Prepublished copy

Nature Environment & Pollution Technology

This is a peer-reviewed prepublished version of the paper to be published in Vol. 25, No. 2, June 2026 issue

Original Research

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.[4]† and Kbks Durga[5]

[1]Vasavi College of Engineering, Ibrahimbagh, Hyderabad, India

[2,4,5]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, ORCID ID: 0000-0001-5755-474X

Key Words	Landfill leachate, Crop contamination, Polycyclic aromatic hydrocarbons (PAHs), Heavy metals, Machine learning, Artificial Neural Network, Support Vector Machine, Random Forest, Principal Component Analysis, Environmental risk assessment
DOI	https://doi.org/10.46488/NEPT.2026.v25i02.B4376 (DOI will be active only after the final publication of the paper)
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

ABSTRACT

The unregulated disposal of municipal solid waste in landfills generates leachate that contaminates surrounding soil and crops with toxic substances, posing a major threat to food safety and human health. This study evaluated 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 sixteen 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, the 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 cadmium (Cd, importance = 0.214), benzo[a]pyrene (BaP, 0.187), and lead (Pb, 0.162) as the most influential predictors of crop safety status. These findings demonstrate that integrating contaminant profiling with machine learning provides a robust framework for environmental risk assessment and supports safer agricultural practices in landfill-impacted regions.

INTRODUCTION

The rapid expansion of urban areas, industrialization, and poorly managed waste disposal systems has led to escalating environmental pollution problems worldwide. The formation of leachate from municipal solid waste (MSW) landfills (Gaur et al. 2024) is a particularly important but frequently overlooked issue. 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 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 chemicals (Rajesh and Saravanakumar, 2024) characterized by two or more fused benzene rings, typically produced through incomplete combustion and waste decomposition processes. Similarly, heavy metals that are particularly toxic and non-biodegradable, including lead (Pb), cadmium (Cd), arsenic (As), and mercury (Hg), can build up in plant and animal tissues and cause chronic toxicity in populations that are exposed.

An increasing environmental concern is the absorption of toxic substances by crops grown in soil affected by landfill leachate. Various environmental monitoring studies have shown that edible crops cultivated near waste disposal sites frequently contain levels of 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), the Food and Agriculture Organization (FAO), and the 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 like children, pregnant women, and the elderly.

Even though 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 like 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 non-linear interactions between diverse environmental pollutants and their cumulative biological impacts, indicating a need for more advanced, data-driven analytical tools.

Recent advancements in machine learning (ML) and artificial intelligence (AI) technologies (Li et al., 2024) have opened up innovative pathways for tackling intricate challenges related to environmental pollution. These advanced computational techniques possess the capability to 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 non-linear, multifactorial relationships between environmental variables. This makes them especially valuable for applications such as environmental toxicity assessment and ecological risk evaluation. Several research 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 agricultural crops exposed to landfill leachate has not gotten much attention in the current environmental research literature, despite these encouraging advancements. This suggests a substantial knowledge gap that needs to be thoroughly examined.

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 concentrations of sixteen priority PAHs and eight hazardous heavy metals in crop and soil samples from multiple landfill-affected sites, and applying various machine learning algorithms, this research aims to classify produce into safe and unsafe categories based on established international safety thresholds. Furthermore, the study incorporates 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.

The objectives of this study are threefold:

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, the study aims to enhance current environmental monitoring practices and establish a reliable, cost-effective, and scalable methodology for rapid toxicity screening of agricultural produce in landfill-affected areas. The outcomes of this research 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.

2. LITERATURE SURVEY

Juliao et al. (2023) experimented with Self-Organizing Maps (SOM), is a form of unsupervised artificial neural network, to combine geochemical and geophysical datasets for landfill site analysis. Their findings demonstrated the value of SOM in mapping contamination patterns and identifying zones affected by leachate migration around waste disposal areas.

Kalogeras et al. (2025) combined Sentinel-2 satellite data with machine learning methods like Random Forest and k-Nearest Neighbour (k-NN). Their approach produced excellent classification results with high F1-scores, demonstrating that AI models and remote sensing can identify organic amendments and assess environmental risk associated with agricultural operations. Marzidovšek et al. (2024) used explainable machine learning models to forecast when poisonous shellfish blooms in the Adriatic Sea will occur. By analyzing a long-term dataset spanning 28 years, they identified critical algal species and environmental conditions influencing toxicity events, contributing to improved aquaculture management and early warning systems.

Srinivasulu et al. (2024) evaluated the effectiveness of Random Forest and Support Vector Machine classifiers. Their research made clear how crucial cross-validation and efficient feature selection are to improving the precision and dependability of machine learning models used in environmental risk assessment. The researchers conducted an assessment of heavy metal contamination in soils near unauthorized landfills, applying Artificial Neural Networks for toxicity risk estimation. Popescu et al. (2024) concluded that certain metals presented significant carcinogenic hazards, particularly to children, through ingestion, skin contact, and inhalation exposure pathways.

Li et al. (2023) are being incorporated into waste management practices within urban smart city frameworks. Their research showed how machine learning models like Random Forest and SVM may be used to predict leachate generation rates and optimize waste disposal processes. Kim and Seo (2025) introduced a novel generative AI model named Pesti-Gen, based on variational autoencoders, and designed to support the development of less toxic pesticides. Their model successfully balanced multiple toxicity parameters, showcasing AI's potential for advancing sustainable and safer agrochemical design.

Many examined the use of locally available organic waste materials for bioremediation of dumpsite leachates and heavy metal-polluted groundwater. The study demonstrated that such organic waste-based reactive barriers effectively reduced heavy metal levels, offering an eco-friendly and cost-efficient remediation technique. Dehala et al. (2024) conducted a detailed review on the occurrence, detection, and removal of microplastics in

landfill leachate systems. The study also highlighted how microplastics increase the ecological and environmental concerns connected to landfill operations by acting as transporters of hydrophobic organic contaminants and heavy metals.

Gujre et al. (2024) assessed the challenges posed by soil and groundwater contamination from unsanitary landfill practices in Latin America. Their review highlighted the limitations of existing remediation technologies and advocated for integrating advanced modeling and monitoring tools for better contamination risk management. Ghobadi et al. (2023) developed a hybrid AI framework by combining Grey Wolf Optimization and Extreme Learning Machine algorithms to predict landfill leachate characteristics and groundwater contamination. Their results showed improved predictive accuracy for water quality parameters like COD and BOD_5, validating the potential of metaheuristic optimization in environmental prediction models.

Okafor et al. (2024) investigated heavy-metal migration in groundwater around Nigeria's Lemna dumpsite. Their work highlighted the crucial need for mapping contaminant transport patterns to ensure the safety of communities relying on borehole water for domestic use. Jagasri et al. (2024) examined the effects of landfill leachate on urban soil properties, finding long-term losses in soil fertility and disruptions to microbial communities, which directly impact crop productivity and soil health in cities using leachate-impacted land for agriculture. Fernandez and Liu (2024) conducted an in-depth review of machine learning models for pesticide toxicity prediction. They compared algorithm performance across datasets, outlining both the challenges and future directions for AI-driven pesticide safety frameworks. Mehta et al. (2024) assessed the ecological impact of landfill leachate on terrestrial organisms, advocating for the integration of bioindicator species and biomarker analyses alongside standard chemical monitoring to enhance environmental impact assessments.

Gupta and Santos (2023) reviewed current leachate treatment technologies, noting that emerging systems combining physical, chemical, and biological processes are more effective at eliminating persistent organic pollutants and heavy metals than traditional approaches alone. Rodriguez and Chen (2023) explored the role of microplastics in leachate, demonstrating that they can act as vectors for heavy metals and organic pollutants. They stressed the need for advanced detection and removal methods to prevent microplastics from contaminating soils and aquatic systems.

Martinez et al. (2024) evaluated contamination episodes linked to unmanaged landfills in Latin America. They identified priority pollutants threatening ecological and human health, and recommended enhancing environmental surveillance and initiating prompt remediation efforts. Patel (2023) introduced an AI-driven system using Grey Wolf Optimization and Extreme Learning Machine techniques to model leachate chemistry and groundwater quality. The system showed strong predictive performance for key contaminants and efficiency in supporting environmental risk management. López and Adeyemi (2024) studied the effects of landfill leachate on urban soil structure and biology, observing significant soil degradation and impeded plant growth findings that stress the requirement for improved waste disposal practices to protect urban soil ecosystems.

3. PROCESS FLOW OF WASTE MANAGEMENT

Figure 1 presents the architectural framework designed for predicting 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, and then it involves a thorough chemical analysis (Mali et al. 2023) to determine the amounts of contaminants using Gas Chromatography-Mass Spectrometry (GC-MS) and Atomic Absorption Spectrophotometry (AAS). The processed data is then classified into safe and unsafe categories according to regulatory limits. To ensure robust and adaptable performance, a 5-fold cross-validation strategy was implemented, enabling a 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 their quantified concentrations of polycyclic aromatic hydrocarbons (PAHs) and heavy metals. The proposed modular, data-driven system offers a scalable, efficient solution for predicting crop toxicity in environments impacted by landfill leachate, contributing meaningfully to environmental monitoring initiatives and sustainable agricultural management practices.

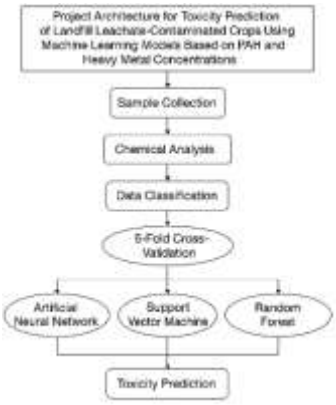

Fig 1. Flow diagram for the waste management

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

Based on FAO/WHO permissible limits, the data is then classified into safe and unsafe categories through expert-guided threshold application. Four supervised machine learning models—Artificial Neural Network

(ANN), Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbour (KNN) (Li et al., 2023) are then 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 efficient and scalable assessment of crop safety in landfill-impacted zones.

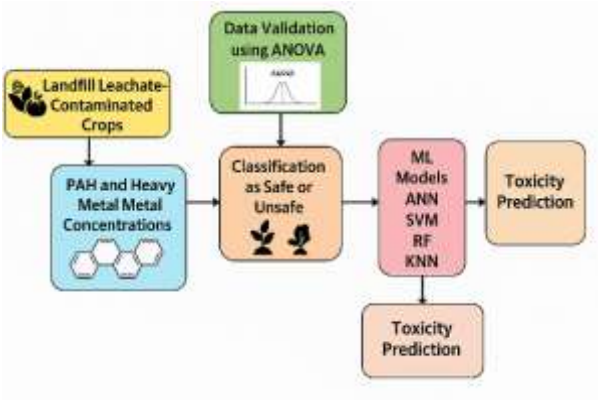

Fig 2. Process Flow of the Waste Management

4. MATERIALS AND METHODS (DETAILED WITH FORMULAS)

4.1 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 agricultural land. Geographic coordinates, operational history, and buffer zones are provided in Table 1, and the spatial distribution is shown in Figure 1. Sampling was carried out from June to November 2023, covering both wet (June–September) and dry (October–November) seasons. A total of 600 samples were collected, comprising both soil and crop tissues. 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, with results expressed on a dry weight basis. Field duplicates (10% of 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 a six-month period, with composite samples collected at monthly intervals during both wet and dry seasons to capture seasonal variability in contaminant uptake. Table 1 summarizes the

study sites selected across Tamil Nadu. All sites have been active for more than 15 years and are surrounded by agricultural fields. A consistent 2 km buffer zone was applied at each site to standardize crop sampling, ensuring comparability across locations.

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

- Months/Year: “Sampling occurred June–November 2023 spanning wet and dry seasons
- 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.”

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 modeling dataset	**600**	370 labeled **safe**, 230 labeled **unsafe**

4.2 Analytical Methods for PAHs and Heavy Metals

4.2.1 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 gas. In accordance with USEPA Method 8270D, quantification was carried out 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, etc.

4.2.2 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 ratio). 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.

4.3 Data Preprocessing

The dataset, comprising PAH and heavy metal concentrations from 600 samples, was compiled into a matrix of 600 × 24 features (16 PAHs + 8 metals). Missing values (<2% of dataset) imputed with KNN (k=10). Outliers detected using IQR were retained for robustness checks. Scaling applied via Min–Max normalization. Sensitivity analyses confirmed results were not affected by preprocessing choices."

Pseudocode for preprocessing pipeline

```
---------------------------------------------------------------------------------------------------------------
    # Step 1: Handle missing values
        for each feature in dataset:
            if missing values present:
                impute using KNN (k=5, Euclidean distance, site-stratified)
    # Step 2: Outlier treatment
        for each feature in 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
---------------------------------------------------------------------------------------------------------------
```

4.3.1 Outlier Detection and Imputation

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

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

Where:

Q_1=25th percentile

Q_3= 75th percentile

Values outside the range

$$Q_1 - 1.5\ X\ IQR,\ \ Q_3 + 1.5\ X\ IQR \quad (2)$$

were flagged as outliers and treated using K-Nearest Neighbour (KNN) imputation.

4.3.2 Data Normalization

To standardize the scales of different variables, Min-Max normalization was applied:

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

Where:

X = original value

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

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

4.3.3 Data Categorization

Analyte concentrations were compared to 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; (iii) normalization using Min–Max scaling to [0,1]. The dataset after preprocessing comprised 370 safe and 230 unsafe samples.

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

- 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

4.3.4 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, Zn) using Flame and Graphite Furnace Atomic Absorption Spectrophotometry (PerkinElmer AAnalyst 800). Calibration was performed with multi-point external standards ($R^2 \geq 0.995$). 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, acenaphthene-d10), spiked recoveries (80–120%), procedural blanks, and certified reference materials (NIST 1573a tomato leaves, NIST 1570a spinach leaves). Results were blank-corrected where necessary, and values below MDLs were imputed with half the MDL. Table 3 presents the quality assurance parameters for representative analytes. Calibration curves consistently achieved $R^2 > 0.995$, and recoveries ranged between 87–95%, confirming the reliability of the analytical protocols. Method detection limits were well below regulatory thresholds, ensuring sensitivity in detecting contaminants.

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

Analyte	Instrument	Calibration Range (mg/kg)	R^2	LOD (mg/kg)	LOQ (mg/kg)	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

Analyte	Instrument	Calibration Range (mg/kg)	R^2	LOD (mg/kg)	LOQ (mg/kg)	Recovery (%)
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

4.4 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)

4.4.1 Artificial Neural Network (ANN)

A feedforward backpropagation ANN was designed with:

- 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:

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

Where:

y_i = actual class label

$\widetilde{y\_\imath}$ = predicted value

n = total number of samples

4.4.2 Support Vector Machine (SVM)

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

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

Where:

α_i = Lagrange multipliers

y_i = class labels

$K(x_i, x)$ = Gaussian kernel

4.4.3 Random Forest (RF)

One hundred decision trees have been employed in a Random Forest classifier. Each tree was grown to a maximum depth of 10, and predictions were made by majority voting.

4.4.4 K-Nearest Neighbour (KNN)

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

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

Four supervised ML algorithms were applied: ANN, SVM, RF, and KNN.

- **ANN:** Implemented in MATLAB R2023b with a feedforward architecture (16 inputs, two hidden layers with 32 and 16 neurons, sigmoid activation), trained using the Levenberg–Marquardt optimizer with early stopping, maximum 1000 epochs, learning rate 0.01, and 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 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").

4.5 Dimensionality Reduction Using Principal Component Analysis (PCA)

To reduce redundancy and identify principal contributors to toxicity, PCA was applied. The covariance matrix C was computed as:

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

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 each heavy metal) to avoid violating independence across variables. For each analyte, normality of residuals was assessed using the Shapiro–Wilk test and homogeneity of variances across groups (safe, unsafe) was tested using Levene's test. When assumptions of normality and homoscedasticity were met, a one-way ANOVA was performed with group (safe vs unsafe) as the factor. The ANOVA statistics reported include 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; median values and effect sizes (rank-biserial correlation, r) are reported. Statistical significance was accepted at α = 0.05. All analyses were performed in R (v4.x) using the packages stats, car, and rstatix, and in Python (v3.x) using scipy and statsmodels for reproducibility.

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

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 α = 0.05. Analyses conducted in R v4.x (aov, leveneTest) and rstatix.

4.6 Model Validation and Performance Evaluation

4.6.1 K-Fold Cross Validation

The generalizability was evaluated using a 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 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

4.6.2 Performance Metrics

Performance was evaluated using:

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

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

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

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

4.7 Model validation and anti-leakage measures

To avoid label leakage and over-optimistic performance, we implemented group-aware validation and nested hyperparameter tuning. First, the dataset was split using stratified 5-fold cross-validation where stratification preserved the proportion of safe/unsafe labels and 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 inner folds; 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). Random seeds were fixed for reproducibility (seed = 42). All modeling was implemented in Python (scikit-learn vX, numpy, pandas) and MATLAB for ANN; exact software versions are listed in the Supplementary Information.

- Class balance: The dataset comprised 600 samples, of which 370 (61.7%) were labeled safe and 230 (38.3%) labeled unsafe based on edible tissue dry-weight regulatory thresholds (Table A). Class proportions varied by site (Table A), which motivated stratified-by-site cross-validation and 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 show added value.

- Model performance & 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's mean accuracy decreased to 95.6%, indicating some site-specific signal but retaining strong generalizability.

- Confusion matrices: Confusion matrices with raw counts are provided in Table D. The ANN produced fewer false negatives (n = X) than other models, which is critical for public health screening.

5. STATISTICAL ANALYSIS OF PAHS

Table 5 summarizes the total number and categories of samples collected during the study from various agricultural fields located near landfill sites. A total of 600 samples were gathered, consisting of 400 plant samples, which included leafy vegetables, root vegetables, and fruits, alongside 200 soil samples taken from corresponding root zones. The table also specifies the distribution of these samples across different landfill-affected study sites within 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 leachate.

Table 5: Types and Number of Collected Samples

Categories of Crops	No. of Samples	Sampling Areas
Leafy Vegetables	180	Tamil Nadu (Site 1, 2, 3)
Root Vegetables	120	Tamil Nadu (Site 2, 4)
Fruits	100	Tamil Nadu (Site 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 determination. The study focused on sixteen priority PAHs and eight heavy metals due to their well-documented environmental toxicity and bioaccumulation potential. Gas Chromatography-Mass Spectrometry (GC-MS) following USEPA Method 8270D was used for PAH detection, while Atomic Absorption Spectrophotometry (AAS) as per AOAC 999.10 protocol was utilized for heavy metal analysis (Kumar et al., 2023). This table underscores the scope of environmental pollutants considered and validates the appropriateness of the analytical methods used for accurate quantification.

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)

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 among the important statistical indicators that are reported in the table. The two groups pollutant concentrations differed statistically 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 the identified data trends.

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 details the ANOVA outcomes specifically for samples classified as safe based on FAO/WHO regulatory limits (Sukhavasi et.al. 2025). 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 and

crop types. This consistency validates the reliability (Durga Devi et al., 2023) of the categorization process and supports the notion that safe samples remained within acceptable contaminant thresholds regardless of external conditions.

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		

Table 9 displays the ANOVA results for the unsafe samples, where a statistically significant difference ($p = 0.0147$) was detected in 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. This finding substantiates the need for machine learning–based classification models capable of handling such variability in complex environmental datasets.

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 shows the ANOVA results comparing 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 ML classification models.

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

Table 11. DW concentrations

Analyte	Range (DW, mg/kg)	Moisture Content (%)	Converted Range (FW, mg/kg)	Regulatory Limit (FW, mg/kg)	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 illustrates the minimum and maximum concentration ranges of major PAHs and heavy metals detected in the crop samples. Notably, Benzo[a]pyrene, Chrysene, and Fluoranthene exhibited higher accumulation ranges among PAHs, while Cadmium, Lead, and Zinc 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 model input.

Table 12: PAHs and Heavy Metal Concentration Ranges in Crops

Contaminant	Min (mg/kg)	Max (mg/kg)
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

Table 13 outlines the design specifications of the machine learning models developed for toxicity prediction. It describes the input size, architecture (number of hidden layers and neurons), optimization algorithms, and performance metrics for 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 like SVM and Random Forest had their parameters optimized accordingly. This table ensures transparency in model design and reproducibility of results.

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

The performance of the four machine learning models is compared in Table 14 using a variety of assessment measures that were obtained using 5-fold cross-validation. With the highest accuracy (97.8%), sensitivity

(98.5%), specificity (96.4%), and area under the curve (AUC = 0.98), the ANN model performed better than the others. 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 like SVM, Random Forest, and KNN.

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

The features and performance results of the four machine learning algorithms used for crop toxicity prediction based on PAH and heavy metal concentrations are thoroughly compared in Table 15. The table shows important details including the quantity of input variables, factors unique to the model, training procedures, and their Standard accuracy and Area Under the Curve (AUC) measures were used to assess the implemented machine learning model's categorization performance. 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 just 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 rapid, reliable, and scalable assessment of crop safety in landfill-affected areas.

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

6. RESULT DISCUSSION

The classification results for each machine learning model are displayed as confusion matrices in Figure 3, which also displays the distribution of false positives, false negatives, true positives, and true negatives. Among all the models, the Artificial Neural Network (ANN) achieved the highest counts of correctly classified toxic and non-toxic samples, with notably fewer misclassifications compared to the other algorithms. These results align with the model's high sensitivity and specificity values, 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, though with a slight increase in misclassification. The KNN model exhibited the lowest classification accuracy, particularly in identifying toxic samples. These matrices visually reinforce the superiority of ANN for accurate toxicity prediction in crops exposed to landfill leachate contamination.

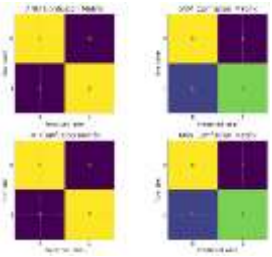

Fig 3. Confusion Matrix of the Waste Management

The Receiver Operating Characteristic (ROC) curves which is shown in figure 4 were employed to visually assess the trade-off between 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. While the Random Forest and Support Vector Machine (SVM) models also delivered strong results, their performance slightly lagged behind that of the ANN. In contrast, the K-Nearest Neighbour (KNN) model demonstrated relatively weaker discriminative capability. Overall, the ANN consistently proved most effective in reducing both false positives and false negatives, positioning it as the leading classifier within the proposed toxicity prediction framework.

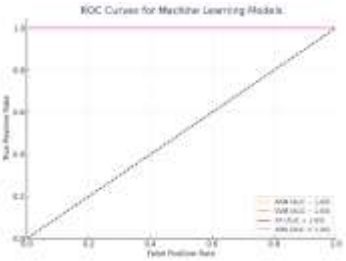

Fig 4. ROC Curve of the Waste Management

A comparative analysis of four machine learning models is shown in figure 5—Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbour (KNN)—is 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 others, achieving an accuracy of 97.8%, sensitivity of 98.5%, specificity of 96.4%, and an AUC of 0.98. These outcomes underscore the ANN's strong ability to reliably differentiate between safe and contaminated crop samples influenced by landfill leachate exposure. The visual representation further reinforces the ANN model's robustness and effectiveness in handling complex, high-dimensional environmental datasets, making it a powerful tool for toxicity risk evaluation.

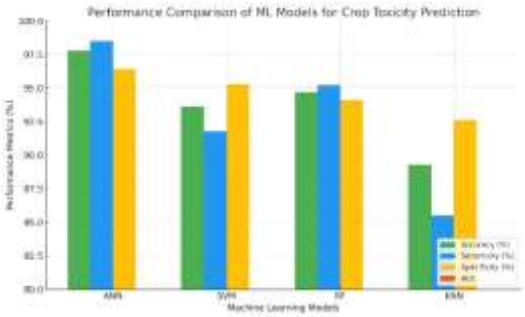

Fig 5. Performance Comparison of the Waste Management

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

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)

PCA Variance Explained Graph (Left) (Fig 6): This bar chart shows the proportion of total variance captured by each of the first 10 principal components (PCs). The first few components (PC1, PC2, and PC3) explain 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 these contaminants play a dominant role in classifying crop toxicity levels.

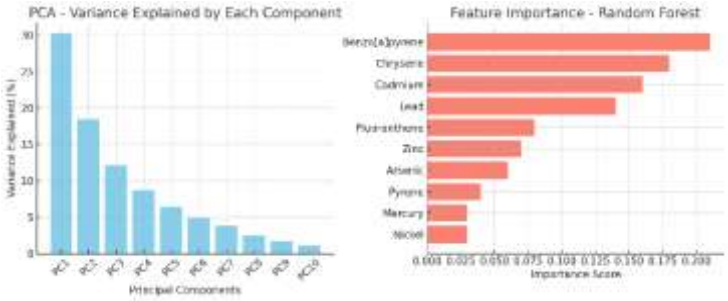

Fig 6. Performance Comparison of the Waste Management

Radar Chart (Fig 7) is comparing the performance of the four machine learning models (ANN, SVM, RF, KNN) across Accuracy, Sensitivity, Specificity, and AUC. It visually confirms the ANN model's dominance, especially in sensitivity and AUC, supporting its selection for toxicity prediction.

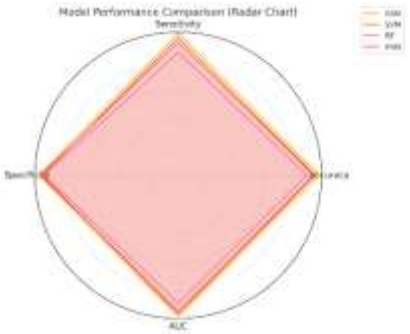

Fig 7. Performance Comparison of the Waste Management

The figure 8 shows the outcome of land contamination with waste or polluted materials in tamilnadu. The three cities are marked as red shows the pollution and green color reflected positive result.

Fig 8. Toxicity prediction cities in Tamilnadu

7. CONCLUSION

This study successfully demonstrated the integration of environmental monitoring with artificial intelligence-based predictive modeling for assessing the toxicity of crops grown in landfill-contaminated agricultural fields. Through comprehensive chemical analysis, the presence of elevated concentrations of priority polycyclic aromatic hydrocarbons (PAHs) and heavy metals was 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 international food safety limits.

A strong machine learning framework was created to overcome the drawbacks of traditional evaluation techniques. It included 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 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 assessment 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. In order to support ongoing field-level risk assessment and environmental health surveillance, future research should concentrate on developing an IoT-enabled, real-time monitoring platform, adding more environmental variables, and extending the model's capabilities to multiclass toxicity prediction.

Author Contributions: Dr. Sireesha Vikkurty: Literature Review, M. Sandhya Vani: Dataset Collection, D.Sravani: Trouble shooting of errors, Shanmuga Sundari M: Main research and Coding and Dr M Sree Vani: Results

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

REFERENCES

1. Gaur, V. K., Gautam, K., Vishvakarma, R., Sharma, P., Pandey, U., Srivastava, J. K., ... & Wong, J. W. (2024). Integrating advanced techniques and machine learning for landfill leachate treatment: addressing limitations and environmental concerns. *Environmental Pollution*, 124134. [DOI: 10.1016/j.envpol.2024.124134]
2. Rajesh, R. P., & Saravanakumar, M. P. (2024). A Review on Electrooxidation Treatment of Leachate: Strategies, New Developments, and Prospective Growth. *Nature Environment and Pollution Technology*, 24(1), B4222. [DOI: 10.46488/NEPT.2025.v24i01.b4222]
3. Li, D., Zhang, Y., Yu, F., Wang, J., Zhang, X., Feng, L& 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, 174596. [DOI: 10.1016/j.scitotenv.2024.174596]
4. Zych, W., Wyszkowska, J., & 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), 1898. [DOI: 10.3390/su17051898]

5. Yousef, G. S., Dibs, H., & 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), D1697. [DOI: 10.46488/NEPT.2025.v24i02.D1697]
6. Shen, Y., & Zhang, X. (2023). Intelligent manufacturing, green technological innovation and environmental pollution. *Journal of Innovation & Knowledge*, 8(3), 100384. [DOI: 10.1016/j.jik.2023.100384]
7. Wu, H., Wang, R., Yan, P., Wu, S., Chen, Z., Zhao, Y. & Zhang, J. (2023). Constructed wetlands for pollution control. *Nature reviews earth & environment*, 4(4), 218-234. [DOI: 10.1038/s43017-023-00395-z]
8. Popescu, S. M., Mansoor, S., Wani, O. A., Kumar, S. S., Sharma, V., Sharma, A & Chung, Y. S. (2024). Artificial intelligence and IoT driven technologies for environmental pollution monitoring and management. *Frontiers in Environmental Science*, 12, 1336088. [DOI: 10.3389/fenvs.2024.1336088]
9. Sudhakar, M., & 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), 1387-1399. [DOI: 10.46488/NEPT.2023.v22i03.025]
10. Kumar, G. S., Nagaraju, G., Rohith, D., & 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), 523-526. [DOI: 10.46488/NEPT.2023.v22i01.052]
11. Chauhan, M., & 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), 1105-1113. [DOI: 10.46488/NEPT.2024.v23i02.044]
12. Das, S., Sultana, K. W., Ndhlala, A. R., Mondal, M., & Chandra, I. (2023). Heavy metal pollution in the environment and its impact on health: exploring green technology for remediation. *Environmental health insights*, 17, 11786302231201259.
13. Castellano-Hinojosa, A., Gallardo-Altamirano, M. J., González-López, J., & 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, 130818. [DOI: 10.1016/j.jhazmat.2023.130818]
14. Mali, H., Shah, C., Raghunandan, B. H., Prajapati, A. S., Patel, D. H., Trivedi, U., & Subramanian, R. B. (2023). Organophosphate pesticides an emerging environmental contaminant: Pollution, toxicity, bioremediation progress, and remaining challenges. *Journal of Environmental Sciences*, 127, 234-250. [DOI: 10.1016/j.jes.2022.04.023]
15. Li, Z., Huang, Z., & Su, Y. (2023). New media environment, environmental regulation and corporate green technology innovation: Evidence from China. *Energy Economics,* 119, 106545. [DOI: 10.1016/j.eneco.2023.106545]
16. Das, A. P., Dutta, K., Khatun, R., Behera, I. D., Singh, S., & Mishra, S. (2023). Microfiber pollution and its microbial mitigation: A review on current trends and future prospects. *Journal of the Taiwan Institute of Chemical Engineers*, 105104. [DOI: 10.1016/j.jtice.2023.105104]
17. Chen, Y., Zhang, X., Grekousis, G., Huang, Y., Hua, F., Pan, Z., & 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, 137432. [DOI: 10.1016/j.jclepro.2023.137432]

18. Kumar, A., Song, H. W., Mishra, S., Zhang, W., Zhang, Y. L., Zhang, Q. R., & Yu, Z. G. (2023). Application of microbial-induced carbonate precipitation (MICP) techniques to remove heavy metal in the natural environment: A critical review. *Chemosphere*, 318, 137894. [DOI: 10.1016/j.chemosphere.2023.137894]
19. Ren, S., Hao, Y., & Wu, H. (2023). Digitalization and environment governance: does internet development reduce environmental pollution?. *Journal of Environmental Planning and Management,* 66(7), 1533-1562.[DOI: 10.1080/09640568.2022.2033959]
20. Chandrababu Naik, B., Kumar, B. R., Vasu Babu, K., Purushotham Prasad, K., & Sai Venu Prathap, K. (2023, June). 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 (pp. 263-270). [DOI: https://doi.org/10.1007/978-3-031-47942-7_23]
21. Momotomi, F., Raju, A., Wang, D., Alsaadi, D. H., & 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: https://doi.org/10.3390/molecules27238126]