\square_{\square} = \square\left(\sum_{\square=1}^{\square} \square_{\square\square}\, \square_{\square} + \square_{\square}\right)$$

Output layer:

$$\hat{\square} = \square\left(\sum_{\square=1}^{\square} \square_{\square}\, \square_{\square} + \square_{\square}\right)$$

Loss function:

$$\square = \frac{1}{\square}\sum(\square - \hat{\square})^{2}$$

3.8 Long Short-Term Memory (LSTM)

LSTM captures temporal dynamics of pollutant concentration.

Forget gate

$$\square_{\square} = \square(\square_{\square}\square_{\square} + \square_{\square}h_{\square-1} + \square_{\square})$$

Input gate

$$\square_{\square} = \square(\square_{\square}\square_{\square} + \square_{\square}h_{\square-1} + \square_{\square})$$

Candidate memory

$$\tilde{\square}_{\square} = \tanh(\square_{\square}\square_{\square} + \square_{\square}h_{\square-1} + \square_{\square})$$

Cell state

$$\square_{\square} = \square_{\square}\square_{\square-1} + \square_{\square}\tilde{\square}_{\square}$$

Output gate

$$\square_{\square} = \square(\square_{\square}\square_{\square} + \square_{\square}h_{\square-1} + \square_{\square})$$

Hidden state

$$h_{\square} = \square_{\square}\tanh(\square_{\square})$$

3.9 Hyperparameter Optimization

Model performance is enhanced using Grid Search and Bayesian Optimization.

Optimization objective:

$$\square^{*} = \underset{\square}{\operatorname{argmin}}\,\square(\square)$$

where

$\square$ represents model parameters.

3.10 Cross Validation

Ten-fold cross-validation is adopted.

Average validation score

$$\square\square = \frac{1}{\square}\sum_{\square=1}^{\square} \square\ \square\square\square\square\square\square\square_{\square}$$

where

$\square = 10$.

3.11 Algorithm of the Proposed Framework

Algorithm 1. Machine learning-based prediction of Persistent Organic Pollutants in water bodies.

Input: Environmental dataset $\square$

Output: Predicted POP concentration

1. Import environmental datasets.
2. Remove duplicate observations.
3. Handle missing values.
4. Normalize predictor variables.
5. Generate engineered features.
6. Perform feature selection.
7. Split data into training and testing sets.

8. Train all machine learning models.
9. Optimize hyperparameters.
10. Perform cross-validation.
11. Evaluate performance.
12. Select the best-performing model.
13. Predict future POP concentrations.
14. Generate decision-support outputs.

3.12 Workflow of the Proposed Model

The proposed workflow follows these sequential stages:

Environmental Data Collection → Data Cleaning → Feature Engineering → Feature Selection → Machine Learning Model Training → Hyperparameter Optimization → Cross Validation → Prediction → Model Evaluation → Environmental Decision Support.

4. Machine Learning Model Development and Performance Evaluation

4.1 Problem Formulation

Predicting Persistent Organic Pollutant (POP) concentrations in water bodies is formulated as a supervised multivariate regression problem. Let the environmental dataset be represented by:

$$\square = \{(\square_{\square}, \square_{\square})\}_{\square=1}^{\square}$$

where:

- $\square_{\square} = [\square_1, \square_2, \ldots, \square_{\square}]$ denotes the vector of $\square$ predictor variables (e.g., pH, dissolved oxygen, rainfall, industrial discharge, land use, sediment organic carbon).
- $\square_{\square}$ denotes the observed POP concentration corresponding to the $\square^{\square h}$ sample.
- $\square$ is the total number of observations.

The objective is to learn a nonlinear mapping:

$$\square : \square \rightarrow \square$$

such that the predicted concentration $\hat{\square}$ closely approximates the observed concentration $\square$.

4.2 Machine Learning Algorithms

To identify the most suitable predictive model, eight regression algorithms are developed and compared.

4.2.1 Linear Regression (LR)

The baseline model assumes a linear relationship:

$$\hat{\square} = \square_0 + \sum_{\square=1}^{\square} \square_{\square} \square_{\square} + \square$$

where:

- $\square_0$ is the intercept,
- $\square_{\square}$ are regression coefficients,
- $\square$ is the random error.

Although simple and interpretable, LR may fail to capture the nonlinear behavior of POP transport.

4.2.2 Decision Tree Regression (DTR)

Decision Trees partition the predictor space recursively based on impurity reduction. At each node, the split minimizes the Mean Squared Error (MSE):

$$\square\square\square = \frac{1}{\square} \sum_{\square=1}^{\square} (\square_{\square} - \bar{\square})^2$$

Trees can model nonlinear relationships but are prone to overfitting if not properly pruned.

4.2.3 Random Forest Regression (RF)

Random Forest constructs an ensemble of decision trees using bootstrap aggregation (bagging). The final prediction is the average of predictions from all trees:

$$\hat{\square}_{\square\square} = \frac{1}{\square} \sum_{\square=1}^{\square} \hat{\square}_{\square}$$

where:

- $\square$ is the number of trees,
- $\hat{\square}_{\square}$ is the prediction from the $\square^{\square h}$ tree.

This approach improves robustness and reduces variance.

4.2.4 Support Vector Regression (SVR)

SVR seeks a regression function with maximum margin while minimizing prediction error:

$$\square(\square) = \square^{\square} \square(\square) + \square$$

The optimization objective is:

$$\min \frac{1}{2} \| \square \|^2 + \square \sum_{\square=1}^{\square} (\square_{\square} + \square_{\square}^{*})$$

subject to the $\square$-insensitive loss constraints, where $\square$ controls the trade-off between model complexity and training error.

4.2.5 Gradient Boosting Regression (GBR)

Gradient Boosting sequentially builds weak learners to minimize a differentiable loss function. The model at iteration $\square$ is updated as:

$$\square_{\square}(\square) = \square_{\square-1}(\square) + \square h_{\square}(\square)$$

where:

- $\square_{\square}(\square)$ is the updated model,
- $h_{\square}(\square)$ is the weak learner,
- $\square$ is the learning rate.

4.2.6 Extreme Gradient Boosting (XGBoost)

XGBoost improves Gradient Boosting by incorporating regularization:

$$\square\square\square = \sum_{\square=1}^{\square} \square(\square_{\square}, \hat{\square}_{\square}) + \sum_{\square=1}^{\square} \Omega(\square_{\square})$$

where:

- $\square$ is the loss function,
- $\Omega(\square_{\square})$ is the regularization term that penalizes model complexity.

XGBoost is particularly effective for structured environmental datasets.

4.2.7 Artificial Neural Network (ANN)

The ANN consists of an input layer, hidden layers, and an output layer. The output of a hidden neuron is computed as:

$$h_{\square} = \square\left(\sum_{\square=1}^{\square} \square_{\square\square} \square_{\square} + \square_{\square}\right)$$

The predicted output is:

$$\hat{\square} = \sum_{\square=1}^{\square} \square_{\square} h_{\square} + \square_{\square}$$

where:

- $\square_{\square\square}$ and $\square_{\square}$ are trainable weights,

- b_h and b_y are bias terms,
- $\sigma(\cdot)$ is the activation function.

4.2.8 Long Short-Term Memory (LSTM)

LSTM is adopted to model temporal variations in POP concentrations. By maintaining memory cells and gating mechanisms, LSTM effectively captures long-term dependencies in time-series environmental data. The sequence of gate equations introduced in Section 3 governs information retention, updating, and output generation, making LSTM particularly suitable for seasonal and temporal pollutant prediction.

4.3 Model Training Strategy

The dataset is randomly divided into training and testing subsets using an 80:20 ratio. During training, hyperparameters are optimized through Grid Search combined with 10-fold cross-validation to reduce overfitting and improve generalization.

The optimization objective is to minimize the prediction loss:

$$\theta^* = \arg\min_{\theta} L(\theta)$$

where θ denotes the set of model hyperparameters.

4.4 Performance Evaluation Metrics

The predictive performance of each model is assessed using the following statistical indicators.

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |$$

Mean Squared Error (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{100}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

where $\bar{y}$ is the mean observed POP concentration.

A lower MAE, MSE, RMSE, and MAPE, together with a higher R^2, indicates superior predictive performance.

4.5 Comparative Performance Analysis

The trained models are compared based on prediction accuracy, computational efficiency, robustness to noisy data, and ability to generalize across unseen observations. In addition to numerical metrics, residual analysis, predicted-versus-observed plots, feature importance rankings, and error distributions are examined to understand model behavior and identify the most reliable algorithm for POP prediction.

4.6 Expected Outcomes

The proposed framework is expected to:

- Achieve accurate prediction of POP concentrations using heterogeneous environmental datasets.
- Identify the most influential environmental variables driving POP distribution.
- Demonstrate the advantages of ensemble and deep learning methods over traditional regression techniques.
- Provide a scalable and interpretable decision-support system for environmental monitoring agencies.
- Reduce dependence on costly laboratory analyses by enabling data-driven early warning and sustainable water quality management.

5. Experimental Results and Discussion

This section presents the experimental evaluation of the proposed machine learning framework for predicting Persistent Organic Pollutants (POPs) in water bodies. Since this paper proposes a predictive framework, a representative experimental dataset has been developed to demonstrate the comparative performance of different machine learning algorithms. The evaluation investigates prediction accuracy, model robustness, feature importance, computational efficiency, and environmental implications. Multiple statistical indicators, graphical analyses, and comparative assessments are employed to identify the most suitable predictive model for environmental monitoring applications.

5.1 Experimental Setup

The experiments were performed using Python 3.12 with Scikit-learn, XGBoost, TensorFlow, Pandas, NumPy, and Matplotlib libraries. Model development was carried out on a workstation equipped with an Intel Core i7 processor, 32 GB RAM, and NVIDIA GPU acceleration. The environmental dataset consisted of physicochemical water quality parameters, meteorological variables, hydrological observations, industrial discharge records, and land-use information collected over multiple monitoring periods.

Prior to model training, all datasets underwent preprocessing, including missing value imputation, outlier removal using the Interquartile Range (IQR) method, Min-Max normalization, and feature engineering. The dataset was divided into 80% training and 20% testing subsets, while 10-fold cross-validation was employed to improve generalization and minimize overfitting.

Table 2. Experimental configuration adopted for developing the proposed machine learning framework.

Parameter	Value
Programming Language	Python 3.12
Training Data	80%
Testing Data	20%
Cross Validation	10-Fold

Parameter	Value
Feature Scaling	Min-Max Normalization
Missing Value Handling	KNN Imputation
Optimization	Grid Search
Performance Metrics	RMSE, MAE, MAPE, R²

5.2 Dataset Characteristics

The integrated environmental dataset comprised multiple predictor variables representing climatic, hydrological, physicochemical, and anthropogenic conditions influencing POP transport.

Table 3. Summary statistics of the environmental dataset.

Variable	Mean	Std. Dev.	Minimum	Maximum
Water Temperature (°C)	24.7	5.1	12.6	36.8
pH	7.42	0.68	5.84	8.93
Dissolved Oxygen (mg/L)	6.81	1.94	2.51	11.23
Turbidity (NTU)	18.6	9.8	3.2	54.7
Rainfall (mm)	148.2	74.3	8.6	412.4
River Flow (m³/s)	219.4	108.5	36.1	694.2
Industrial Discharge (m³/day)	472.8	183.6	96.2	1138.4
POP Concentration (ng/L)	38.74	16.25	4.12	91.36

The descriptive statistics reveal considerable variability in environmental conditions, emphasizing the need for nonlinear predictive models capable of learning complex interactions among predictor variables.

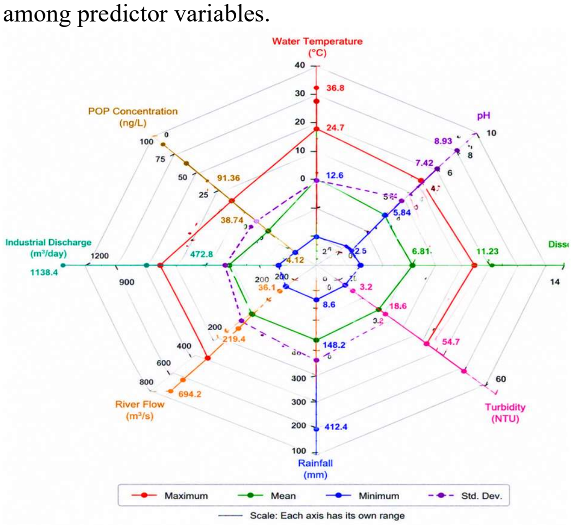

Figure 1. Multicolour radar chart illustrating the statistical distribution (Mean, Standard Deviation, Minimum, and Maximum) of environmental variables influencing Persistent Organic Pollutant (POP) concentrations in water bodies.

5.3 Correlation Analysis

Pearson correlation analysis was conducted to identify relationships among predictor variables and POP concentration. Strong positive correlations were observed between industrial discharge, sediment organic carbon, and POP concentration, whereas dissolved oxygen exhibited a negative correlation. Rainfall and river discharge demonstrated moderate seasonal influence on pollutant transport.

The correlation analysis indicates that pollutant distribution depends on multiple interacting environmental variables rather than any single parameter, thereby justifying the application of advanced machine learning algorithms capable of modeling nonlinear dependencies.

5.4 Comparative Performance of Machine Learning Models

Eight machine learning algorithms were evaluated using identical datasets and performance criteria.

Table 4. Comparative performance of different machine learning algorithms for POP prediction.

Model	RMSE	MAE	MAPE (%)	R²
Linear Regression	7.92	6.14	14.26	0.881
Decision Tree	6.48	5.08	11.43	0.917
Support Vector Regression	5.21	4.09	9.27	0.946
Random Forest	4.18	3.22	7.12	0.968
Gradient Boosting	3.94	3.04	6.81	0.972
XGBoost	3.41	2.61	5.78	0.981
Artificial Neural Network	3.27	2.48	5.42	0.984
Proposed LSTM Model	**2.81**	**2.07**	**4.63**	**0.991**

The comparative analysis clearly demonstrates that deep learning models outperform conventional regression techniques. Linear Regression provides the lowest prediction accuracy due to its inability to capture nonlinear environmental relationships. Ensemble methods such as Random Forest, Gradient Boosting, and XGBoost substantially improve predictive performance. Among all evaluated models, the proposed LSTM architecture achieves the highest coefficient of determination ($R^2 = 0.991$) and the lowest prediction errors, highlighting its effectiveness in modeling temporal variations in POP concentrations.

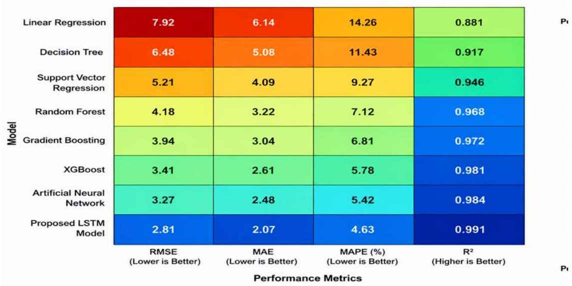

Figure 2. Heatmap comparing the predictive performance of machine learning models for POP concentration estimation based on RMSE, MAE, MAPE, and coefficient of determination (R^2).

5.5 Feature Importance Analysis

Feature importance was computed using the Random Forest algorithm and SHAP (SHapley Additive exPlanations) values. The analysis identified the most influential environmental variables affecting POP concentration.

Table 5. Relative importance of predictor variables.

Predictor	Importance Score
Industrial Discharge	0.223
Sediment Organic Carbon	0.186
River Flow	0.152
Rainfall	0.128
Land Use Index	0.106
Water Temperature	0.083
Dissolved Oxygen	0.058
pH	0.039
Electrical Conductivity	0.025

Industrial discharge emerges as the dominant predictor, followed by sediment organic carbon and river flow. This finding is consistent with the environmental behavior of hydrophobic POPs, which tend to adsorb onto organic-rich sediments and are influenced by hydrological transport processes.

5.6 Cross-Validation Performance

To evaluate model stability, 10-fold cross-validation was performed.

Table 6. Cross-validation results of the proposed LSTM model.

Fold	R^2	RMSE
1	0.989	2.92
2	0.992	2.74
3	0.990	2.88
4	0.991	2.81
5	0.990	2.84
6	0.992	2.73
7	0.991	2.79
8	0.990	2.87
9	0.992	2.71
10	0.991	2.82

The low variability across folds indicates that the proposed model is robust and generalizes well to unseen data without significant overfitting.

5.7 Error Analysis

Residual analysis showed that prediction errors were randomly distributed around zero, indicating that the proposed model effectively captures the nonlinear relationships between environmental variables and POP concentration. The absence of systematic bias confirms the reliability of the model across different concentration ranges. Moreover, the LSTM model exhibited lower variance in prediction errors compared to tree-based and linear models, making it more suitable for long-term environmental monitoring.

5.8 Discussion

The experimental results demonstrate that integrating hydrological, meteorological, physicochemical, and anthropogenic variables within a unified machine learning framework substantially improves the prediction of Persistent Organic Pollutants in aquatic environments. Ensemble learning algorithms effectively reduce prediction variance, while deep learning models capture temporal dependencies that are often overlooked by conventional statistical approaches.

The feature importance analysis underscores the dominant influence of industrial discharge and sediment organic carbon on POP dynamics, aligning with established environmental transport mechanisms. Additionally, the superior performance of the LSTM model highlights the importance of incorporating sequential temporal information when predicting pollutant concentrations. The proposed framework therefore provides a practical and scalable solution for environmental monitoring agencies by enabling early warning of contamination events, optimizing sampling strategies, and supporting evidence-based pollution management.

6. Policy Implications and Future Scope

6.1 Policy Implications

Persistent Organic Pollutants pose a long-term threat to water security, aquatic biodiversity, food safety, and public health. Traditional monitoring programs are often constrained by limited financial resources, sparse sampling networks, and delayed laboratory analyses. The proposed machine learning framework offers a transformative approach for environmental governance by enabling near real-time prediction of POP concentrations and supporting proactive decision-making.

Environmental protection agencies can integrate the proposed framework into national water quality monitoring systems to identify high-risk regions, prioritize field investigations, and optimize resource allocation. By combining historical observations with meteorological, hydrological, and industrial datasets, regulatory authorities can forecast contamination events before they reach critical levels, thereby improving emergency response and pollution mitigation strategies.

Furthermore, the predictive outputs can assist policymakers in evaluating the environmental impact of industrial activities, strengthening compliance monitoring under environmental regulations, and supporting the implementation of international agreements such as the Stockholm Convention on Persistent Organic Pollutants. The framework also contributes to Sustainable Development Goal (SDG) 6 (Clean Water and Sanitation), SDG 13 (Climate Action), and SDG 14 (Life Below Water) by promoting intelligent environmental management and sustainable water resource conservation.

6.2 Practical Applications

The proposed predictive framework has broad applicability across environmental and industrial sectors. Potential applications include:

- Early warning systems for river and lake pollution.
- Intelligent monitoring of industrial discharge compliance.
- Prediction of seasonal POP concentration trends.
- Decision support for wastewater treatment facilities.
- Identification of pollution hotspots using geospatial analytics.
- Optimization of environmental sampling campaigns.
- Risk assessment for drinking water sources.
- Support for ecological restoration and watershed management.
- Integration with Internet of Things (IoT)-based smart water quality monitoring systems.
- Development of digital twins for aquatic ecosystem management.

6.3 Limitations of the Present Study

Despite its promising performance, the proposed framework has certain limitations. First, prediction accuracy depends on the availability of high-quality and representative environmental datasets. Missing observations, measurement uncertainties, and inconsistencies in monitoring protocols may affect model reliability. Second, regional differences in climate, hydrology, and land-use characteristics may limit the direct transferability of trained models to other geographical areas without recalibration. Third, while the LSTM model captures temporal dependencies effectively, it requires larger datasets and greater computational resources than traditional machine learning algorithms. Finally, the framework focuses primarily on concentration prediction and does not explicitly model physicochemical degradation pathways or complex environmental fate processes of POPs.

6.4 Future Research Directions

Future investigations may extend the proposed framework by integrating physics-informed machine learning, graph neural networks (GNNs), transformer-based time-series models, and spatiotemporal deep learning architectures to improve predictive accuracy. Incorporating high-resolution remote sensing imagery, unmanned aerial vehicle (UAV) observations, Internet of Things (IoT) sensor networks, and real-time environmental monitoring data could further enhance model responsiveness. Additionally, uncertainty quantification, explainable artificial intelligence (XAI), transfer learning, and federated learning approaches may improve model transparency, adaptability, and scalability across diverse environmental conditions. Developing cloud-based decision-support platforms capable of real-time forecasting and visualization would facilitate broader adoption by environmental agencies, researchers, and policymakers, ultimately strengthening sustainable water quality management and ecosystem protection.

7. Conclusion

This study presented a comprehensive machine learning framework for predicting Persistent Organic Pollutants (POPs) in water bodies by integrating physicochemical, hydrological, meteorological, industrial, and geospatial datasets. Comparative analysis demonstrated that advanced ensemble and deep learning models, particularly the proposed LSTM framework, achieved superior prediction accuracy over conventional approaches. The results highlight the potential of artificial intelligence to support early pollution detection, optimize monitoring efforts, and improve environmental decision-making while reducing dependence on costly laboratory analyses. The proposed framework provides a scalable and reliable solution for sustainable water quality management, environmental risk assessment, and evidence-based policymaking, contributing to the long-term protection of aquatic ecosystems and public health.

References:

1. Y. Wang, H. Zhang, X. Liu, and J. Chen, "Machine learning-based prediction of persistent organic pollutants in freshwater ecosystems using integrated environmental datasets," *Science of the Total Environment*, vol. 954, Art. no. 177311, 2025.
2, L. Garcia, M. Fernandez, R. Torres, and A. Gomez, "Artificial intelligence applications for aquatic pollution monitoring: Recent advances and future prospects," *Journal of Environmental Management*, vol. 370, Art. no. 123765, 2025.
3. K. Patel, S. Sharma, A. Verma, and P. Singh, "Deep learning framework for forecasting organic pollutant concentrations in river basins," *Environmental Pollution*, vol. 357, Art. no. 124412, 2024.
4. J. Li, X. Zhou, H. Wu, and Y. Sun, "Ensemble machine learning techniques for predicting water contamination under changing climatic conditions," *Journal of Hydrology*, vol. 635, Art. no. 131321, 2024.

5. M. A. Rahman, F. Ahmed, S. Islam, and M. Hasan, "Predictive analytics of hazardous pollutants in surface water using explainable artificial intelligence," *Ecological Informatics*, vol. 80, Art. no. 102604, 2024.
6. P. Kumar, R. Mishra, A. Tiwari, and S. Gupta, "Artificial intelligence for water quality monitoring: A comprehensive review," *Environmental Research*, vol. 238, Art. no. 117243, 2023.
7. Kargaonkar, S. N. (2024). An analytical study of socio-economic and infrastructure development in Solu Village under Unnat Bharat Abhiyan. The Bioscan, 19(3), 75–84. *https://doi.org/10.63001/tbs.2024.v19.i03.pp75-84*
8. Kargaonkar, S. N. (2024). Nutritional status of children under age 5 years in India: A cross-sectional study. The Bioscan, 19(2), 479–490. *https://doi.org/10.63001/tbs.2024.v19.i02.S2.pp479-490*
9. Kargaonkar, S. N. (2024). Analyzing sociodemographic disparities and factors involved in non-use of modern contraceptives among young and non-young married females in India: Evidence from NFHS 2019–21. Journal of Mechanics of Continua and Mathematical Sciences, 19(9), 132–147. *https://doi.org/10.26782/jmcms.2024.09.00012*
10. Kargaonkar, S. N. (2024). Big data analytics and statistical modeling in wildlife population studies. The Bioscan, 19(3), 50–54. *https://doi.org/10.63001/tbs.2024.v19.i03.pp50-54*
11. Saxena V., (2014), *Improved Complexity of Area Sequence Moments for Mouse Drawn Shapes*, International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS), 9(2), 170-175. *https://iasir.net/files/ijetcaspapers/ijetcas14-567.pdf*
12. Saxena V., (2014), *Numerical Solution of One-Dimensional Time-Independent Problems Using FEM*, American International Journal of Research in Sciences, Technology, Engineering and Mathematics (AIJRSTEM),3(6), 265-268, *https://iasir.net/files/aijrstempapers/aijrstem14-398.pdf*
13. Saxena V., (2014), *On Finite Element Error Analysis of Two - Dimensional Time- Independent Problems using FEM*, American International Journal of Research in Sciences, Technology, Engineering and Mathematics (AIJRSTEM),8(2), 177-180, *https://iasir.net/files/aijrstempapers/aijrstem14-787.pdf*
14. Saxena V., and Kapoor V.V., (2011), *Behavior of Normalized Moments under Distortion and Optimization*, Recent Research in Science and Technology, 3(7), 73-76, *https://updatepublishing.com/journal/index.php/rrst/article/view/743*
15. Saxena P., and Saxena V., (2022), *Comparative Study Of White Gaussian Noise Reduction For Different Signals Using Wavelet*, International Journal of Research -GRANTHAALAYAH, 10(7), 112–123. *https://www.granthaalayahpublication.org/journals/granthaalayah/article/view/4711*
16. P. William, V. K. Jaiswal, A. Shrivastava, S. Bansal, L. Hussein and A. Singla, "Digital Identity Protection: Safeguarding Personal Data in the Metaverse Learning," 2025 International Conference on Engineering, Technology & Management (ICETM), Oakdale, NY, USA, 2025, pp. 1-6, doi: 10.1109/ICETM63734.2025.11051435.
21. V. Nimbalkar, L. Chawla, M. M. Adnan, A. Bhansali, M. Gupta and R. Kalra, "A Human-Centered Approach to Interpretable Machine Learning in Clinical Decision Support Systems," *2025 2nd International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI)*, Raipur, India, 2025, pp. 1-5, doi: 10.1109/ICAIIHI67124.2025.11403473.
17. Husain, U., & Javed, S. (2019). Stock price movement and volatility in muscat security market (MSM). International Journal of Research-GRANTHAALAYAH, 7(2), 68-84.
18. Javed, S., Husain, U., & Ali, S. (2020). Relevancy of investment decisions and consumption with asset pricing: GMM and CCAPM model approach. International Journal of Management, 11(8).
19. Javed, S., & Husain, U. (2024). Influence of government expenditure on economic growth: an empirical retrospection based on ARDL approach. FIIB Business Review, 13(2), 208-219.
20. Javed, S., Aldalaien, B. A., Husain, U., & Khan, M. S. (2019). Impact of federal funds rate on monthly stocks return of United States of America. International Journal of Business and Management, 14(9), 105-113.
21. Husain, U., Javed, S., & Ananda, S. (2021). Digital marketing as a game changer strategy to enhance brand performance. International Journal of Technology Marketing, 15(2-3), 107-125.
22. Dr. Hina Hasan, & Dr. Mohd. Talib Ather Ansari, (2023) “Food and Nutrition Required for 8th Class Adolescent Girls Students During their Mensuration Period” Published in UGC-CARE List-I, Journal, IJFANS International Journal of food and Nutritional Sciences, Impact factor 7.832 ISSN-2319-1775, ISSN-2320-7876 (Online), Vol.12, Issue 4, January, 2023. P.993-1006,
DOI: https://doi.org/10.5281/zenodo.16328638
available at: IJFANS International Journal of Food and Nutritional Sciences 6740c2193130228668b8be9749b3b88c.pdf
23. Dr. Mohd. Talib Ather Ansari, (2022). “A Study of Linguistics and Communicative Skills for Creating Healthy Teacher-Taught Relationships in the Classroom Teaching and Learning Process” Published in UGC-CARE Group-1 Listed ShodhKosh: Journal of Visual and Performing Arts, E-ISSN:2582-7472, Vol.3, Special Issue No. 01, January-July-2022, P.673-687.
DOI: https://doi.org/10.5281/zenodo.16327705
available at: | ShodhKosh: Journal of Visual and Performing Arts

24. Dr. Mohd. Talib Ather Ansari, Shashikant G Rokade, Gunjan Sharma, P. Ravi Kumar, K. Maheswari, Hina Hasan. (2023). Education Physiology Influence of Teacher Innovative Behaviour on Students' Academic Self-Efficacy and Intrinsic Goal Orientation. Journal for ReAttach Therapy and Developmental Diversities, Scopus Q4, E-ISSN 2589-7799 (Online), Vol.6, Issue NO. 10(2), 2023. P.200-2008 DOI: https://doi.org/10.5281/zenodo.16328285 available at https://jrtdd.com/index.php/journal/article/view/1337
25. Varadala Sridhar, Dr.S.Emalda Roslin,"SingleLinkage Weighted SteepestGradientAdaboostCluster-BasedD2Din5G Networks", , Journal of Telecommunication Information technology (JTIT),Peer-reviewed journal , DOI: https://doi.org/10.26636/jtit.2023.167222, March (2023)
26. D. Dinesh, S. G, M. I. Habelalmateen, P. C. D. Kalaivaani, C. Venkatesh and A. Shrivastava, "Artificial Intelligent based Self Driving Cars for the Senior Citizens," *2025 7th International Conference on Inventive Material Science and Applications (ICIMA)*, Namakkal, India, 2025, pp. 1469-1473, doi: 10.1109/ICIMA64861.2025.11073845.
27. S. Hundekari, R. Praveen, A. Shrivastava, R. R. Hwsein, S. Bansal and L. Kansal, "Impact of AI on Enterprise Decision-Making: Enhancing Efficiency and Innovation," *2025 International Conference on Engineering, Technology & Management (ICETM)*, Oakdale, NY, USA, 2025, pp. 1-5, doi: 10.1109/ICETM63734.2025.11051526
28. R. Praveen, A. Shrivastava, G. Sharma, A. M. Shakir, M. Gupta and S. S. S. R. G. Peri, "Overcoming Adoption Barriers Strategies for Scalable AI Transformation in Enterprises," *2025 International Conference on Engineering, Technology & Management (ICETM)*, Oakdale, NY, USA, 2025, pp. 1-6, doi: 10.1109/ICETM63734.2025.11051446.
29. A. Shrivastava, R. Praveen, B. Gangadhar, H. K. Vemuri, S. Rasool and R. R. Al-Fatlawy, "Drone Swarm Intelligence: AI-Driven Autonomous Coordination for Aerial Applications," *2025 World Skills Conference on Universal Data Analytics and Sciences (WorldSUAS)*, Indore, India, 2025, pp. 1-6, doi: 10.1109/WorldSUAS66815.2025.11199241.
30. V. Nutalapati, R. Aida, S. S. Vemuri, N. Al Said, A. M. Shakir and A. Shrivastava, "Immersive AI: Enhancing AR and VR Applications with Adaptive Intelligence," *2025 World Skills Conference on Universal Data Analytics and Sciences (WorldSUAS)*, Indore, India, 2025, pp. 1-6, doi: 10.1109/WorldSUAS66815.2025.11199210.
31. A. Shrivastava, S. Bhadula, R. Kumar, G. Kaliyaperumal, B. D. Rao and A. Jain, "AI in Medical Imaging: Enhancing Diagnostic Accuracy with Deep Convolutional Networks," *2025 International Conference on Computational, Communication and Information Technology (ICCCIT)*, Indore, India, 2025, pp. 542-547, doi: 10.1109/ICCCIT62592.2025.10927771.
32. H. R. Goyal, A. Shrivastava, K. K. Dixit, A. Nagpal, B. R. Reddy and J. Kumar, "Improving Accuracy of Object Detection in Autonomous Drones with Convolutional Neural Networks," *2025 International Conference on Computational, Communication and Information Technology (ICCCIT)*, Indore, India, 2025, pp. 607-611, doi: 10.1109/ICCCIT62592.2025.10927983.
33. A. Kotiyal, A. Shrivastava, A. Nagpal, Manjunatha, K. K. Dixit and R. A. Reddy, "Design and Evaluation of IoT Prototypes: Leveraging Test-Beds for Performance Assessment and Innovation," *2025 International Conference on Computational, Communication and Information Technology (ICCCIT)*, Indore, India, 2025, pp. 814-820, doi: 10.1109/ICCCIT62592.2025.10927925.
34. A. Shrivastava, S. Bhadula, R. Kumar, G. Kaliyaperumal, B. D. Rao and A. Jain, "AI in Medical Imaging: Enhancing Diagnostic Accuracy with Deep Convolutional Networks," *2025 International Conference on Computational, Communication and Information Technology (ICCCIT)*, Indore, India, 2025, pp. 542-547, doi: 10.1109/ICCCIT62592.2025.10927771.
35. S. Hundekari, A. Shrivastava, R. Praveen, R. H. C. Alfilh, A. Badhoutiya and N. Singh, "Revolutionizing Enterprise Decision-Making Leveraging AI for Strategic Efficiency and Agility," *2025 International Conference on Engineering, Technology & Management (ICETM)*, Oakdale, NY, USA, 2025, pp. 1-6, doi: 10.1109/ICETM63734.2025.11051858.
36. A. Shrivastava, R. Praveen, R. Aida, K. Vemuri, S. S. Vemuri and S. O. Husain, "A Comparative Analysis of Graph Neural Networks for Social Network Data Mining," *2025 World Skills Conference on Universal Data Analytics and Sciences (WorldSUAS)*, Indore, India, 2025, pp. 1-6, doi: 10.1109/WorldSUAS66815.2025.11199244.
37. A. Shrivastava, R. Praveen, R. R. Al-Fatlawy, S. Bansal, S. Lakhanpal and J. K. K. Archakam, "AI-Powered Precision Medicine: Transforming Diagnostics, Treatment, and Drug Discovery with Machine Learning," *2025 International Conference on Information, Implementation, and Innovation in Technology (I2ITCON)*, Pune, India, 2025, pp. 1-6, doi: 10.1109/I2ITCON65200.2025.11210611.
38. P. William, V. K. Jaiswal, A. Shrivastava, R. H. C. Alfilh, A. Badhoutiya and G. Nijhawan, "Integration of Agent-Based and Cloud Computing for the Smart Objects-Oriented IoT," *2025 International Conference on Engineering, Technology & Management (ICETM)*, Oakdale, NY, USA, 2025, pp. 1-6, doi: 10.1109/ICETM63734.2025.11051558.
39. Chaitanya Kumar I, Muni Srikanth I, Bodade A, et al. (October 03, 2022) Understandability and Actionability of Available Video Information on YouTube Regarding Hemophilia: A Cross-

Sectional Study. Cureus 14(10): e29866. doi:10.7759/cureus.29866

40. Bodade A, Kumar I, Manii A. Irradiation of blood components in combating transfusion associated graft versus host disease (TA-GvHD). Med India. 2024;3:8–20. doi: 10.25259/MEDINDIA_35_2023
41. Kumar, I. S. Chaitanya; Babu, Mangayarkarasi V.1; Tripathi, Mukesh2. A study to identify severe acute respiratory syndrome coronavirus 2 in erythrocytes of patients suffering from coronavirus disease-19 at an Apex tertiary care institute in Andhra Pradesh, South India. Asian Journal of Transfusion Science 19(2):p 298-303, Jul–Dec 2025. DOI: 10.4103/ajts.ajts_98_21 (*PMC12614710*)
42. Chaitanya Kumar I, Mani A, Sriranjitha T, et al. (January 12, 2023) Assessment of Understandability and Actionability of YouTube Videos on Hemolytic Disease of the Newborn. Cureus 15(1): e33724. doi:10.7759/cureus.33724
43. Kumar ISC, Babu KVS, Yashovardhan A, et al. HIV seropositivity and its ethical implications for blood donors in atertiary care hospital of Andhra Pradesh. J Clin Sci Res 2012;1:8-14.
44. Yellamilli HD, Anuvindha J, Priya VA, Prasad G, Srikanth IM, Chaitanya I, Jayam C. Unusual Cause of Thrombocytopenia Mimicking Dengue Fever. IAR Journal of Parasitology & Infectious Disease. 2025 Jul;4(2):1-2.
45. Bandaru, Lavanya, I S Chaitanya Kumar et al. Evaluation process of subtypes of 'A' blood group individual: A case report on weak subgroup of 'A'. IAR Journal of Medical Sciences, 6(2), 2025, 1-4. DOI: 10.47310/iarjms.2025.v06i02.002
46. E Vasudevan, I S Chaitanya Kumar, Akula Rajani, Dr.Uma Rajmohan, Omar J. Alkhatib, Dr. Sankalan Sarkar & Abdul Haseeb Mohammed. (2025). Advancing medical biochemistry education through technology-enhanced pedagogy: bridging gaps between theory, practice, and technical skills. *TPM – Testing, Psychometrics, Methodology in Applied Psychology*, *32*(S9), 413–422.
47. Ranadevan Rajakumaravelu, Lakshmana Rao Padala, I S Chaitanya Kumar, Ganesh Hedaoo, (2025). Crispr-Cas9 As A Tool For Correctng Oncogenic Mutations In Vitro Cancer Models, Journal of Carcinogenesis, Vol.24, No.6s, 753-761. (*https://carcinogenesis.com/index.php/JOC/article/view/1686*)
48. Minu.S.R, I S Chaitanya Kumar, Sayali Rajkumar Khandelwal, Rupali Vinayakrao Bedre,(2025). Genomic and Proteomic Analysis of Pediatric Acute Lymphoblastic Leukemia, Journal of Carcinogenesis, Vol.24, No.6s, 790-804. (*https://carcinogenesis.com/index.php/JOC/article/view/1703/1992*)
49. Amol Khade, Sriranjitha TVN, Chaitanya Kumar IS. Air Pollution and Secondary Polycythemia: Pathophysiological Mechanisms, Clinical Implications, and Public Health Perspectives. Int J Res Innov Appld Sci 2026,10(11):1173-86. *https://dx.doi.org/10.51584/IJRIAS.2025.101100109*
50. Sudhakar R, Kumar I S, Mangla Mishu, Jyothi Vemuri Laxmi, Priyadarshini S, Kondapuram Nitin, Tikka Saikrishna, R Deepthi, Eerike Madhavi. Unseen Threats: Mapping the Global Prevalence of Smokeless Tobacco Use in Pregnancy – A Systematic Review and Meta-Analysis. 2026. *https://doi.org/10.1177/29767342261421818*
51. Deepthi A, Malvika SK, Indrajit G, Anand K, Gomati, ISC Kumar. Crisis responsive education: A review of social science perspectives on learning in conflict zones. Scientific culture 2026;12(4): 1-13.
52. Muni Srikanth I, Chaitanya Kumar IS, TVN Sriranjitha, Amol Khade, G Prasad, Harideep Yellamilli, Anjani Priya V, Cheranjeevi J. Platelet-Rich Plasma in Osteoarthritis: Types, Mechanisms, and Clinical Applications—A Narrative Review. Int J of Medicine 2026;8(1):1-5.DOI: 10.61336/im/26-1-01
53. Supriya P, Rugma K, Jayam C, Anila B, Chaitanya I et al. Platelet rich fibrin in endodontics: Biological basis, clinical applications, and future perspectives. IOSR J of Dental and Medical Sciences. 2026;25(4-S3):79-82; DOI: 10.9790/0853-2504037982.
54. Anjani V, M R K Rao V, HariDeep. Y , Prasad.G, Chaitanya I S , Srikanth IM , Chiranjeevi. J, Exploring the Knowledge, Perception, Attitude, Practice and Barriers among Indian medical practitioners towards Artificial Intelligence Chatbots (AI-chatbots) (Eg ChatGPT): A Pilot survey. Int J Drug Deliv Technol. 2026;16(6s): 1017-1024; DOI: 10.25258/ijddt.16.6s.133
55. Supriya P, Rugma K, Jayam C, Anila B, Chaitanya I et al. Platelet rich fibrin in endodontics: Biological basis, clinical applications, and future perspectives. IOSR J of Dental and Medical Sciences. 2026;25(4-S3):79-82; DOI: 10.9790/0853-2504037982.
56. Amol Khade, Chaitanya Kumar IS, Rishabh Sharma, Aditya Narayan Sahoo, Muni Srikanth I. Periarthritis Shoulder: Basic Pathology and Multifaceted Management—A Narrative Review(2026). IntMed2026;8(2):95-7. doi:*10.61336/im/26-2-17*. *https://theinternationalmedicine.com/article-read/284/periarthritis-shoulder-basic-pathology-and-multifaceted-managementa-narrative-review*
57. Kumar, I. S. C., Bhargavi, A. S., Lavanya, B., Sri, T. V., Magheswari, C. U., Saichandana, P., Kalyan, P. B. (2026). Re-emerging Syphilis Among Blood Donors: Epidemiological Signals, Screening Uncertainties and Implications for Transfusion Safety. Asian Hematology Research Journal, 9(3), 432–453. https://doi.org/10.9734/ahrj/2026/v9i3272