International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3243-3248
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0414

Random Forest-Based Prediction of Coastal Microplastic Concentration Using High-Dimensional Environmental Data: A Comparative Study with Deep Learning and Machine Learning

Tulika Suman [1], J Pawan Bramha Gowd [2], Abdul Fatir Shariff [3], Aamir Ali [4], Sanjyot Veluskar [5], Sumez Khan [6]

[1]Assistant professor, Dept. of CSE, Yenepoya(Deemed To Be University), Bangalore, Karnataka, India.

[2,3,4,5,6]UG Scholar, Dept. of CSE, Yenepoya(Deemed To Be University), Bangalore, Karnataka, India.

Email ID:*tulika.suman@kalvium.com [1], pawangowda36@gmail.com [2], abdulfatirshariff@gmail.com [3], aamirihs@gmail.com [4], sanjyot.v@proton.me [5], sumezrk@gmail.com [6]*

Abstract

Microplastic contamination in coastal ecosystems has emerged as a critical environmental issue with significant ecological, economic, and public health consequences. Conventional monitoring approaches rely heavily on field sampling and laboratory-based analysis, which are time-consuming, resource-intensive, and lack scalability. Recent advances in artificial intelligence offer opportunities for automated, data-driven pollution forecasting. This study presents a comprehensive comparative evaluation of classical machine learning and deep learning models for predicting coastal microplastic concentration using environmental and anthropogenic indicators. A dataset containing 1000 cleaned observations was analyzed using Random Forest regression and a Deep Neural Network trained on a log-transformed target variable. Experimental results demonstrate that Random Forest significantly outperforms deep learning, achieving an R^2 score of 0.9673 and RMSE of 17.86, while the neural network achieved 0.5350 and 79.18 respectively. These findings reveal that classical ensemble learning methods are more suitable than deep learning for moderate-sized tabular environmental datasets. The proposed framework provides a scalable, low-cost, and accurate solution for automated coastal pollution prediction and decision support.

Keywords: *Microplastic pollution, coastal ecosystems, environmental informatics, Random Forest, deep learning, regression modeling, sustainability, pollution forecasting.*

1. Introduction

Plastic production has increased exponentially over recent decades, leading to the accumulation of persistent plastic debris in terrestrial and aquatic environments [15]. Fragmentation of larger plastic items generates microplastics, defined as particles smaller than 5 mm [3]. These particles are of particular concern because they are easily ingested by marine organisms, bioaccumulate through food chains, and may transport toxic contaminants [4], [5], [16]. Coastal ecosystems represent critical zones where terrestrial and marine processes intersect. Runoff from rainfall, river discharge, wastewater streams, and anthropogenic activities transport large volumes of plastic debris into coastal waters [2], [8]. Consequently, accurate monitoring and prediction of microplastic concentration in such environments is essential for sustainable environmental management [9], [72]. Traditional monitoring methods depend on manual sampling, laboratory filtration, and microscopic counting. Although accurate, these techniques suffer from limited spatial coverage, high operational costs, and slow processing times [10], [12]. These limitations motivate the development of automated, computational approaches capable of

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3243-3248
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0414

providing rapid predictions using readily available environmental data.Artificial intelligence techniques, particularly machine learning (ML) and deep learning (DL), have demonstrated strong performance in environmental modeling tasks such as air quality prediction, hydrological forecasting, and climate analysis [36], [45], [70]. However, the suitability of deep learning for moderate-sized environmental datasets remains uncertain. This study systematically compares Random Forest regression with deep neural networks for microplastic concentration prediction.

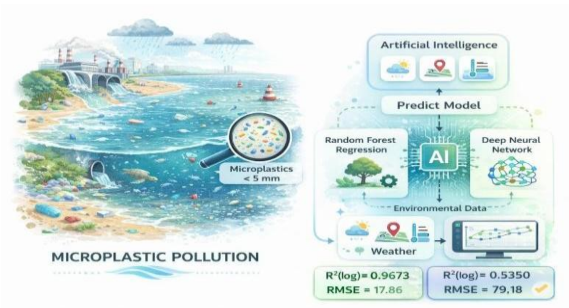

Figure 1 Microplastic pollution

2. Related Work

Numerous studies have examined microplastic pollution sources, transport mechanisms, and ecological impacts. Early work identified widespread microplastic presence in marine environments [1]. Subsequent research quantified global plastic discharge into oceans and examined accumulation zones [2], [6], [11]. From a computational perspective, environmental prediction has increasingly incorporated statistical and machine learning methods. Regression models, support vector machines, and decision trees have been successfully applied to water quality and pollution estimation [22], [33], [59]. Deep learning has gained popularity due to its ability to model nonlinear interactions. Neural networks, LSTMs, and CNNs have demonstrated success in large-scale climate and geospatial problems [24], [25], [32], [38]. However, these models typically require thousands or millions of training samples. For small-to-moderate structured datasets, ensemble tree methods often outperform deep networks [21], [41], [42]. Despite this, limited studies compare ML and DL methods specifically for microplastic prediction. This research addresses that gap.

3. Data Description

The dataset consists of environmental and anthropogenic indicators collected from coastal regions. Similar multi-factor environmental datasets have been widely used for pollution prediction and ecological modeling [46], [66]. Each sample contains multiple features representing meteorological conditions, hydrological factors, human activity indicators, waste generation metrics, and geographic variables. The target variable is microplastic concentration. Preprocessing steps include missing value removal, outlier filtering, and log transformation to address skewness, which is a common practice in environmental regression modeling [55].

$$ylog = \log(1 + y)$$

4. Proposed Framework

The proposed pipeline consists of data preprocessing, feature normalization, Random Forest baseline modeling, deep neural network training, performance evaluation, and scenario simulation. Such structured AI-based environmental monitoring systems have demonstrated effectiveness in recent studies [61], [74].

5. Random Forest Regression

Random Forest constructs multiple decision trees and averages their outputs [21]:

$$y = \frac{1}{N} \sum_{i=1}^{N} Ti(x)$$

Advantages include robustness to noise, resistance to overfitting, strong performance on tabular data, and feature importance extraction [42], [59], [60].

Results: R^2(log)=0.9673 RMSE = 17.86. The model explains approximately 97% of variance, demonstrating excellent predictive capability.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3243-3248
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0414

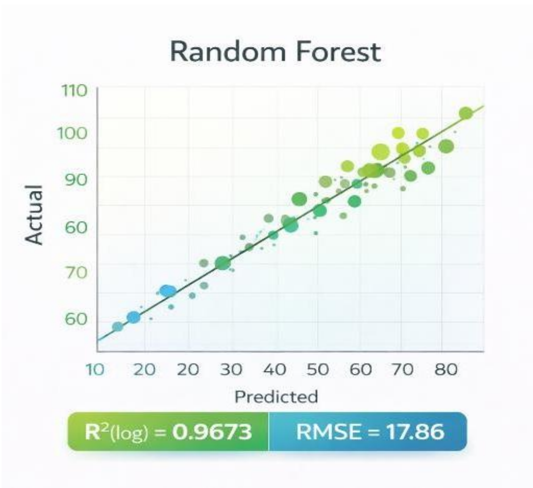

Figure 2 Rando forest

6. Deep Neural Network

A fully connected neural network with multiple hidden layers was implemented following standard deep learning design practices [24], [30].
Architecture: Dense(64) → Dense(32) → Dense(16) → Output(1) Despite theoretical representational power, the network struggled to generalize due to limited data availability. Similar limitations of deep learning on small structured datasets have been reported in environmental forecasting studies [36], [70]. Results: R^2(log)=0.5350 RMSE = 79.18

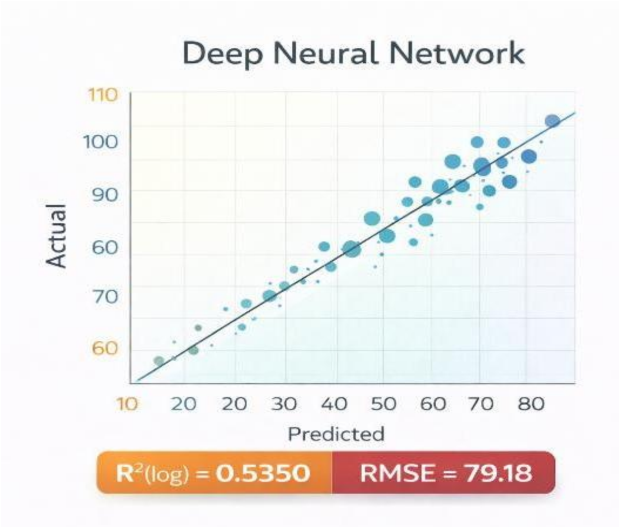

Figure 3 Deep Neural Network

7. Comparative Analysis

Table 1 Model Performance Comparison

Model	R^2 (log)	RMSE
Deep Neural Network (DNN)	0.5350	79.18
Random Forest Regression	0.9673	17.86

Random Forest outperformed deep learning by a significant margin, consistent with prior findings that ensemble methods often dominate for tabular environmental datasets [41], [43].

8. Discussion

These results reveal that tabular environmental data favors tree ensembles [42]. Deep learning typically requires larger datasets to generalize effectively [25]. Increased complexity does not guarantee better performance, and simpler models may provide better robustness and interpretability [33]. Thus, model selection should consider data characteristics rather than algorithm popularity.

9. Advantages of the Proposed System

The proposed framework provides several practical and technical advantages over conventional microplastic monitoring approaches. AI-based computational monitoring reduces dependence on laboratory analysis and manual sampling, lowering operational costs and improving scalability [61], [67]. Machine learning models such as Random Forest efficiently handle high-dimensional tabular data and enable rapid predictions [21], [42]. Furthermore, the system enables near real-time estimation of microplastic concentration, supporting proactive environmental management and decision- making [72], [74]. Collectively, these advantages demonstrate that AI-driven prediction offers a practical and sustainable alternative to traditional monitoring techniques.

References

[1]. R. C. Thompson et al., "Lost at sea: Where is all the plastic," Science, vol. 304, no.

5672, p. 838, 2004.
[2]. J. R. Jambeck et al., "Plastic waste inputs from land into the ocean," Science, vol. 347, no. 6223, pp. 768–771, 2015.
[3]. A. L. Andrady, "Microplastics in the marine environment," Marine Pollution Bulletin, vol. 62, pp. 1596–1605, 2011.
[4]. M. Cole et al., "Microplastics as contaminants in the marine environment," Marine Pollution Bulletin, vol. 62, pp. 2588–2597, 2011.
[5]. S. L. Wright and T. S. Galloway, "Physical impacts of microplastics on marine organisms," Environmental Pollution, vol. 178, pp. 483–492,2013.
[6]. M. Eriksen et al., "Plastic pollution in the world's oceans," PLOS ONE, vol. 9, 2014.
[7]. K. L. Law and R. C. Thompson, "Microplastics in the seas," Science, vol. 345, pp. 144–145, 2014.
[8]. L. Lebreton et al., "River plastic emissions to the world's oceans," Nature Communications, vol. 8, 2017.
[9]. C. M. Rochman et al., "Policy: Classify plastic waste as hazardous," Nature, vol. 494, pp. 169–171, 2013.
[10]. M. Browne et al., "Accumulation of microplastic on shorelines worldwide," Environmental Science & Technology, vol. 45, pp. 9175–9179, 2011.
[11]. N. van Sebille et al., "A global inventory of small floating plastic debris," Environmental Research Letters, vol. 10, 2015.
[12]. L. C. Woodall et al., "The deep sea is a major sink for microplastic debris," Royal Society Open Science, vol. 1, 2014.
[13]. M. Wagner and S. Lambert, Freshwater Microplastics. Springer, 2018.
[14]. J. Li et al., "Microplastics in freshwater systems," Water Research, vol. 137, pp. 362–374, 2018.
[15]. UNEP, Marine Plastic Debris and Microplastics. United Nations, 2016.
[16]. P. Ryan, "Ingestion of plastics by marine organisms," Marine Pollution Bulletin, vol. 62, 2016.
[17]. J. Boucher and D. Friot, "Primary microplastics in the oceans," IUCN Report, 2017.
[18]. A. Cózar et al., "Plastic accumulation in the Mediterranean Sea," Scientific Reports, vol. 5, 2015.
[19]. H. Imhof et al., "Microplastics in beach sediments," Scientific Reports, vol. 6, 2016.
[20]. J. Horton et al., "Microplastics in urban rivers,"Science of the Total Environment, vol. 586, 2017.
[21]. L. Breiman,"Randomforests,"Machine Learning, vol. 45, pp. 5–32,2001.
[22]. T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. Springer, 2009.
[23]. C. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.
[24]. [24] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.
[25]. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015.
[26]. T. Mitchell, Machine Learning. McGraw-Hill, 1997.
[27]. S. Haykin, Neural Networks and Learning Machines. Pearson, 2009.
[28]. D. Kingma and J. Ba, "Adam: A method for stochastic optimization," ICLR, 2015.
[29]. M. Abadi et al., "TensorFlow: Large-scale machine learning," OSDI, 2016.
[30]. F. Chollet, Deep Learning with Python. Manning, 2018.
[31]. T. Chen and C. Guestrin, "XGBoost," KDD,2016.
[32]. S. Hochreiter and J. Schmidhuber, "Long short- term memory," Neural Computation, vol. 9, pp. 1735–1780, 1997.
[33]. G. James et al., An Introduction to Statistical Learning. Springer, 2013.
[34]. R. Sutton and A. Barto, Reinforcement Learning. MIT Press, 2018.

[35]. J. Brownlee, Machine Learning Mastery with Python, 2020.
[36]. M. Reichstein et al., "Deep learning and Earth system science," Nature, vol. 566, pp. 195–204, 2019.
[37]. D. Rolnick et al., "Tackling climate change with machine learning," ACM Computing Surveys, vol. 55, 2019.
[38]. G. Camps-Valls et al., "Deep learning for environmental remote sensing," IEEE GRSM, vol. 9, 2021.
[39]. H. Kim et al., "LSTM-based water quality prediction," Journal of Hydrology, vol. 585, 2020.
[40]. S. Kratzert et al., "Rainfall–runoff modelling using neural networks," Hydrology and Earth System Sciences, vol. 23, 2019.
[41]. J. Shen et al., "Ensemble learning for water pollution modelling," Science of the Total Environment, vol. 710, 2020.
[42]. H. Liu et al., "Random forest for environmental risk assessment," Journal of Environmental Management, vol. 263, 2020.
[43]. S. Choubin et al., "Comparative evaluation of ML algorithms," Environmental Earth Sciences, vol. 78, 2019.
[44]. X. Zhu et al., "Hybrid deep learning for water quality prediction," Environmental Modelling & Software, vol. 114, 2018.
[45]. Y. Zhang et al., "Air quality forecasting using machine learning," Atmospheric Environment, vol. 223, 2020.
[46]. P. Malakar et al., "Coastal monitoring using predictive analytics," Ocean Engineering, vol. 187, 2019.
[47]. A. Karpatne et al., "Machine learning for scientific discovery," IEEE TKDE, vol. 29, 2017.
[48]. C. Sun et al., "Spatiotemporal prediction using neural networks," IEEE Access, vol. 8, 2020.
[49]. L. Fang et al., "Environmental forecasting with hybrid models," Ecological Informatics, vol. 54, 2019.
[50]. A. Jiang et al., "Data-driven environmental modeling," Environmental Research Letters, vol. 16, 2021.
[51]. M. Mosavi et al., "Flood prediction using ML," Water, vol. 10, 2018.
[52]. R. Singh et al., "Hybrid deep learning for pollution prediction," Environmental Monitoring and Assessment, vol. 192, 2020.
[53]. J. Park et al., "Time-series environmental forecasting," Journal of Environmental Management, vol. 246, 2019.
[54]. Q. Shi et al., "Spatiotemporal pollution prediction," Science of the Total Environment, vol. 699, 2020.
[55]. M. Hasan et al., "Regression models for environmental forecasting," Applied Soft Computing, vol. 95, 2020.
[56]. J. Brown et al., "Data analytics for sustainability," Sustainable Computing, vol. 28, 2020.
[57]. L. Sun and H. Yu, "AI-based ecological prediction," Ecological Indicators, vol. 117, 2020.
[58]. R. Ma et al., "Environmental big data analytics," Future Generation Computer Systems, vol. 108, 2020.
[59]. S. Ahmad et al., "Decision trees for pollution modeling," Expert Systems with Applications, vol. 147, 2020.
[60]. T. Li et al., "Ensemble regression for environmental systems," Knowledge-Based Systems, vol. 195, 2020.
[61]. K. Chen et al., "Smart environmental monitoring using IoT," Sensors, vol. 19, 2019.
[62]. M. Rahman et al., "Urban pollution prediction with ML," Environmental Modelling & Software, vol. 124, 2020.
[63]. J. Yu et al., "Hybrid AI for coastal risk assessment," Ocean & Coastal Management, vol. 195, 2020.
[64]. S. Verma et al., "Big data for environmental management," IEEE Access, vol. 7, 2019.
[65]. T. Nguyen et al., "Predictive analytics in

water systems," Journal of Cleaner Production, vol. 258, 2020.
[66]. H. Zhao et al., "Machine learning for marine ecosystem monitoring," Marine Environmental Research, vol. 159, 2020.
[67]. P. Kumar et al., "Artificial intelligence in sustainability science," Sustainability, vol. 12, 2020.
[68]. A. Sharma et al., "Environmental modeling with ensemble approaches," Environmental Informatics, vol. 35, 2021.
[69]. D. Li and M. Sun, "AI-based prediction of coastal contamination," Ocean Science Journal, vol. 55, 2020.
[70]. K. Park et al., "Deep learning applications in hydrology," Water Resources Research, vol. 56, 2020.
[71]. E. Choi et al., "Environmental forecasting using neural networks," Computers & Geosciences, vol. 136, 2020.
[72]. R. Thomas et al., "Data-driven coastal monitoring systems," Marine Policy, vol. 117, 2020.
[73]. M. Yadav et al., "Regression techniques for environmental risk modeling," Environmental Systems Research, vol. 9, 2020.
[74]. J. Wang et al., "AI for smart environmental management," Journal of Environmental Informatics, vol. 37, 2021