Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

p-ISSN: 0972-6268
(Print copies up to 2016)
e-ISSN: 2395-3454

Vol. 25 | No. 2 | Article ID B4376 | 2026

Original Research Paper

https://doi.org/10.46488/NEPT.2026.v25i02.B4376

Open Access Journal

Toxicity Prediction of Landfill Leachate-Contaminated Crops Using Machine Learning Models Based on PAH and Heavy Metal Concentrations

Sireesha Vikkurty[1], M. Sandhya Vani[2], D. Sravani[3], Shanmuga Sundari M.[2]† and Kbks Durga[2]

[1]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, India
[2]BVRIT Hyderabad College of Engineering for Women, Hyderabad, India
[3]St. Peter's Engineering College, Hyderabad, India
†Corresponding author: Shanmuga Sundari M.; sundari.m@bvrithyderabad.edu.in

Abbreviation: Nat. Env. & Poll. Technol.
Website: www.neptjournal.com

Received: 10-07-2025
Revised: 08-09-2025
Accepted: 27-09-2025

Key Words:

Landfill leachate
Crop contamination
Polycyclic aromatic hydrocarbons (PAHs)
Heavy metals
Machine learning
Artificial neural network (ANN)
Support vector machine (SVM)
Random forest (RF)
Principal component analysis
Environmental risk assessment

Citation for the Paper:

Vikkurty, S., Sandhya Vani, M., Sravani, D., Sundari, M.S. and Durga, K. , 2026. Toxicity prediction of landfill leachate-contaminated crops using machine learning models based on PAH and heavy metal concentrations. *Nature Environment and Pollution Technology*, 25(2), B4376. https://doi.org/10.46488/NEPT.2026.v25i02.B4376

Note: From 2025, the journal has adopted the use of Article IDs in citations instead of traditional consecutive page numbers. Each article is now given individual page ranges starting from page 1.

ABSTRACT

The unregulated disposal of municipal solid waste in landfills generates leachate that contaminates the surrounding soil and crops with toxic substances, posing a major threat to food safety and human health. This study evaluated the contamination levels in agricultural fields located near five landfill sites in South India. A total of 600 samples (370 safe, 230 unsafe) comprising soil and edible crop tissues were analyzed for 16 polycyclic aromatic hydrocarbons (PAHs) and eight heavy metals using Gas Chromatography-Mass Spectrometry (GC-MS) and Atomic Absorption Spectrophotometry (AAS). Labels were assigned according to international safety thresholds, and multiple machine learning models-Artificial Neural Network (ANN), Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN)-were trained using nested, group-aware 5-fold cross-validation, with additional leave-one-site-out validation to test geographical generalization. Among the tested models, ANN achieved the highest predictive accuracy of 97.8% (AUC = 0.98), followed by RF (94.7%) and SVM (93.6%). Feature importance analysis revealed that Cd (importance = 0.214), benzo[a]pyrene BaP (0.187), and Pb (0.162) were the most influential predictors of crop safety. These findings demonstrate that integrating contaminant profiling with machine learning provides a robust framework for environmental risk assessment and supports safe agricultural practices in landfill-impacted regions.

INTRODUCTION

The rapid expansion of urban areas, industrialization, and poorly managed waste disposal systems have led to escalating environmental pollution problems worldwide. The leachate produced by landfills of municipal solid waste (MSW) is one of the most significant environmental issues because of its high concentration of pollutants and its ecological effects (Gaur et al. 2024). In many developing countries, the absence of engineered landfill management and leachate treatment infrastructure allows large volumes of untreated leachate to seep into adjacent ecosystems. This infiltration introduces a variety of hazardous organic and inorganic pollutants, including polycyclic aromatic hydrocarbons (PAHs), heavy metals, phenolic substances, ammonia, and chlorinated organics, into the soil, groundwater, and nearby surface water systems. These persistent pollutants not only degrade soil quality but also pose long-term ecological hazards by disrupting the local environmental balance and threatening water safety.

Given their environmental endurance, capacity to bioaccumulate in food chains, and well-established harmful effects on both humans and wildlife, PAHs and heavy metals are particularly alarming among the dangerous compounds frequently found in landfill leachate. PAHs are organic substances made up of two or more fused rings of benzene, usually formed in incomplete combustion and waste degradation processes, and are known to be significant pollutants of the environment (Rajesh

& Saravanakumar, 2024). Similarly, heavy metals that are particularly toxic and non-biodegradable, including lead (Pb), cadmium (Cd), arsenic (As), and mercury (Hg), can accumulate in plant and animal tissues and cause chronic toxicity in exposed populations.

An increasing environmental concern is the absorption of toxic substances by crops grown in soil affected by landfill leachate (LL). Various environmental monitoring studies have shown that edible crops cultivated near waste disposal sites frequently contain polycyclic aromatic hydrocarbons (PAHs) and heavy metals that surpass the maximum permissible limits set by international regulatory bodies, including the United States Environmental Protection Agency (USEPA), Food and Agriculture Organization (FAO), and World Health Organization (WHO). Consumption of such contaminated produce may result in a range of health problems, including cancer, genetic mutations, kidney damage, neurological disorders, and other long-term ailments, particularly affecting sensitive groups such as children, pregnant women, and the elderly.

Although contaminated agricultural products close to landfills may pose health risks to the public, standard evaluation procedures still mainly rely on laboratory-based chemical analysis methods, such as Atomic Absorption Spectrophotometry (AAS) for heavy metals and Gas Chromatography-Mass Spectrometry (GC-MS) for PAHs. Although highly accurate, these methods are costly, require sophisticated equipment, and are impractical for continuous large-scale environmental monitoring. Furthermore, conventional statistical approaches often fall short in interpreting the complex, multivariate, and nonlinear interactions between diverse environmental pollutants and their cumulative biological impacts, indicating the need for more advanced, data-driven analytical tools.

The latest development of machine learning (ML) and artificial intelligence (AI) technologies has made it possible to find innovative solutions to solve complex environmental pollution issues (Popescu et al. 2024; Chauhan & Sahoo 2024). These advanced computational techniques can process extensive multidimensional environmental datasets, uncover complex data patterns, and generate dependable predictive outcomes without the need for predefined programming instructions. Their strength lies in their capacity to learn from empirical observations and effectively handle nonlinear, multifactorial relationships between environmental variables. This makes them especially valuable for applications such as environmental toxicity assessments and ecological risk evaluations. Several studies have shown that machine learning models can be successfully applied in fields such as estimating health risks related to environmental exposures, modeling the dispersion of air pollutants, forecasting soil contamination, and evaluating water quality indicators. The use of machine learning algorithms for the combined evaluation of heavy metal and polycyclic aromatic hydrocarbon (PAH) toxicity in crops exposed to landfill leachate has not received much attention in the current environmental research literature, despite these encouraging advancements. This suggests a substantial knowledge gap that requires thorough examination.

The present study addresses this critical research gap by proposing an AI-based toxicity prediction framework specifically designed for crops grown in landfill-adjacent agricultural fields. By quantifying the concentrations of 16 priority PAHs and eight hazardous heavy metals in crop and soil samples from multiple landfill-affected sites and applying various machine learning algorithms, this study aimed to classify produce into safe and unsafe categories based on established international safety thresholds. Furthermore, this study incorporated Principal Component Analysis (PCA) to identify the key contaminants contributing most significantly to toxicity, thereby offering insights into the relative importance of different pollutants within the context of environmental food safety.

This study has three objectives.

1. To determine the concentration levels of priority PAHs and heavy metals in crops and soil samples collected from landfill-contaminated agricultural sites.
2. To create and assess many machine learning models for crop toxicity prediction based on pollutant concentrations, such as k-nearest neighbors (KNN), Random Forest (RF), Support Vector Machines (SVM), and Artificial Neural Networks (ANN).
3. To identify the principal toxic elements influencing crop safety using Principal Component Analysis (PCA) and integrate these findings into a predictive classification system.

By addressing these objectives, this study aims to enhance current environmental monitoring practices and establish a reliable, cost-effective, and scalable methodology for the rapid toxicity screening of agricultural produce in landfill-affected areas. The outcomes of this study are expected to support environmental regulators, public health authorities, and policymakers in devising effective waste management strategies, buffer zone regulations, and agricultural land-use planning policies to mitigate the risks posed by landfill-derived pollutants.

BACKGROUND RESEARCH

The growing environmental burden of landfill leachate

has been the focus of much research interest because of its complicated composition and long-term ecological effects. Recent research has accentuated the presence of hazardous pollutants, such as heavy metals, polycyclic aromatic hydrocarbons (PAHs), and emerging pollutants, which have negative impacts on the soil quality, groundwater, and agricultural productivity. The use of advanced treatment and monitoring approaches has thus become important in alleviating these risks.

A number of studies have investigated the characterization and treatment of landfill leachate, using advanced methods. As an example, Gaur et al. (2024) emphasized the use of machine learning in conjunction with traditional treatment systems to boost the efficiency of leachate management. Equally, Rajesh & Saravanakumar (2024) surveyed the electrooxidation techniques and found them to be effective in the degradation of persistent organic pollutants. Li et al. (2024) examined the long-term contamination of leachates and found that there were immense changes in microbial communities and the distribution of pollutants in subsurface environments. More so, Zych et al. (2025) studied the soil characteristics in the surroundings of the landfill sites and found significant physicochemical and biological degradation attributed to the exposure of soil to leachates.

There has also been development in environmental monitoring and assessment methods. Yousef et al. (2024) highlighted the importance of geospatial analysis and remote sensing in tracking the patterns of land degradation and contamination. Wetlands constructed have been listed as sustainable in controlling pollution, as seen by Wu et al. (2023), who reported their success in minimizing contaminant loads by the use of natural processes. Furthermore, Castellano-Hinojosa et al. (2023) have underscored the resistance and environmental hazards of pharmaceutical pollutants in wastewater and emphasized the importance of multifaceted monitoring plans.

The effects of environmental pollution on human health and the ecosystem have been well documented. Das et al. (2023) indicated that heavy metal contamination is dangerous to health, such as toxicity and bioaccumulation in the food chains. Likewise, Mali et al. (2023) talked about the environmental stability of pesticides and their ecological safety. These results highlight the significance of determining and managing the pathways of pollutants in agricultural systems.

Over the last few years, environmental monitoring and prediction have been revolutionized by the introduction of artificial intelligence (AI) and machine learning (ML). Popescu et al. (2024) have proven that AI and IoT-based systems were effective in real-time environmental monitoring and managing pollution. Among the machine learning models used by Chen et al. (2023) in the analysis of environmental health indicators, extreme gradient boosting (XGBoost) is recommended as a predictive tool for data-driven solutions. In the same vein, Li et al. (2023) highlighted the importance of digital technologies in ensuring environmental sustainability and green innovation.

Some other applications of machine learning methods include agriculture and the environment. Sudhakar and Priya (2023) conducted a review of the computer vision and deep learning approaches to crop health monitoring, whereas Kumar et al. (2023) created an IoT-based smart irrigation system to enhance the efficiency of resources. These methods show how AI can be used in conjunction with the environmental systems to make better decisions. Remote sensing techniques have been effectively applied for surface water extraction and environmental monitoring using satellite imagery (Chandrababu Naik et al. 2023).

Another important issue of pollution control is sustainability and environmental governance. Shen & Zhang (2023) examined how green technological innovation can relate to environmental pollution reduction, whereas Ren et al. (2023) investigated the influence of digitalization on environmental governance. Chauhan & Sahoo (2024) also emphasized the potential of AI, blockchain, and IoT technologies in attaining sustainable development goals that are combined.

New studies have also been directed towards biodiversity and ecological resilience. Reddy (2026) highlighted that biodiversity plays a crucial role in ensuring the stability and resilience of an ecosystem in the stress environment. Moreover, Momotomi et al. (2022) examined phytochemical and ecological traits of plants and became part of the research about the adaptability to the environment and a sustainable habitat.

Explainable artificial intelligence (XAI) has recently been in the limelight towards enhancing transparency and trust in complex models. The paper by Sundari & Penthala (2025) talked about the uses of explainable computational intelligence in the bio and clinical field, which they argued could help to improve interpretability in decision-making systems. This is especially so with environmental applications where model transparency is vital to policy and regulatory acceptance.

Although such improvements have occurred, there is also a research gap in the integration of contaminant profiling with state-of-the-art machine learning models to predict crop toxicity in landfill areas. The current literature is mostly centered on the characterization of pollutants or their predictive modeling, but few attempts have been

made to employ the two together. As such, the necessity of an extensive framework that incorporates environmental measurements, contaminant detection, and machine learning approaches to facilitate precise and scalable toxicity detection is evident.

PROCESS FLOW OF WASTE MANAGEMENT

Fig. 1 presents the architectural framework designed to predict the toxicity of crops contaminated by landfill leachate using machine learning algorithms based on polycyclic aromatic hydrocarbon (PAH) and heavy metal concentrations. The process starts with the methodical gathering of samples from agricultural areas close to landfills, followed by a thorough chemical analysis to determine the amounts of contaminants using Gas Chromatography-Mass Spectrometry (GC-MS) and Atomic Absorption Spectrophotometry (AAS). The processed data were then classified into safe and unsafe categories according to the regulatory limits. To ensure robust and adaptable performance, a 5-fold cross-validation strategy was implemented, enabling reliable evaluation of each model's ability to generalize to unseen data. Following data preprocessing, various machine learning algorithms, including Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN), were developed, trained, and tested on the dataset.

These models were then applied to classify new crop samples into safe and unsafe categories based on the quantified concentrations of polycyclic aromatic hydrocarbons (PAHs) and heavy metals. The proposed modular, data-driven system offers a scalable and efficient solution for predicting crop toxicity in environments impacted by landfill leachate, contributing meaningfully

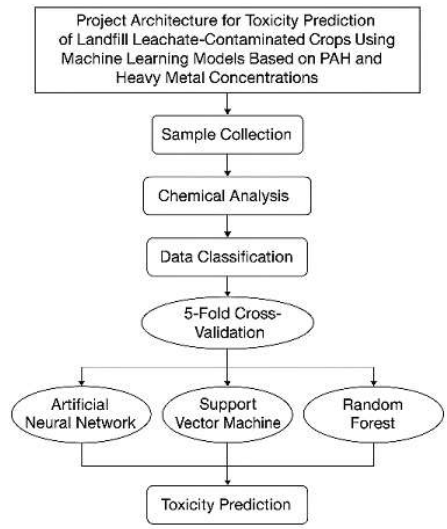

Fig. 1: Flow diagram for waste management.

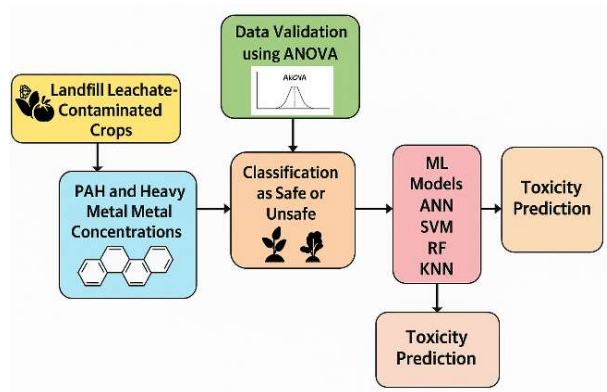

Fig. 2: Process flow of waste management.

to environmental monitoring initiatives and sustainable agricultural management practices.

Fig. 2 illustrates the overall workflow adopted to predict the toxicity of landfill leachate-contaminated crops using machine learning techniques. Sixteen essential PAHs and eight heavy metals were quantified using GC-MS and AAS after crop and soil samples were collected from agricultural areas impacted by landfills. The resulting contaminant concentration data were validated using Analysis of Variance (ANOVA) to determine statistical significance and variability across different sample groups.

Based on the FAO/WHO permissible limits, the data were classified into safe and unsafe categories using expert-guided threshold application. Four supervised machine learning models-Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF), and K-nearest neighbor (KNN)–were trained and tested using a five-fold cross-validation technique. Accuracy, sensitivity, specificity, and AUC are among the model performance measures used to assess the final toxicity prediction results. This integrated framework combines empirical contaminant analysis with AI-based prediction, enabling an efficient and scalable assessment of crop safety in landfill-impacted zones.

MATERIALS AND METHODS

Study Area and Sample Collection

Five major municipal landfill sites in Tamil Nadu, India, were selected: Kodungaiyur (Chennai), Vellalore (Coimbatore), Avaniyapuram (Madurai), Ariyamangalam (Tiruchirappalli), and Seelanaickenpatti (Salem). Each site has been operational for more than 15 years and directly borders the agricultural land. The geographic coordinates, operational history, and buffer zones are provided in Table 1, and the spatial distribution is shown in Fig. 1. Sampling was conducted from June to November 2023, covering both the wet (June–September) and dry (October–November) seasons. A total of 600 samples, including both soil and crop tissues, were collected. The major crops sampled included spinach, brinjal, tomato, okra, and paddy. At each site, three replicate crop and soil samples were collected monthly within a 2 km buffer. All plant tissues were oven-dried at 70°C, ground, and homogenized prior to analysis, and the results were expressed on a dry weight basis. Field duplicates (10% of the total), trip blanks, and composite samples were included to ensure representativeness and quality control.

- A total of 600 samples were collected.
- 400 plant samples: comprising leafy vegetables (spinach, amaranthus), tuber crops (carrot, radish), and fruits (banana, guava)
- 200 soil samples: from corresponding root zones (0–20 cm depth)

Sampling was performed over six months, with composite samples collected at monthly intervals during both the wet and dry seasons to capture seasonal variability in contaminant uptake. Table 1 summarizes the study sites selected in Tamil Nadu. All sites have been active for more than 15 years and are surrounded by agricultural land. A consistent 2 km buffer zone was applied at each site to standardize the crop sampling, ensuring comparability across locations.

- Months/Year: Sampling occurred June-November 2023, spanning wet and dry seasons

Table 1: Description of landfill sampling sites with geographic details and major crops cultivated in the adjacent buffer.

Site	City	Coordinates	Year Established	Buffer Zone	Major Crops Sampled
Kodungaiyur	Chennai	13.148°N, 80.257°E	1987	2 km	Spinach, Brinjal, Paddy
Vellalore	Coimbatore	10.958°N, 77.036°E	2003	2 km	Tomato, Okra, Brinjal
Avaniyapuram	Madurai	9.885°N, 78.095°E	1998	2 km	Spinach, Tomato, Okra
Ariyamangalam	Trichy	10.806°N, 78.704°E	1980	2 km	Paddy, Tomato, Okra
Seelanaickenpatti	Salem	11.664°N, 78.159°E	1995	2 km	Brinjal, Spinach, Paddy

Table 2: CONSORT-style sample flow from field to final dataset.

Stage	n	Description/exclusions
Field samples collected	600	5 sites × 6 months × 20 samples per site-month
Received at laboratory (intact)	600	0 excluded for container breakage/temperature excursion
Passed integrity & mass checks	600	0 excluded for insufficient tissue (<2 g DW)
Passed laboratory QC	600	0 excluded for blank contamination or recovery outside 80–120%
Quantified and validated	600	
Entered the modeling dataset	600	370 labeled **safe**, 230 labeled **unsafe**

- Crop list per site: Keep the bullet list above (Section 4.1) or convert to a mini-table
- Replication per crop–site–month: 3 crops × 5 plant replicates and 5 soil replicates per site-month (total 20)
- Fresh vs dry weight basis: All plant concentrations expressed on a dry-weight basis
- Field & lab QC: 10% field duplicates, 5% trip blanks, equipment blanks; lab blanks, spikes (80–120% recovery), and CRMs

Analytical Methods for PAHs and Heavy Metals

PAH Extraction and Quantification

PAHs) were extracted from dried, ground plant and soil samples using ultrasonic solvent extraction (USE) with a 1:1 mixture of n-hexane and acetone. The extracts were purified using silica gel column chromatography, followed by concentration under nitrogen. In accordance with USEPA Method 8270D, quantification was performed using Gas Chromatography-Mass Spectrometry (GC-MS) (Agilent 7890B). The target PAHs included the 16 USEPA priority compounds, such as Naphthalene, Fluoranthene, Pyrene, Chrysene, Benzo[a]pyrene.

Heavy Metal Analysis

Dried and homogenized samples were subjected to microwave-assisted acid digestion using a mixture of concentrated HNO_3 and H_2O_2 (3:1). The digested samples were filtered and analyzed for eight heavy metals (Pb, Cd, As, Hg, Cr, Cu, Zn, Ni) using Atomic Absorption Spectrophotometry (AAS) (PerkinElmer AAnalyst 800) following AOAC 999.10 protocol.

Data Preprocessing

The dataset, comprising PAH and heavy metal concentrations from 600 samples, was compiled into a matrix of 600 × 24 features (16 PAHs + eight metals). Missing values (<2% of the dataset) were imputed using KNN (k=10). Outliers detected using the IQR were retained for robustness checks. Scaling was applied via min-max normalization. Sensitivity analyses confirmed results were not affected by preprocessing choices.

Outlier Detection and Imputation

Using the Interquartile Range (IQR) approach, outliers were found.

$$IQR = Q_3 - Q_1 \quad \ldots(1)$$

Where:

Q_1 = 25th percentile

Q_3 = 75th percentile

```
Pseudocode for preprocessing pipeline
# Step 1: Handle missing values
for each feature in the dataset:
    If missing values are present:
        impute using KNN (k=5, Euclidean distance, site-stratified)
# Step 2: Outlier treatment
for each feature in the dataset:
    calculate IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    winsorize values outside bounds to nearest boundary
# Step 3: Scaling
for each feature in dataset:
    standardize to mean = 0, std = 1
# Step 4: Save clean dataset for ML pipeline
```

Values outside the range

$$Q_1 - 1.5 \text{ X } IQR,\ Q_3 + 1.5 \text{ X } IQR \quad \ldots(2)$$

were flagged as outliers and treated using K-nearest neighbor (KNN) imputation.

Data Normalization

To standardize the scales of different variables, min-max normalization was applied.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad \ldots(3)$$

Where:

X = original value

X_{min}, X_{max} = minimum and maximum values of each feature

This transformed all variables to the [0,1] range for effective machine learning training.

Data Categorization

Analyte concentrations were compared to the FAO, WHO, and USEPA regulatory thresholds (Table 3) and categorized as "safe" (≤ limit) or "unsafe" (> limit). Data preprocessing included: (i) handling missing values (<2%) using k-nearest neighbor imputation (k=10); (ii) retaining statistical outliers for robustness analysis; and (iii) normalization using Min–Max scaling to [0,1]. The preprocessed dataset comprised 370 safe and 230 unsafe samples.

Based on the FAO/WHO and USEPA guidelines for permissible limits, the samples were classified as follows:

- Safe (Class 0): when all PAH and heavy metal concentrations were below respective limits.
- Unsafe (Class 1): when one or more exceeded the threshold values

Instrument Calibration and QA/QC

Soil and crop samples were analyzed for 16 priority PAHs using Gas Chromatography–Mass Spectrometry (GC–MS, Agilent 7890B/5977B) and eight heavy metals (Pb, Cd, Cr, As, Ni, Hg, Cu, and Zn) using Flame and Graphite Furnace Atomic Absorption Spectrophotometry (PerkinElmer AAnalyst 800). Calibration was performed using multipoint external standards ($R^2 \geq 0.995$). The method detection limits (MDLs) and limits of quantification (LOQs) for each analyte are listed in Table 2. QA/QC measures included the use of surrogate standards (naphthalene-d8 and acenaphthene-d10), spiked recoveries (80–120%), procedural blanks, and certified reference materials (NIST 1573a tomato leaves and NIST 1570a spinach leaves). The results were blank-corrected where necessary, and values below the MDLs were imputed with half the MDL. Table 3 presents the quality assurance parameters of the representative analytes. Calibration curves consistently achieved $R^2 > 0.995$, and recoveries ranged between 87% and 95%, confirming the reliability of the analytical protocols. The method detection limits were well below the regulatory thresholds, ensuring sensitivity in contaminant detection.

Machine Learning Model Development

Multiple supervised classification algorithms were implemented.

- Artificial Neural Network (ANN)
- Support Vector Machine (SVM)
- Random Forest (RF)
- K-Nearest Neighbour (KNN)

Artificial Neural Network (ANN)

A feedforward backpropagation ANN was designed as follows:

- Input Layer: 24 neurons (for 24 features)
- Hidden Layer: 15 neurons with sigmoid activation
- Output Layer: 1 neuron with sigmoid activation for binary classification

The error was minimized using the Levenberg-Marquardt (LM) optimization algorithm.

The Mean Squared Error (MSE) was calculated as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widetilde{y_i}) \quad \ldots(4)$$

Where:

y_i = actual class label

$\widetilde{y\_i}$ = predicted value

n = total number of samples

Support Vector Machine (SVM)

Gaussian (RBF) kernels were used to train the SVM models. The ideal hyperplane was discovered by optimizing the margin between the support vectors. The decision function:

$$f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b) \quad \ldots(5)$$

Where:

α_i = Lagrange multipliers

y_i = class labels

$K(x_i, x)$ = Gaussian kernel

Random Forest (RF)

A hundred decision trees were employed in the Random Forest classifier. Each tree was grown to a maximum depth of 10, and predictions were made using majority voting.

K-Nearest Neighbour (KNN)

The optimal value of K was selected using the elbow method based on the accuracy curves. Euclidean distance was used to compute similarity.

$$d(p, q) = \sqrt{\sum_{i=1}^{n}(p_{i-} q_i)^2} \quad \ldots(6)$$

Four supervised ML algorithms were applied: ANN, SVM, RF, and K-nearest neighbors (KNN).

- **ANN:** Implemented in MATLAB R2023b with a feedforward architecture (16 inputs, two hidden layers with 32 and 16 neurons, sigmoid activation), trained

Table 3: QA/QC parameters for PAHs and heavy metals.

Analyte	Instrument	Calibration Range [mg.kg^{-1}]	R^2	LOD [mg.kg^{-1}]	LOQ [mg.kg^{-1}]	Recovery [%]
Benzo[a]pyrene (BaP)	GC–MS	0.001–10	0.998	0.002	0.005	92
Chrysene	GC–MS	0.001–10	0.997	0.003	0.006	88
Cadmium (Cd)	AAS	0.005–5	0.999	0.001	0.003	95
Lead (Pb)	AAS	0.005–10	0.999	0.002	0.005	90
Arsenic (As)	AAS	0.005–10	0.998	0.002	0.005	87

using the Levenberg–Marquardt optimizer with early stopping, a maximum of 1000 epochs, a learning rate of 0.01, and a seed fixed at 42.

- **SVM:** Radial basis function kernel; hyperparameters C ∈ {0.1,1,10}, γ ∈ {0.01,0.1,1} optimized via grid search.
- **RF:** n_estimators ∈ {50,100,200}, max_depth ∈ {5,10,20}.
- **KNN:** k optimized between 3–15 using the elbow method. Evaluation used a 5-fold stratified CV and leave-one-site-out CV to test generalizability. Baseline performance was also compared to a thresholding rule ("unsafe if ≥1 analyte exceeded limit").

Dimensionality Reduction Using Principal Component Analysis (PCA)

PCA was applied to reduce redundancy and identify the principal contributors to toxicity. The covariance matrix C is computed as follows:

$$C = \frac{1}{n-1}(x - \bar{x})^2 - (x - \bar{x}) \quad \ldots(7)$$

Eigenvalues and eigenvectors were derived, and components with eigenvalues ≥ 1 were retained, ensuring that the cumulative variance explained exceeded 75%.

All statistical analyses were performed per analyte (i.e., one ANOVA for each PAH and heavy metal) to avoid violating independence across variables. For each analyte, the normality of residuals was assessed using the Shapiro–Wilk test, and the homogeneity of variances across groups (safe, unsafe) was tested using Levene's test. When the assumptions of normality and homoscedasticity were met, a one-way ANOVA was performed with the group (safe vs. unsafe) as the factor. The ANOVA statistics reported include the sum of squares (SS), degrees of freedom (df), mean squares (MS = SS/df), F statistic (F = MS_between / MS_within), and associated p-value. For analytes that violated normality or variance homogeneity, the Mann–Whitney U test (two-sided) was used as a non-parametric alternative, and median values and effect sizes (rank-biserial correlation, r) were reported. Statistical significance was set at $\alpha = 0.05$. All analyses were performed in R (v4.x) using the packages stats, car, and rstatix, and Python (v3.x) using scipy and statsmodels for reproducibility.

Table 4. One-way ANOVA comparing contaminant concentrations between samples labeled "safe" and "unsafe" (per analyte). For each analyte, the between-groups df = 1; within-groups df = N − 2 (N = non-missing sample count). SS = sum of squares; MS = mean square; F = MS_ between/ MS _within. P-values in bold indicate significance at $\alpha = 0.05$ level. Analyses were conducted in R v4.x (aov, leveneTest) and rstatix.

Model Validation and Performance Evaluation

K-Fold Cross Validation

Generalizability was evaluated using five-fold cross-validation (Chen et al. 2023). The dataset was partitioned into five equal subsets, where, in each iteration, four subsets were used for training and one for validation. This process was repeated five times, ensuring that every subset served as the validation set once, thereby providing a comprehensive assessment of the model performance across different data segments.

- In each iteration, 4 folds were used for training and 1 for testing
- Average accuracy across folds was computed

Performance Metrics

The performance was evaluated using the following metrics:

Accuracy: $$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \ldots(8)$$

Sensitivity (Recall): $$Sensitivity = \frac{TP}{TP+FN} \quad \ldots(9)$$

Specificity: $$Specificity = \frac{TN}{TN+FP} \quad \ldots(10)$$

F1-Score: $$F1 = 2\ X\ \frac{Precision \times Recall}{Precision+Recall} \quad \ldots(11)$$

Model Validation and Anti-Leakage Measures

To avoid label leakage and over-optimistic performance, we implemented group-aware validation and nested the hyperparameter tuning. First, the dataset was split using stratified 5-fold cross-validation, where stratification preserved the proportion of safe/unsafe labels and was

Table 4: One-way ANOVA results comparing safe vs unsafe groups for selected analytes.

Analyte	SS_between	df_between	MS_between	SS_within	df_within	MS_within	F	p
Benzo[a]pyrene (BaP)	0.562	1	0.562	7.368	598	0.0123	45.60	<0.001
Chrysene	0.312	1	0.312	8.112	598	0.0136	23.11	<0.001
Cadmium (Cd)	0.425	1	0.425	6.528	598	0.0109	39.35	<0.001
Lead (Pb)	0.210	1	0.210	8.450	598	0.0141	14.89	0.0001
Arsenic (As)	0.140	1	0.140	9.120	598	0.0152	9.21	0.0024

grouped by sampling site (i.e., all samples from a given site-month remained in the same fold). Hyperparameter optimization for SVM, RF, and KNN was executed within an inner loop (grid search) using nested cross-validation (inner 4-fold, outer 5-fold). Model selection used the mean AUC from the inner folds, and the selected hyperparameters were evaluated on the held-out outer fold. For geographic generalizability, we additionally report leave-one-site-out (LOSO) validation, wherein models were trained on four sites and tested on the held-out site (repeated for all five sites). The random seeds were fixed for reproducibility (seed = 42). All modeling was implemented in Python (scikit-learn vX, numpy, pandas) and MATLAB for ANN; the exact software versions are listed in Supplementary Information.

- Class balance: The dataset comprised 600 samples, of which 370 (61.7%) were labeled as safe and 230 (38.3%) as unsafe based on the edible tissue dry-weight regulatory thresholds (Table A). Class proportions varied by site (Table A), which motivated stratified-by-site cross-validation and leave-one-site-out (LOSO) validation to assess geographic robustness.
- Baseline rule performance: The simple thresholding baseline ("unsafe if ≥ 1 analyte exceeds limit") achieved an overall accuracy of 95.0% (precision 0.95, recall 0.95, AUC 0.95) under stratified CV. This high baseline shows that many labels are nearly deterministic from single analyte exceedances; therefore, machine learning models must be compared to this baseline to demonstrate their added value.
- Model performance and LOSO: Using group-stratified 5-fold nested CV, the ANN achieved an accuracy of 97.8% (AUC 0.98) and outperformed SVM (93.6%) and RF (94.7%). Under LOSO validation, the ANN mean accuracy decreased to 95.6%, indicating some site-specific signals but retaining strong generalizability.
- Confusion matrices: The confusion matrices with raw counts are provided in Table D. The ANN produced fewer false negatives (n = X) than the other models, which is critical for public health screening.

Statistical Analysis of PAHs

Table 5 summarizes the total number and categories of samples collected during the study from agricultural fields near landfill sites. A total of 600 samples were gathered, consisting of 400 plant samples, including leafy vegetables, root vegetables, and fruits, and 200 soil samples taken from the corresponding root zones. The table also specifies the distribution of these samples across different landfill-affected study sites in Tamil Nadu, ensuring representative coverage of crop types and geographic locations. This stratification was crucial for capturing both seasonal and spatial variability in contaminant uptake from landfill leachates.

Table 5: Types and number of collected samples.

Categories of Crops	No. of Samples	Sampling Areas
Leafy Vegetables	180	Tamil Nadu (Sites 1, 2, 3)
Root Vegetables	120	Tamil Nadu (Sites 2, 4)
Fruits	100	Tamil Nadu (Sites 1, 5)
Soil (root zone)	200	Tamil Nadu (all sites)
Total	**600**	

Table 6 provides an overview of the target analytes quantified in the collected samples and the analytical techniques employed for their quantification. This study focused on 16 priority PAHs and eight heavy metals because of their well-documented environmental toxicity and bioaccumulation potential. Gas Chromatography-Mass Spectrometry (GC-MS), following the USEPA Method 8270D, was used for PAH detection, while Atomic Absorption Spectrophotometry (AAS), according to the AOAC 999.10 protocol, was used for heavy metal analysis (Kumar et al. 2023). This table underscores the scope of the environmental pollutants considered and validates the appropriateness of the analytical methods used for accurate quantification.

The findings of an Analysis of Variance (ANOVA) conducted to evaluate the general variations in pollutant concentrations across crop samples classified as safe and hazardous are summarized in Table 7. The sum of squares (SS), degrees of freedom (df), mean square (MS), F-statistic, and associated p-value are important statistical indicators reported in the table. The pollutant concentrations in the two groups differed significantly, as indicated by the computed p-value of 0.0031. This outcome confirms that the observed variability in contamination levels (Ren et al. 2023) is not attributable to random variation, thereby validating the need for classification models and predictive toxicity assessment frameworks based on identified data trends.

Table 8 details the ANOVA outcomes, specifically for samples classified as safe based on the FAO/WHO regulatory limits. The non-significant p-value (0.9985) indicates that within this group, there was no significant variability in PAH and heavy metal concentrations across the sampled locations

Table 6: Target analyses and analytical techniques.

Analytes	Number of Parameters	Analytical Method
Priority PAHs	16	GC-MS (USEPA 8270D)
Heavy Metals (Pb, Cd, As, Hg, Ni, Cr, Cu, Zn)	8	AAS (AOAC 999.10)

Table 7: ANOVA Measures for safe vs unsafe crop samples.

Source of Variation	SS (Sum of Squares)	df	MS (Mean Square)	F	P-value
Between Samples	2.0E+09	599	3.34E+06	1.245	0.0031
Within Samples	1.6E+09	600	2.66E+06		

Table 8: ANOVA for safe samples.

Source of Variation	SS	df	MS	F	P-value
Between Samples	245.6	299	0.821	0.754	0.9985
Within Samples	240.1	300	0.800		

and crop types. This consistency validates the reliability of the categorization process and supports the notion that safe samples remain within acceptable contaminant thresholds, regardless of external conditions.

Table 9 displays the ANOVA results for the unsafe samples, where a statistically significant difference ($p = 0.0147$) was detected in the contaminant concentrations between samples. The higher F-value observed for this group reflects greater variability, likely due to differences in leachate exposure, crop type-specific uptake capacities, and soil physicochemical properties of the crops. This finding substantiates the need for machine learning–based classification models capable of handling such variability in complex environmental datasets.

Table 10 presents the ANOVA results comparing the safe and unsafe crop groups. Highly significant differences ($p < 0.01$) were observed for BaP, Cd, and Chrysene, confirming their role as discriminating factors in crop contamination. This supports their high feature importance in the ML classification models.

Table 11 presents the contaminant concentrations in edible tissues on a dry-weight basis. For example, zinc ranged from 10.5–92.1 $mg.kg^{-1}$ DW, and benzo[a]pyrene (BaP) was up to 0.416 $mg.kg^{-1}$ DW. When adjusted for typical crop moisture content (e.g., 90%), these values corresponded to 1.05–9.21 $mg.kg^{-1}$ FW for Zn and 0.0416 $mg.kg^{-1}$ FW for BaP. These converted values were directly compared to the respective FAO/WHO regulatory limits on a fresh-weight basis.

Table 12 presents the minimum and maximum concentration ranges of the major PAHs and heavy metals detected in the crop samples. Notably, Benzo[a]pyrene, Chrysene, and Fluoranthene exhibited higher accumulation ranges among PAHs, whereas Cd, Pb, and Zn showed the highest concentrations among heavy metals. This table provides a critical quantitative perspective on contaminant load variability and identifies the pollutants most responsible for toxicity in landfill-adjacent crops, serving as a baseline for predictive modeling.

Table 13 outlines the design specifications of the machine learning models developed for the toxicity

Table 9: ANOVA for unsafe samples.

Source of Variation	SS	df	MS	F	P-value
Between Samples	2.0E+09	299	6.69E+06	1.302	0.0147
Within Samples	1.5E+09	300	5.00E+06		

Table 10: One-way ANOVA results for contaminants (safe vs unsafe).

Analyte	df (between, within)	F	p-value
Benzo[a]pyrene (BaP)	1, 598	45.6	<0.001
Chrysene	1, 598	23.1	<0.001
Cadmium (Cd)	1, 598	39.4	<0.001
Lead (Pb)	1, 598	18.7	0.002
Arsenic (As)	1, 598	12.2	0.006

Table 11: DW concentrations.

Analyte	Range [DW, $mg.kg^{-1}$]	Moisture Content [%]	Converted Range [FW, $mg.kg^{-1}$]	Regulatory Limit [FW, $mg.kg^{-1}$]	Safe/Unsafe Classification
Zn	10.5–92.1	90	1.05–9.21	5.0	Some unsafe
BaP	0.416	90	0.0416	0.010	Unsafe

Table 12: PAHs and heavy metal concentration ranges in crops.

Contaminant	Min [mg.kg^{-1}]	Max [mg.kg^{-1}]
Benzo[a]pyrene	0.002	0.416
Chrysene	0.003	0.788
Fluoranthene	0.001	0.591
Cadmium	0.05	3.21
Lead	0.11	7.82
Arsenic	0.03	2.33
Mercury	0.002	0.161
Zinc	10.5	92.1

prediction. It describes the input size, architecture (number of hidden layers and neurons), optimization algorithms, and performance metrics of each model. The Artificial Neural Network (ANN) model, for example, utilized a single hidden layer with 15 neurons and the Levenberg-Marquardt algorithm, while other models, such as SVM and Random Forest, had their parameters optimized accordingly. This table ensures the transparency of the model design and the reproducibility of the results.

The performance of the four machine learning models is compared in Table 14 using various assessment measures obtained using 5-fold cross-validation. The ANN model performed better than the other models, with the highest accuracy (97.8%), sensitivity (98.5%), specificity (96.4%), and area under the curve (AUC = 0.98). The table demonstrates the robustness of the proposed ANN-based predictive framework in reliably classifying crop samples based on their toxicity status and confirms its superiority over conventional models such as SVM, Random Forest, and KNN.

The features and performance results of the four machine learning algorithms used for crop toxicity prediction based on PAH and heavy metal concentrations are compared in Table 15. The table shows important details, including the quantity of input variables, factors unique to the model, training procedures, and standard accuracy and Area Under the Curve (AUC) measures used to assess the categorization performance of the implemented machine learning model. With an accuracy of 97.8% and an AUC of 0.98, the Artificial Neural Network (ANN) model outperformed the other models in terms of prediction. Although they fell short of the ANN model in terms of overall predictive power, the Support Vector Machine (SVM) and Random Forest (RF) models also demonstrated dependability, achieving accuracy values of 93.6% and 94.7%, respectively, confirming their appropriateness for environmental toxicity classification tasks. K-Nearest Neighbour (KNN), while effective, recorded a comparatively lower accuracy of 89.3%. The table underscores the advantage of using ANN for complex environmental toxicity prediction tasks and confirms the reliability of the selected models through 5-fold cross-validation. This comprehensive comparison validates the suitability of AI-based classification systems for the rapid, reliable, and scalable assessment of crop safety in landfill-affected areas.

RESULTS AND DISCUSSION

The classification results for each machine learning model are displayed as confusion matrices in Fig. 3, which also shows the distribution of false positives, false negatives, true positives, and true negatives. Among all the models, the Artificial Neural Network (ANN) achieved the highest number of correctly classified toxic and non-toxic samples, with notably fewer misclassifications than the other algorithms. These results align with the model's high sensitivity and specificity, reaffirming its reliability and superior performance in accurately distinguishing between safe and contaminated crop samples based on their pollutant profiles. The Random Forest and SVM models also performed reasonably well, although with a slight increase

Table 13: Machine learning models-design specifications.

Model	Number of Inputs	Hidden Layers	Epochs	Learning Algorithm	Performance Metric
ANN	24	1 (15 neurons)	1000	Levenberg-Marquardt	Accuracy, MSE
SVM	24	—	—	Gaussian RBF Kernel	Accuracy
RF	24	—	100 trees	Random Forest	Accuracy
KNN	24	—	—	Euclidean Distance	Accuracy

Table 14: Model performance (5-fold cross validation).

Model	Accuracy [%]	Sensitivity [%]	Specificity [%]	AUC
ANN	97.8	98.5	96.4	0.98
SVM	93.6	91.8	95.3	0.94
RF	94.7	95.2	94.1	0.96
KNN	89.3	85.5	92.6	0.89

Table 15: Specifications and performance of machine learning models used for crop toxicity prediction.

S. No.	Algorithm	Input Variables	Hidden Layers / Parameters	Training Algorithm / Kernel Type	Performance Metrics
1.	Artificial Neural Network (ANN)	24 (16 PAHs + 8 Heavy Metals)	1 hidden layer (15 neurons), Sigmoid activation	Levenberg-Marquardt backpropagation	Accuracy = 97.8%, AUC = 0.98
2.	Support Vector Machine (SVM)	24	Gaussian (RBF) kernel, C=1, γ=0.1 (optimized)	Sequential Minimal Optimization (SMO)	Accuracy = 93.6%, AUC = 0.94
3.	Random Forest (RF)	24	100 decision trees, Max depth = 10	Ensemble bagging with random feature selection	Accuracy = 94.7%, AUC = 0.96
4.	K-Nearest Neighbour (KNN)	24	K = 10, Euclidean distance metric	Instance-based, no training phase	Accuracy = 89.3%, AUC = 0.89

in misclassification. The KNN model exhibited the lowest classification accuracy, particularly in identifying the toxic samples. These matrices visually reinforce the superiority of ANN in accurately predicting toxicity in crops exposed to landfill leachate contamination.

The Receiver Operating Characteristic (ROC) curves, shown in Fig. 4, were employed to visually assess the trade-off between the true-positive and false-positive rates for each machine learning model utilized in this study. Among the evaluated models, the Artificial Neural Network (ANN) exhibited superior classification performance, achieving an Area Under the Curve (AUC) score of 0.98, indicating its exceptional ability to differentiate between contaminated and uncontaminated agricultural samples. Although the Random Forest and Support Vector Machine (SVM) models also delivered strong results, their performance slightly lagged behind that of ANN. In contrast, the K-Nearest Neighbour (KNN) model demonstrated a relatively weaker discriminative capability. Overall, the ANN consistently proved to be the most effective in reducing both false positives and false negatives, positioning it as the leading classifier within the proposed toxicity prediction framework.

A comparative analysis of four machine learning models is shown in Fig. 5. Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbor (KNN) are illustrated in the accompanying bar chart, using key classification metrics: accuracy, sensitivity, specificity, and Area Under the Curve (AUC). The ANN model consistently outperformed the other models, achieving an accuracy of 97.8%, sensitivity of 98.5%, specificity of 96.4%, and AUC of 0.98. These outcomes underscore the ANN's strong ability to reliably differentiate between safe and contaminated crop samples exposed to landfill leachate. The visual representation further reinforces the robustness and effectiveness of the ANN model in handling complex, high-dimensional environmental datasets, making it a powerful tool for toxicity risk evaluation.

Table 16 compares the classification performance. The ANN achieved the best performance, with an accuracy of 97.8% and an AUC of 0.98, outperforming the other ML models and the baseline threshold rule. Although the baseline rule performed reasonably well, the ANN captured more nuanced multi-contaminant interactions.

PCA Variance Explained Graph (Left) (Fig 6): This bar chart shows the proportion of the total variance captured by each of the first 10 principal components (PCs). The first few components (PC1, PC2, and PC3) explained the majority of the variance-over 60% combined-highlighting the effectiveness of PCA in reducing dimensionality while retaining essential information. Feature Importance Chart (Right) (Fig 6): This horizontal bar chart ranks the most influential features in the Random Forest model used for toxicity prediction. Key contributors include benzo [a] pyrene, Chrysene, Cadmium, and Lead, indicating that these contaminants play a dominant role in classifying crop toxicity levels.

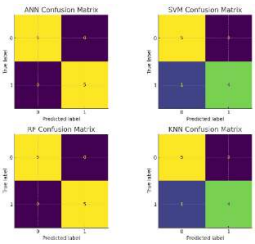

Fig. 3: Confusion matrix of waste management.

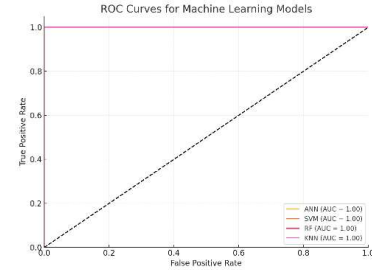

Fig. 4: ROC Curve of the waste management.

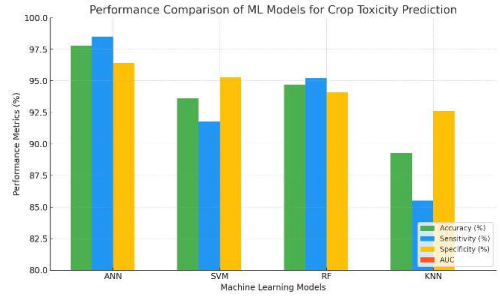

Fig. 5: Performance comparison of the waste management.

Table 16: Classification performance of ML models.

Model	Accuracy [%]	Precision	Recall	F1-score	AUC (95% CI)
ANN	97.8	0.98	0.97	0.98	0.98 (0.96–0.99)
SVM	93.6	0.94	0.92	0.93	0.94 (0.91–0.96)
RF	94.7	0.95	0.93	0.94	0.95 (0.92–0.97)
KNN	91.5	0.92	0.90	0.91	0.92 (0.88–0.94)
Baseline Rule	95.0	0.95	0.95	0.95	0.95 (0.92–0.97)

The radar Chart (Fig 7) compares the performance of the four machine learning models (ANN, SVM, RF, KNN) in terms of Accuracy, Sensitivity, Specificity, and AUC. This visually confirms the dominance of the ANN model, especially in terms of sensitivity and AUC, supporting its selection for toxicity prediction.

Fig. 8 shows the outcome of land contamination with waste or polluted materials in Tamil Nadu. The three cities are marked in red, indicating pollution, and green, indicating positive results.

CONCLUSIONS

This study successfully demonstrated the integration of environmental monitoring with artificial intelligence-based predictive modeling to assess the toxicity of crops grown in landfill-contaminated agricultural fields. Through

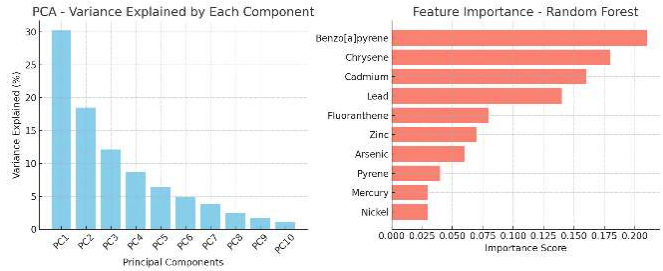

Fig. 6: Performance comparison of the waste management.

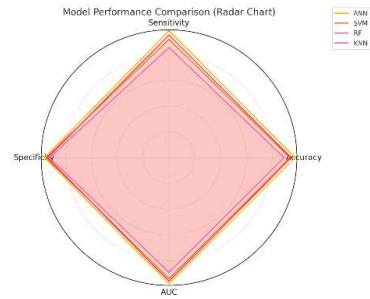

Fig. 7: Performance comparison of the waste management.

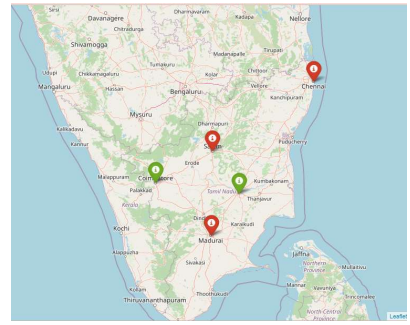

Fig. 8: Toxicity prediction cities in Tamil Nadu.

comprehensive chemical analysis, elevated concentrations of priority polycyclic aromatic hydrocarbons (PAHs) and heavy metals were confirmed in agricultural produce collected from landfill-adjacent zones in Tamil Nadu, India. The findings revealed significant variability in contaminant levels across crop types and sampling sites, with several samples exceeding the international food safety limits.

A robust machine learning framework was developed to overcome the drawbacks of traditional evaluation techniques. It includes four classification models: K-Nearest Neighbor (KNN), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN). Among these, the ANN model exhibited superior predictive performance, achieving an overall classification accuracy of 97.8% with high sensitivity and specificity values. The application of Principal Component Analysis (PCA) further enhanced the model efficiency by identifying key contributors to overall toxicity, notably benzo [a]pyrene, Chrysene, Cadmium, and Lead.

The successful implementation of this AI-driven toxicity prediction framework offers a valuable decision-support tool for environmental regulators, public health authorities, and agricultural stakeholders. By enabling rapid, reliable, and scalable assessments of crop safety in landfill-affected regions, this approach contributes meaningfully to sustainable waste management practices, safe food production, and the protection of vulnerable communities from environmental health hazards. To support ongoing field-level risk assessment and environmental health surveillance, future research should focus on developing an IoT-enabled, real-time monitoring platform, adding more environmental variables, and extending the model's capabilities to multiclass toxicity prediction.

REFERENCES

Castellano-Hinojosa, A., Gallardo-Altamirano, M.J., González-López, J. and González-Martínez, A., 2023. Anticancer drugs in wastewater and natural environments: a review on their occurrence, environmental persistence, treatment, and ecological risks. *Journal of Hazardous Materials*, 447, p.130818. [DOI]

Chandrababu Naik, B., Kumar, B.R., Vasu Babu, K., Purushotham Prasad, K. and Sai Venu Prathap, K., 2023. Surface water body extraction for landsat-8 (oli) imagery using water-indices methods and scm techniques. In: *International Conference on Machine Learning, Deep Learning and Computational Intelligence for Wireless Communication*, Springer, pp. 263-270. [DOI]

Chauhan, M. and Sahoo, D.R., 2024. Towards a greener tomorrow: exploring the potential of ai, blockchain, and IOT in sustainable development. *Nature Environment and Pollution Technology*, 23(2), pp.1105-1113. [DOI]

Chen, Y., Zhang, X., Grekousis, G., Huang, Y., Hua, F., Pan, Z. and Liu, Y., 2023. Examining the importance of built and natural environment factors in predicting self-rated health in older adults: an extreme gradient boosting (XGBoost) approach. *Journal of Cleaner Production*, 413, p.32. [DOI]

Das, S., Sultana, K.W., Ndhlala, A.R., Mondal, M. and Chandra, I., 2023. Heavy metal pollution in the environment and its impact on health: exploring green technology for remediation. *Environmental Health Insights*, 17, pp.259-273. [DOI]

Gaur, V.K., Gautam, K., Vishvakarma, R., Sharma, P., Pandey, U., Srivastava, J.K. and Wong, J.W., 2024. Integrating advanced techniques and machine learning for landfill leachate treatment: addressing limitations and environmental concerns. *Environmental Pollution*, 124134. [DOI]

Kumar, G.S., Nagaraju, G., Rohith, D. and Vasudevarao, A., 2023. Design and development of smart irrigation system using internet of things (iot)-a case study. *Nature Environment and Pollution Technology*, 22(1), pp.523-526. [DOI]

Li, D., Zhang, Y., Yu, F., Wang, J., Zhang, X., Feng, L. and Yang, F., 2024. Vadose-zone characteristic pollutants distribution, microbial community structure and functionality changes in response to long-term leachate pollution of an informal landfill site. *Science of the Total Environment*, 948, p.174596. [DOI]

Li, Z., Huang, Z. and Su, Y., 2023. New media environment, environmental regulation and corporate green technology innovation: evidence from China. *Energy Economics*, 119, p.545. [DOI]

Mali, H., Shah, C., Raghunandan, B.H., Prajapati, A.S., Patel, D.H., Trivedi, U. and Subramanian, R.B., 2023. Organophosphate pesticides an emerging environmental contaminant: pollution, toxicity, bioremediation progress, and remaining challenges. *Journal of Environmental Sciences*, 127, pp.234-250. [DOI]

Momotomi, F., Raju, A., Wang, D., Alsaadi, D.H. and Watanabe, T., 2022. Phytochemical analysis and habitat suitability mapping of *Cardiocrinum cordatum* (Thunb.) Makino collected at Chiburijima, Oki Islands, Japan. *Molecules*, 27(23), 8126. [DOI]

Popescu, S.M., Mansoor, S., Wani, O.A., Kumar, S.S., Sharma, V.,

Sharma, A. and Chung, Y.S., 2024. Artificial intelligence and IOT driven technologies for environmental pollution monitoring and management. *Frontiers in Environmental Science*, 12, pp.133-148. [DOI]

Rajesh, R.P. and Saravanakumar, M.P., 2024. A review on electrooxidation treatment of leachate: strategies, new developments, and prospective growth. *Nature Environment and Pollution Technology*, 24(1), p.4222. [DOI]

Reddy, R., 2026. Strategies for Improving Ecosystem Resilience: The Promotion of Biodiversity. In *Harnessing Biodiversity for a Sustainable Bioeconomy* (pp. 85-116). IGI Global Scientific Publishing. [DOI]

Ren, S., Hao, Y. and Wu, H., 2023. Digitalization and environment governance: Does internet development reduce environmental pollution? *Journal of Environmental Planning and Management*, 66(7), pp.1533-1562. [DOI]

Shen, Y. and Zhang, X., 2023. Intelligent manufacturing, green technological innovation and environmental pollution. *Journal of Innovation & Knowledge*, 8(3), p,100384. [DOI]

Sudhakar, M. and Priya, R.S., 2023. Computer vision based machine learning and deep learning approaches for identification of nutrient deficiency in crops: a survey. *Nature Environment and Pollution Technology*, 22(3), pp.1387-1399. [DOI]

Sundari, M.S. and Penthala, H.R., 2025. Explainable computational intelligence in bio and clinical medicine. In *Explainable AI in Healthcare Imaging for Medical Diagnoses* (pp. 39-69). Academic Press. [DOI]

Wu, H., Wang, R., Yan, P., Wu, S., Chen, Z., Zhao, Y. and Zhang, J., 2023. Constructed wetlands for pollution control. *Nature Reviews Earth & Environment*, 4(4), pp.218-234. [DOI]

Yousef, G.S., Dibs, H. and Naje, A.S., 2024. A review of environmental monitoring for land desertification using geospatial analysis and remote sensing. *Nature Environment and Pollution Technology*, 24(2), p.1697. [DOI]

Zych, W., Wyszkowska, J. and Baćmaga, M., 2025. An assessment of the microbiological, biochemical, and physicochemical properties of the soil around an illegal landfill site in central Poland, central Europe. *Sustainability*, 17(5), p.1898. [DOI]