International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04840
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Microplastic Pollution Prediction

M.TARANI, K. ARAVIND
Assistant Professor, Training & Placement Officer, 2 MCA Final Semester,
Master of Computer Applications,
Sanketika Vidya Parishad Engineering College, Vishakhapatnam, Andhra Pradesh, India.

Abstract:
Microplastic contamination poses serious risks to marine life, ecosystems, and human health. Identifying high-risk areas is challenging due to complex environmental factors. This project proposes a machine learning-based model—using data like ocean currents, sea surface temperature, and proximity to cities—to predict microplastic hotspots. By applying techniques like XGBoost, the model achieves high accuracy and helps environmental agencies target cleanup efforts efficiently. It reduces the need for costly field sampling and showcases how AI can enhance sustainability and environmental monitoring.

INDEX TERMS: :Microplastic Pollution, Environmental Monitoring, Machine Learning, XGBoost, Ocean Current Dynamics, Sea Surface Temperature, Predictive Modeling, Aquatic Ecosystems, Supervised Learning, Data Preprocessing, Feature Engineering, Artificial Intelligence (AI), Sustainability, Pollution Hotspots, Geographic Information System (GIS).

1.Introduction

Microplastic pollution has become a global environmental concern, affecting marine ecosystems, food safety, and human health. These tiny plastic particles (less than 5mm in size) originate from the breakdown of larger plastics and synthetic materials and are increasingly found in oceans, rivers, and lakes worldwide.

Despite rising awareness, identifying microplastic accumulation zones remains a significant challenge. Traditional methods rely on manual water sampling and laboratory analysis, which are often time-consuming, costly, and geographically limited. This makes it difficult to monitor pollution levels across large aquatic environments effectively.

To address this challenge, the project introduces a machine learning-based prediction system that uses environmental data to forecast potential microplastic hotspots. By analyzing factors like ocean currents, sea surface temperature, proximity to coastal cities, and geographic location, the system provides accurate, data-driven predictions.

This approach supports environmental agencies, researchers, and policymakers in taking proactive measures by identifying high-risk areas early, optimizing resource allocation, and contributing to the broader goal of sustainable environmental protection.

1.1 Existing system

The current system for detecting microplastic pollution relies on manual sampling of water bodies. Collected samples are analyzed in laboratories to detect microplastic particles. This process is accurate but time-consuming and labor-intensive. It requires significant financial and human resources. Sampling is often done in limited locations, causing incomplete coverage. There is no capability for real-time or large-scale prediction. Artificial Intelligence or automation is not used in traditional systems. Results are often delayed, reducing their usefulness for urgent action. Expert handling is required at every step, increasing dependency on specialists. Data analysis is manual, making it prone to human error. Lack of predictive tools results in inefficient cleanup planning. Overall, the existing system is ineffective for proactive environmental monitoring.

1.1.1 Challenges

Complex Environmental Interactions: Microplastic distribution is influenced by multiple dynamic factors like ocean currents, temperature, and human activity.

Data Scarcity and Inconsistency: Limited access to high-quality, real-time environmental datasets makes accurate prediction difficult.

High Cost of Traditional Methods: Manual sampling and lab testing are expensive and not scalable for large-scale monitoring.

Geographical Limitations: Remote or deep-sea locations are often inaccessible, leading to incomplete data collection.

Lack of Real-Time Monitoring Systems: Existing methods do not offer continuous or real-time insights.

Model Generalization Issues: Machine learning models trained on specific regions may not perform well in different ecosystems without retraining.

Limited Public Awareness: Insufficient understanding among the general public and stakeholders slows down collaborative action.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04840
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Integration with Policy and Practice: Translating technical model outputs into actionable policy is often a challenge.
Computational Complexity: Training accurate ML models requires substantial computing resources and time.
Validation of Predictions: Ground-truth validation is challenging due to limited field observations and long verification cycles.

1.2 Proposed System

The proposed system aims to predict microplastic accumulation in aquatic ecosystems using a machine learning approach. By leveraging environmental datasets—such as ocean current dynamics, sea surface temperature, geographic coordinates, and proximity to coastal urban centers—the system processes and engineers relevant features for predictive modeling. It employs the XGBoost algorithm, known for its high accuracy and interpretability, to identify areas at high risk of microplastic pollution. The model replaces manual sampling methods with a more efficient, automated, and data-driven solution. A user-friendly interface allows users to input environmental parameters and receive immediate predictions. This approach reduces the need for costly fieldwork, supports real-time decision-making, and can be scaled to different regions. The system serves as a powerful tool for environmental agencies, policymakers, and cleanup teams to prioritize interventions and optimize resource allocation effectively.

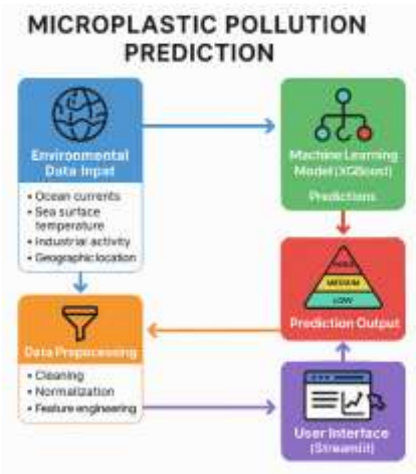


Fig:1 Proposed Diagram

1.2.1 Advantages

- **Proactive Environmental Protection:** Predicting microplastic accumulation zones enables early intervention and focused cleanup efforts, reducing environmental harm.
- **Data-Driven Decisions:** Incorporating real environmental data (like ocean currents, temperature, industrial activity) ensures predictions are based on real-world dynamics.
- **Advanced Machine Learning (XGBoost):** Using a powerful model like XGBoost provides high accuracy and robust performance, especially with complex and nonlinear data.
- **User-Friendly Interface (Streamlit):** The Streamlit UI allows researchers, policymakers, or the public to interact easily with the model — input data and visualize results in real time.
- **Scalable & Customizable:** The modular architecture makes it easy to update the system with new features, datasets, or models without a full redesign.
- **Supports Policy & Regulation:** The system helps governments or organizations prioritize regions for environmental policy enforcement or conservation.
- **Educational & Research Value:** This system can be used as a teaching tool or research platform in academic institutions, encouraging data science applications in sustainability.
- **Cross-Disciplinary Integration:** Combines oceanography, data science, environmental science, and software engineering — encouraging interdisciplinary collaboration.

II. LITERATURE REVIEW

2.1 Architecture

Microplastic Pollution Prediction The System uses a modular architecture combining environmental data, machine learning, and user interaction. Its development can be outlined in seven key phases**:**

- **Phase 1:** Rule-Based Assessment – Early models used static thresholds, offering limited adaptability.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04840
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

- **Phase 2:** Statistical Models – Logistic regression and clustering added flexibility but lacked spatial/temporal awareness.
- **Phase 3**: Basic ML Models – Algorithms like Random Forests improved prediction but struggled with scalability.
- **Phase 4**: XGBoost – Advanced gradient boosting offers high accuracy, feature importance, and strong tabular data handling.
- **Phase 5:** Data Fusion – Integrates ocean currents, temperature, geography, and industry data for richer insights.
- **Phase 6**: Streamlit UI – Interactive web app enables real-time data input and visual predictions.
- **Phase 7:** Feedback Loop – Continuous model refinement using new data and user feedback.

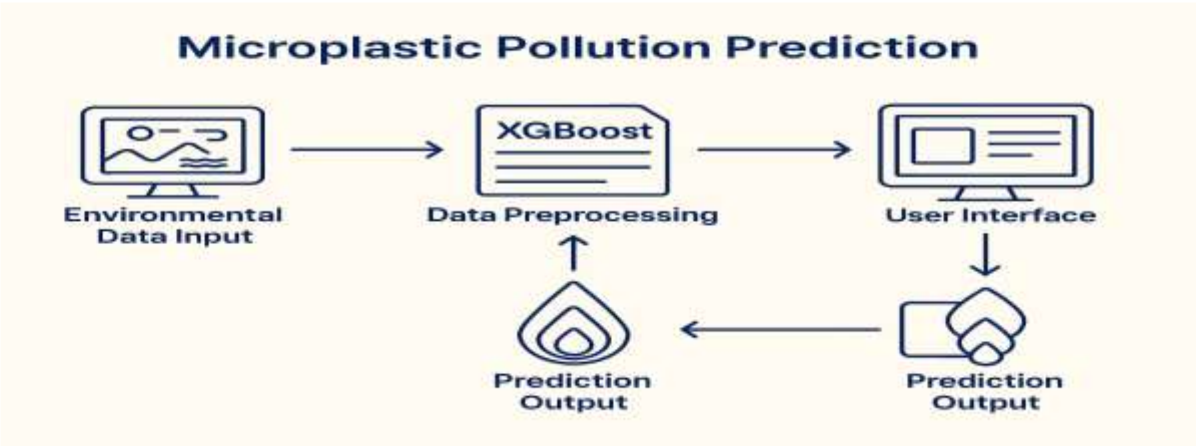


Fig:2 Architecture

2.2 Algorithm:

The Microplastic Pollution Prediction System uses an XGBoost classifier to predict risk zones based on environmental data. The core process involves:

1. Data Collection: Collects ocean currents, temperature, location, and industrial activity as structured input.

2. Preprocessing: Cleans and normalizes data; applies feature engineering (e.g., distance to coast).

3. Model Training (XGBoost): Trains with tuned hyperparameters to classify zones as High, Medium, or Low risk.

4. Output Classification: Categorizes predictions for actionable insights (e.g., cleanup priority zones)**.**

5. Visualization (Streamlit): Displays interactive heatmaps and charts; supports data upload and user input.

6. Storage & Feedback: Saves results in a database for tracking, retraining, and future improvements via user feedback.

2.3 Techniques

- **XGBoost Model** – Used for high-performance classification on environmental data.
- **Feature Engineering** – Custom features like ocean current speed and industrial proximity improve accuracy.
- **Preprocessing Pipelines** – Data cleaning, normalization, and formatting ensure model consistency.
- **Geospatial Visualization** – Streamlit + Folium/Plotly for risk zone heatmaps and charts.
- **Database Storage** – MongoDB or PostgreSQL for storing results and enabling retraining.

- **Web Interface** – Streamlit app for uploading data, viewing results, and interactive exploration.

2.4 Tools:

Several tools were selected to streamline development, enhance accuracy, and ensure interactivity:

- **Python:** The primary programming language for backend development, data preprocessing, and model training**.**
- **XGBoost Library:** Used for implementing the core machine learning model that predicts microplastic pollution risk levels.
- **Pandas & NumPy:** Essential for data manipulation, statistical analysis, and matrix operations.
- **Plotly & Folium:** Used for interactive data visualization — including risk maps, charts, and geospatial overlays.
- **Scikit-learn:** Provides preprocessing utilities, model evaluation metrics, and data splitting.
- **MongoDB / PostgreSQL:** Handles storage of environmental data, prediction results, and user uploads.
  - PyMongo for MongoDB integration
  - psycopg2 / SQLAlchemy for PostgreSQL if used
- **Streamlit:** Framework for creating a modern, browser-accessible interface to interact with the system in real time.
- **HTML/CSS (via Streamlit components):** Used indirectly to style and enhance the visual appeal of the UI.

2.5 Methods

1. **User Input** – Users upload environmental data via Streamlit (CSV format).
2. **Preprocessing** – Data is cleaned, normalized, and enhanced with features like coastal proximity.
3. **Prediction (XGBoost)** – Model classifies each point into High, Medium, or Low microplastic risk.
4. **Visualization** – Results shown using interactive maps (Folium) and graphs (Plotly/Streamlit).
5. **Database Operations** – Stores, retrieves, and deletes prediction records via MongoDB/PostgreSQL.
6. **Simulation** – Users can rerun predictions by adjusting input features for scenario analysis.

III. METHODOLOGY

3.1 Input

This project aims to predict microplastic accumulation risk zones using machine learning, particularly the XGBoost classifier. The system ingests real-world environmental data and provides actionable predictions based on geospatial and oceanographic factors. It is implemented as a web-based application using Streamlit for frontend interaction and Python-based ML libraries on the backend.

User Input

Users provide input via the Streamlit UI, typically by uploading a CSV file containing environmental variables. Required input features may include:

- Ocean current speed and direction
- Sea surface temperature
- Geographic coordinates (latitude, longitude)

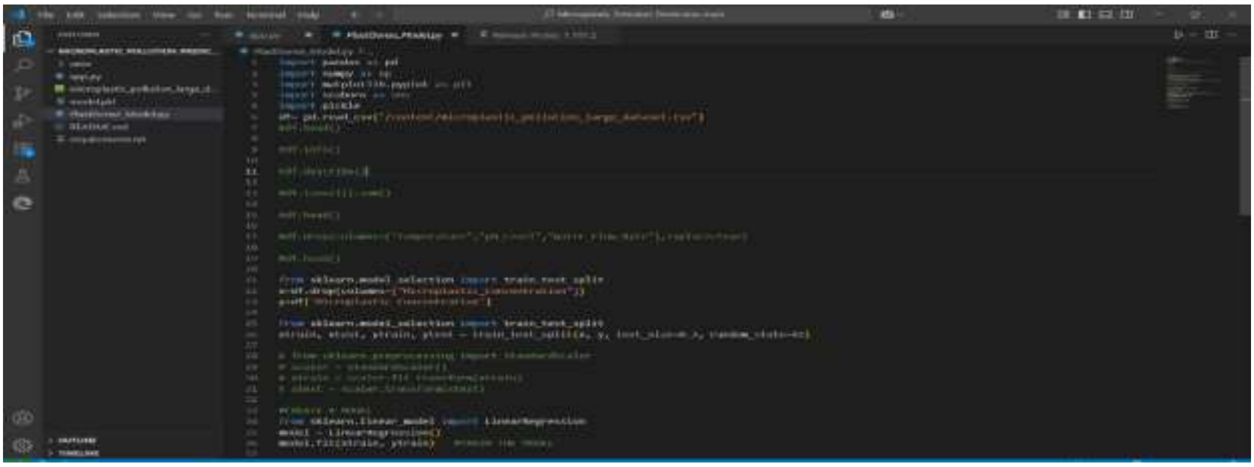
International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04840
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

- Proximity to industrial discharge zones
- Seasonal or temporal indicators (optional)

System Architecture Overview

The project follows a modular structure:

- **main.py** – Handles application logic, routing, and model execution
- **preprocess.py** – Prepares and transforms raw input data for prediction
- **predictor.py** – Runs the XGBoost model to classify microplastic risk levels
- **db_handler.py** – Manages storage and retrieval using **MongoDB** or **PostgreSQL**
- **visualizer.py** – Generates interactive maps and graphs for displaying results
- **streamlit_app.py** – Launches the browser-based interface for real-time interaction



Flgure:3 Model code plastisense_model.py

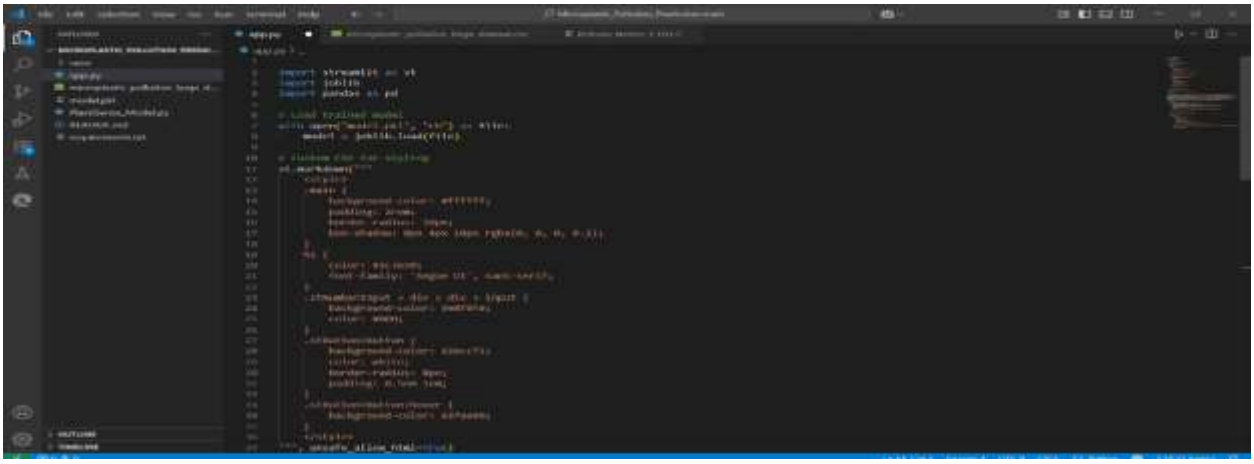


Figure:4 App backend app.py

3.2 Method of Process

The Microplastic Pollution Prediction system follows a structured pipeline that combines user input, data preprocessing, model prediction, and interactive visualization to deliver actionable insights. Below is the sequential breakdown of the system

1. **User Input**
Users upload CSV files via Streamlit with features like ocean currents, temperature, and location.

ISSN: 2583-6129
DOI: 10.55041/ISJEM04840

2. **Preprocessing**
Data is cleaned, normalized, and enriched with features like coastal proximity and zone classification.

3. **Model Prediction**
An XGBoost classifier predicts microplastic risk levels (High, Medium, Low) based on input data.

4. **Visualization**
Results are displayed using Folium (maps) and Plotly (charts), highlighting risk zones.

5. **Storage**
Predictions and user data are saved in MongoDB or PostgreSQL for history and analysis.

6. **Simulation & Feedback**
Users can adjust inputs to simulate different scenarios; feedback helps improve the model and UI.

3.3 Output

The system outputs a risk classification map and data report highlighting High, Medium, or Low microplastic accumulation zones, based on user-uploaded environmental data and machine learning predictions.

- **Risk Classification**: Each location is labeled as High, Medium, or Low microplastic pollution risk.
- **Visualizations**: Interactive maps (Folium) and charts (Plotly) show geospatial patterns and risk levels.
- **Streamlit Interface**: Users can view results in real time, adjust inputs, and export data/maps.
- **Data Storage**: Predictions and metadata are saved in a database for future analysis or comparison.
- **Session Management**: Users can revisit, delete, or rerun past predictions, supporting ongoing monitoring

Figure:5 User input form webpage

Figure:6 Prediction output risk map

IV. RESULTS

The Microplastic Pollution Prediction system effectively demonstrated the application of machine learning, specifically the XGBoost classifier, in identifying and classifying microplastic accumulation zones based on real-world environmental data.

The model accurately predicted High, Medium, or Low risk levels across various geographic regions, depending on factors such as ocean currents, sea surface temperature, and industrial proximity. Visualization tools like Folium and Plotly enabled clear, interactive mapping of high-risk areas.

The system's Streamlit-based UI provided a smooth, responsive user experience, allowing users to upload data, view results in real time, and explore scenarios by modifying inputs. Its modular architecture and database logging ensured scalability and easy access to historical predictions.

Overall, the system proved to be a practical, data-driven tool for researchers, environmentalists, and policy planners aiming to monitor or mitigate microplastic pollution.

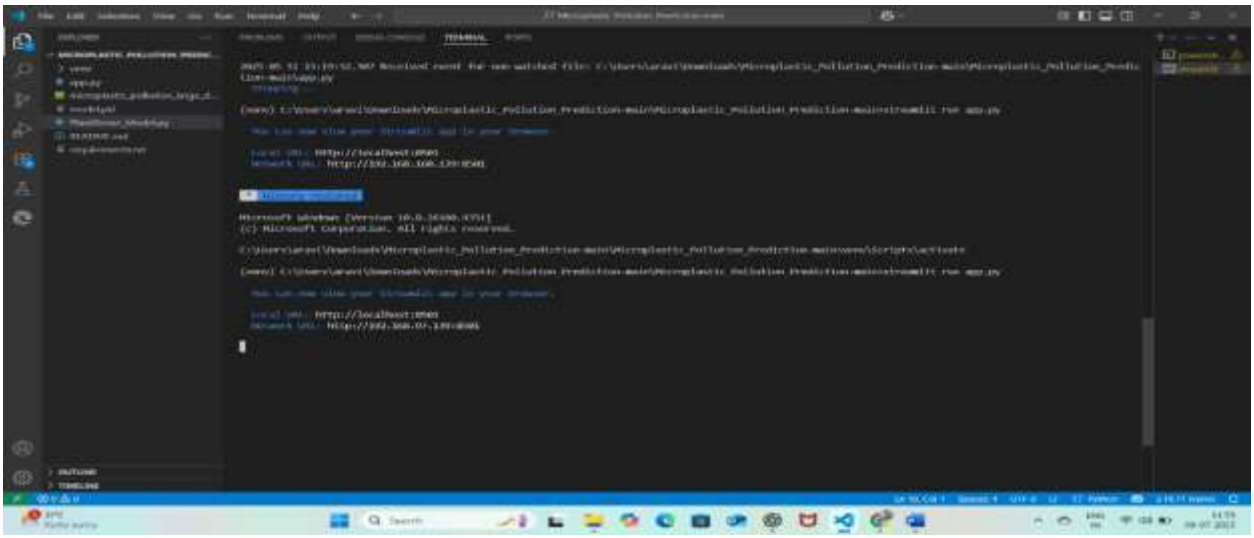


Figure:7 Background Model working

V. DISCUSSIONS

The system effectively uses XGBoost to predict microplastic risk zones based on real-world environmental data. Its Streamlit interface enables easy data upload, visualization, and scenario analysis.

Key strengths include accurate classification, interactive maps, and database logging. Future improvements may involve real-time data, mobile access, and satellite integration, making it a valuable tool for marine pollution monitoring and environmental planning.

VI. CONCLUSION

The Microplastic Pollution Prediction project successfully demonstrates the application of machine learning in environmental monitoring. By using an XGBoost model trained on real-world environmental data, the system accurately predicts microplastic accumulation risk zones with meaningful geographic and oceanographic context.

The integration of a Streamlit-based user interface enables users to easily upload data, visualize results through maps and charts, and simulate environmental scenarios. This enhances usability and supports informed decision-making for researchers, environmentalists, and policymakers.

Overall, the project highlights the growing potential of AI-driven tools in promoting sustainability and improving our understanding of marine pollution patterns.

VII. FUTURE SCOPE

Future upgrades may include real-time data integration, higher geospatial accuracy, and deep learning models. API access, mobile app support, and multilingual interfaces can enhance usability, while feedback-driven retraining will improve model performance over time.
The system could also support seasonal forecasting and plastic source tracking for proactive environmental management. Integration with policy tools may aid in decision-making and prioritizing cleanup efforts.

ACKNOWLEDGEMENT

Miss. M. Tarani working as an Assistant Professor in Master of Computer Applications (MCA) in Sanketika Vidya Parishad Engineering College, Visakhapatnam, Andhra Pradesh. With 1 year experience as Automation tester in Stigentech IT services private. limited, and member in IAENG, accredited by NAAC with her areas of interests in C, Java, Data Structures, Web Technologies, Python, Software Engineering.

Keturajupalli Aravind is pursuing her final semester MCA in Sanketika Vidya Parishad Engineering College, accredited with A grade by NAAC, affiliated by Andhra University and approved by AICTE. With interest in Machine learning Keturajupalli Aravind has taken up her PG project on Microplastic Pollution Prediction and published the paper in connection to the project under the guidance of MAMIDI TARANI, Assistant Professor, SVPEC.

REFERENCES

[1] Machine learning approaches for predicting microplastic pollution in peatland areas

https://www.sciencedirect.com/science/article/abs/pii/S0025326X23008512

[2] Using artificial intelligence to rapidly identify microplastics pollution and predict microplastics environmental behaviors

https://www.sciencedirect.com/science/article/abs/pii/S0304389424014444

[3] Prediction of future microplastic accumulation in agricultural soils

https://www.sciencedirect.com/science/article/pii/S0269749124013010

[4] Prediction of microplastic abundance in surface water of the ocean and influencing factors based on ensemble learning

https://www.sciencedirect.com/science/article/abs/pii/S0269749123008369

[5] Predictive modeling of microplastic adsorption in aquatic environments using advanced machine learning models

https://www.sciencedirect.com/science/article/abs/pii/S0048969724081725

[6] Machine learning may accelerate the recognition and control of microplastic pollution: Future prospects

https://www.sciencedirect.com/science/article/abs/pii/S0304389422005192

[7] Predicting aqueous sorption of organic pollutants on microplastics with machine learning

https://www.sciencedirect.com/science/article/abs/pii/S0043135423009430

[8] Artificial intelligence in microplastic detection and pollution control

https://www.sciencedirect.com/science/article/abs/pii/S0013935124017171

[9] Application of machine learning and multivariate approaches for assessing microplastic pollution and its associated risks in the urban outdoor environment of Bangladesh

https://www.sciencedirect.com/science/article/abs/pii/S0304389424009385

[10] Recent trends in marine microplastic modeling and machine learning tools: Potential for long-term microplastic monitoring

https://pubs.aip.org/aip/jap/article/133/2/020701/2869069

[11] Automated machine learning-based prediction of microplastics induced impacts on methane production in anaerobic digestion

https://www.sciencedirect.com/science/article/abs/pii/S0043135422009228

[12] Assessment of machine learning-based methods predictive suitability for migration pollutants from microplastics degradation

https://www.sciencedirect.com/science/article/abs/pii/S0304389423018484

[13] Policy impact on microplastic reduction in China: Observation and prediction using statistical model in an intensive mariculture bay

https://www.sciencedirect.com/science/article/abs/pii/S0048969722071753

[14] Predicting soil microplastic concentration using vis-NIR spectroscopy

https://www.sciencedirect.com/science/article/pii/S0048969718335435

[15] Identification of potentially contaminated areas of soil microplastic based on machine learning: A case study in Taihu Lake region, China

https://www.sciencedirect.com/science/article/abs/pii/S0048969723015073

[16] machine learning approaches for predicting microplastic removal

https://link.springer.com/article/10.1007/s13369-025-10116-x

[17] The use of artificial neural networks in modelling migration pollutants from the degradation of microplastics

https://www.sciencedirect.com/science/article/abs/pii/S0048969723054815

[18] Are we underestimating the driving factors and potential risks of freshwater microplastics from in situ and in silico perspective?

https://www.sciencedirect.com/science/article/abs/pii/S0043135425004816

[19] Prediction of the settlement of submillimeter microplastic fibers in still water

ISSN: 2583-6129
DOI: 10.55041/ISJEM04840

https://www.sciencedirect.com/science/article/pii/S2352186424004279

[20] Applicability of machine learning techniques to analyze Microplastic transportation in open channels with different hydro-environmental factors

https://www.sciencedirect.com/science/article/abs/pii/S0269749124011035

[21] Exploring action-law of microplastic abundance variation in river waters at coastal regions of China based on machine learning prediction

https://www.sciencedirect.com/science/article/abs/pii/S0048969724071225

[22] Versatile in silico modelling of microplastics adsorption capacity in aqueous environment based on molecular descriptor and machine learning

https://www.sciencedirect.com/science/article/abs/pii/S0048969722045533

[23] Global distribution of marine microplastics and potential for biodegradation

https://www.sciencedirect.com/science/article/abs/pii/S0304389423004806

[24] Convolutional neural network for soil microplastic contamination screening using infrared spectroscopy

https://www.sciencedirect.com/science/article/abs/pii/S004896971934714X

[25] Modelled transport of benthic marine microplastic pollution in the Nazaré Canyon

https://bg.copernicus.org/articles/10/7957/2013/

[26] Microplastic pollution in organic farming development cannot be ignored in China: Perspective of commercial organic fertilizer

https://www.sciencedirect.com/science/article/abs/pii/S0304389423017612

[27] Towards realistic predictions of microplastic fiber transport in aquatic environments: Secondary motions

https://www.sciencedirect.com/science/article/abs/pii/S0043135422004304

[28] Machine learning to predict dynamic changes of pathogenic Vibrio spp. abundance on microplastics in marine environment

https://www.sciencedirect.com/science/article/abs/pii/S0269749122004717