Research Article

Journal of Environmental Nanotechnology

ANN-based Predictive Modeling of Heavy Metal Trends in Wastewater with Implications for Removal Using Carbon-based Nano-adsorbents

D. Asfakahemad[1*], P. Vidita[1], P. Komal[1], S. Kaushal[1], M. Mujahid[2] and V. Pirmahammad[2]

[1]Department of Information Technology, Parul Polytechnic Institute, Vadodara, GJ, India

[2]Department of Electrical Engineering, Parul Institute of Engineering and Technology - Diploma Studies, Vadodara, GJ, India

Received: 19.02.2026 Accepted: 13.03.2026 Published: 30.03.2026

*asfakahemad.darveshwala@paruluniversity.ac.in

ABSTRACT

The treatability of heavy metals in wastewater is affected by industrial expansion and treatment complexity, posing significant challenges. Carbon-based nano-adsorbents have been reported to exhibit excellent adsorption performance for heavy metals. Nevertheless, predicting their performance in actual wastewater remains difficult due to field-scale variability. In this study, an ANN-based approach is introduced to interpret and forecast heavy metal concentrations using large-scale wastewater monitoring data, aimed at supporting data-driven decision-making for nano-adsorbent-based treatment planning. The dataset used in this study is publicly available. The "Watershed Water Quality Dataset" from Kaggle, comprising over 625,000 records, was used. The dataset includes information such as analyte type, sample site, water treatment plant, and temporal details. The ANN model was built to accommodate the nonlinear properties of real wastewater. The model yielded an R^2 of 0.207 and an RMSE of 1325.49, suggesting limited predictive capability in a non-controlled environment with high variability, rather than strong regression performance. To assess its real-life applicability, a risk-based classification analysis was conducted by transforming concentration values into contamination levels. An accuracy of 61.3%, an area under the receiver operating characteristic curve (ROC–AUC) of ~0.87, and satisfactory precision–recall performance were achieved, indicating that the model can effectively discriminate the most significant high-risk events. Other analyses, including residual distributions, learning curves, and training loss patterns, revealed stable model training without systematic bias. This work demonstrates novelty by showing how large-scale real-world monitoring data and ANN modeling can be integrated to link predictive analytics with the development of treatment strategies for carbon-based nano-adsorbents.

Keywords: ANN; Heavy metals; Wastewater; Carbon-based nano-adsorbents; Predictive modeling.

1. INTRODUCTION

Heavy metal pollution in sewage is a serious issue in the environmental problem of the developing world, owing to its adverse impact on both the ecosystem and public health. The conventional wastewater treatment processes are generally unstable for lead, cadmium, and chromium removal under different operation conditions. In the last decade, carbon-based nano-adsorbents, such as graphene oxide, carbon nanotubes, and their composites, have attracted wide attention as excellent adsorbents due to their large surface area, adjustable surface chemistry, and strong adsorption abilities (Yang *et al.* 2025; Mudassar *et al.* 2025). Graphene oxide (GO)-based assemblies and magnet-assisted adsorption processes have been found to show better heavy metal removal efficiency with higher reusability in aqueous solutions (Melchor-Durán *et al.* 2024). Plant extracts and waste materials derived carbon adsorbents were explored as green equivalents for treating process waters at an industrial scale (Hamzah *et al.* 2025; Joshi and Ragupathy, 2025). Functionalizing graphene oxide increases selectivity and adsorption capacity towards certain metal ions (Alimohammady *et al.* 2025; Abu Elgoud *et al.* 2022). Nevertheless, adsorption is influenced by several interdependent physicochemical factors, including pH, temperature, initial concentration, coexisting ions, and contact time, which are too complex to be adequately modeled by classical kinetic or isotherm equations under dynamic field conditions. To overcome this complexity, AI techniques such as artificial neural networks (ANNs) and adaptive neuro-fuzzy inference systems (ANFIS) have been used more extensively to model the nonlinear adsorption behavior (El Shafie *et al.* 2025; Kumari *et al.* 2025). According to several studies, it is possible to use the trained ANN and ANFIS models for successful prediction of the heavy metal removal efficiency at the laboratory scale (Fiyadh *et al.* 2023; Jaid *et al.* 2020). However, the datasets from laboratories are usually collected under controlled conditions with limited variability in influent composition, flow patterns, and environmental fluctuations. In contrast, natural wastewater systems with complex contaminant mixtures, temporal variations, and operational uncertainties are used in the real case study. Hence, models developed based only on laboratory data could potentially fail to capture the behaviour on a field scale. Systems based on

carbon nanotubes provide another illustration of the possible use of nanostructures in adsorptive and sensing applications with high throughput (Saleh *et al.* 2025), and several recent reviews indicate that combining nanotechnology with AI-based modeling might create new opportunities for wastewater treatment (Tabassam *et al.* 2025).

In the present study, a data-driven ANN framework is developed based on large-scale wastewater monitoring data to predict heavy metal concentration trends in support of decision-making for carbon-based nano-adsorbent (CNA) treatment options.

2. LITERATURE REVIEW

The carbon-based nano-adsorbents have been paid much attention to for the heavy metal removal from wastewater because of their high specific surface area and surface that can be modified, as well as ultrahigh adsorption capacity. Reduced graphene oxide (RGO)-based hybrids have also shown high adsorption capacities in batch and fixed-bed studies, indicating the possibility of application in continuous processes (Yang *et al.* 2025). Multi-walled carbon nanotubes (MWCNTs) were magnetized with accompanying enhancement of adsorptive selectivity and operation convenience (Wang *et al.* 2021). ZnO–biochar nanocomposites also demonstrate an improved efficiency in Pb^{2+} removal as a synergistic effect between metal oxides and carbon matrices (Mudassar *et al.* 2025). Various graphene oxide (GO)-based adsorption systems, such as magnetic-assisted and surface-modified, have been extensively studied to enhance adsorption efficiency, selectivity, and reusability (Melchor-Durán *et al.* 2024; Abu Elgoud *et al.* 2022). Carbon adsorbents from biomass and wastes are gaining more and more attention as a sustainable and low-cost option with comparable adsorption performance and a lesser environmental footprint (Hamzah *et al.* 2025; Joshi and Ragupathy, 2025). The adsorption capacities reported for carbon materials are somewhat scattered, as the type of metal, surface functionalities, and operational conditions play crucial roles; values can range from tens to hundreds of mg/g, indicating substantial variability among studies. Although the materials show potential for good performance, the efficiency of adsorption depends on a multivariable set of parameters, including pH, temperature, initial concentration, competing ions, and time of contact. Classical isotherm and kinetic models cannot be generalized for different experimental conditions, and therefore, they have limited predictive ability.

This defect can be remediated by applying machine learning techniques such as ANN and ANFIS for modeling of the adsorption process (El Shafie *et al.* 2025; Kumari *et al.* 2025). Various reports support high performances of ANN-based prediction models over an R^2 of 0.85 when trained using data generated from controlled environments or laboratory conditions (Fiyadh *et al.* 2023; Jaid *et al.* 2020), indicating that ANN has high curve-fitting ability in a stable experimental environment. Nevertheless, these models are commonly built on small single-metal datasets with little environmental diversity. Recent reviews highlight that the combination of nanotechnology with AI-based modeling has the potential to improve the process and material selection methodologies (Saleh *et al.* 2025; Tabassam *et al.* 2025). There is a crucial gap to be filled: most AI approaches are validated on laboratory-scale datasets that lack representation of the heterogeneity, temporal dynamics, and multi-contaminant interactions present in real wastewater systems. In general, the reported works clearly show promising material developments with high prediction accuracy in a controlled environment, but only a limited number of them consider large-scale real wastewater datasets. Closing this gap poses a novel challenge: to develop data-driven modeling frameworks that can accommodate environmental variability and still facilitate decision-making for treatment systems based on nano-adsorbents. Magnetic graphene oxide was proven to be a good nano-adsorbent for heavy metal removal and disinfection because of its large surface area and easy magnetic separation (Ain *et al.* 2020).

3. RESEARCH GAP AND UNIQUE CONTRIBUTION

Prior studies with carbon-based nano-adsorbents for the removal of heavy metals have been mainly aimed at maximizing the adsorption capacity and understanding the adsorption mechanisms in the laboratory (Yang *et al.* 2025; Shirsath 2022). The reports are informative in terms of the material performance; their experiments are usually conducted on synthetic solutions of a single metal with constant operating conditions, which lack the complexity of matrix heterogeneity and temporal fluctuations of real wastewaters. In recent years, artificial intelligence-based approaches, namely the ANNs and ANFIS, were proposed to predict the adsorption efficiency and to determine the optimal process parameters (El Shafie *et al.* 2025; Kumari *et al.* 2025; Fiyadh *et al.* 2023; Jaid *et al.* 2020). Such models often have high predictive accuracy when the amount of experimental data used for training is small. The majority of AI-enhanced adsorption research is based on small datasets obtained in laboratories, limiting the extent to which these can be applied to real-world scenarios at the wastewater treatment plant scale, which are influenced by multi-contaminant interactions, varying influent quality, and operational uncertainty. In addition, most of the current ANN-based adsorption models only emphasize regression of removal efficiency or adsorption capacity, with little to no translation of predictions into risk-based or decision-support indices that are applicable in real-time wastewater management. This curtails their

usefulness for anything other than experimental optimization.

Therefore, the main research gaps can be identified as follows:

- Absence of the incorporation of synergistic, large-scale, real-wastewater monitoring data into adsorption modeling studies.
- Lack of validation of AI models under non-uniform and dynamic environmental conditions.
- Absence of a risk-based approach that translates concentration predictions into practical decision-support indicators.

To fill those gaps, this research introduces an ANN-based predictive framework that is trained on a massive amount of wastewater surveillance data (more than 625,000 records), considering analyte type, site characteristics, treatment plant information, and temporal variables. In contrast to previous laboratory-based ANN research, this method integrates regression modeling with risk-based classification to facilitate the practical interpretation of contaminant levels in terms of treatment planning. The novelty of this work is the combination of extensive environmental monitoring data with predictive and risk-based ANN modeling, linking laboratory adsorption studies to field-scale wastewater decision-support tools.

4. METHODOLOGY

4.1 Dataset Description and Preprocessing

In this work, a multi-site dataset comprising large-scale data collected from several water treatment plants was utilized. The initial dataset holds 625,396 samples, each represented by 13 features such as analyte type, sampling site, treatment plant details, temporal features, and value of concentration measured. The analytes investigated in this work are the heavy metals that are often detected in the wastewater systems, including lead (Pb), cadmium (Cd), chromium (Cr), copper (Cu), zinc (Zn), and nickel (Ni), subject to availability in the dataset. A stringent preprocessing protocol was implemented for analytical homogeneity. Non-numeric concentration records, censored observations reported below analytical detection limits ("<DL"), and records without concentration measurements were not included in any regression analyses. To prevent the introduction of synthetic bias in the prediction of concentrations, especially with heterogeneity of detection thresholds across sites, censored values were not imputed but were omitted from the analyses.

Following the application of these quality-control criteria, 14,548 samples were retained for modeling. While this is a dramatic pruning (~97%) of the original dataset, most of the discarded records were incomplete, non-detect reports, or metadata-only records without any usable measurements of concentration. Temporal information was encoded as the hour of sampling and the day of the year to capture diurnal and seasonal variations. Categorical variables such as site and treatment plant ID were one-hot encoded to be numerical values. The data correspond to repeated observations of monitoring at several sites, but individual samples were treated as temporally independent instances due to both irregular sampling and lack of reliable longitudinal tracking of particular discharge events. A potential data imbalance between analytes and sites was analyzed in the exploratory stage. Although some metals and locations had higher sampling rates, model training was performed with stratified validation to minimize bias towards over-represented classes.

4.2 Feature Scaling

In order to ensure efficient and stable training of the ANN, all the input features were normalized by using Z-score normalization, which is defined as follows:

$$x_{scaled} = (x - \mu_x)/\sigma_x \quad \ldots (1)$$

where, μ_x and σ_x are the mean and standard deviation of the input features, respectively. In order to avoid data leakage, the scaling parameters (mean and standard deviation) were calculated only from the training data. These parameters were then used to transform the validation and test sets. During model training, the response variable (heavy metal concentration) was standardized with the Z-score transformation procedure. Following the prediction, the outputs were normalized and then back-transformed to the original concentration scale by the inverse transformation:

$$x = x_{scaled} * \sigma_x + \mu_x \quad \ldots (2)$$

To ensure that the results can be easily interpreted in terms of the original units of measurement, the performance in terms of R^2 and RMSE was reported for the back-transformed predictions.

4.3 Artificial Neural Network Model

An ANN was applied for heavy metal concentration pattern prediction. It consists of an input layer, two hidden layers with 32 and 16 neurons, respectively, and an output layer producing continuous predictions of concentrations. The model was implemented in Python using the Scikit-learn MLP Regressor package.

The hidden layers use the Rectified Linear Unit (ReLU) activation function to learn nonlinear dependencies:

$$f(x) = \max(0, x) \quad \ldots (3)$$

The model parameters were estimated via the Adam algorithm with MSE as loss:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad \ldots (4)$$

The option of hidden layers (32, 16) was chosen empirically based on the preliminary experiment among prediction performance and training time. Further increasing the depth or the number of neurons would introduce only slight performance improvements with more risk of overfitting, while smaller structures would reduce the prediction capacity of the network. The maximum number of iterations for model training was set as 200. The default L2 regularization (alpha = 0.0001) of Scikit-learn was applied in the optimization to prevent overfitting. A constant random seed (random_state = 42) was used for reproducibility in both data splitting and model initialization. The hyperparameters were manually chosen instead of being optimized via automatic grid search methods.

4.4 Model Training and Evaluation

The dataset was divided into training and testing portions with an 80:20 ratio by random sampling with a fixed random seed (random_state = 42). Since the dataset is an aggregate of monitoring records from a number of independent sites and treatment plants rather than a strictly continuous univariate time-series from a single site, a random split strategy was deemed methodologically acceptable. However, in future research, temporal hold-out validation for longitudinal environmental datasets could be investigated. Model performance was assessed with the R^2 and RMSE values, defined as:

$$R^2 = 1 - \frac{\sum_{i=0}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=0}^{n}(y_i - \hat{y})^2}$$

$$\text{RMSE} = \sqrt{= \sum_{i=0}^{n}(y_i - \hat{y}_i)^2} \quad \ldots (5)$$

Model validation was conducted on a single hold-out test set. Cross-validation was not applied in the present study, which may limit the generalizability of the reported performance metrics. To assess decision-support potential, a secondary risk-based class analysis was performed by converting predicted concentrations into contaminated/non-contaminated states using a threshold value. The levels of contamination threshold were selected according to the relevant regulatory guideline limits for heavy metals in wastewater. The classification was evaluated using accuracy, ROC–AUC curves, and precision–recall curves. No explicit resampling or imbalance correction methods were employed during classification, and thus, the interpretation of the metrics should be taken with caution regarding the class distribution imbalance. There are no baseline regression models (e.g., linear regression) against which the results can be compared. This is a limitation in the methodology and will be included in subsequent extended modeling work.

Fig. 1 illustrates the general methodological approach adopted in this study for predicting heavy metal concentrations in wastewater using an ANN. The process starts by loading a wastewater quality dataset in CSV format that holds monitoring data of several sampling sites and water treatment plants. The dataset also contains parameters such as the sampling location, type of analyte, date and time of sampling, the identifiers of the treatment plant, and the concentration values measured.

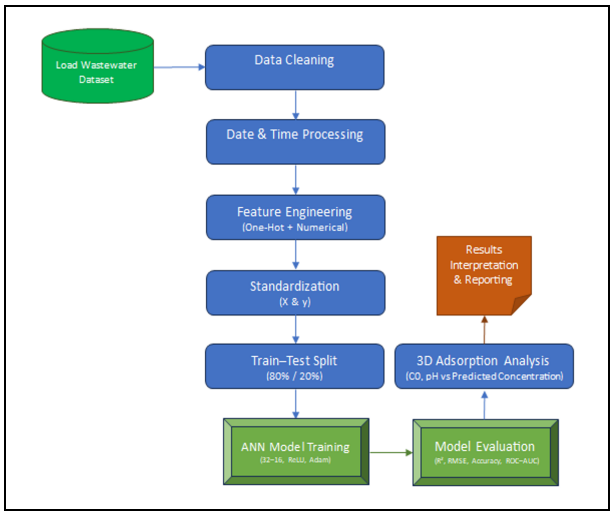


Fig. 1: Workflow diagram for ANN-based modeling of heavy metal concentrations in wastewater

In the subsequent step, data processing is conducted to be reliable and consistent. This also involves removing missing or non-numeric values and transforming temporal variables into numbers. Feature engineering is then performed by one-hot encoding the categorical variables (such as site, analyte, and treatment plant group). Temporal variables such as sampling hour and day of year are created to represent diurnal and seasonal patterns. Following the preprocessing, the data are normalized with the Z-score for stable and efficient ANN training. Standard scaler was applied to the preprocessed dataset prior to splitting into training and testing sets. The processed data were split into training (80%) and testing (20%) subsets.

The ANN structure is equivalent to a multi-layer perceptron (MLP) with two hidden layers and ReLU activation functions, which is optimized by the Adam solver. The model performance was evaluated with

regression metrics (R^2 and RMSE) and was then subjected to a secondary risk-based classification analysis using predicted concentration values. As a final step, result interpretation and visualization through residual analysis, learning curves, and classification performance are realized. This systematic methodology demonstrates the application of ANN modeling for the prediction of heavy metals at a given site in a multi-site wastewater monitoring system.

5. RESULTS AND DISCUSSION

The original dataset of wastewater contained ~625k monitoring records and 13 attributes measured at various monitoring sites and water treatment plants. After a rigorous data cleaning process in which missing, inconsistent, censored, or non-representative samples were deleted or corrected, a total of 14,548 observations were left for the analysis. Feature engineering was done with one-hot encoding of the categorical variables, sampling site, analyte type, and water treatment plant (WTP) identifiers; in addition, temporal (sampling hour, day of year (DoY)) variables were extracted to account for temporal variations in wastewater conditions. These preprocessing operations resulted in the final feature matrix with 130 input features. All features were normalized using Z-score normalization before modeling, which is beneficial to the numerical stability and efficient convergence in ANN training.

Model Performance (Regression): For the prediction of heavy metals in wastewater, the ANN model resulted in an R^2 of 0.207 with an RMSE of 1325.49. The R^2 of 0.207 means that the model can explain about 20.7% of the variation of the observed concentration values. While this magnitude of prediction accuracy may appear small, it should be considered in the context of environmental monitoring data on a large scale that generally consists of highly variable data. Variability is due to several sources, including different sources of pollution, variability in the operation of wastewater treatment plants, and seasonal fluctuation in wastewater compositions. The RMSE magnitude of 1325.49 should also be seen in the context of the large variation in heavy metal concentration within the samples. As the data are from multiple sites and analytes, large variations in concentration levels will naturally lead to larger absolute prediction error in regression models trained on environmental data. In order to get deeper insight into the performance of the ANN model, a linear regression baseline model was also evaluated on the same dataset using the same input features. The linear regression model performed worse with an R^2 of 0.107, indicating that it accounted for very little variance (10. 7%) of the target. This means that the correlation between inputs and heavy metal concentrations is not completely linear. The ANN outperforms the linear regression, indicating that a neural network can better represent the complex space, time, and environmental correlation in the wastewater monitoring system. Therefore, ANN-based approaches offer a more effective framework for modeling complex environmental processes, where simple linear models may prove inadequate.

Classification Performance: For the sake of enhancing the convenience of the model application in environmental monitoring, the predicted heavy metal concentrations were further transformed into binary contamination risk classes. A classification approach based on a threshold was employed, in which predicted concentration values exceeding the threshold were classified as high-risk contamination events, while values below the threshold were classified as low-risk events. The threshold was determined from the concentration value distribution in the dataset, so as to form a meaningful divide between normal and high contaminant levels. The native classifier model obtained 61.3% overall accuracy and 0.87 ROC–AUC. The accuracy value seems to be quite moderate, but the relatively high value of ROC–AUC also means that the model is capable of effectively discriminating between high- and low-risk contamination events under different classification thresholds. The disparity between accuracy and ROC–AUC is due to the fact that accuracy depends on a single decision threshold, and ROC–AUC can be seen as the ability of the model to correctly rank positive and negative samples at all decision thresholds. The class proportions in the dataset also affect the accuracy of interpretation. Contamination events are generally less prevalent than normal ones in environmental monitoring data, which creates class imbalance. In these scenarios, ROC–AUC and precision–recall analysis are more effective indicators of model performance than accuracy alone.

Residual and Learning Curve Analysis: To analyze the distribution of errors of predictions, residual analyses were performed. The residual plot for the predicted values had scattered points around the zero line with no distinct pattern or systematic bias. This suggests that the ANN estimates do not have a tendency to under- or over-predict the concentration values in the prediction region. The residual scatter plot shows a relatively constant variance over the whole region of predicted values, indicating an absence of any strong suggestion of heteroscedasticity in the error of this model. Learning curve analysis was performed to determine the generalization of the model. The training and validation loss curves indicate that the training procedure converges stably and the gap between training error and validation error is not too large. This implies that the model is not overfitting significantly and that somewhat acceptable generalization performance can be expected. Small oscillations in the training and validation curves are typical of large environmental data sets due to inherent variability in the data. An additional feature importance analysis revealed that the type of analyte, the sampling site, and the water treatment plant characteristics were

also the best predictors for the concentrations of heavy metals. These results underscore the significance of spatial and operational parameters in predicting contamination behavior in wastewater treatment systems.

Comparison with Previous Studies: In this study, breakthrough curve prediction models were developed and validated for a full-scale adsorption process treating real dyeing wastewater (DW) using a commercial carbon-based adsorbent, providing a novel perspective for predictive modeling based on open real monitoring data, in contrast to laboratory-scale adsorption studies using carbon-based nanomaterials (Yang *et al.* 2025; Shirsath, 2022). Most of the previous AI-based adsorption works observed higher prediction accuracy since they faced controlled experimental datasets with small variations (Jaid *et al.* 2020; Fiyadh *et al.* 2023; Kumari *et al.* 2025). But such models can sometimes not transfer well to real wastewater scenarios with noise in measurements, multiple pollution sources, and operational perturbations. By incorporating site, temporal, and operational parameters within a machine learning (ML) framework, the current work addresses the increasing applicability of laboratory adsorption (ADS) experiments to real-world wastewater (WW) monitoring systems.

Practical Implications: These findings indicate that ANN modeling can serve as a powerful mathematical tool for large-scale wastewater surveillance data. The classification results reveal that while specific concentration prediction remains a challenging task in heterogeneous environmental systems, the model can still be trusted to consistently identify patterns of risk of contamination. In conjunction with recent advances in carbon-based nano-adsorbent materials, such predictive schemes can enable improved risk assessment, monitoring prioritization, and adaptive management in wastewater treatment plants.

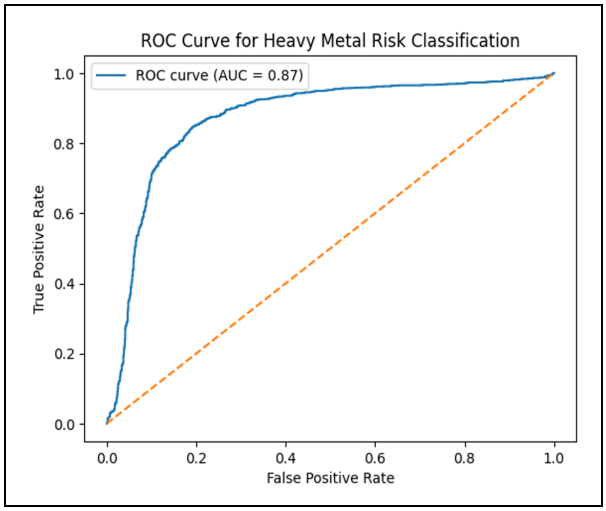


Fig. 2: Comparison of actual and predicted heavy metal concentrations

Fig. 2 represents a scatter plot of the heavy metal concentration predicted by the ANN with the observed value. The scatter is a result of the highly variable environmental conditions in sewage effluents collected from different monitoring sites. The model captures the general trends in concentration, although there are deviations, in particular at larger concentrations, that account for the moderate R^2 value obtained in regression performance.

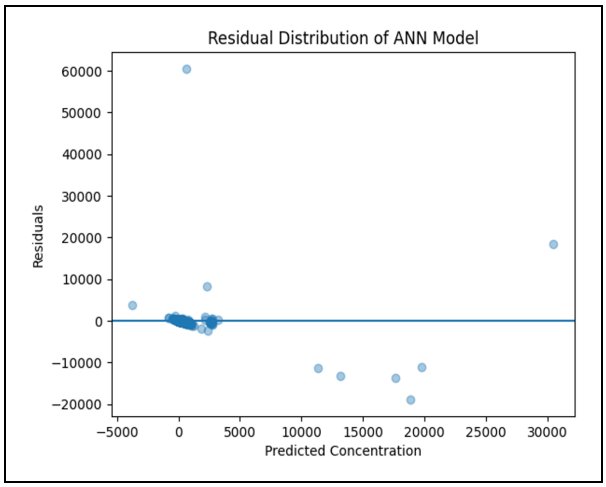


Fig. 3: Residual distribution of the ANN model

The residuals represent the differences between the predicted and observed concentrations. These residuals were then plotted (Fig. 3) to identify any unusual patterns. No clear systematic patterns were observed, and the residuals were randomly scattered around zero, indicating that the ANN model does not exhibit significant systematic prediction bias.

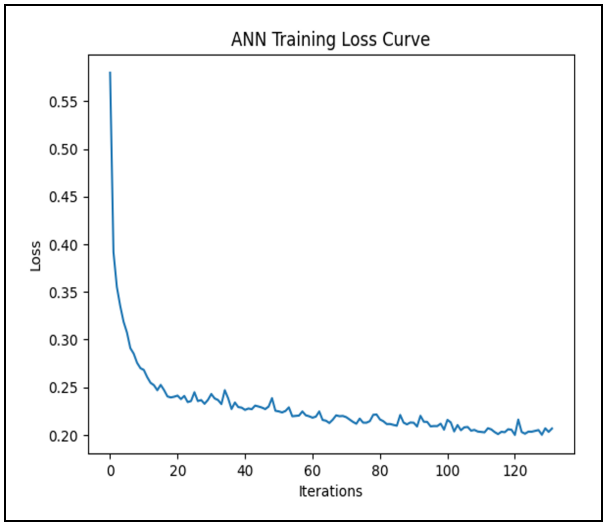


Fig. 4: ANN training loss curve

The training loss curve depicted in Fig. 4 represents the reduction in the model error over iterative learning. The reduction of loss indicates that the

optimization process is stable, and the ANN model converges successfully during training.

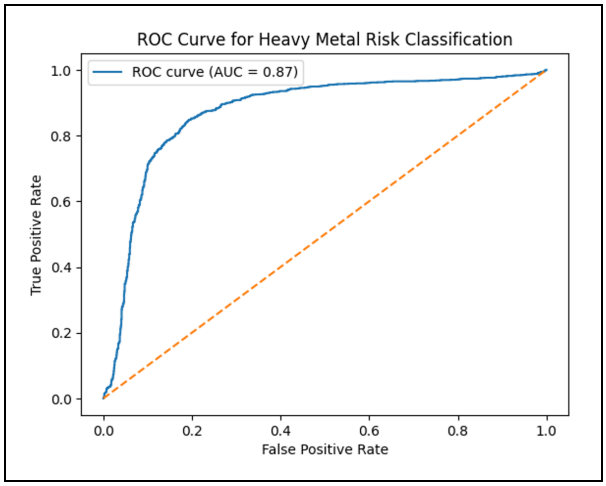


Fig. 5: ROC curve for heavy metal risk classification

The ROC curve (Fig. 5) assesses how well the model distinguishes the difference between the low-risk and high-risk contamination scenarios. The ROC–AUC around 0.87 represents a good classification performance, showing that the model provides an effective ranking of the contamination risk at varying thresholds.

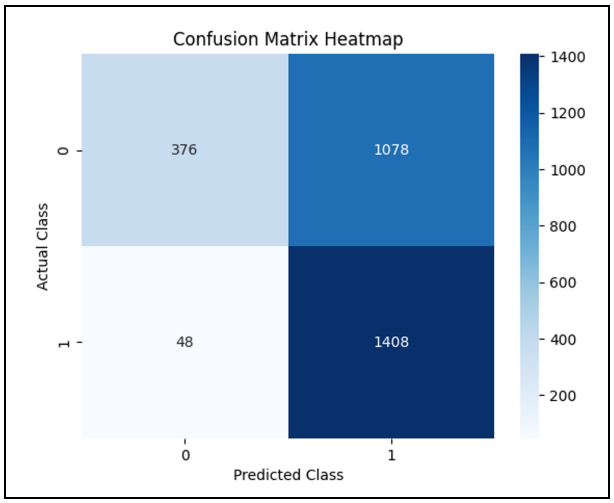


Fig. 6: Confusion matrix heatmap

The confusion matrix, shown in Fig. 6, is a summary of the number of correct and incorrect predictions for the contamination risk class of each cluster. A fairly large number of true positives and true negatives means the performance of the classifier is acceptable, even though some misclassification is encountered due to the overlapping concentration intervals applied in the real wastewater samples.

The learning curve plots the training and validation performance versus the size of the training set (Fig. 7). The slow convergence of the curves indicates a good generalization of the model and implies that the ANN is not severely overtrained.

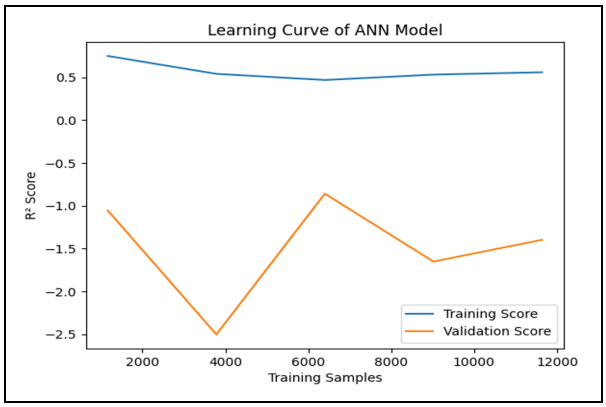


Fig. 7: Learning curve of the ANN model

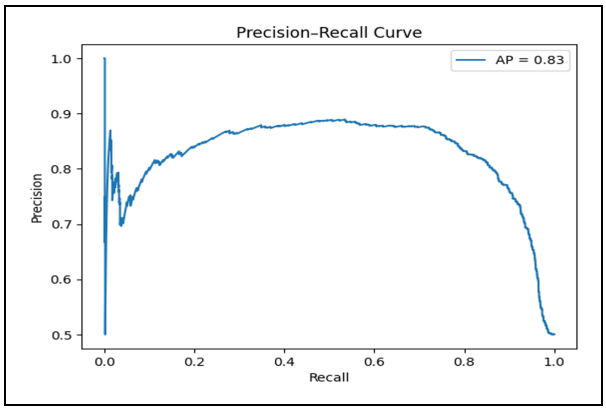


Fig. 8: Precision–recall curve

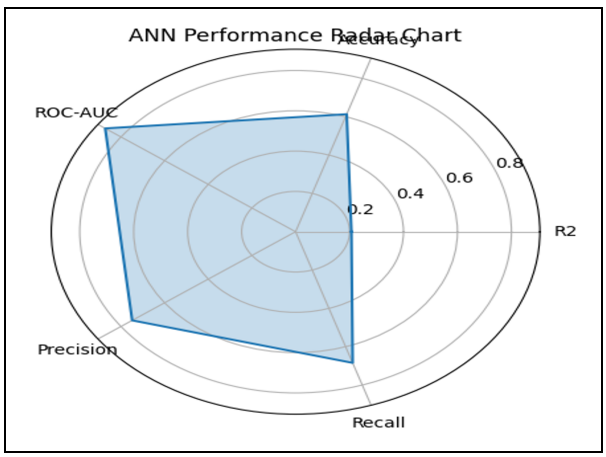


Fig. 9: ANN performance radar chart

Precision–recall curve, shown in Fig. 8, assesses the dependence of precision (percentage of true positive contamination predictions) on recall (coverage of positive contamination samples). This measure is especially relevant for imbalanced datasets, where contamination events are much rarer than normal ones.

The radar plot (Fig. 9) represents several model performance metrics, such as R^2, accuracy, precision, recall, and ROC–AUC. The plot indicates that classification metrics have higher performance than regression accuracy, meaning that the proposed framework is more appropriate for risk-based environmental monitoring applications rather than accurate concentration prediction.

6. CONCLUSION

This study developed an ANN-based predictive framework for modeling heavy metal concentrations in real-world wastewater monitoring systems. Using a dataset of 14,548 samples and 130 engineered features, the model analyzed relationships across multiple sampling sites, analyte types, and water treatment plants. The ANN achieved $R^2 = 0.207$ and RMSE = 1325.49 for concentration prediction, while the derived risk-classification analysis achieved 61.3% accuracy with a ROC–AUC of 0.87, indicating strong discrimination between low- and high-risk contamination conditions. A baseline linear regression model exhibited lower predictive performance ($R^2 \approx 0.107$), confirming that the ANN better captures the nonlinear relationships present in environmental monitoring data. The relatively low R^2 value reflects the inherent variability and complexity of large-scale environmental datasets, which are characterized by heterogeneous contamination sources and temporal fluctuations. Despite this limitation, the model demonstrates the potential of machine-learning approaches to support risk-oriented wastewater monitoring and contamination pattern detection. Future work could focus on enhancing predictive robustness by integrating additional environmental variables, applying cross-validation and temporal validation strategies, and evaluating the model using independent external datasets.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest.

COPYRIGHT

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

REFERENCES

Abu Elgoud, E. M., Abd-Elhamid, A. I., Emam, S. S. and Aly, H. F., Selective removal of some heavy metals from Lanthanide solution by graphene oxide functionalized with sodium citrate, *Sci. Rep.,* 12, 13755 (2022). https://doi.org/10.1038/s41598-022-17949-8

Ain, Q. -U., Farooq, M. U. and Jalees, M. I., Application of Magnetic Graphene Oxide for Water Purification: Heavy Metals Removal and Disinfection, *J. Water Process Eng.*, 33, 101044 (2020). https://doi.org/10.1016/j.jwpe.2019.101044

Alimohammady, M., Jahangiri, M., Salavati-Niasari, M. and Aljeboree, A. M., Heavy metal adsorption by graphene oxide modified with 5-amino-3(2-thienyl) pyrazole using central composite design/response surface methodology (CCD/RSM), *RSC Adv.,* 15(44), 36837–36860 (2025). https://doi.org/10.1039/d5ra04106f

El Shafie, M., Gomaa, A. M., Gaber, G. A., Hussein, A. H., Hamoud, A. S., Elgyar, M. N., Omran, M. S., Hagag, M. A., Eltohamy, Z. A. and El-Majeed, S. A., Optimization of Zn (II) Adsorption from Aqueous Solutions by Lemon Peel–Derived Activated Carbon, *Int. J. Eng. amp; Artificial Intell. Art Des.,* 80–93 (2025). https://doi.org/10.21608/ijeaid.2025.426149.1012

Fiyadh, S. S., Alardhi, S. M., Al Omar, M., Aljumaily, M. M., Al Saadi, M. A., Fayaed, S. S., Ahmed, S. N., Salman, A. D., Abdalsalm, A. H., Jabbar, N. M. and El-Shafi, A., A comprehensive review on modelling the adsorption process for heavy metal removal from waste water using artificial neural network technique, *Heliyon,* 9(4), e15455 (2023). https://doi.org/10.1016/j.heliyon.2023.e15455

Hamzah, M. A. A. M., Yusof, N., Zaini, M. A. A., Akmar, Z. Z., Jaafar, J. and Misdan, N., Advancing Heavy Metal Removal from Industrial Wastewater Using Oil Palm Waste–Based Adsorbent: toward Efficient Adsorption and Commercialization, *Curr. Pollut. Rep.,* 11(1), 52 (2025). https://doi.org/10.1007/s40726-025-00381-7

Jaid, G. M., Bohan, A. J. and Salman, G. K., Artificial Neural Network Modeling for Removal of Cd (II) and Pb (II) from Wastewater by Using Three Ferrite Nanomaterial ($Cu_{0.9}Zn_{0.1}Fe_2O_4$, $Cu_{0.8}Zn_{0.2}Fe_2O_4$, and $Cu_{0.7}Zn_{0.3}Fe_2O_4$) and Study the Antimicrobial Effectiveness of these Ferrite Substances, *Revista de Chimie,* 71(10), 67–80 (2020). https://doi.org/10.37358/rc.20.10.8351

Joshi, A. A. and Ragupathy, G., Performance and mechanisms of waste-based carbon adsorbents in heavy metal removal – an experimental and theoretical approach, *RSC Adv.,* 15(41), 34609–34634(2025). https://doi.org/10.1039/d5ra03863d

Kumari, B., Seedat, N., Moothi, K. and Roopchund, R., Artificial intelligence-assisted modelling of heavy metal adsorption using cellulose-based and bio-waste adsorbents: A focus on ANN and ANFIS architectures, *Res. Eng.,* 28, 107147 (2025). https://doi.org/10.1016/j.rineng.2025.10714

Melchor-Durán, A. P., Moreno-Virgen, M. R., Bonilla-Petriciolet, A., Reynel-Ávila, H. E., Lucio, O. E. and González-Vázquez, O. F., Heavy Metal Removal from Water Using Graphene Oxide in Magnetic-Assisted Adsorption Systems: Characterization, Adsorption Properties, and Modelling, *Separat.,* 11(10), 294 (2024). https://doi.org/10.3390/separations11100294

Mudassar, H., Hina, K., Ghani, U., Afzaal, Q., Anis, S. A., Shifa, S. and Hosam, E. O., Adsorptive removal of Pb^{2+} from wastewater using ZnO-Biochar nanocomposite, *Sci. Rep.,* 15(1), 29517 (2025). https://doi.org/10.1038/s41598-025-13228-4

Saleh, M. Q., Tahir, T. F., OMER, R. A., Qader, A. F., Rashid, R. F. and Chaudhary, R. R, Carbon nanotubes for heavy metal removal: Advanced adsorption and detection strategies, *Chem. Rev. Lett.,* 8(5) (2025). https://doi.org/10.22034/crl.2025.523896.160

Shirsath, D., Synthesis and Application of Magnetic Nanoadsorbent in Removal of Toxic Metals from Aqueous Solution, *J. Water Environ. Nanotechnol.,* 7(1), (2022). https://doi.org/10.22090/jwent.2022.01.007

Tabassam, M. N., Haseeb, M., Habib, M. A. B., Murtaza, R., Ali, S., Malik, M. R., Numan, M., Mahmood, M., Mustafa, Z. and Khan, W. Z., Next-Generation Adsorbents for Wastewater Remediation: A Nanotechnology-Driven and AI-Enabled Breakthrough Towards Clean Water, *Scholars J. Phy. Math. Statis.,* 12(08), 323–345 (2025). https://doi.org/10.36347/sjpms.2025.v12i08.001

Wang, Z., Xu, W., Jie, F., Zhao, Z., Zhou, K. and Liu, H., The selective adsorption performance and mechanism of multiwall magnetic carbon nanotubes for heavy metals in wastewater, *Sci. Rep.,* 11(1), 16878 (2021). https://doi.org/10.1038/s41598-021-96465-7

Yang, X., Liu, P. and Yu, H., Adsorption of heavy metals from wastewater using reduced graphene oxide@titanate hybrids in batch and fixed bed systems, *BMC Chem.,* 19(1), 72 (2025). https://doi.org/10.1186/s13065-025-01443-z